

Zpráva ze zahraniční služební cesty

Jméno účastníka cesty	Jan Hutař
Pracoviště – instituce, adresa	ODO
Pracoviště – zařazení	1.5
Důvod cesty	návštěva IBM (Mainz), NK Německa (Frankfurt n. M.) a Univerzitní knihovny v Goettingen – seznámení a analýza s LTP systémem DIAS a s procesy digitalizace
Místo – město	Frankfurt n. Mohanem, Goettingen
Místo – země	Německo
Datum (od-do)	15-17.6.2009
Podrobný časový harmonogram	15. 6. odpoledne příjezd do Frankfurtu, jednání 16. 6. celý den, jednání v prostorách Deutsche Nationalbibliothek, večer přejezd do Göttingenu 17. 6. návštěva digitalizačního pracoviště Univerzitní knihovny v Göttingenu, odjezd do Prahy
Spolucestující z NK	Jiří Polišenský, Tomáš Foltýn, Jiří Vrbický
Finanční zajištění	IOP-NDK
Cíle cesty	seznámení a analýza s LTP systémem DIAS a s procesy digitalizace
Plnění cílů cesty	splněno
Program a další podrobnější informace	viz níže
Přivezené materiály	
Tištěné přílohy a elektronické dokumenty	
Datum předložení zprávy	25.7.2009
Podpis předkladatele zprávy	

1. den – školící středisko IBM, Mainz

přítomen Wilhem Stoll – odpovědný za projekt KOPAL ze strany IBM

nový zákazník si pořídil DIAS, společnost Dassault, Francie – budou mít DIAS verze 2.x – budou v tom ukládat 3D modely součástí letadel, kt. je nutno ze zákona uložit min. na 80 let

- vytvářejí web interface (GUI) pro ingest
- preservation manager bude součástí DIASu
- řešení ve Francii bude komplet přepracovaný DIAS, nepoužijí KOPAL ani nic z toho co je v KB – jen jádro DIASU zůstane stejné (tj. ta černá skříňka)
- chystají behaviour map – pro vyjádření rights
- snad vznikne pre-ingest s možností zapojení více serverů (R i Tessella již dávno mají)
- DIAS bude umět PREMIS 2.0 a METS 1.8 –reálně
- chystají i feedback pro producenta o tom, jak přišlo jeho SIP a zda prošlo až do archivu (opět R a T už mají)
- problémem je, že Dassault má zcela jiné nároky než knihovny (stávající zákazníci)- nechtějí žádné napojení na DROID/PRONOM apod. -> to bude až v DIAS verze 3 > tj. vypadá to, že IBM to dělí na 2 části – první normální zákazník, kt. stávající DIAS v podstatě vyhovuje a druhá větev jsou knihovny, pro kt. mají v plánu verzi 3, ale až se někdo další chytne. Můj názor je, že

klasický zákazník je pro ně lepší market než knihovny... takže s verzí 3 bych to růžově neviděl...

- celý systém v Dassault bude certifikován od Evropské vesmírné agentury ESA
- start systému v Dassault je plánován na Q1 2010
- IBM dává specifikace systému a programování se odehrává v Indii – tak mi to řekl Raymond van Diesen
- jedinou starostí Dassault je prokázat, že se s DO nehýbalo (něměnil se, není poškozený apod.) – problém zpřístupnění neřeší
- v Dassault bude mít AIP stále jako .tar- takže změna metadat v AIP je komplikovaná, v. 3 by už měla mít AIP ve file systému (jako např. R a Tessella ;-)
- pro Dassault vyvíjejí web server administer interface – lze oddělit pro 2 instituce, takže dias asi bude mít „lidské“ rozhraní

post box – součást diasu, něco jako preingest bez funkcí, umí jen MD5 a kontrolu názvu souboru, v ingestu není možnost oprav (ani dat ani metadat), musí se to vrátit producentům (automatická oprava ne, ručně velmi omezeně)

moje otázky:

- API – ingest ani access nemá open API, tj. v podstatě na něj nelze zapojit služba třetí strany bez toho, aby to dělala IBM (trvat na tom ve specifikaci)
- přístup – jaký je formát zpřístupnění? umí jen surový DIP, kt. je v .zip nebo .tar
- kolik dig. objektů dias zvládně za den? více než 80 tis.
- bude se implementovat něco z projektu Planets (test bed, preservation planning)? NE
- kolibri není součástí diasu
- lze vynechat v rámci ingestu nějaké kroky? šlo by to ;-) na process serveru, tj. ne přes rozhraní
- m

následovaly prezentace o storage DS5000, páskových knihovnách apod. – prezentace budou k dispozici

