Národní knihovna
České republiky
National Library
of the Czech Republic

Zpráva ze zahraniční služební cesty

Jméno a příjmení účastníka cesty	**Marek Melichar**
Pracoviště – dle organizační struktury	ODODD
Pracoviště – zařazení	. . .
Důvod cesty	**IIPC fórum – účast na Preservation Working Group**
Místo – město	**HAAG**
Místo – země	**Nizozemí**
Datum (od-do)	- . .
Podrobný časový harmonogram	10. 5. Cesta do Haagu 11. 5. 9:30 až 17:00 Jednání v Haagu 12. 5. 9:30 až 13:30 Jednání v Haagu pak cesta do Prahy
Spolucestující z NK	0
Finanční zajištění	136
Cíle cesty	Seznámení s činností Preservation WG IIPC
Plnění cílů cesty (konkrétně)	Návrhy zapojení NK do činnosti Preservation WG IIPC – viz. přiložená zpráva
Program a další podrobnější informace	10. 5. - 11:50 Odlet z Prahy, 16:30 Příjezd do Haagu 11. 5. - 9:30 Zahájení jednání (prezentace v General Assembly) 10:00 Jednání pracovní skupin 13:30 Pokračování jednání pracovních skupin 12. 5. - 9:30 Zahájení jednání (Prezentace v General Assembly) 10:00 jednání pracovních skupin. 13:00 Ukončení 19:00 Odlet z Amsterdam Schiphol
Přivezené materiály	
Datum předložení zprávy	25. 5. 2011
Podpis předkladatele zprávy	
Podpis nadřízeného	
Vloženo na Intranet	
Přijato v mezinárodním oddělení	

❖ ZPRÁVA Z JEDNÁNÍ PRACOVNÍ SKUPINY PRESERVATION GROUP, IIPC, 11. A 12. 5. 2011, HAAGUE

Marek Melichar , 13. 5. 2011.

OBSAH

INFORMACE ČINNOSTI SKUPINY IIPC WG PRO PRESERVATION V MINULÉM ROCE:

JHOVE 2

- Nástroje pro extrakci technických MD a charakterizaci – BNF ve spolupráci ve firmou ATOS Origin vyprodukovala Jhove2 modul pro validaci gzip komprese a formátu ARC
- Nastroj, umí extrahovat technická metadata a validuje také objekty uložené uvnitř ARCu
- Jhove 2 zatím nemůže zcela nahradit Jhove 1, zatím nepodporuje některé důležité formáty jako PDF nebo JP2. Vývoj většiny modulů dělá CDL – a nemá už peníze na další vývoj, jako jsou moduly pro WARC nebo html. Bude se jednat o tom, aby z financi IIPC byl znovu osloven ATOS origin. Na stránkách Jhove 2 jsou vyjmenovány formáty, které ještě stihnou udělat, dál se neví.
- Jhove 1 – umí validovat HTML soubory, zatím neumí ani jhove 2

NÁVRH PRO NK

Měli bychom udělat testovací projekt, který by poskytl zpětnou vazbu vývojovému týmu a umožnil nam ziskat s jhove2 zkušenosti. ta by byla užitečná zdaleka nejen pro data z webarchivu. Tzn. Nainstalovat jhove2, konfigurovat modul pro ARCy, připravit testovací balík dat – v řádech x desítek gb – a provést dokumentovanou validaci a různým nastavením

Bude třeba přidělit IT – infrastrukturu a vypracovat dokumentaci (anglicky) o tom, co jsme udělali. Pokud bychom byli schopni proces validace monitorovat z hlediska náročnosti na výkon procesu a času bylo by to výborné

MIGRACE ARC > WARC

- Debata o migraci ARC do WARCu – je to preservation action-přibude místo pro metadata, problém – zatím není standard jak metadata do WARCů plnit.

NÁVRH PRO NK

Vyzkoušet migraci ARC > WARC a zjistit, jestli WARC tools přidávají do těla WARCu nějaká metadata. Porovnat ARC a WARC a podle toho se rozhodnout, co dál, jestli ma cenu v našem kontextu migrovat nebo ne. Bude vyžadovat plán, infrastrukturu a zapojení IT, ODO, i WA.

NEJDŮLEŽITĚJŠÍ TÉMATA PRESERVATION GROUP

PRESERVATION METADATA PRO WEBARCHIV

BNF pracuje na implementaci PREMISu pro WA – bude navrhovat update PREMISu. BNF pošle ke komentáři návrh metadatového schématu pro WA – část bude mimo schéma PREMISu zatím. Během pár měsíců začne BNF toto schéma používat. Editorial Comitee PREMISu bude vyzvána k updatu. Schéma umožní zaznamenat „event" sklizně, nastavení crawleru, seedy, ale také zapsat do xml seznam všech objektu z ARCu nebo WARCU. (také počítají s jinými balícími formáty jako ZIP, TAR, RAR), to pouze volitelně.

NÁVRH PRO NK

Oponovat jejich metadatové schéma. Tak dalece, jak budeme schopni konfrontovat jejich návrh polí s našimi představami pro IOP.

JAK MÁ VYPADAT INFORMAČNÍ BALÍČEK PRO WEB ARCHIV

Mezi členy pracovní skupiny není shoda o tom, co má byt základní jednotkou pro archivaci. Většina archivuje a přidává metadata především na úroveň „harvesting instance" – čili sklizně, další metadata mají pro jednotlivé balíky, a jiná metadata pro collections. Velikosti balík, technologie sklizeni i organizace sklízení se mezi institucemi dost liší, někdo má. jako my, malé ARCy se smíšeným obsahem, LOC.gov má WARCy velikosti GB a vice, s metadaty třeba 4GB.

NÁVRH PRO NK

NK – znovu oživit debatu o tom, jak by mel vypadat AIP pro WA, naším cílem, je jako všech instituci, nakonec vložit data do sdíleného repositáře, udělat reálný ingest a přidat MD. Dohoda o tom, jak AIP balit a kam přidat jaká MD je třeba před implementaci LTP systému v NK. Bylo by dobře oživit use case připravovaný pro IOP – ingest dat z WA.

BUDOVÁNÍ DATABÁZE RIZIK, SOFTWARU A FORMÁTŮ

- Probíhající úkol – nástroje jako PRONOM nejsou pro WA právě ideální je třeba je doplnit. Skupina pracuje na databázi rizik a všech souvisejících technologií (formátů, rendering aplications a browserů, dokumentace k harvestorům atd.)

NÁVRH PRO NK

Měli bychom vyzkoušet projít jejich nástroj na risk assessment. Je to podobné jako DRAMBORA. Udělat pracovní skupinu (ODO + WA + IT) a vyzkoušet si hodnocení jejich nastrojem. Poslat zpětnou vazbu, opakovat za rok. Nástroj bude dostupný na novém webu IIPC.

http://www.ignaciogc.com/netpreserve/risks.php

(zatím tady, bude na novem webu IIPC, pak budou lepe reguovany pristupy)

PREZENTACE KB O NOVÉM E-DEPOTU

E-depot jako základní infrastruktura celé knihovny:

- jsou v něm všechna data (journals, digitized masters and papers i WA)

- ne všechno bude dostupné stejně rychle – digitized master za 20s z pasky

- spojují archiv, digitální knihovnu a katalog do jednoho kusu.

- všechno modulární, z různých komponentů, co vyvinou zveřejní pro ostatní

- Národní archiv NL má od Tessellu SDB, KB se velmi líbí systém workflow, tak budou vypadat i jejich workflow

- trvalý vývoj...počítají s tím, že systém bude TRVALE vyvíjen, protože to má být systém pro trvalé uložení a zpřístupnění

- před tím než začali s novým e-depotem si doma pěkně zametli – sloučili všechny digitální projekty a lepe je provázali, aby se nedělaly zbytečnosti nebo věci , nesmyslné, nespojitelné s jinými komponenty. Vše směřuje k naplnění jasné definované strategie. Chtějí mít jeden integrovaný celek, ne projekty kolem každého SW a dig knihovny, a výzkum – ale instituci, který má mission a cíle, a jednotlivé projekty je pomáhají naplnit.

Během měsíce zveřejní svoje definitivní požadavky na nové e-depot a budou k dispozici v AJ vsem.

e-depot – jejich verze transformačního modulu bude zpracovávat 28 kanálů vstupujících dat s různým nastavením ingestu, další musí byt možné kdykoli přidat. Vše postaveno na rule based workflows, workflow budou řídit vše, včetně accessu třeba.

Nástroje pro charakterizaci a podobné služby - jsou volně použitelné, zapojitelné do workflow všude v E-depotu – jakýkoli nástroj je možné zapojit v archivu, accessu nebo ingestu, kdekoli, nástroj/service je tam ale jen jednou. Týká se všech nástrojů i pro validaci MD například-

Mají nějaké komponenty systemu nad OAIS, především tzv. „Proces data store“ – dokumentace procesu v celém e-depotu pro reporting, management IT, cost analyses, služby vydavatelům atd atd.

V květnu 2012 první ingest –nejdříve migrují journals, převod starých dat, pak přepnou všechny dodavatele obsahu (několik desítek velkých vydavatelů odborných časopisů) na nový E-depot. 2013 začnou do E-depotu dávat data z digitalizace, v létě 2013 websites – single systém, single point of Access.

Systém E-depot se dotkne všeho v knihovně, každý dokument jim projede.

Hlavní jejich snaha je vyhnout se „company lock-in" a také „black box solution" – to už měli dost dlouho od IBM. Momentálně dělají RFI na storage management systém, je pravděpodobné, že tohle možná bude komerční řešení – chtějí mít poměrně složité uložiště, s řadou pravidel a policies pro různé typy dat. Do E-depotu nepůjdou jen data archivní, ale i data krátkodobého významu.

Systém jako celek obsahuje jak databázi pro zpřístupnění (asi i indexy atd), tak archivní databázi a navíc ještě onu „process data store" databázi pro management a CRM.

NÁVRH PRO NK

Sledovat jak budou pokračovat, v blízké době zveřejni (mail konference, web atd.)

