

Zpráva ze zahraniční služební cesty

Jméno účastníka cesty	Jan Hutař
Pracoviště – instituce, adresa	ODO
Pracoviště – zařazení	1.5
Důvod cesty	návštěva firmy Tessella a Welcome Trust Library
Místo – město	Abingdon a Londýn
Místo – země	Velká Británie
Datum (od-do)	26-28.7.2009
Podrobný časový harmonogram	26.7 odlet Praha – Londýn, vlak Londýn Oxford 27.7. návštěva firmy Tessella – Abingdon 28.7. návštěva Welcome Library Londýn- odlet do Prahy
Spolucestující z NK	
Finanční zajištění	IOP-NDK
Cíle cesty	seznámit se s řešením Safety Deposit Box od firmy Tessella
Plnění cílů cesty	splněno
Program a další podrobnější informace	viz níže
Přivezené materiály	
Tištěné přílohy a elektronické dokumenty	
Datum předložení zprávy	10.8.2009
Podpis předkladatele zprávy	

Tessella a SDB + jednání ve Wellcome Trust Library London

Jan Hutař/Petr Žabička

Tessella - pondělí 27.7.2009

– Ann Keen, Rob Sharp, James Carr (developer)

James Carr nám předvedl plnohodnotné DEMO stávající verze SDB systému, které se z mého pohledu velmi vyrovná tomu, co jsme viděli v Rosettě na NZ.

- současná verze SDB 3.1 – od podzimu bude verze 4 – některé součásti jsme již viděli na testovacím serveru...
- hlavní rozdíl – stávající verze není jeden kus SW, je to několik komponentů (framework), SDB 4 bude produkt (core SW)
- verze 4 – zásadní změna workflow na lineární + možnost si nadefinovat vlastní workflow (v současné verzi musí udělat Tessella), založeno na DROOLS
- součástí 4 je i emulace
- verze 4 bude mít možnost dát data do archivu a až zpětně je charakterizovat

- verze 4 má podstatně vylepšené vyhledávání a zobrazování v archivu – ne pro uživatele, ale pro pracovníky (SDB explorer) – lze tam vidět veškerá data i metadata, jejich vlastnosti apod + možnost downloadu dat, metadat, odděleně nebo i dohromady apod.

-

INGEST

- snaha napojit na digitalizační linku a katalog (wellcome trust library)
- proveden benchmark na normálním PC – SDB je schopno loadovat 10.000 souborů během pár hodin (použito normální PC ne server)
- mají tam účty depositorů
- integrace s katalogem ano, závisí i na druhé straně
- normalizace a vytvoření uživatelské kopie možné během ingestu (má to Wellcome trust library)
- je tam fronta, takže se upload neděje najednou, ale po kouskách

manuální

- pomocí XIP creatoru, kt. vytváří SIP
 - XIP creator lze upravovat na přání
 - lze tam dát vlastní XML – tj. na ingestu obrazová data + xml s metadaty > bude převedeno do XIP formátu, kt. používá SDB
 - xml lze v creatoru upravovat
 - je to vlastně na preingest – pak následuje vlastní ingest v rámci SDB
 - funguje např. s ftp, z webu apod.

automatický ingest umí SDB samozřejmě také – je potřeba nastavit workflow (raději pragmatický přístup, ať to není příliš komplikované)

Ingest ARC/WARC – nebo .zip, .tar apod. – uložení tak jak to je, ale SDB je schopné vidět a charakterizovat soubory a formáty „uvnitř“ – to platí pro všechny podobné případy (WORD dokument s jpeg obrázkem uvnitř apod.)

migrace a ingest lze libovolně škálovat na několik serverů

PRESERVATION PLANNING

- v SDB lze testovat různé postupy před vlastní akcí
- vše visí v podstatě na PRONOMU a na pathways, kt. navrhuje
- lze si navrhnout vlastní pathways- vlastně co a jak se bude dělat a pomocí jakých tools (i externích)
- SDB nespolupracuje výlučně jen s PRONOM – vnímají pozitivně i GDFR, wellcome library používá i NZ metadata extractor
- PRONOM local – lze doplnit interní věci a risky! nebo také jen vybrat požadovaný formát k migraci (transformaci) na základě filtrů
- 2 typy defaultních transformací (preservation a presentation type)
- PLANETS PLATO není součástí SDB, ale něco hodně podobného – není
- v samotném sdb lze i plánovat i testovat preservation action
- využívá JHOVE, ale lze napojit i jiné nástroje
- po migracích se věci automaticky kontrolují (validace formátů apod.)

