

Esej pro předmět PV070 Digitální knihovny



Evropská digitální knihovna

<<http://www.europeana.eu>>

Tomáš Drusa

9. 12. 2009

3. ročník

Login: **xdrusa**

UČO: **256167**

Úvod

Europeana [1] je projekt Evropské unie s cílem je vytvořit webový portál, ze kterého by bylo možné procházet digitální dokumenty všech typů vztahující se k Evropskému kulturnímu dědictví a vyhledávat v nich. Nejde tedy o digitální knihovnu v plném slova smyslu, ale o specifický typ vyhledavače, který pracuje nad centrální databází metadat objektů. Objekty samy se ale nacházejí v on-line repozitářích mnoha jednotlivých evropských kulturních institucí, které jsou do projektu zapojeny. Projekt je stále ve stádiu betaverze. Spuštěn byl v listopadu 2008, kdy pracoval nad dvěma miliony digitálních objektů; plné spuštění bylo určeno na rok 2010, s cílem dosáhnout v té době deseti miliónů objektů.

Zázemí

Projekt je podporován Evropskou unií, která jej financuje a zaštiťuje. Z programů eContent-Plus (do r. 2008) a Information and Communications Technologies Policy Support Programme (ICT-PSP) [2] do ní je investováno kolem 2 milionů eur ročně; další miliony jsou vyčleněny na podporu digitalizace a zpřístupňování obsahu on-line v zemích Unie. Formální sdružení stojící za Europeanou je Evropský spolek digitálních knihoven (European Digital Library Foundation), založený 8. listopadu 2007. Fakticky Europeana sídlí v prostorách Holandské národní knihovny v Haagu.



Problémy se spuštěním

Krátce po spuštění betaverze ve čtvrtek 20. 11. 2008 okolo 11. hodiny web Europeany zkolaboval z důvodu přílišné zátěže. Tři v Nizozemí umístěné servery, na kterých služba běžela, byly testovány na zátěž až pět milionů přístupů za hodinu a nepředpokládalo se, že by jí mělo být dosaženo. V den spuštění se ale zátěž pohybovala okolo deseti milionů přístupů v hodině, ve špičce pak dosahovala asi třinácti milionů, tj. přibližně 3700 požadavků za vteřinu. Během čtyř hodin tak byla předstihnuta návštěvnost stránek Evropské knihovny za celý rok. Ani nouzové navýšení kapacity na cca 8 milionů přístupů nepomohlo a portál musel být odstaven kvůli rozšíření serverové kapacity. Znovuspuštěna byla Europeana až koncem roku 2008, nyní se čtyřnásobnou kapacitou – měla by tedy zvládat přes pět a půl tisíce požadavků za vteřinu [3].

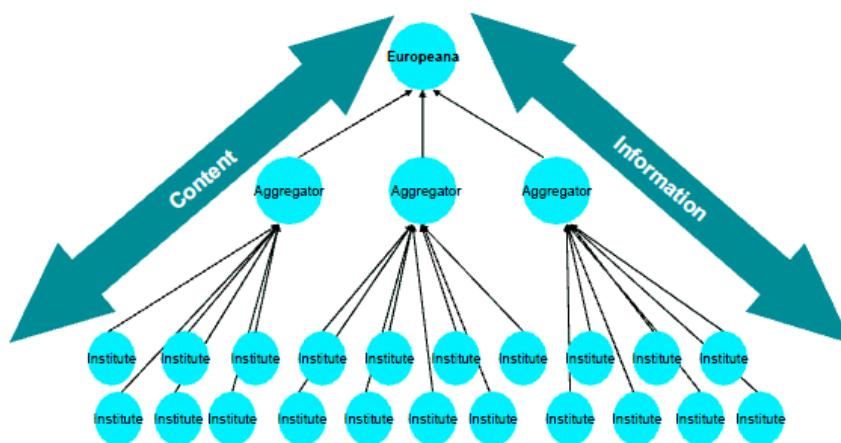
Principy, myšlenky, cíle

Jedním z principů je postavení prostoru metadat Europeany jako sémantického webu, ve kterém se kromě klasického katalogového vyhledávání dají sledovat i vztahy a podobnosti objektů. To by, spolu s navázáním objektů na časové a zeměpisné údaje, mělo umožnit přirozené vyhledávání pomocí otázek kdo, kde, kdy a co. Provázání má být realizováno technologií RDF a návrh je v souladu s iniciativou Open Linked Data, která se snaží přiblížit vizi Tima Berners-Leeho o budoucím webu.

