

Zpráva ze zahraniční služební cesty

Jméno účastníka cesty	Jan Hutař
Pracoviště – instituce, adresa	ODO
Pracoviště – zařazení	1.5
Důvod cesty	jednání pracovní skupiny národních knihoven k LTP systémům
Místo – město	Frankfurt n. Mohanem
Místo – země	Německo
Datum (od-do)	22-24.4.2009
Podrobný časový harmonogram	22.4.2009 let Praha – Frankfurt n. Mohanem 23.4.2009 workshop v DDB 24.4.2009 jednání na workshopu; odlet do Prahy
Spolucestující z NK	Ing. Jiří Vrbický
Finanční zajištění	IOP-NDK
Cíle cesty	účast na workshopu, který měl za úkol vytvořit obecnou koncepci LTP systému k dalšímu použití a jako podklad pro výběrové řízení pro LTP systém v různých zemích
Plnění cílů cesty	splněno
Program a další podrobnější informace	viz níže
Přivezené materiály	
Tištěné přílohy a elektronické dokumenty	
Datum předložení zprávy	4.5.2009
Podpis předkladatele zprávy	

Den první 23.4.2009

Ingest

PSP – publisher submission package – používá se namísto SIPU – náš případ nechťejí diktovat vydavatelům co a jak jim mají dávat – psp > sip > format normalisation
ltp systém pro všechny digi data nebo jen pro ty, které se mají ochránit (tj. například vyloučit uživatelské kopie) nebo přístup před DK který sahá do repozitáře?

data management – obsahuje všechna metadata, měla by probíhat synchronizace mezi katalogem a data managementem, na obě strany ale jen popisná metadata (všechna ta metadata jsou pak v AIP)

DDB migrace dat na ingestu –napřed uloží originální data a pak je v rámci ingestu migrují do preferovaných nebo aktuálních formátů (to samé Švýcaři)

kde končí ingest? administrace je součástí ingestu v podstatě končí doručení AIP

data management

má metadata o DO

není tam žádný přímý link do archival storage, vždy to jde přes něco – jiný modul aip obsahuje DO a metadata, jsou ta metadata i v data managementu? ano

vyřizování dotazů na DO- na jeden, na množiny omezené limity apod.
udržovat metadata v aip? ano, když ztratíte vše a zbyde aip, máte v podstatě všechno (redundance)
co je minimální množina metadat, kt. uchovávat s aip?

Archival storage

dias ukládá aip jako .tar nebo .zip
NZ ve file systému – lepší pro update metadat apod.

LTP založený na OAIS v podstatě není pro přístup uživatelům, jen pro přístup v rámci long term preservation > na tom závisí množství metadat, která chceme ukládat – tj. na NZ nemají moc metadat protože to mají jen jako LTP systém, my chceme dávat přístup? tj. hodně metadat, tohle by se mělo rozmyslet na začátku – budeme potřebovat pro LTP systém takové množství metadat jaké momentálně máme?

BL plánuje mít 3 sites synchronizované, protože si nemohou dovolit, aby jedna nehoda na jednom místě ohrozila chod DK

otázka rychlých disků – měly by být všude, protože pokud se překročí určitý počet DO, byl by problém na velkém množství dat dělat jakékoliv operace (Norsko to tak má, BL taky protože to na pomalých prostě nešlo)
manage storage hierarchy- není třeba –HW, to samé replace media
ne všechny části OAIS jsou automatické procesy, něco musí dělat i lidi

Access

- vše co jde z archivu; převod AIP do DIP
- věci okolo autorizace
request od uživatele, access modul to postoupí data management modelu, access kopie do staging area a pak uživateli?
kde končí access – dodáním dipu nebo i jeho prezentace?
Norové mají billing, ale je to v rámci katalogu, mimo LTP – například práva k použití fotografie do publikace
v Německu mají přístup do archivu jen z pár počítačů, normální uživatel nemá přístup – tj. drží DK mimo systém asi
přístup admina, uživatele, systému samotného k dalším akcím (DIP do staging>úprava) a pak B2B z archivu do archivu

napsat a specifikovat jak si představujeme access v oaiss, jen doručení dipu nebo kompletní end user přístup? katalog, dk, primo

Administration

pomohlo by jim, kdyby viděli Rosettu, mluví o něčem o čem nemají potuchu, že to už někde funguje

nastudovat si všechny moduly oais, hlavně administrativ modul

hodně fcí administrace bude muset být manuálních, viz NZ
administrace kontroluje workflow, jednotlivé procesy,

kontroly v ingestu a jak bychom to chtěli v budoucnu, jak vidíme ty chyby, které se stanou a když je vidíme, jak se na ně dá reagovat? automaticky nebo manuálně? tj. vlastně jaké jsou naše nároky na administration module – něco z KB pošlou, my doplnit ...

