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		

Přijato v mezinárodním oddělení

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.

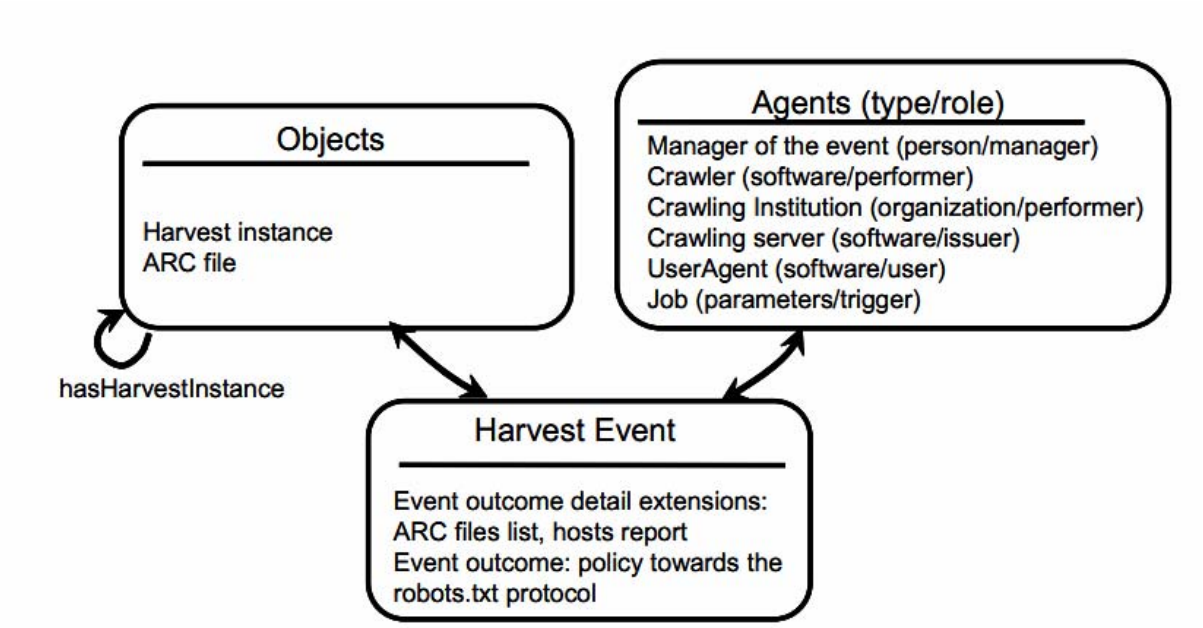

Figure 3. Aligning web archiving concepts with PREMIS

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.