- *Requirements na svůj systém*
- *Data model - !*
- *Svoje představu o storage management systému*
- *Sledovat OPF – dají tam všechno, co půjde, z toho co vyvinou*
- *Ředitel NK by tam měl jet, a měl by mluvit s Hilde a lidmi nad ní a kolem ní.*

DALŠÍ ZAJÍMAVÉ PREZENTACE:

- Implementace SORLU do hledání ve WA /dobré nápady – filtrování podle typu dokumentu – formáty pdf, ppt, filtry podle času sklizně nebo full text- nevím kolik toho dovede náš Nutchwax/
- Projekty SCAPE a KEEP – obzvlášť SCAPE je mamuti projekt pokračovaní Planets. Budou dělat 1000 člověku-měsíců – vývoj z vetší části – v Planets se ukázalo, ze nástroje, které mají jsou omezené na menší pocty dat, nelze je použít na miliony objektů. Scape je asi největší projekt na DP v EU ever.
- ISO standard pro WA –digitální strategii to nepůjde. Koordinovaně říct co chceme dělat v čem je přínos (WA digitalizace), marketing a komunikace se stakeholdery. Kvalita – kolik obsahu WA v CR už není online? Lze to zjisti?
- COST – jak vyjádřit náklady na archivaci WA, jak měřit
- Recollection tool loc gov – SEE WEB
- Quality assurence v kontextu WA..

NÁVRHY PRO NK

Marketing pro WA – různé definovat cílové skupiny a na ně zaměřit komunikaci-

Popularizace WA u badatelů (politologové, novináři, studenti, lingviste atd.) – obsah, lepší pohledy na archiv, lepší vyhledávání, filtry, highlights, special collections, text dumps analýzy google NGgram atd. Stanovit cíle – počet UV pro rok 2012, sledovat live kolik toho má NK jak jedina, kolik už není on-line…

Popularizace WA mezi techniky a IT komunitami – analýza obsahu WA – formáty, objemy, technologie sklizeni, emulace a zpřístupnění

Změna technologie sklízení WA

Zvážit – proč nesklízet větší dokumenty (na 100MB, videa atd)

Implementovat efektivní deduplikaci do workflow – uvolněné místo je až 80 % při opakovaném targeted harvestu. Jinak alespoň 30%.

NK CR Obecně

NLCR by si měla udělat jasnou strategii, jak a co chce dělat v digitálním světě: Vytvořit srozumitelnou collection policy, digital strategy,– jasně říct, že WA chceme dělat a dáme na to peníze, resp. budeme je hledat…..stejně tak ostatní projekty, bez jasné strategie a plánu se budeme potácet v chaosu daleko ze KB a BNF, BL atd.. Například rozhodnutí, že nesklízíme WA některé dokumenty (videa z webu) měli bychom mít jasně argumentované ve collection strategii – že se nás to netýká – a nebo týká? Kdo jiný sklidí ten nadměrný obsah?

Podle mě je třeba zřídit koordinátora pro přechod knihovny do digitálního světa, který ale nebude mít jen poradenskou roli, ale i exekutivní: Člověk s vizí, který udělá strategii přizpůsobení digitálnímu světu a pak ji bude postupně prosazovat (Národní knihovna v pohybu), navzdory zájmům různých skupin v NK, firem, který z ní žijí, navzdory organizační strnulosti atd. Inspirovat se organizačními změnami v KB – všechny projekty směřují k naplnění jasně formulované strategie….

-

Zpráva ze zahraniční služební cesty

Jméno a příjmení účastníka cesty	**Jan Hutař, Bedřich Vychodil**	
Pracoviště – dle organizační struktury	Odbor dlouhodobé ochrany digitálních dat 1.5	
Pracoviště – zařazení	Vedoucí 1.5	
Důvod cesty	**návštěva konference Archiving 2011**	
Místo – město	**Salt Lake City**	
Místo – země	**USA**	
Datum (od-do)	- . .	
Podrobný časový harmonogram	Jan Hutař: 14.5. odlet Praha-Paříž-SLC, přílet 14.5. 16.5 workshop PREMIS-+ 17.-19.5. konference Archiving2011 20.5. odlet SLC-Paříž-Praha, přílet 21.5. Bedřich Vychodil: 15.5. odlet Washington-Deitroid-Salt Lake City 15.5 workshop T1D Color in Image Capture, Archiving, T2A Scanner&Camera Imaging Performance: Benchmarking, compliance, and Workflow Monitoring 17.-19.5. konference Archiving2011	
Spolucestující z NK	Bedřich Vychodil – konference Archiving	
Finanční zajištění	Jan Hutař: IOP-NDK Bedřich Vychodil: 0136	
Cíle cesty	účast na konferenci, workshopu, návštěva pracoviště dlouhodobého uchování dat v knihovně FamilySearch	
Plnění cílů cesty (konkrétně)	úkoly splněny – viz níže	
Program a další podrobnější informace	viz níže – poznámky k zajímavým prezentacím	
Přivezené materiály	NA	
Datum předložení zprávy	6.6.2011	
Podpis předkladatele zprávy		
Podpis nadřízeného	Datum: 6.6.2011	Podpis:
Vloženo na Intranet	Datum:	Podpis:
Přijato v mezinárodním oddělení	Datum:	Podpis:

pondělí – tutorial PREMIS – viz příloha zprávy

úterý . . ----------

Implementation of a High Performance Architecture for Managing and Storing Web-Harvested Collections
Michael Smorul and Joseph JaJa, University of Maryland (USA)
text příspěvku viz https://wiki.umiacs.umd.edu/adapt/images/6/6b/Archiving11-smorul.pdf

Zpráva je pracovníkem do mezinárodního oddělení předložena nejpozději při vyúčtování cesty do 2 týdnů po jejím ukončení. Bez cestovní zprávy nebude provedeno vyúčtování. Při výjezdu více pracovníků na tutéž služební cestu s týmž programem lze odevzdat společnou cestovní zprávu.

- Příspěvek se věnoval správě a indexaci uložených WARC souborů z archivace webu.
- Aplikace warc manager – ukazuje co arcy a warcy obsahují, umožňuje rychle vyhledávat, procházet a analyzovat obsah warců. Má GUI.
- Jde o malou databázi, která ukládá data o sklizených stránkách.
- Ukazuje jaká URL jsou uvnitř Warců, kolik kopií jedné stránky máme i kolik zabírá jeden web warců.
- Aplikace vznikla, protože LOC dostala 9TB dat a měli je indexovat, neměli potuchu, co je uvnitř warců a arců.
- bude publikováno jako open source během následujících měsíců

- simple webaccessible preservation systém (SWAP)
- systém uložení dat na různá místa, použitelné nejen pro WA– slices na různá místa a servery

Using Tape for Large-Scale Digital Preservation
Gary Wright, FamilySearch (USA)

- digital records preservation systém (DPRS) – mohou používat všechny části LDS (The Church of Jesus Christ of Latter-day Saints), kromě familyhistory, ty mají vlastní systém, protože mají obrovské množství dokumentů a nechtěli to míchat.
- DRPS je vlastně Rosetta obohacená o různé aplikace před a za systémem, podobně jak je plánováno v NDK.
- Mají vytvořený vlastní DRPS Ingest Tools. Typ storage řešení – storage grid NetApp FAS3170. Původně vyvinuto pro medicínské záznamy – information lifecycle management

Rosetta tedy ukládá data do gridu, ten běží na HDD a má zálohu na páskách IBM TS3500

optimalizace –storage layer mezi Grid a rosettu

očekávají, že budou mít 100+ PB AV dokumentů
archival storage medium mají pásky

- analýza nákladů, 1/3 ceny oproti HDD za deset let (interní průzkum, kompletní náklady)
- na objemy typu PB je nejlevnější
- Kazdý rok prochází celý archiv a kontroluji integrity

pásky mají i své stinné stránky

- human error
- životnost pásek – odvíjí se od zacházení a prostředí
- kvalita pásek
- chyby HW – pásky nebo mechaniky
- validace integrity dat – dělají jednou ročně
- data transfer náklady
- optimalizace zápisu
- rychlost přístupu
- maximalizace využití pásky, není vždy využit všechen storage pásky
- verifikace integrity – jak se zjistí kvalita zápisu? zápis bez chyby….

Rosetta běží na discích – metadata, content data na páskách

Moving On: When it is Time to Re-Archive
Michael Selway, Quantum Corporation (USA)

Cena za HW klesá, ovšem velikost jednotlivých filů stoupá, tím pádem stoupá i cena za uložení (za HW). Data se budou muset migrovat na jiné úložiště, dříve nebo později, z disku na disk není problém, páska na pásku je problém, je handlováno archivním systémem – jinak to nejde, u HDD přímo.
Migrace dat z dig. archivu je ještě problémovější, arch. systémy nejsou kompatibilní. Jak se na to připravit –migrace a příprava zabere čas.