Vyhledávání má dva režimy: jednoduchý fulltextový, á la Google, a rozšířený, který umožňuje navolit několik parametrů a vyhledávání přesněji vymezit. Ambicí je autorů je postavit jednotný vyhledávací jazyk, který bude sloužit jak uživatelům, tak i aplikacím třetích stran a vnitřním rutinám systému. Tato část v demoverzi nijak dobře nefunguje a i v dokumentaci se autoři vyjadřují nejasně – mluví o indexování obsahů, podobnostním vyhledávání a dotazování sémantické sítě pomocí SPARQL, konkrétní popis však chybí. Zřejmě tedy jde o dosud nedořešené téma.

Celý portál má být ze zadání přístupný ve všech jazycích Unie. Není ovšem dosud jisté, nakolik hluboko půjde tato mezijazykovost – vývojáři prezentují čtyři navrhované úrovně „nadjazykovosti“ portálu, od prostého překladu uživatelského rozhraní (což je současný stav portálu), přes překlad vyhledávaných termínů (nahlížený jako rozumně realizovatelný kompromis) až po úplné překládání včetně nalezených metadat či dokonce dokumentů (které je sledováno náročným a v praxi dosud nepříliš použitelným).

Dalším principem je Europeana jako mocná platforma, jejíž obsah je poskytován a udržován institucemi. Europeana podporuje myšlenku sdružování primárních zdrojů pod agregační autority, například na národní úrovni, které by určily data vhodná k zařazení a předpřipravily je např. vyloučením duplicit, zjednoznačněním apod. Až tito agregátoři by pak měli svá data poskytovat do evropské digitální knihovny (mimo jiné právě na vznik takových center jsou určeny dotace programu ICT-PSP). Zejména u poskytovatelů větších, dobře spravovaných kolekcí lze data poskytovat i přímo.



Základním kamenem je protokol OAI-PMH, který je od poskytovatelů vyžadován. Europeana využívá jeho možnosti, sklízí periodicky a inkrementálně, s občasným úplným refreshem – navzdory úporné práci se mi nepodařilo zjistit na jaké intervaly jsou tyto procesy nastaveny. Systém má ovšem být schopen zpracovat celou řadu dalších používaných metadatových schémat, doplnit jejich údaji případně mezery a zejména pomocí nich vyhledávat (TEL, METS, EAD, EBU Core, Im-mix, CIDOC CRM, MODS, MARCXML, MPEG-21, CDWA, Dismarc, museumdat a Moreq2). Zde je požadováno po poskytovatelích obsahu, aby vypracovali jednoznačný popis převodu metadat ze svých systémů zápisu do ESE. Komponenta Europeany SPM poté vybere příslušný algoritmus podle zdroje příchozích dat a výstupem jsou metadata ve formátu ESE v databázi Europeany. Europeana sama podporuje celou řadu externích služeb a UDDI registr, např. protokoly OAI-PMH, XML-RPC, SRU-CQL.

Europeana Semantic Elements a ESE XML Schema

Vývojáři Europeany nazývají její vnitřní metadatové schéma Europeana Semantic Elements, zkráceně ESE [4]. Není to nic jiného než Dublin Core rozšířený o několik nových značek. Z Dublin Core je podporováno všech 15 značek, navíc je převzata značka provenance ze skupiny dcterms. Kromě těchto jsou definovány nové značky ve jmenném prostoru europeana.

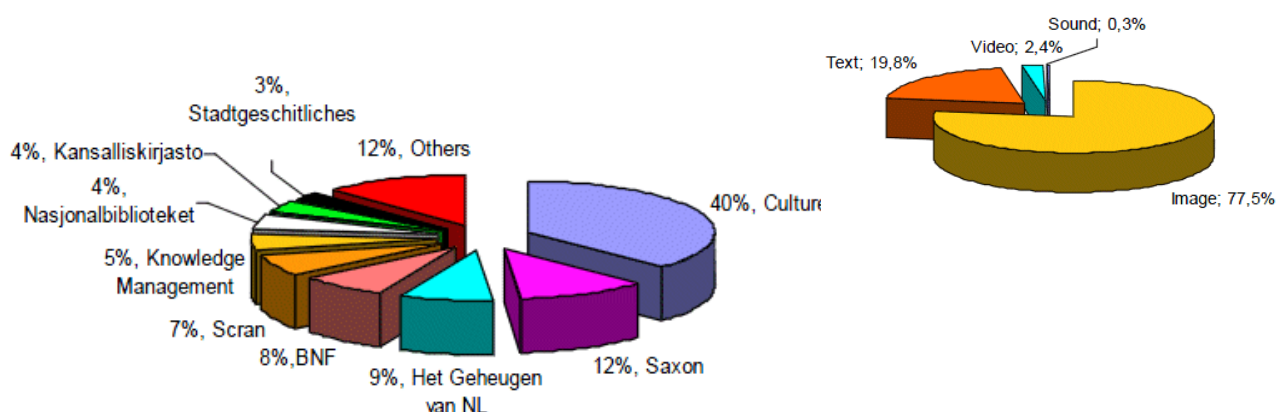
Source	Element	Element Refinement(s)
DC	Title	Alternative
DC	Creator	
DC	Subjekt	
DC	Description	tableOfContents
DC	Publisher	
DC	Contributor	
DC	Date	created; issued
DC	Type	
DC	Format	extent; medium
DC	Identifier	
DC	Source	
DC	Langure	
DC	Relation	isVersionOf; hasVersion; isReplacedBy; replaces; isRequiredBy; requires; isPartOf; hasPart; isReferencedBy; references; isFormatOf; hasFormat; conformsTo
DC	Coverage	spatial; tempoal
DC	Rights	
DC terms	Prohnance	
Europeana	Unstored	
Europeana	Object	
Europeana	Provider	
Europeana	Type	
Europeana	isShownBy;	isShownAt
Europeana	Country	
Europeana	hasObject	
Europeana	Langure	
Europeana	userTag	
Europeana	Uri	
Europeana	Year	