ACCESS

- user access v podstatě nemají, počítají s tím, že zákazník má své aplikace a do těch mu to s radostí připojí
- DIP lze vyexportovat se spoustou nastavení (data, metadata, různě osekáné verze) buď přímo nebo lze vyexportovat někam- systém pak pošle určité osobě mail s linkem pro stažení dokumentu (ideální pro HQ dokumenty z Manuskriptoria)
- přístup do SDB aplikace omezený na role pracovníků apod.

Demo Swiss Archive

- mají vlastní formát dat a metadat (ARELDA)
- SDB je schopná to zpracovat – tj. my máme urč. formát dat a metadat také...
- ukládají i to původní co přijde + tu obohacenou a upravenou kopii
- zcela jiný workflow – nastavení proběhlo bez problémů
- urč. fce a zodpovědnosti lze přiřadit jednotlivých osobám (stejně jako Rosetta)
- když se něco pokazí během ingestu, lze řešit, opakovat apod. – jednotlivé osoby
- během ingestu je k AIP přidávána signatura papírového dokumentu, to pak oba dokumenty pojí

BL

- také uchovávají původní dokumenty
- ukládají i CD/DVD image

obecné

- testování škálovatelnosti ve spolupráci se SUN
- SDB je schopné kontrolovat životnost pásek, MD5 průběžně, storage adapters, integrity check mezi daty a metadaty (tj. úložiště vs. databáze apod.)

Wellcome Trust Library (Dave Thompson) dále WTL

BL a Wellcome Trust Library – berou všechny formáty

WTL má primárně zájem na uchování digital born věcí (články, zprávy, výzkumná data, maily apod.)

- digitalizace je pro ně sekundární
- SDB mají nainstalovanou od března, velmi dlouho jim trvá nastavení workflow
- pracují na tom 4 archiváři, ORACLE outsourcovali Tesselle – nikoho nového na SDB nenabírali
- mají tam už image CDček
- SDB mají jen jako back-end pro objekty a metadata (user copies mimo systém a přístup přes stávající aplikace a katalog)

-

SDB je dle Dave Thompsona velmi jednoduché na použití, méně nastavení než v Rosettě, což jim vyhovuje

- jsou s SDB max. spokojení (vypadal, že to myslí vážně)

dočasné úložiště – mají tam věci v karanténě

XIP creator –

- budou tam přidávat spojení s katalogem (přes record ID)¹
- jednoduše stažitelná metadata
- mělo by to umět i strukturu dokumentů (indigo umí)

momentálně pro workflow zkoumají Digital forensic toolkit na pre-ingest digital born dokumentů

<http://www.forensics.ie/software/product/forensic-toolkit-ftk/>

dokumenty

veškeré plány a policies na DP před spuštěním nákupu probrali s mezinárodní peer review skupinou

- pak bylo daleko lehčí sehnat na SDB peníze od vedení (opora odborníků)
- hodně to pomohlo řešení a plánům

¹ record ID je checksum!

- <http://library.wellcome.ac.uk/node289.html>
- <http://library.wellcome.ac.uk/assets/wtx043369.pdf>

digitalizační projekt

budou používat JPEG2000 – kvůli úspoře místa (získat analýzu)

- digitalizace 1mil knih (dle jejich výpočtů v tiffu 3,5 PB, v jpeg2000 0,5PB)
- bibliografická metadata jsou v katalogu a nebudou se ukládat nikde jinde, ani v archivu

o rossettě

- jednali s nimi, cena bylo 1.000.000 liber
- SDB mají za čtvrtinu ceny! tj. 250.000 liber
- na Rossettu je potřeba min. 1 osoba na udržování UNIX serveru na full time
- na SDB nic takového, jejich IT to zvládá v původním obsazení (pozor- zatím nezačali s tou digitalizací)