- HDD – týdny
- pásky – měsíce
- arch. systém – roky

indikátory, že arch. systém začíná zastarávat a mohl by s ním nastat problém

- žádná roadmapa
- převezetí jinou firmou
- nové funkcionality se opozdí nebo nejsou vyvíjeny vůbec
- technická podpora je špatná a trvá čím dále více času
- …viz sborník

co s tím?
split migration (full split migrace)
stará data jsou neaktivní

- co vše spočítat
- na co si dát pozor
- na co se soustředit
- co vše vzít v potaz

velmi dobrá přednáška o migracích, k využití v IT

FamilySearch: An End-to-End Process for Scanning, Characterizing, Preserving and Providing Access to Very Large Collections of Vital Records
Tom Creighton, FamilySearch (USA); Jonathan Tilbury, Tessella plc (UK); and Mark Evans, Tessella Inc. (USA)

- Ve FamilySearch začali napřed s mikrofilmy, od poloviny 19 st.
- 3,3 miliony mikrofilmů, genealog. záznamy ze 105 zemí světa, 12 miliard jmen
- zdigitalizovat vše a uložit zabere 90TB storage
- mikrofilmy mají uloženy v granitové skále
- skenování mikrofilmů
- s tím se pojí problematika DP
- 2025 budou mít 300PB dat

- DPS je jejich file systém archiv na páskách, SDB (Tessela) s tím pracuje
- vyvíjejí storage adapter mezi HW a SDB
- chtějí mít ještě jeden LTP systém pro ta samá data, bude ingestovat DIPy toho prvního, může to klidně být Rosetta nebo něco jiného

The Audit and Certification of FDsys
David Walls, US Government Printing Office (USA) _ Kate Zwaard přednášela

- FDsys – CMS systém, search engine a preservation repository v jednom, prošel self-auditem TRAC, dodržuje standardy vedoucí k dlouhodobé ochraně, odpovídá OAIS, http://www.gpo.gov/fdsys/

- ukládají data přicházející z administrace US vlády, kongresu, federálních úřadů, weby apod.
- TRAC, hathi, UNT, Portico, metaarchive, chronopolis v procesu prošly TRAC certifikací

- stupnice compliancy 1-5, ani jedno není, že by audit byl neúspěšný,
- chtějí projít externím auditem

How Long is Long-Term Data Storage?
(Focal), Barry M. Lunt, Brigham Young University, and Douglas Hansen, Wayne Rust, and Mark Worthington, Millenniata, Inc. (USA)
o životnosti CD, DVD, USB a disků

- Pěkné prezentace o mediích degradace, mikrosnímky, koroze
- FLASH technologie 10-13 let se informace ztrácí /díky tomu že je kanál plny elektronů obklopeny izolantem, který kousek po kousku ztrácí náboj, tím dojde k odlivu elektronu a odpojeni tranzistoru, tedy ztrátě informace/
- MTTF by mělo byt schopno uchovat informaci na 100let

Test archivních disků, disky byly uchovány v perfektních podmínkách a presto se objevily chyby:

- Library 1 2,1%
- Library 2 1,8 %

Souhrn průměrné životnosti:

- Pásky 10.50 let
- HDD 1-7 let
- Viz sborník

Quality Assurance of Digital Information in Long-Term Digital Preservation
Margarita Korenkova and Ann Hägerfors, Lulea University of Technology (Sweden)

- celý příspěvek se věnoval tzv. significant properties digitálních objektů
- průzkum co jsou signif. properties v dokumentech z různých projektů
- shrnutí relevatních SP pro dlouhodobou ochranu dat v archivech a knihovnách, které zaručí, že bude budoucí uživatel datům rozumět

Towards Interoperable Preservation Repositories: Repository Exchange Package Use Cases and Best Practices
Joseph Pawletko, New York University, and Priscilla Caplan, Florida Center for Library Automation (USA)

- TIPR – testování a vývoj exchange formátu pro výměnu metadat

- různé repozitáře mají různé formáty SIP, AIP i DIP
- jak udělat interoperabilitu mezi heterogenními repozitáři?
- TIPR – dip ven do RXP, ostatní repozitáře musí znát RXP formát a normálně si to přeberou, tj. každý repo zná 2 formáty, svůj a RXP
- ne že DIP jde do hub a ten překládá metadata pro každý repozitář zvlášť, to by to musel někdo udržovat apod.

- RXP – mets a premis semantika – je to mets plus další soubory metadat, ne v metsu, ale vedle v balíčku

- na co je to dobré? na co to můžeme potřebovat? succession, disaster recovery,
- succession - při migraci na nové úložiště
- jako výměnný formát
- export aip jako rxp balíků
- disaster recovery – uložíme někam rxp a v případě potřeby je nacpeme zpátky do repozitáře jako aip
- migrace SW
- starý repozitář exportuje aip jako rxp, nový systém vezme rxp a přemění je na sip
- diversifikace – uložení dat v různých formátech, tj. např. záloha jako rxp
- migrace dat v obskurních formátech – dát do rxp – čistě pro přenos k dalšímu zpracování
- vrátí se pak zase nové, obohacené rxp do archivu

SARKK—Comprehensive Digital Archive Services for Finnish Municipalities
Katariina Ryhänen, Etelä-Savon Tietohallinto Oy (Finland)

- firma, kterou vlastní 300 finských měst, poskytuje služby úřadům státní a městské správy v ICT, long-term a storage služby, elektronická spisová služba
- plán na poskytování digitalizačních služeb
- speciální GUI pro odeslání, vyhledání a zobrazení dat
- počítají i s preserv. planning funkcionalitou,
- je to vlastně datové úložiště pro města, je tam poplatek měsíční, ale firma je nezisková ze zákona
- obdoba našich krajských projektů, jen ve Finsku se domluvili, a dělají to všichni dohromady na jednom řešení pro východní část Finska, všichni zákazníci jsou vlastně majiteli té firmy a majiteli těch svých dat

Magnetic Tape Technology – economic advantages for preservation
Gary Francis, Oracle USA

- Oracle prezentace
- TCO, škálovatelnost, redukce risků
- clipper group 2010 – in search for the long-term archiving solution – tape delivers significant TCO advantage over disk
- 5TB na 1 cartridge – sun/oracle

středa 18.5.2011 ----------

Color In Digital Preservation,
Robert Buckley University of Rochester/NewMarket Imaging;
Steven Puglia, National Archives and Records Administration; and Michael Stelmach, Library of Congress (USA)

- Podle Buckleho ICC profil není potřeba pro naše účely /knihy, noviny/. I tak je třeba vložit informaci o barevném prostoru, v kterém je informace /obrázek/ uložen. Např. sRGB, RGB, atd.
- Je tedy ICC profil důležitý?
    - Důležité je vybrat správný barevný prostor /Color Space/ s dostatečným počtem barev /GAMUTem/
        - sRGB /často je do defaultní nastavení skenerů/
        - AdobeRGB
        - Pro FotoRGB
- Je důležitý pro barevné dokumenty, kde je kladený důraz na přesnou reprezentaci /uchování a reprodukci/ barvy.

Multispectral Image Archiving of Watermarks in Historical Papers,
Peter Meinlschmidt, Wilhelm-Klauditz-Institut, Fraunhofer-Institute for Wood Research, and Volker Märgner, Technische Universität Braunschweig (Germany)

Požívá se pro osvětlení
- Reflected light
- Transmited light

Ale pod tímto světlem není filigrán /vodoznak/ vidět, proto využívají multi-spectral imaging
- Testují termal imaging
    - Nahřátá měděná deska /3000nm a 5000nm/, přes ní je daná černá fólie aby se obrázek nedotýkal originálu
    - Tato technika funguje pro různé tloušťky a různé druhy pigmentů a inkoustů
    - Experimentují i s jiným spektrem aby tuto techniku zdokonalili
    - Dosahují vyšší rozlišení než jiných alternativních technik
    - Problém je s teplotou papíru, záleží na době expozice, když je papír černý, tak se ohřívá rychleji

Implementing a Quality Assurance Program for Monitoring Scanner Performance, Michael J. Horsley and John T. Berezich, National Archives and Records Administration (USA) DAITSS

- Zabývají se nastavením image quality pro proces digitalizace a mikrofilmování
- Různé kopie originálu musí mít zaručenou kvalitu zpracování /originál skelněný negativ, mikrofilm, sken a různé kopie/
- NARA 2004 Guidelines https://docs.google.com/viewer?url=http%3A%2F%2Fwww.archives.gov%2Fpreservation%2Ftechnical%2Fguidelines.pdf
- Metamorfose
- Atd. viz sildes
- Analizují data – Quantitive performance /slides/
- Web based database /sharepoint/ – pomáhá při definování problémů v procesu digitalizace
- Chtějí se stát součástí ISO 9000

Preservation in a Digital Age
Jay Verkler, FamilySearch (USA)

- vývoj dalších částí s Tessellou pro SDB, budou k dispozici všem, open community[1]
- AIP není statické (doplňování, nové verze atd.)
- 15TB ingest /denně mají – dobrovolníci posílají data z archivů, plus digitalizace
- data loss is intrinsic > manage and mitigate that loss > periodická validace bitů, stále dokola
- 6 bodů o čem je DP – validace, charakterizace, formáty apod.
- počítají s tím, že budou mít data na více typech médií
- zajímavé body o udržování prostředí pro pásky a provoz serverů – ochrana proti požáru, ztrátě energií apod.
- chtějí nabízet „preservation as a service" – pro menší organizace, mnoho výhod, zatím jen nápad

Curation of the End-of-Term Web Archive: Classification and Metrics
Kathleen Murray Lauren Ko, and Mark Phillips, University of North Texas (USA)

- sklízení webů US administrativy od roku 2009
- eotcd archiv http://research.library.unt.edu/eotcd/wiki/Main_Page
- používají WARC, zpřístupnění přes aplikace
- pro vyhledávání nutno znát URL, nejsou tam popisná metadata ani věcný popis
- 16TB dat
- věcná klasifikace jednotlivých domén (1500), SuDocs klasifikační systém
- 16tisíc domén, rozsekali domény (ne Warcy) na hlavní a poddomény, ty pak klasifikovali a pak to pospojovali dohromady, každá subdoména měla svou váhu

DAITSS Grows Up: Migrating to a Second Generation Preservation System (Focal)
Priscilla Caplan and Carol Chou, Florida Center for Library Automation (USA)

- Daitss systém verze 1 se nedal nainstalovat jinde, sice open source, ale nešlo to ;-)
- cíle pro D 2.0 – podpora formátů, flexibilita, interoperabilita, možnost použít jinde… lepší práce s procesy, používá externí nástroje, action plans are standalone XML with processing instructions
- DAITSS 2 je řetězec web servisů (RESTful web services) – lze použít kdykoliv kdekoliv v procesech
- implementuje kompletně PREMIS data model
- workspace information package (WIP), to samé jako naše PSP

- refresh funkce – provedení service na již uložených datech AIP, vznikne nové AIP
- disseminate- do a refresh and export new AIP as DIP
- withdraw – remove AIP from storage, retaining provenance

