

KRAMERIUS – VÝVOJ APLIKACE PRO ZPŘÍSTUPNĚNÍ

Ivan Ljubka, Národní knihovna ČR Praha

V průběhu roku 2002 bylo zcela jasné – po zkušenostech z dosavadního vývoje – že musí dojít k oddělení prezentační vrstvy od archivační a výrobní, která do té doby byla zabezpečována jedním nástrojem (AiP safe). Obrazně řečeno: posledním impulzem pro toto oddělení byl „cracking“ našeho systému, jehož výsledkem byla nucená odstávka a zastavení poskytovaných služeb z digitalizovaných periodik.

Pro úplnost výčtu podmínek pro vznik Krameria jako aplikace byly i události zcela nelaskavé, dá se říci, že byly často až tragické. Těmi nešťastnými událostmi byly na sklonku léta roku 2002 rozsáhlé povodně. Na jejich likvidaci byly uvolněny značné prostředky, které umožnily na jedné straně bezprostřední pomoc postiženým subjektům, a na straně druhé napomohly dynamickému vývoji v oblasti digitalizace. Do tohoto výčtu patří

- přechod od starého formátu popisných metadat DOBM/SGML k novému formátu, založenému na standardu XML
- modernizace datového úložiště NK ČR, bez kterého by digitalizace skončila nejpozději v r. 2004
- vývoj aplikace, která umožní digitalizovanou náhradu primárních dokumentů zpřístupnit na webu, později pojmenovanou jako Kramerius. Jedním z hlavních požadavků bylo, aby aplikace byla vyvíjena jako tzv. Open Source (licence GNU GPL)¹.

Ve své přednášce se budu zabývat – jak je patrné již ze samotného názvu mé prezentace – aplikací Kramerius.

V počátcích vývoje a implementace Krameria figurovala Národní knihovna ČR. Zvláště bych chtěl uvést jména dvou kolegů, a to pana Adolfa Knolla jako autora nové specifikace [document type definition – DTD]², a Jiřího Polišínského jako hlavního oponenta prováděných analýz. Záhy se však do vývoje systému zapojily i další instituce a osoby. Jedním z prvních byl M. Lhoták z Knihovny AV ČR, který přispěl k vývoji projektu nejdříve osobní oponenturou a konzultacemi; později se KN AV ČR spolupodílela

¹ viz: http://cs.wikipedia.org/wiki/GNU_GPL

² Viz: <http://www.digit.nkp.cz/techstandards.html>

na financování projektu a na jeho rozvoji podáním několika výzkumných záměrů, které dál posunuly možnosti jejího využití. Postupně začaly spolupracovat další významné instituce, jako je MZK Brno (zjm. P. Žabička), SVK Olomouc (zjm. M. Vojnar) a řada dalších institucí a osob.

Základní technologie, které aplikace Kramerius využívá, jsou: JAVA, Linux, Apache, Tomcat, Postgres SQL, Lucene. Tedy technologie, které jsou rovněž zdarma a nepředstavují pro případného zájemce o instalaci žádné dodatečné náklady.

Aplikace má uživatelské rozhraní, o jehož přívětivosti z dnešního pohledu již mohou zaznívat kritické připomínky, ale svou jednoduchostí a klasičtým knihovnickým přístupem je srozumitelná široké obci uživatelů. Dostupnost je bez přihlašovacích hesel a restrikcí. Problémem rovněž zůstává proprietární obrazový formát djvu, který vyžaduje, aby si uživatel při prvním přístupu na svůj počítač nainstaloval Pluginu. Administrátorské rozhraní je srozumitelné a s dostatečnou škálou práv pro různé činnosti.

Současný stav:

- umožňuje ukládat dokumenty za předpokladu dodržení specifikací dle DTD pro periodika a monografie
- provádí importy XML, TXT a obrazových souborů u těchto formátů:
 - DjVu
 - JPG
 - PNG
 - PDF
- se vyznačuje integrací nástroje pro fulltextové vyhledávání (Lucene), který usnadňuje práci s dokumenty
 - současně je ponechána možnost využít propojení na další alternativní vyhledávací systémy – například v NK ČR je to aplikace Convera (dříve známá pod názvem Excalibur)
- umožňuje provádění replikací dat mezi jednotlivými instalacemi velmi jednoduchým způsobem
- od r. 2005 probíhala postupná implementace podpory protokolu The Open Archive Initiative Protocol for Metadata Harvesting (OAI-PMH):
 - Aplikace umožňuje zpřístupnit popisná metadata prostřednictvím verze 2.0.
 - Specifikace protokolu je dostupná na <http://www.openarchives.org>.
 - Zprovozněnou instancí aplikace Kramerius je vhodné zaregistrovat [zdarma] jako **poskytovatele dat** na internetových stránkách OAI (<http://www.openarchives.org/data/registerasprovider.html>), kde jsou popsána pravidla registrace, v jejím rámci je ze

strany OAI provedeno automatické testování, zda implementace odpovídá specifikaci. Testování je poté periodicky opakováno.