Německá NK Frankfurt nad Mohanem

Kopal – 2004-2007

od r. 2006 zákon sbírat e-publikace, tj. musí najít řešení pro DP

- mají novou verzi systému, testy dobíhají a bude nasazena během měsíce, je to jen vylepšená verze 2
- KOPAL byl pilot, chtěli zjistit, jaké jsou možnosti, až nyní zapojují DIAS do knihovních systémů naostro – zatím to nefunguje ;-)
- KOPAL byl vývoj praktického konceptu – nešlo úplně o to zprovoznit na ostro běžící systém
- webarchiving zatím nemají vůbec, takže neřeší uložení, začínají se selektivními sklizněmi
- do diasu dávají jen – these + e-journals+ digitalizovaný materiál (minimum)
- hosting celého systému dias je outsourcovaný, kolibri běží přímo v knihovně
- nikdy nepočítali s tím, že by dias byl jejich LTP systémem
- přístup k diasu přes kolibri, dias nemá žádné user friendly rozhraní
- kolibri má více modulů – viz web kolibri – v podstatě si udělali svůj ingest a access a migration manager (v diasu je taky migration manager, ten ale s kolibri nespolupracuje)
- LMER metadata – vznikla před PREMISem, přejdou na PREMIS
- UOF – viz níže – vzniklo v rámci kolibri, aby mohli použít LMER a METS
- UOF nemá deskriptivní metadata, protože ta nešla změnit v rámci AIP (např. při změně v katalogu) – DO má spojení do katalogu přes URN a systémové číslo)

- preservation planning v diasu není, jejich názor je, že by to mělo být mimo systém (můj názor je, že ano – mimo systém je nějaké plánování, rozhodnutí apod, ovšem systém musí tohle podporovat – mít plánovací nástroj, kt. bude propojený s PRONOMEM nebo něčím podobným a nabídne několik možností , hodnocení risků, lokální formát registr apod.)
- pro user copies mají externí repozitář- tj. data z diasu nezpřístupňují

celkově vzato, o spoustě věcí zatím v DNB jen přemýšlejí, ale R a Tessella už to umějí (což se zdá nevědí ...)

DP data streamu není problém (řešeno skrz HW), problémem je to nad tím (formáty, SW apod.)

dias core – má definovaná rozraní přímo pro DNB

verze 2 diasu - vlastnosti

- multi user podpora
- UOF – vlastní datový model DNB – universal object format (v podstatě tak obešli tragickou podporu metadat v diasu, udělali si svůj data model a drží ho mimo dias – ta metadata)
- query interface pro data management
- administrative functions – nově možnost smazat DO z archivu (úplné vymazání nebo se zachováním metadat)

vývoj po Kopalu (od 2007) v podstatě minimální, Frankfurt dosud nemá běžící systém, jen v Goettingen to jakž takž běží

ostatní knihovny chtějí v DNB uložit svá data, vzniká nový projekt, kt .by to umožnil v rámci diasu (tj. bude se muset dodělat nový modul)

tohle si protiřečí s tím, co jsme slyšeli v dubnu ve Frankfurtu – že nebudou do diasu už dávat žádné peníze a taky to logicky odporuje aktivitě DNB v LTP skupině

SUB Goettingen 17.6

gdz.goobi.org

mají 8 projektů digitalizačních ze kterých vzniká 10.000 DO/den

- copyright- vše do roku 1900 je volně
- mají 30 zaměstnanců
 - SW development
 - digitalizace
 - quality assurance odd.
 - image improvement odd.
 - library management odd.

všechna oddělení jsou na jedné chodbě – pro lepší komunikaci, přes telefon spousta věcí nejde řešit OCR dělají externě – nějaká firma, kt. to seká na 10 PCčkách, 18stran/minuta

- nedělají novogotické písmo pro OCR – příliš drahé a nedokonalé > počkají pár let až to bude levnější a přesnější a pak to udělají naráz

-
workflow SW Goobi

prezentace tady v profilu R. Stockmana na SlideShare <http://www.slideshare.net/rstockm>

Goobi je propojené se všemi myslitelnými systémy v knihovně + obsahuje modul na access, přístup k diasu atd.

- prezentační vrstva digitální knihovny SUB je vytvořena přímo v Goobi, kt. používá CMS typo3 <http://typo3.org/>
- příklad hlavního webu GDZ <http://gdz.sub.uni-goettingen.de/> a knihovny v typo3 tady <http://gdz.sub.uni-goettingen.de/dms/load/img/?IDDOC=143>
- to celé je udělané pomocí Goobi a typo3, kt. obsahuje
 - mají print on demand
 - používají PURL pro DO
 - PNG pro user copy- dělá se on the fly z tiffů
 - imagemagic SW
 - při vytváření metadat se bibliogr. metadata tahají přes sysno z katalogu (proč je u nás sysno nespolehlivé!?) > vytvoří se kopie bibl. záznamu knihy, kt. se digitalizuje v katalogu a pak má digitální kopie knihy svůj vlastní záznam v katalogu! to celé automaticky

SIP obsahuje- resp. SIP vzniká z:

- METS – včetně OCR, které není jako text vedle jak je tomu u nás, ale je přímo v METSu
- image (master copy + upravená preservation copy)
- DC metadata pro vyhledávání

user copy vzniká on the fly

V Goettingen ukládají reálně do DIASu data, ale mají tam jen 70GB nějakého archivu, zbytek (několik TB) je zatím vedle na úložišti, zatím neví jestli se to tam nasype – měl jsem pocit, že spíše ne, že čekají co bude

- používají kolibri pro vytvoření SIPu – stojí mezi goobi a diasem, jen pro vznik SIPu z toho co dodá goobi, aby to odpovídalo tomu co chce dias
- kolibri používá JHOVE
- opět – při změně AIP se to musí re-ingestovat včetně dat

POZOR – dias má licenční omezení podle množství dat!!!!!!!!!!!!