- file identifikátory se změnily z interních na PRONOM, UDFR

[1] FamilySearch má LPT systém SDB od firmy Tessella

- všechna AIP se musela změnit na nový typ oproti DAITSS 1
- 300tis AIPs, 30 mil. souborů, 80TB
- refresh funkce lehce upravená – bere D1 AIP, vytvoří D2 AIP a tím se doplní i databáze

A Community Driven Micro Services Architecture Supporting Long Term Digital Preservation
Mark Evans and Bill Steel, Tessella Inc. (USA), and Robert Sharpe, James Carr, Alan Gairey, and Jonathan Tilbury, Tessella plc (UK)
pilotní projekt nový – s NDIIPP v LoC , SDB pro několik institucí menších

micro-services
- spojeno do procesu- velká komplexní funkcionalita se rozbije do malých procesů/služeb a ty lze pak libovolně spojovat
- jednoduché na definování, na údržbu, vývoj
- levné je přidat další funkcionalitu, snižuje závislost na jedné technologii
- lze sdílet mezi systémy i komunitami
- typy: custom code (pro zákazníky), wrapped COTS aplikace, wrapped open source nástroje, webové formuláře
- mikroservicy dávají SDB obrovskou flexibilitu úprav a doplnění funkčnosti pro všechna workflow
- i storage adaptér mají jako micro službu
- všechny web services jsou přístupné přes API

- aktivní a pasivní DP – to samé jako logická a ochrana bit-streamu
- **během 6 měsíců do SDB chtějí přidat emulátor!!**
- SDB lze mít lokálně, nebo mít hostované, storage lze mít v cloudu (S3 http://aws.amazon.com/s3/), SaaS funkcionalitu bude mít SDB brzy
- systém podporuje multi tenancy – každý vidí svou část, může mít dohled nad obsahem, funkcionalitou, policy apod.

Pozn:
- z debat apod. se zdá, že Tessella s SDB začíná mít navrch nad Rosettou, hlavně v USA
- např. KB se líbí více SDB pro svou flexibilitu a workflow možnosti, to samé Familysearch…

v příspěvku dále zazněly další skutečnosti o nasazení SDB ve FamilySearch
- požadavek ingestovat 20TB za den
- 4.4TB SIP v testu (10MB JP2 soubory)
- uložení na páskách

komunita SDB
- velký vliv na vývoj
- vznik 2008
- sdílení zkušeností, micro services, workflows, nástroje atd.
- možnost ovlivnit roadmap

Automated Metadata Creation to Enhance Search Capabilities in GPO'S Federal Digital System

Lisa LaPlant and Blake Edwards, US Government Printing Office (USA)

- popisná metadata
- FDsys – OAIS dig. archiv pro vládní dokumenty, dostupné vyhledávání na GPO.gov
- 50 kolekcí, každá různá metadata, jak vytvořit schéma?
- tisíce elementů
- MODS
- ruční vkládání do šablon, ty jsou uloženy v xml, dle xml se otevřou okénka k vyplnění
- automatické parsování - z názvu titul x, část 2 vznikne xml <název>x</název>, <part>2</part>

- to celé vyjádřeno v DMD – data model definition – to parsování, xml šablon apod.
- z DL lze stáhnout PREMIS, MODS metadata, PDF pro obrazy – viz např. http://www.gpo.gov/fdsys/search/pagedetails.action?st=prague&granuleId=&packageId=DCPD-200900228

čtvrtek 19. . ----------

Preservation Starts from the Beginning, Michael Wash, US Department of Transportation (USA)

George Eastman: „Kodak doesn't sell film It sells memories"
Autographic Kodak 1916 – bylo možné po vyfocení obrázku napsat stylusem na druhou stranu filmu zapsat informaci o vyfoceném snímku (metadata)

"The No. 1A Autographic Kodak Special of 1917 was the variant with coupled rangefinder, following the No. 3A Autographic Kodak Special of 1916 which was the first rangefinder camera. It had a Kodak Anastigmat f.6.3 lens and a Kodamatic shutter with speeds from 1/2 to 1/200 sec. plus bulb and time mode."
http://camerapedia.wikia.com/wiki/No._1A_Autographic_Kodak_Special

Kodak Advantix
Poslední fotoaparát který byl schopen zaznamenávat metadata na mnoho let, se tento vývoj zastavil

Colorite: A Flexible Cross-Platform Software Solution for Automatic Image QualityAnalysis Using Arbitrary Targets, Henrik Johansson, National Library of Sweden (Sweden)

- Jde o Color management a detekování a analýzu targets
- Vytvořili software COLORITE
- Staví na projektech FADGI a Metamorfoze
    - Target se detekuje automaticky, snaží se aby tento proces nezatěžoval stávající wokflow. Je možné využívat jakýkoli target!
    - Podporované formát TIFF, JPEG, JP2, PNG
    - Operátor nemusí mít žádné znalosti o ICC profilech ani o teorii záznamu
    - Používají state-of-art feature based image matching algoritm /ImageMagic/. ImageMagic ma v sobě funkci.
    - Algoritmus najde Color target i přesto, že je položen pod úhlem a dokonce, když jeho část je překryta nebo poškozena!
    - Výsledek testu je uložen ve formát XML, generováno automaticky
    - Mají GUI, které se bude v budoucnu vypnout aby se proces urychlil pro BATCH proces. Pracují na tom.
    - Jsou ochotni sdílet informace a nabízí spolupráci v této oblasti. Je možné kontaktovat Henrika Johanssona.

What if the Image Quality Analysis Rates My Digitization System a " No Go"?, Dietmar Wueller, Image Engineering (Germany)

- Golden thread
- UTT targer
- Interně UTT používá výrobce skenerů Zeutschel
    - Problém není v současné době s rozlišením, ale se zaostřením /focus/. Pokud není kvalitně provedeno, tak na reprodukci není dostatek detail a pak je vysoké rozlišení k ničemu. Některé kamery je velmi těžké zaostřit, např. DSLR (digital single-lens reflex cameras). Není možné se spolehnout na autofocus!

Dobrý rozpis parametrů, které je třeba hlídat v procesu digitalizace, viz sborník.

Establishing Resolution Requirements for Digitizing Transmissive Content: A Use Case Approach, Michael Stelmach, Library of Congress; Don Williams, Image Science Associates LLC; and Steven Puglia, National Archives and Records Administration (USA)

- Vyberou hrany na originálu a vypočítají SFR a podle toho se dá pak odhadnout rozlišení obrázku, tedy rozlišení, které je vhodné na zachování všech informací
- Based on 10% SFR limiting resolution criteria, how much the image information will be captured
    - 1. polovina 20 stol 1200-1600 PPI
    - 2. polovina 20 stol up to 2800 PPI
    - Dufaycolor méně než 750 PPI
    - Autochrom více než 2500 PPI

Digitise More, Pay Less: Optimising the Workprocess for both Heritage Institute and Imaging Provider, Olaf Slijkhuis, Pictura Imaginis (the Netherlands)

- Velmi propracovaný tender dokument /kvalita, parametry, atd./ viz sborník
- Kontrolují míru prachu, chlupů, artefaktů na diapozitivech, fotografiích
- Zjistili pomocí praxe, že na tisk 1:1 stačí I pouhých 100 PPI!!! Jedná se o minimální možné rozlišení, ale postačuje pro základní účely. Je to levnější a tím mohou zdigitalizovat vice materiálu.
- Viz reference /použitá literatura/ ve sborníku. Zde jsou reference na nástroje pro kontrolu obrazu.

Návštěva /debata s familysearch 20.5.2011 -------

- debata o budoucí spolupráci na problematice Digital Preservation
- dodělávají nějaké části k SDB, bude zveřejněno, podle smloluvy dají tesselle, která tyto části bude muset poskytnout zdarma všem uživatelům systému SDB
- mají oba systémy Rosettu i SDB, v obou stejná data, řeší wrapper okolo, kt. to bude celé řídit a synchronizovat
- mají po celém světě 250.000 dobrovolníků, kt. digitalizují, 80tis z nich aktivní, dělají pro ně i částečné ocr, aby rukopisy šly prohledávat a klíčová slova
- snaha spolupracovat na národní úrovni, s archivy a knihovnami
- ukládají tiffy v digitalizaci, jpg pro uživatele
- možná spolupráce jestli budeme mít tessellu, zájem z jejich strany
- debata o možnosti nabízet DP služby jako službu – ostatním, komukoliv – nechtěli by na tom vydělat, ale pomoct ostatním, kt. nemají peníze na nákup LTP – chtěly by to ostatní knihovny? přistoupily by na to, že nemají data u sebe? musely by mít nějakou kontrolu nad daty i nad procesy…šlo by to?