Europeana pracuje se čtyřmi typy digitálních objektů (europeana:type) – textové, obrazové, zvukové a videa. Informace europeana:isShownAt obsahuje URL stránky, na které je objekt vystaven svou domovskou institucí, zatímco isShownBy odkazuje přímo na soubor s daty (obrázek, zvukový soubor apod.). Prvky object a booleovský přepínač hasObject se používají pro práci s náhledovými miniaturami. Usertag je určen pro data registrovaných uživatelů, se kterými se počítá do budoucna. Někdy by se mohlo zdát, že se metadata překrývají, ale není tomu tak – např. dc:language a europeana:language. V těchto případech je striktně řečeno, co který obsahuje. Tag unstored pak slouží pro uložení informací, které nespádají do žádné škatulky. Součástí dokumentace je i detailní návod, který dělí metadata dle důležitosti a popisuje, co do kterého elementu patří [5].

Pro validaci XML je k dispozici XML Schema soubor [6] na adrese
 <http://www.europeana.eu/schemas/ese/ESE-V3.2.xsd>.

Obsah

Při spuštění portálu autoři zveřejnili statistiky; obsah ohodnotili slovy „více než dva milióny digitálních objektů“. Zdaleka největším přispěvatelem byla Francie, odkud pocházela více než polovina obsahu. Následovaly Velká Británie a Nizozemí s deseti procenty, Finsko s osmi a Švédsko se sedmi. Mnoho států se pak pohybuje mezi půl procentem a procentem, a velká skupina států (kam patří i ČR) nepřispívá ani desetinu procenta [1]. V současnosti se autoři chlubí již 4.6 miliony objektů, bohužel bez podrobností. Jedinou zapojenou organizací z České republiky je Národní knihovna, a to nikoliv jako přispěvateľ, ale partner a národní autorita [7]. Díla, která do Evropské digitální knihovny přispěla (např. ilustrace z Vyšehradského kodexu nebo Labyrintu světa a ráje srdce) mají jako poskytovatele uvedenu Evropskou knihovnu.



Názor autora

Projekt evropské digitální knihovny Europeana má veliké ambice, které se mu zatím nedaří příliš naplnit. Autoři si vytkli mnoho obtížných cílů, které budou jen s obtížemi plnit. Při čtení technických „deliverables“ máte pocit, že vše musí být skvělé; o to tvrdší je potom realita.

Současná betaverze disponuje zejména nepřívětivým vyhledáváním a špatnou češtinou, která je plná špatně užitých výrazů i prostých překlepů a působí dojmem na poslední chvíli psaného referátu. Také pochybuji, že by se snad podařilo dodržet termín a během jediného roku rozrůst kolekci na všude proklamovaných deset miliónů děl.

Přesto, pokud se autorům podaří splnit jen třetinu z vytčených cílů, půjde bezesporu o úspěch. Jednotlivé myšlenky jsou přínosné a funkční, jen – jak už to s vyšlapáváním nových cest bývá – uvést je do praxe bude stát více dřiny, času a peněz, než se čekalo.

Reference

- Rozcestník oficiální technické dokumentace projektu:

<<http://version1.europeana.eu/web/europeana-project/technicaldocuments/>>.)

- Hlavní zdroj:

Europeana Outline Functional Specification For development of an operational European Digital Library (Deliverable 2.5). Dostupné jako PDF na:

<http://version1.europeana.eu/c/document_library/get_file?uuid=a9e29cb4-a9b3-462a-a43d-0b480c677088&groupId=10602>.