Den druhý 24.4.2009 – prezentace jednotlivých knihoven

DNB

jaké mají moduly

JCR repository mimo DIAS – pro DO, na kterých se pracuje – work space
co jde do AIP musí být zkatologizováno> kolibri vezme objekt z JCR a pošle ho do DIASu (v podstatě ingest phase 1)

vše co se dává do archivu má URN

DNB chtějí mít dias maximálně další 2 roky

KB musí měnit železo a nechce už IBM- soud řekl, že není možné to brát od jedné firmy, takže musí řešit nový systém, aby byl nezávislý na HW – DIAS neběží na ničem jiném, musí to vyřešit hodně rychle, DNB zatím nic nepálí – umí bit preservation, ale mají problémy s komplexními objekty,

nemají validaci formátů DIAS (jhove apod.)

tesela je spíše pro archivy, BL s tím měla hodně problémů, ingest je pro archivy, musí se upravovat

dias nemá access – ani to nikdy nechtěl mít

DNB se na diasu naučila spoustu věcí, nikdy to v podstatě nefungovalo úplně jak by chtěli, ale pomohlo jim to se zorientovat a hlavně přeorganizovat organizaci práce v knihovně

NORSKO

dsm digital safe store, mají asi 5 PB dat

používají všechny části OAIS na různých úrovních

homemade systém, mají napojení na jhove, droid, apod.

dogitalizují do tiffu, všechny procesy probíhají na tom tiffu až nakonec jej převedou na jpeg2000, kt. mají jako preservation format a tiff mažou (image magic with jasper na JPEG2000 lossles)

SUN FS storage

2 geografické lokality

všechny sw php, java, c++ co si vytvořili sami

docworks používají na ocrko, mets alto (pak zabudované v každém xmlku ke každé stránce)

vše pak ukládají v metsu uvnitř archivu- pro celý titul novin

ale mají i mets pro každou stránku
jmenná konvence souborů, lze z toho vidět název, číslo apod. – pak se jim automaticky dělá
struktura (podobné jako to dělá Indigo) – tohle bychom měli promyslet, abychom se zbavili
číselských zakázek, které se špatně dohledávají a neříkají nic o struktuře dokumentu
systém je 7 let starý a nedá se už dál škálovat – potřebují nový

SPANĚLSKO

mají digiTool, začínají s masovou digitalizací, budou mít více než 5 mil DO, což digiTool
nezvládne

mají nabídku přejít na rosettu, ta nemá omezení ohledně počtu DO, ale museli by platit za
migraci... zatím se rozhodují, ale myslím z toho co říkali, že je to pro ně dobrá šance

KB

vše IBM, nemají od toho kód, takže to nemůžou upravovat, dias je v podstatě poskládaný
z normálních produktů IBM, poslepovaných dohromady, což moc nefunguje

dias skutečně neběží na jiném HW, je potřeba aix middleware

mají dias 1.3, Němci verzi 2, letos bude verze 3 – výrazné zlepšení

nemají preservation planning – ani Němci

všechny změny jsou šíleně drahé i pro KB – závislost na IBM není možné ukočírovat

je tam preservation manager, ale nedá se použít- testování v KB

získat prezentaci od Larse a Jos Pijpers

KB ani DDB tu verzi asi nebudou ani chtít – tím LTP chtějí dias donutit aby něco dělali, když vše
splní tak to bude IBM kdo postaví nový systém, ale doufají že ne

v 3 monitoring a reporting...

ale nikdo neví jestli to bude, čeká se na zákazníka

KB si nechá dias ale vedle toho by chtěli nový jiný systém

PORTUGALSKO

zajímají se o LTP zatím jen rok, mají nějaký systém for masters – zatím vývoj
jsou na tom jako my, všechno probíhá při digitalizaci, pak nemají nástroje na kontrolu,
charakterizaci apod.

v podstatě mají DK s metadaty jako my, kt. systém není schopen použít jako u nás

BL

mají aleph, i primo

na deposit mají tessellu

hodně pdfs

velký sound archive

digitalizace s CCS

pdfka jako user copy

5mil stran novin (nejsou zatím v úložišti), stěhují sbírku novin do Yorkshiru, hodně

digitalizovat + digital on demand

povinný výtisk novin – 2000 titulů, chodí jim to v pdfku, spousta problémů (claimování
apod.)

access cache – jednou to už někdo chtěl z repozitáře > nechá se ta kopie a třeba se použije znovu (cache)

rules pro omezení práce a zpřístupnění DO – mají to na ingestu, myslím že NZ to má v access modulu – tím pádem ale nemusí držet všechny smlouvy v metadatach, protože to takto vyspecifikují

3 lokace, nezrcadlené

Technical restrictions

- 25 mil euro ze strukturálních fondů je náš rozpočet

systém bychom rádi měli příští rok 2010/2011

CMS, content management systém – aby se pod něj dalo dát jiný dokument systém (ECM – enterprise content management) /libovolný > CMIS content management interface standard

storage, networking, computing, security, OS level maintenance – KB chce tohle všechno outsourcovat: musí to být na holandské půdě aby to podléhalo holandským zákonům

musí to být multiplatformní a podporovat virtualizaci

rychlá škálovatelnost je potřeba – nahoru i dolu

certifikace nutná co nejdříve pro KB

non functional requirements – důležité z různých hledisek, co vše chceme akceptovat, minimální nároky na procesy, které se týkají např. počtu lidí kteří to budou muset opatrovat, kolik to bude stát apod.

je nutné vlastnit kod systému, nemusí to být open source, ale nesmí to být black box

Portugalci chtějí systém koupit, nevyvíjet, core musí být koupen, napojení na jejich systémy jsou schopni si udělat – to stejně musí udělat všichni sami

Němci mají hodně open source sw, ale taky z toho nejsou nadšení – mají ale dost silný tým developerů

udělat a zjistit-----

otázky na NZ

NZ – migrace je součást ingestu ale originály neuchovávají?

co mají za metadata v aip? všechna metadata nebo jen část a zbytek je v data managementu? nebo se to dubluje? přemýšleli o tom?

mají nějaké rozkreslení rosetty na oais? nebo seznam modulů?

kdo rozhoduje, že nějaký formát je zastaralý? mají nějakou strategii k tomu?