SPOLUFINANCOVÁNO ZE STRUKTURÁLNÍCH FONDŮ EU (EVROPSKÉHO FONDU PRO REGIONÁLNÍ ROZVOJ) PROSTŘEDNICTVÍM IOP

Zpráva ze služební cesty

Projekt „Vytvoření Národní digitální knihovny“

CZ 1.06/1.1.00/07.06386

Jméno a příjmení účastníka cesty	Jan Hutař
Pracoviště – dle organizační struktury	ODF 8.1
Pracoviště – zařazení	vedoucí odboru
Důvod cesty	návštěva konference iPRES 2011
Místo - město	Singapur
Místo – země	Singapur
Datum (od – do)	30.10-5.11.2011
Podrobný časový harmonogram	30-31.10. let Praha-Dubaj>Singapur 1.11. – začátek konference - tutorialy 2.11-4.11 konference 4-5.11 návrat – let Singapur >Dubaj >Praha
Spolucestující z NK	Mgr. Marek Melichar (hrazeno z projektu 0136)
Finanční zajištění	IOP „Vytvoření Národní digitální knihovny“
Vztah k projektu	získání nových informací o problematice digital preservation; o projektech v ostatních knihovnách; konzultace s kolegy a firmami
Cíle cesty	viz vztah k projektu, využít veškeré výstupy pro plánování a chod projektu NDK; využít pro budoucí řešení problematiky digital preservation v NK/NDK
Plnění cílů cesty	splněno – viz podrobný zápis níže a sborník na SPS

Další podrobnější informace	SHRNUTÍ A PŘÍNOS K PROJEKTU NDK - znatelný nástup řešení dlouhodobé ochrany pomocí emulace (v minulých letech migrace) > oba přístupy se zdá se budou doplňovat - posun k ochraně komplexních dat – databáze apod. /NK zatím neřeší/ - spousta příspěvků použitelná i do NK a NDK (webarchivace a ochrana v NK Francie, audity, emulace, info o SDB systému (Tessella) a o systému RODA; certifikace -viz Rouchon apod.) - info o problémech a řešení využití v reálném prostředí nástrojů typu JHOVE, PRONOM aj. - 2 příspěvky o zálohování optických disků – aktuální problém i v NK – ideálně následovat popsané postupy ve sborníku! - jasná potřeba mezinárodní spolupráce a dodržování standardů tak, aby taková spolupráce byla možná podrobněji viz níže
Podpora publicity projektu	NA

Související materiály	
Materiál	***Místo uložení***
sborník z konference	SPS složka se zprávami z SC

Datum předložení zprávy	15.11.2011
Podpis předkladatele zprávy	

	Datum	***Podpis***
Podpis nadřízeného	15.11.2011	
Vloženo na intranet		

Seamus Ross digital curation and preservation

- preserving data sets – využití statistických dtb a výzkumných dat vs. ochrana textových informací
- personal data už nejsou jen fotky v krabici (Flicker apod.)
- banky – spousta osobních dat – využití v budoucnu pro historiky, nutno uchovat – instituce to také dělají
- viz mckinsey.com big data full report pdf
- proč tedy dělat DP? slide 16 – budoucí generace to očekávají; pro historiky, vědce aby měli nějaké zdroje; odkaz o současnosti pro budoucnost – information ecosystem; to enable storytelling
- důraz od ochrany textových informací na ochranu komplexních databází

A capability model for DP – Ch. Becker et al.

- výzkum v rámci projektů shaman a scape projektu
- sos – systéms of systems
- 3 druhy systémů ¨
- DPS – jako funkční requirement
- SoS – business systém – systém v systému – data se pak sypou do DPS
- DPS – kde DP není funkční requirement, ale přesto to dělá (DP ready systém) – business systém s DP funkcionalitou
- jak ale do enterprise systémů DP dostat? model pro implementaci DP do jakéhokoliv systému, v rámci projektu shaman - capability-based reference architecture
- governance, business and technical? (operation) capability – podklad pro rozhodnutí a posouzení stavu
- capability maturity model CMM – procesy posouzení a zlepšení s SW vývoji

Olivier Rouchon – certification and quality at Cines

- ukládají these, digitalizované věci, multimédia dokumenty, data sets vědecké
- datové centrum pro celou Francii
- mají odborníky na formáty, xml, 11 lidí
- 15TB dat
- certifikace – národní zákon – cines je národní centrum pro DP thesí – mají na to oddělení, lidi, peníze – postup a přípravy viz níže
- **příprava na certifikaci – testování drambory, DSA, TRAC, ISO 16363 a ISO 16919**
- **krok - 2009 drambora audit, 2 kontroly risků za rok, jak se postupuje s jejich řešením!**
- **krok – formalizace business procesů, 14 procesů dle ISO 9001**
- **management, operational a support processes (presentováno na ipres2010)**

- **2009 – externí pre- audit, 2 lidi, 19 man days- založeno na všech dostupných standardech, pomocí konltroly dokumentace, rozhovorů**
- **2010 – SIAF audit – 4 měsíce, dělá to NA Francie, pro každý archiv, kt. ukládá veřejná data delají audit každé 3 roky, zpráva měla 800 stran**

- **2010 – data seal of approval - součást EU framework for audit and certification of trusted repositories (MoU mezi třemi aktivitami na certifikaci)**
- **2011 – v rámci projektu aparsen dělali také ISO 16363, spolu s DANS a UKDA procházeli tím auditem**
- napřed internal leden až duben 2011 (60 man days), pak external- 12 odborníků (KB, BL, NASA apod.) v červnu 2011 (3 dny)

MoU – DSA>ISO 16363 jako druhý krok (internal) > ISO 16363 extended
audit je rychlejší tím, čím víc jich děláte, tj. pokud je to pravidelně, není to tak časově náročné

NK Nového Zélandu prošla certifikací TRAC na podzim 2011

Andreas Rauber - dopad preservation actions na repozitáře

- co se děje se samotným repozitářem?
- simulace repozitáře RepoSim
- kvůli analýze, na testování migrace – co se stane, když fily se budou zvětšovat, co když v repu budeme mít více typů formátů apod.
- RepoSim – simulátor, flexibilní, irregular patterns
- zatím interní verze, hibernate, java, mysql
- jde naspecifikovat jaké formáty přijímá, jejich popis, ingest nastavení, hypotetické nástroje (hlavně na migraci), nastavení pravidel na ochranné aktivity (migrace do jakého formátu, jaké verze, jaké soubory, kolikrát, pravidla + filtry)
- možnost spustit virtuální migraci – vzniknou grafy, kt. řeknou jak to bude dlouho trvat apod.
- co, jak, na co a po jakou dobu migrovat, proběhne virtuálně – uvidíme výsledek
- dobré na plánování – pro IT a HW
- **dobré na plánování různých scénářů, porovnání s předpokládaným vývojem, plánování rozvoje HW a investic**
- musí dodělat ještě možnost zadat deletion policies, reporty apod.

José Barateiro Risk assessment in DP of e-science data and processes

- DP as risk management
- ISO 31000 – definice risk managementu
- podobné jako drambora
- k risk managementu je mnoho standardů
- rozvedení metodiky iso 31000 na jednotlivé kroky
- TIMBUS project http://timbusproject.net/ - jedním z partnerů je i SAP (Německo)

mad talks

- open source SW pro LTP – RODA - je zpátky, rozvíjí se v rámci SCAPE projektu, nové funkce, plány na rozvoj a vznik uživatelské komunity
- 4 postery o emulaci! Emulace v rámci KEEP, emulace pro studovny v knihovnách, OPF eco systém registry
- TOTEM – metadatový standard pro popis technického prostředí pro emulaci

ANDS – Ross Wilkinson

- datová centra v Austrálii, min. 3 pro různé oblasti života
- ANDS – existuje skoro 3 roky, peníze od aus. vlády
- obrovské množství dat – nikdy nebudou využita/čtena člověkem – jen automatické procesy vytěžení
- nutnost ukládat a ochraňovat research data, protože už nemusí být možné je znovu vytvořit – tak, aby je šlo znovu použít, aby bylo možné z nich vyvodit nové závěry, aby je měli k dispozici vědci
- nutno dělat ve spolupráci, nelze pouze z titulu jedné instituce
- kdo řeší uložení vědeckých dat v ČR? Akademie věd? CESNET?
- podobná datová centra jsou I ve Velké Británii

Rob Sharpe - Considerations for High Throughput Digital Preservation

Prezentace firmy Tessella. Jejich testování výkonu ingestu do SDB ve Family Search.

- SDB vzniká od roku 2002, kdy prvním zákazníkem byl National Archive, UK
- nový zákazník – UK parlament
- test s FamilySearch
- 20TB ingest za den, skenované materiály – workflow s antivirem, charakterizací (PRONOM, JHOVE) apod.
- 1 package je zhruba 1GB, 20tis. balíčků za den!
- 2 servery dell poweredge R710, cena dohromady max. 20.000 Liber
- ukázalo se, že limitující je rychlost čtení disků, na kt. jsou na počátku ingestu uložena data, potřebovali tedy 130 paralelnich disků (50tis liber)
- uloženo na pásky, taky pomalé, potřebovali tedy 8 paralelních zápisů na pásky (30tis. liber)
- uložení stojí 100 liber za TB
- 7.3peta za rok
- závěr – zápis a čtení je pomalé, nástroje jako jhove a pronom dostatečně rychlé, vysoké náklady I na uložení se ukázaly

Pro ingest dat z projektu Family Search potřebovali zajistit prostupnost 20TB dat denně, při zachování dostatečných procedur pro zpracování dat podle požadavků OAIS a zadavatele. V projektu šlo o to identifikovat úzká hrdla ingestu velkého množství dat.

Procesy jako generování hashů nebo jejich kontrola, identifikace formátů a extrakce technických metadat vyžadují obvykle velký při velkých objemech rychlý storage systém. V projektu family search chtějí do SDB ingestovat (content aquisition, content preparation, ingest:fixity check, content metadata integrity check, charakterizace, tj. identifikace a validace formátů a extrakce tech MD) max 700MB za sekundu.

Řešili jak takové masivní workflow efektivně paralelizovat při minimalizaci nákladů. Podle jejich zjištění paralelizace umožňuje obejít problémy s výkonem nástrojů jako DROID a JHOVE, celkově výkon softwaru nebyl oproti jejich očekávání problém. Větší problémy jsou v HW – aby byl schopen dostatečně rychle zapisovat.

Tj. úzké hrdlo bylo v HW a přesunech dat z místa na místo, spiš než ve výkonu nástrojů pro digital preservation

Přínos pro NK:

Nebát se výkonu SW jako DROID nebo JHOVE.