[1] – MEMO/08/724. EUROPEANA – Europe’s Digital Library: Frequently Asked Questions. *EUROPA – PRESS RELEASES* [on-line]. 20. listopadu 2008 [cit. 2. 12. 2009]. Dostupné jako PDF na:

<<http://europa.eu/rapid/pressReleasesAction.do?reference=MEMO/08/724&format=PDF&aged=0&language=EN&guiLanguage=cs>>.

[2] – eContentPlus [on-line]. [cit. 2. 12. 2009]. Dostupné na:

<http://ec.europa.eu/information_society/activities/econtentplus/index_en.htm>.

[3] – MEMO/08/733. Europeana website overwhelmed on its first day by interest of millions of users. *EUROPA – PRESS RELEASES* [on-line]. 21. listopadu 2008 [cit. 2. 12. 2009]. Dostupné jako PDF na:

<<http://europa.eu/rapid/pressReleasesAction.do?reference=MEMO/08/733&format=PDF&aged=0&language=EN&guiLanguage=cs>>.

[4] – Europeana Semantic Elements specifications v3.2.1. 06/11/2009. Dostupné on-line na:

<http://version1.europeana.eu/c/document_library/get_file?uuid=c56f82a4-8191-42fa-9379-4d5ff8c4ff75&groupId=10602>.

[5] – Metadata Mapping & Normalisation Guidelines for the Europeana Prototype v1.2. 07/08/2009.

Dostupné on-line na: <http://version1.europeana.eu/c/document_library/get_file?uuid=58e2b828-b5f3-4fe0-aa46-3dcbc0a2a1f0&groupId=10602>.

[6] – About the ESE v3.2 XML Schema. 07/08/2009. Dostupné on-line na:

<http://version1.europeana.eu/c/document_library/get_file?uuid=104614b7-1ef3-4313-9578-59da844e732f&groupId=10602>.

[7] – Europeana – Partners (seznam partnerů a příspěvatelů portálu). [cit. 9. 12. 2009] Dostupné na:

<<http://www.europeana.eu/portal/partners.html>>.

Metadata (RDF, Dublin Core):

```
<?xml version="1.0"?>
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:dcq="http://purl.org/dc/documents/rec/dcmes-qualifiers-
20000711.htm">
  <rdf:Description rdf:about="Europeana">
    <dc>Title>
Europeana
    </dc>Title>
    <dc:Creator>
    <rdf:Bag>
    <rdf:li>
Tomáš Drusa
    </rdf:li>
    <rdf:li>
    <rdf:Description>
    <dcq:CreatorType>personalName</dcq:CreatorType>
    <rdf:value>Drusa, Tomáš</rdf:value>
    </rdf:Description>
    </rdf:li>
    <rdf:li>
    <rdf:Description>
    <dcq:CreatorType>corporateName</dcq:CreatorType>
    <rdf:value>Faculty of Informatics, Masaryk University,
Brno</rdf:value>
    </rdf:Description>
    </rdf:li>
    <rdf:li>
    <rdf:Description>
    <dcq:CreatorType>address</dcq:CreatorType>
    <rdf:value>256167@mail.muni.cz</rdf:value>
    </rdf:Description>
    </rdf:li>
    </rdf:Bag>
    </dc:Creator>
    <dc:Subject>
Europeana
    </dc:Subject>
    <dc>Date>
    <rdf:Description>
    <dcq:DateType>created</dcq:DateType>
    <dcq:DateScheme>W3C-DTF</dcq:DateScheme>
    <rdf:value>2009-12-09</rdf:value>
    </rdf:Description>
    </dc>Date>
```

```
<dc:Type>
<rdf:Description>
<dcq:TypeScheme>DCMIType</dcq:TypeScheme>
<rdf:value>Text</rdf:value>
</rdf:Description>
</dc:Type>
<dc:Format>
<rdf:Description>
<dcq:FormatScheme>IMT</dcq:FormatScheme>
<rdf:value>application/pdf</rdf:value>
</rdf:Description>
</dc:Format>
<dc:Format>
<rdf:Description>
<dcq:FormatType>medium</dcq:FormatType>
<rdf:value>computerFile</rdf:value>
</rdf:Description>
</dc:Format>
<dc:Format>
<rdf:Description>
<dcq:FormatType>extent</dcq:FormatType>
<rdf:value>7</rdf:value>
</rdf:Description>
</dc:Format>
<dc:Identifier>
<rdf:Description>
<dcq:SourceScheme>URL</dcq:SourceScheme>
<rdf:value>http://www.europeana.eu</rdf:value>
</rdf:Description>
</dc:Identifier >
<dc:Language>
<rdf:Description>
<dcq:LanguageScheme>RFC3066</dcq:LanguageScheme>
<rdf:value>cze</rdf:value>
</rdf:Description>
</dc:Language>
<dc:Rights>CC by</dc:Rights>
</rdf:Description>
</rdf:RDF>
```