- Úspěšně otestovaný poskytovatel dat je zařazen do veřejně přístupného seznamu poskytovatelů dat (viz printscreen) na adrese: <http://www.openarchives.org/Register/BrowseSites>. Před registrací musí aplikace obsahovat alespoň jeden záznam přístupný přes OAI-PMH (**viditelný uživatelům**), aby mohla splnit podmínky testů.

Record	Identify	Name	URL	Contact
Record	Identify	National Library of Australia Digital Object Repository	http://www.nla.gov.au/np/oa/oa/identify	nla.gov.au
Record	Identify	National Library of Czech Republic - Digitized documents	http://krampus.slp.cz/krampus/oa/	krampus.slp.cz
Record	Identify	Naturalis Publications	http://data.wra.nl/cgi/oa/oa/naturalis	
Record	Identify	Nature Precedings	http://precedings.nature.com/oa2/	
Record	Identify	NCSA-ROT Teachers Resource Page	http://es2003.nca.usc.edu/oa2.php	es2003.nca.usc.edu
Record	Identify	NDAD - UK National Archive of Datasets	http://ndad.uk.ac.uk/cgi-bin/oa/OAIZ0	ndad.uk.ac.uk
Record	Identify	NECTAR	http://nectar.northampton.ac.uk/oa2/	nectar.northampton.ac.uk
Record	Identify	NELCO Legal Scholarship Repository	http://ir.nelco.org/oa2.cgi	nelco.org
Record	Identify	NetMob Units OAI repository	http://www.science.uu.nl/0080netmob/oa2/oa2.php	UUITM Netmob
Record	Identify	New Zealand Electronic Text Centre	http://www.nzetc.org/cgi-bin/gateway/gateway.cgi	www.nzetc.org/oa-ig/htu-0vc.xml

- v roce 2007 byla dokončena implementace podpory formátů METS a MARC21 a začlenění formátu METS do OAI-PMH
- v roce 2007 bylo integrováno fulltextové vyhledávání pomocí Lucene
- perzistentní identifikátor na všech úrovních dokumentu (UUID)
- využití handle systému pro perzistentní URL
- zpřístupnění všech elementů přes OAI-PMH
- podpora hromadných importů a exportů dat. Zejména automatizace prováděných importů nabývá na významu při robotickém skenování, kdy se předpokládá zpracování velkého množství dokumentů (možností nastavení různých úrovní předvoleb pro zpřístupňování dokumentů, jejich spojování-slučování a validace názvů obrazových a textových souborů podle definovaných pravidel)
- kontroluje správnost kódování OCR a v případě korektních údajů ve hlavičce souboru provádí konverzi do UTF-8, v opačném případě nepovolí import nekorektních dat a oznámí tuto skutečnost ve výpisu logů (přehled importů)
- disponuje propojením do nejrůznějších portálů:

- The European Library³
- národní digitální knihovna⁴
- do SK ČR – link z Krameria na základě ISSN, link ze SK ČR na el. dokument
- podpora perzistentních URL umožňuje stabilní prolinkování do knihovního OPACu

Na první pohled by se mohlo zdát, že aplikace toho zvládá hodně, že splňuje požadavky, které na ni byly kladeny. Přesto je z hlediska narůstajících potřeb evidentní, že Kramerius jako systém musí projít zásadní změnou, aby si udržel pozici, jakou má dnes u svých uživatelů. Proto další vývoj musí být nasměrován na:

- využití systému FEDORA jako jádra
 - hlavním důvodem FEDORA je robustní stroj s podporou složených objektů; je vhodný i z hlediska dlouhodobé archivace
- Kramerius pod FEDOROU by měl v první fázi zajistit minimálně stávající funkčnost a měl by podporovat stejné typy dokumentů (monografie a periodika)
- další vývoj zaměřit na využití vlastností FEDORY
- rozšíření administrace
- rozšířená správa uživatelských a přístupových práv
- dávkové operace s digitálními dokumenty
- nové typy dokumentů
- využití vlastností FEDORY pro dlouhodobé uchování dat

Na závěr příspěvku uvádím částečný výčet neaktivnějších institucí, které aplikaci Kramerius využívají:

- Národní knihovna ČR, Moravská zemská knihovna v Brně, Státní technická knihovna, Knihovna Akademie věd ČR
- Vědecké knihovny: České Budějovice, Havlíčkův Brod, Hradec Králové, Olomouc, Ostrava, Zlín
- Muzejní knihovny: UPM Praha, ŽM Praha, DA Praha, MVČ Hradec Králové
- Archivní knihovny: Národní filmový archiv

³ viz: <http://www.theeuropeanlibrary.org>

⁴ viz: <http://www.ndk.cz/>