Ross King –Evolving domains, problems and solutions for LT DP

- info o projektech SCAPE apod.
- programme, http://cordis.europa.eu/fp7/ict/telearn-digicult/report-research-digital-preservation_en.pdf, Stephan Strodl, Vienna University of Technology, AustriaPetar Petrov, Vienna University of Technology, Austria, Andreas Rauber, Vienna University of Technology, AustriaPěkný Timeline for preservation projects, whitepaper about the past of european dp
- Finance vydané na výzkum DP postupně rostou. Projekty a finance nic nevyřeší
-
- ARCOMEM – archivace webarchivů, socially driven web preservation model
- social web analysis
- archive enrichment
- ENSURE – evaluation between cost and value, automatizace ochranného cyklu, testbeds – healthcare, clinical trials, financial services
- SCAPE
- presevation planning and action workflows – jak je udělat škálovatelné
- vytvoření infrastruktury pro škálovatelné akce ochrany
- vývoj policy-based preservation planning nástroje s automatickou preservation watch
- 3 testbeds – wa, larg-scale repositories, research data sets
- všechny projekty vytvoří prototypní SW
- digital lifecycle approach
- preservatin planning hraje roli ve všech těchto projektech, spolu s virtualizací
- slide s trendy v DP za poslední roky
- Research on Digital Preservationwithin projects co-funded bythe European Union in the ICT
- Ensure,
- Scape
- Wf4Ever http://www.wf4ever-project.org/about
- Timbus – sw nestaci, soustredi se na kontext, organizaci LTP není o objektech jen, ale o službách atd
- Totem

Přínos pro NK:

Sledovat projekty v oblasti dlouhodobé ochrany digitálních dat. Poslední projektu EU jako SCAPE povedou k urychlení vývoje konkrétních nástrojů pro dlouhodobou ochranu digitálních dat.

Record keeping in temporary command settings, Erik Borglund

- ochrana dokumentace ke krizovým situacím vzniklých z činnosti policie apod.
- jak zachytit kontext? lze uchovat flipcharty, videa, zápisy ale kontext?
- u analogových dokumentů není problém, problém je s digitálními věcmi a rozhovory
- měl by se o to starat národní archiv, ten ovšem bere jen papírové dokumenty nebo např. fotky z místa jednání- otázka – archivace spisového materiálu je to samé jako archivace průběhu jednání v digitální podobě?

----- webarchiving session--------

BnF – 200TB webarchivovaných dat
1.5 milionů ARCů, musí je charakterizovat, validovat – časově náročné
ukládají v shared repository
SPAR (LTP systém Francouzské NK) má kapacitu 16PB!
používají jhove2 na charakterizaci, vytvářejí modul na arcy
nechtějí dělat charakterizaci a validaci pro obsah arců, jen identifikaci formátů

- PREMIS v METSu, by byl příliš dlouhý, budou tedy zapisovat jen metadata na úrovni informačního balíku (AIP), kt. jsou stejná pro celý balík – resp. 1 vlastnost se vyjádří a pak se k tomu jen přidá informace o tom, kt. fily tomu odpovídají, namísto opakování té infromace pro každý file
- vytvořili speciální metadatový formát
- tj. jsou schopni se LTP systému zeptat: dej mi všechny informační balíčky, které obsahují formát XY apod.- není ale třeba indexovat metadata těch obsahů, to by trvalo dlouho – stejný přístup mají I pro digitalizované knihy
- různé DP policy a úrovně validace pro různé typy wa dat – kompletní sklizně vs. tématické sklizně

NL NZ

- 2 sklizně, 20 TB dohromady
- řeší metadata, kolik metadat je hodně a kolik málo,
- policy knihovny říká, že se musí ukládat co nejvíce metadat, to by byl ovšem z hlediska velikosti metadat problém
- pro selektivní webarchvest mají hotové workflow, WCT, vše se katalogizuje

IA

- 1.6 miliard URL
- nejstarší z roku 1996
- 3TB za den, 1PB za rok je přírustek

Euan Cochrane, Dirk von Suchodoletz - Replicating Installed Application and Information Environments onto Emulated or Virtualized Hardware

- zachycení, uchování celkového prostředí na emulovaný HW
- např. vzít prostředí desktopu předsedy vlády a uložit v archivu
- problémy se zobrazením
- computer forensic
- možnost pro ochranu vědeckých dat a záznamů
- celé je to o tom, jak replikovat HDD a pustit prostředí, kt. na něm je ve virtuálním prostředí
- řešení:
- vykuchali HDD z několika starých PC > identifikovat nároky na HW (analýza HDD > odhad nároků automaticky – je to součást každého PC prostředí) > vybrat emulační/virtualizační SW (tool registry jako např. TOTEM z projektu KEEP) > úprava HDD na disk image vhodný pro emulaci > zkusit nabootovat image disku na emulovaném HW > přidat drivery
- problémy s licencemi, ochranou osobních dat, autenticitou (20% věcí se změní – barvy apod.)
- QEMU sparc processor emulator

Klaus Rescher - Remote Emulation for Migration Services in a Distributed Preservation Framework

použití emulace jako nástroje pro migraci

- mnohdy nejsou dostupné nástroje pro migraci určitých formátů
- Dig. objekt vložíme do emulovaného prostředí (virtuálního stroje) – pak ho vidíme v prostředí emulovaného systému, můžeme ho otevřít v původní nebo vhodné aplikaci, uložit jako jiný formát a uložit opět do virtuálního stroje

Bram Lohman - Emulation as a Business Solution: the Emulation Framework

Keep projekt

emulation framework – 7 emulátorů, 6 platforem (x86, Amiga aj.), 23 file formátů

- řešení pro správu emulačních nástrojů
- setup emulačních procesů
- prostředí, kt. obsahuje emulátory a pokud do něj nahrajeme aplikaci nebo soubor, měl by se spustit jako v původním prostředí
- prostředí obsahuje I nástroj, kt. u souborů ukáže jaký je to formát a jaké prostředí je potřeba pro jeho spuštění – na základě PRONOMu – rovnou lze to prostředí připravit a soubor v něm spustit – načte SW image z databáze aplikací OPF, která se buduje

Geofrey Brown - Developing Virtual CD-ROM Collections: The Voyager Company Publications

- publikace konkrétního vydavatelství na CD ,interaktivní aplikace pro Mac, z 200 vydaných je nyní dostupných pouze asi 50
- emulace do dnešních systémů
- hdd snapshot přímo v emulátoru, tj. je to na jedno kliknutí a velmi rychlé
- sheepshaver emulátor

Evaulation of danish large migration project

- Před rokem 1998 neměli formáty stanovené zákonem
- Mezi rokem 2005 a 8 zavedli standardy
- Hodnoceni se týká stanovených standardů a migrace do nich v národním archivu
- Hodnoceni dělali pro toho, kdo to financoval
- Mezi rokem 2005 a 8 strávili 30 person years na migraci, měli 10=15 lidi na to, investovalo 190 tis USD, celkové náklady 2,6 milionu USD

Není to moc dat reálné, co migrovali, asi 1.777GB

Různé části archivu – tapes data o populaci, data na cd-r, registries a data elektronicky plněna

- Nemohli přečíst všechny soubory, zvlášť na páskach
- 5 různých typu pasek
- Některé museli za drahé peníze zachraňovat

- Celkové náklady na vyrobení preservation standardu 10 men years, 12 tis USD – včetně manuálu a implementačních doporučeni

Pilot – plánovaní a management projektu, a ověřit informační balíčky

Cílem bylo v pilotu získat lepší budget a plán of projekt

Některá problematická data ve starých formátech, jako staré databáze atd. potřebují chytré lidi který dělají repetitivní prací, trvající dlouho…potřebovali dobry knowledge management, aby to bylo efektivní

Způsob migrace – napsali požadavky na nastroj, a popis toho, jak by se mela dělat manuální migrace

Příprava dat (restructure data a registrovat metadata of IP) a příprava dokumentace těch migrovaných IP

Vývoj softwaru – inhouse development.

Potřebovali 50 person years na 1

Závěry,

- migrace standardních dat je levnější☺ migrace z některých pasek standardních je levnější atd.
- Většina 80 % nakladu padla na nestardizovaná data – při výrobě softwaru – na migraci. Vývoj nástroje na migraci heterogenních dat nebo nestandardních dat, je nejdražší.
- Co se naučili – neměli dostatečné analyzovaná stara data!
- Projekt management měli loose, ztratili peníze☺
- Knowledge management – dobry popis starých dat a všech jejich typu, generaci umístěni atd. – u nás neexistuje, a budeme s tím mít potíže – migrace starých dat v \NK bude problematická☺

Angela Dappert – rubust migration workflow – pro offline media

- Co je archival object - hezky slide, cd není archive object, je to pro ne hand held carrier – lepší je bit stable object, ten může mít backup atd. až k archivnímu objektu, který má další metadat – logical preservation.
- Cd není searchable, nedá se snadno replikovat, ma large manual overhead, rendering technology zastarává velmi rychle,
- Projekt endangered archives: optical disks, cdr, external HD, tapes, celkem 67 terrabytes
- OFFLINE hand held nosiče byly v tom projektu endangered archives velmi variabilní, obsahovaly data s drm, pod copyrightem a radou těch problémů.
- Moznosti mezi kterými se rozhodovali u každého zdroje dat –
- Disk image – jeden soubor, který obsahuje všechno, co na něm je
- Nebo extrakce jen některých souboru
- Jak důležitý je ten vlastní nosič? Potřebujeme o něm mít nějaké informace, můžou tam byt stopy po smazaní nějakých dat a chceme je třeba mít? Disk image dělali ze všeho možného – hybridní dvd. Zvuky kde byla i data atd.
- Jaký disk image byl měli použít? Ne jen jeden formát disk image pro všechna data – pro každé speciální disk image formát
- Dělali to robotama, disk copying robots někdy – large scal disk copying robots – nešlo použít, umí dobře vyrábět cd, ale ne ripovat data z cd

- Udělali si svoji aplikaci s diska stacks a nějaké menší roboty používali LIFO nebo FIFO, nakonec použili fifo, lifo mel problémy se zvedačkou CD

Table 1. 4-category digital object status progression

Unsatis-factory object status	Bit-stable object status	Content stable object status	Archival ob-ject status
Hand-held carriers	Content has been transferred onto man-aged hard disk stor-age. Storage is backed up. Checksums have been calculated.	Content has been QA'ed. Metadata has been pro-duced and QA'ed. File formats have been identified. Representa-tion Informa-tion has been deposited.	Automatic check for cor-ruption via checksums. Automatic replication over remote locations. Digital signa-tures. Integration with the cata-logue.
	Step 1	Step 2	Step 3

- V Kb promysleli poměrně složité workflow, jak to popsat atd.
- U každého robota měli PC
- Problémy měli s radou věci, see presentation.
- Nenašli doby sw pro management imagu, jen command liny, ale netechnicky staff by nasekal radu bot
- Je to hodně lidi, než se to dostane na online
- Musí byt dobře vychovaní, flexibilně, ale taky umet dělat tidieus jobs, systematic, patient

POZOR – důležité pro NK, kde se převod dat z disků bude také řešit a už i řešil

Keep propjekt - Antonio Cuiffreda- towards integrated migration environment

- **Disk transfer tool**. Převede disk na image file. Obsahuje tady další metadata – o file systému, a md5 souboru atd.
- Keep vyrábějí **Transfer tool Framework**
- **Magnetic media** – disk transfer tools for flopy disk komerční a opensource
- *Disk2FDI* – komerční – DOS tool, velmi přesný image floppy disku, trvá mu to 1 hodinu, a celý to je pak velmi velký, desetkrát vetší než byl vlastní floppy disk. Testoval asi 2260 disku, testovali emulaci.
- *Catweasel* komerční nastroj , je to PCI card , bezi na linuxech a win xp, ma gui. Velka chybovost, ale rychlejsi image file kvalita byla nizka
- *Nibtools* – free tool, G64 a D54 – covers ony C64 , dos, win, linux, ale to command lin. Potřebuje commodor disk drive a special cables. Testovali par disku asi půlka nefungovala pak v emulátoru.
- **Optical media** – použili 5 transfer tools, u všech stejny cd a dvd a games.

1. Alcohol 120, komerční, umí obcházet drm atd. support Win systems

   Ze 13 fungovalo 12, 2MB za sekundu
2. Deamon tools – commercial, několik typu image files, ISOP, MDS, MDF, support win, tri ze 13 nbefungovaly
3. CloneCD . commercial – používá IMG nebo ISO, obchází safedisk3 protection, support Win, ma gui . 11 bylo ok ze 13
4. Blindwrite – commercial, podporuje dvd, blue ray, WM, Xbox a další speciální disky

   Generuje ISO a nějaký proprietární formáty iso imagu, jeden nefungoval, rychlost stahovaní
5. ImgBurn – nečte do image file subchannel informace (nelze posouvat film atd.) je to opensource generuje dvd, bin, cue, img, win a linux 4 nefungovali, je rychlý

Závěry.

Pro magnetická media – komerční a nekomerční – výkon není rozdílný, disk2FDi je přesný, ale velmi pomalý, Keep použije NibTools.

Optical – myslet na ochranu proti kopírovaní, mají podobny výkon, vždycky budou chyby v těch images, mezi 30 a 10 proc, blindWrite umi herní disky xbox atd. Keep použije ImgBurn protože je to open source.

Pro komplex images je lepší Blindwrite

Přínos pro NK:

Zvážit, zda by v NK nebylo vhodné opravdu udělat projekt na migraci obsahu CD a DVD na online media. Zde prezentují konkrétní zkušenosti s robotickým zpracovávání, a ukazují jaké problémy měli s vymýšlením workflow, volbou typu ISO image atd.

BNF – archivace webu

Mají tri vrstvy:

Harvest definition - collection

Harvest instance –crawling metadata

ARC files

Collection bude naprilad selektivni volby 2000, pak jednotlivé harvest instances, a pak arcy

- Sbírají logy, config, a report
- Tohle skladuji také v arcu – specialni arc metadata pro každý crawling instance

Premis:

Object, agent event.

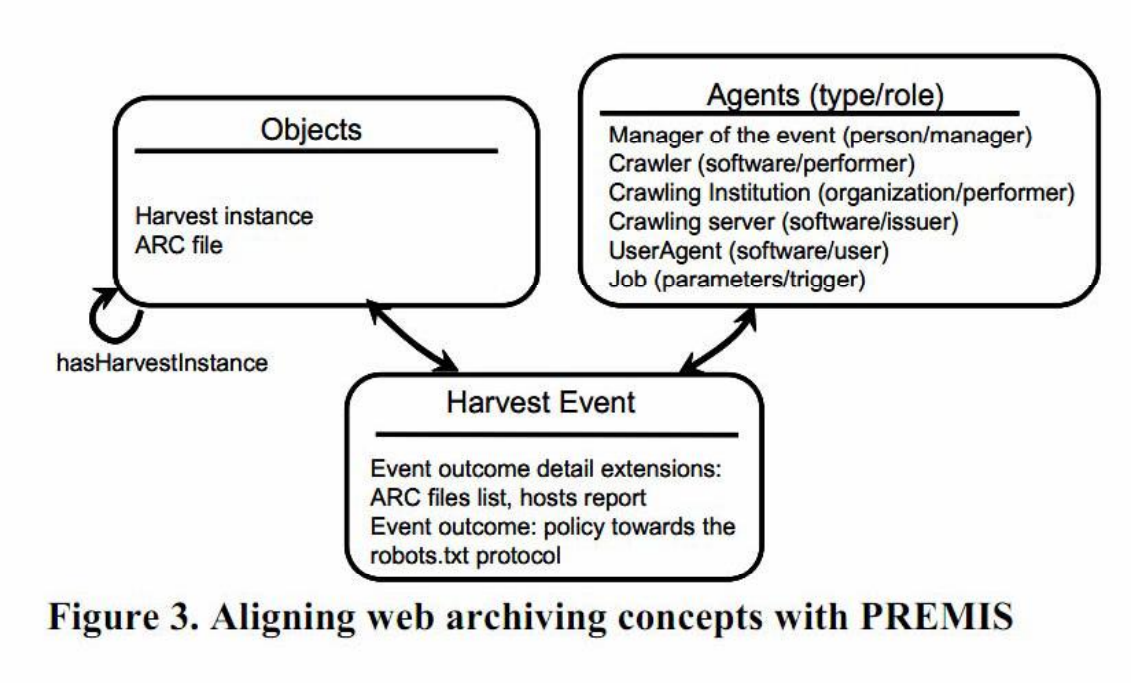

Figure 3. Aligning web archiving concepts with PREMIS

Objects:

1. arc files a metadata arcy
2. harvest instances

Harvest event. – in premis event. – creation of content files

Events – reporty jako extense eventu – host report a harvest report

Agents – afdministator, sw, instituitiolns, organizations, který perfomujou harvrst

ContainerMD

http://bibnum.bnf.fr/containerMD-v1/documentation/containerMD-v1.html

zvláštní metadata pro věci z Web Archivu

http://bibnum.bnf.fr/containerMD-v1/

odlisny SLA pro ruzny typy materialu, pro ruzná data z ruznych sklizni, shared repository ocekavaji ruzne benefity – sklils pro ruzne formaty není třeba v instituci dublovat

pristi rok by merl existovsat taky jhov2 modul pro warc

memento – Meta vyhledavač

Přínos pro NK:

Jejich model archivace webu by se dal využít v NK.

Cost models – dánská NK + TU Wien

- Stephan strodl – TU Viden, mají svůj cost model – ale jen small scale automated preservation action cost se zda
- Dánska národní knihovna – dělali svůj model, který by mel byt univerzální a použitelný kdekoli
- Měřili cost of submission podle standardu paimas
- Při počítáni cost používají oais a paimas, mapuji aktivity na tyhle modely, a pak podle toho odhaduji ceny procesu
- Costmodelfordigitalpreservation.dk

Přínos pro NK:

K projektu 0136, tam se řešily možnosti odhadování nákladů na dlouhodobé uložení.

Meet RODA, a Full-Fledged - Digital Repository for Long-Term Preservation

- Původně projekt Portugalského národního archivu sledujeme až několik let. Teď systém RODA podporuje nezávislá firma a částečně ho také dále vyvíjí. Zatím RODA podporuje pouze archivní formát metadat (EAD) ale další vývoj by měl zahrnout i knihovnické formáty.

- RODA je teď součástí projektu SCAPE, kde bude možné systém dále vyvíjet a škálovat pro použití v masivní produkci.

- http://redmine.keep.pt/projects/roda-public

Přínos pro NK:

Sledovat další vývoj, možná i pro projekty INCAD + KNAV pro vývoj LTP pro menší instituce by tohle mohla být v budoucnu zajímavá alternativa.

Národní knihovna
České republiky
National Library
of the Czech Republic

Zpráva ze zahraniční služební cesty

Jméno a příjmení účastníka cesty	**Mgr. Andrea Fojtů (AF)**
Pracoviště – dle organizační struktury	1.5 Oddělení dlouhodobé ochrany digitálních dat (ODODD)
Pracoviště – zařazení	1.5.2 Oddělení správy obsahu digitálního repozitáře
Důvod cesty	**Účast na konferenci „Aligning National Approches to Digital preservation“**
Místo – město	**Tallinn**
Místo – země	**Estonsko**
Datum (od-do)	**22. - 26.5. 2011**
Podrobný časový harmonogram	Pondělí 23.5. – registrace na konferenci, komentovaná prohlídka Národní knihovny, Keynote Address by Laura Campbell – Kongresová knihovna USA **Panel 1: Technical Alignment** **Panel 2: Organizational Alignment** Úterý 24.5. – Keynote Address by Gunnar Sahlin – Národní knihovna Švédska **Panel 3: Standards Alignment** **Panel 4: Legal Alignment** Breakout Sessions for panels 3 & 4 Středa 25.5. **Panel 5: Education Alignment** **Panel 6: Economic Alignment** Breakout Sessions for panels 5&6 Synthesis/Closing remarks
Spolucestující z NK	**PhDr. Bohdana Stoklasová (BS)** **Ing. Tomáš Svoboda (TS)**
Finanční zajištění	**IOP NDK**
Cíle cesty	Přítomnost na konferenci s mezinárodní účastí, získání kontaktů pro oblast dlouhodobé ochrany digitálního dokumentů a povinného elektronického výtisku, podrobnější vhled to problematiky dlouhodobé ochrany digitálních dokumentů (zejména) v národních knihovnách.
Plnění cílů cesty (konkrétně)	Závěry konference „Aligning National Approaches to Digital Preservation“ vesměs kopírují závěry wokshopu „The Future of the Past – Shaping new visions for EU-research in digital preservation“ (zpráva dostupná na http://cordis.europa.eu/fp7/ict/telearn-digicult/future-of-the-past_en.pdf), např. v případé chybějící ekonomického modelu pro komerční sféru, která by dlouhodobou ochranu vnímala jako neoddělitelnou součást všech svých procesů. Byl navázán kontakt s pracovníky národních knihoven Estonska a Finska (pracovnící pro archivaci webu a dlouhodobou ochranu obecně).

Program a další podrobnější informace	Hlavním cílem konference bylo sjednotit národní postupy v oblasti dlouhodobé ochrany digitálních dokumentů napříč všemi oblastmi, od technických, organizačních, vzdělávacích až po standardizační, ekonomické a finanční.
Přivezené materiály	konferenční program, letáky vystavujících firem (Tessella, Equella, Guardtime) + další materiály, zápisky
Datum předložení zprávy	8.6.2011
Podpis předkladatele zprávy	
Podpis nadřízeného	
Vloženo na Intranet	
Přijato v mezinárodním oddělení	

Příloha k této zprávě: Poznámky z konference v anglickém jazyce

Příloha: Poznámky z konference v anglickém jazyce

Exploring What We Can Do Together – Strategic Alliance for International Collaboration / Laura Campbell

- 185 digital preservation partners in more than 25 countries (education, research, LAM)
- strategic goals: National Content Stewardship Network (national digital collection, technical architecture, public policy outreach
- NDIIPP – Content Domain Map – a mind map of geospatial, audiovisual, image&text and web content
- then & now: cognitive surplus vs. digital libraries/digital preservation
- solution: framework, actively working together, special interest groups, establishing a common index, international digital collection (freely available)

PANEL 1
Technical Alignment (The role of testing) / Prof. Dr. Michael Seadle (Panel Head)

- to collaborate on requiring and implementing rigorous and independent tests

DNB Contribution to the Tallinn Alignment / Sabine Schrimpf

- key theme is infrastructure
- network of hard- and software that permits operation of application of SW
- question of interoperability is crucial (standards, technical specifications)
- SW elements
- Components of the DP infrastructure was compared to the pallets at railroads
    - Source: PARSE.Insight Roadmap 2020
- PersistentID resolvers, certification process
- kopal (ingest KoLIBri)
- nestor (German Network of Expertise)
- DP4Lib – Digital Preservation as a Service; reduce dependency between components
    - redundant storage at different locations
    - KOLiBRI Modules
- LUKII – set up as an economical LOCKSS network in Germany
- SHAMAN
- APARSEN – wants to bring coherence and cohesion to the digital preservation research
    - trends in DP research projects
    - modular DP systems
    - distributed as SOA
    - elimination of technology dependencies

EDINA THE UK LOCKSS Alliance / Adam Russbridge

- EDINA offers underlying technical support & coordination
- threats to digital stewardship:
    - failure (media, HW, SW, network, format obsolesce, natural disaster, economic/organization failure)
    - attack (insider/outsider)
    - operator error
        - source: Requirement to Digital Preservation
- projects PEPRS and PECAN help identify coverage and requirements for DP

Public testing / Michael Seadle

- traditional physical archiving relies heavily on trusted institutions
- distrust, not trust, need to be the basis of digital archiving – testing therefore plays a key role
- goals of testing: demonstrate functionality, reveal weaknesses, provide data for planning improvements
- key issues for testing – integrity, authenticity (can the origin or geniuses be shown?), usability (can migration/emulation be demonstrated?), access, financial integrity
- Dr.Who (drwho1.com)
- bit stream testing is the most important – authenticity and usability may be impaired
    - the type of storage media, the number of copies + frequencies of checking and replacement get us to the relevant results
    - no reliable metrics exists, however (what is an acceptable loss, etc.?)
- without well-documented, peer reviewed, publicly available test results, librarians are buying archiving systems on faith

Presentation without a title /Andy Rauber

- evaluation vs. testing vs. benchmarking
- DP testing and testing – evaluation rather than testing, far from benchmarking (few tests, but not near a definition of benchmarks)
    - existing evaluations are not repeatable
    - focus on the simple things
    - building the frameworks before having clear test scenarios
- necessary to move towards comparative benchmarking
- what is needed: commit that we want a culture of benchmarking and comparative evaluation, understanding of what we want to benchmark, benchmark data + ground truth, measurement scales and measures that remain constant, knowledge bases to collect these

Organizing digital preservation on an international level / Michelle Gallinger (NDIIPP)

- focus on an national DP agenda
- community – driven, action-oriented (National Digital Stewardship Alliance)
    - present a distributed national digital collection for the benefit of citizens

The European Research Arena / David Giaretta

- technologies – GEANT, EGEE/EGI
- EU research projects: TIMBUS, BLOG4EVER, SCAPE, ENSURE, APARSEN, ARCOMEM, WF4EVER
    - SCIDIP – ES (2011 – 2014)
- Alliance for Permanent Access (APA) – formed as a legal entity 3 years ago
    - opportunities for networking
- ISO 16363 – Audit and Certification of Trustworthy Digital repositories
- ISO 16919 – Requirements for Bodies Providing Audit and Certification

Observation from the MetaArchive Cooperative program / Martin Halbert

- distributed DP programs, different from other programs:
    - replication of content, distribution of these replicated copies to distinct geographical locations and network organization to connect these replicated copies
- MetaArchive – established in 2003, funded by NDIIPP
    - seeks to foster broader awareness to digital preservation issues
- IIPC – members are all institutions that focus on WA
    - 39 members - national, university libraries + other organizations (Internet Archive)

    - ISO standard WARC format for web archives + Heretrix and Nutchwax
    - growing membership (Africa & South America)

DAY 2

Keynote address
International and National Collaboration in the digital age / Gunnar Sahlin

- 2012 – a new law for e-deposit
- Samsök (search together) – in 2005 (upgrade – new system)
- Swepub and long-term preservation
- Consortium of the Swedish research libraries for licensing e-journals and databases (ICOLC)
- Open Access and e-publishing (all universities have their repositories for e-pubs)
- NL: aggregator for the Europeana, TEL, Apres, Athena, EU-screen
    - common system for the preservation of digital materials
    - common search portal for materials from the Swedish National Library and Swedish National Archive

Raivo Ruusalepp

- standard RAC, DSA, CIDOC (CRM), PAIMAS, ISA (DG), DDI
- use of information security standards for digital preservation
- information security: administrative and technical (physical = data security vs. IT = communication)
- company implemented security measurement with typical cyber crime scenarios
- survey of security:
    - provision for information security in national legislation and development plane (1/2 of the respondents - ISO 27000 series, only 2 formal audits, the rest are „looking into it"; ½ of the respondents do not use standards or formal measurements to control information security)
    - IT & disaster plan – 65% (data recovery from the off site location tested – 0%)
- alignment:
    - better use of community standards for information security and preservation
    - agreement on security requirements

Standards-based approach to preservation planning / Matthew Woollard

- ISO 27001 – very expensive implementation, 100 000 £
- Basic Data Seal of Approval Guidelines (helps understand your business better)
- Audit And Certification of Trustworthy Digital Repositories
- Memorandum of Understanding to Create a European Framework
- ISO 16363 – external / DSA or ISO 16363 – self-audit

Best Practices & Standards / Bram van der Werf

- self-assessment ← trust → audits
- certification → trust
- ISO 30300 (draft) Record Management

PANEL 4 – Legal Alignment

Legal deposit & Web Archiving / Adrienne Muir

- legal deposit provisions: purpose, scope, deposit mechanisms, roles & responsibilities & liability, access provisions, sanctions
- implementation definitions: scope – offline/online, freely available/pay well, technology neutral/incremental
- legal deposit vs. voluntary (interim/hybrid approach, model agreements and licenses, flexibility)
- other legal issues: intellectual property rights, preservation, access, unlawful material, privacy/data protection
- voluntary approaches have disadvantages but maybe necessary and can be useful

Breakout Session

- standards bring along alignment by themselves; but only if you use the to the full, not half-way
- depends on community (users) – enforceable or voluntary compliance (standards)
- next step for standard alignment:
    - corpora as benchmarks
    - export-import completeness
    - educational standards
    - validation tools
    - accredited training courses to accredit auditors
    - framework standards

DAY 3
PANEL 5 Educational Alignment

- key elements of the DCC Curation Lifecycle Model
- Framework for the Education Alignment (USA – grads programs for digital preservation; new models for grad programs / internship programs (diverse knowledge) / workshops)
    - sharing tools
    - national programs
- related issues to digital preservation:
    - nature of costs and business models
    - strategies for selection & appraisal
    - ground roles and responsibilities
    - effectiveness and demand for services
- focus of the panel: factors influencing the actual sustainability of a digital archive
- 2 considerations – collaboration + user demand
- challenges + gaps – span national boundaries, public + private funding, education, exportation, DP certification, competition, funding gaps, policy, selection criteria, roles & responsibilities standards
- Magazzini Digitali – e-legal deposit in Italy
- PADI – a failed project (discontinued in 2010)

Presentation without a title / Neil Grindley

- how much does it cost to manage information?, what institutional financial strategies are required to facilitate effective preservation?, what general economic frameworks are required to enable information to persist and be accessible?

- JISC – 2010 – Infrastructure for Education and Research Programme
- archival storage and preservation activities are constituting a very small proportion of the overall costs – 15%
- 31% - access, 55% - outreach, acquisition, ingest
    - approx. 333 Euros for a set of 1000 records
- KRDS2, p. 83 – future tool development supporting automation of ingest

Sustainable Preservation in North America: ADPNet & Friends / Aaron Trehub

- solution – distributed digital preservation (in at least 3 copies vs. LOCKSS – 6 copies)
- DPP + LOCKSS = PLN – open SW developed at Stanford
- MetaArchive, COPPUL: Canada, ADPNet – www.adpn.org
  #