

Konfigurace OAI-PMH v DSpace

Vlastimil Krejčíř, krejcir@ics.muni.cz

Ústav výpočetní techniky, Masarykova univerzita, Brno

CZDSUG 2009, Ostrava

Obsah přednášky

- Jemný úvod do protokolu OAI-PMH.
- OAI-PMH v DSpace (oai_dc etc.).
- Konfigurace OAI-PMH (Crosswalks).
- Řešení pro DML-CZ (dml_dc).

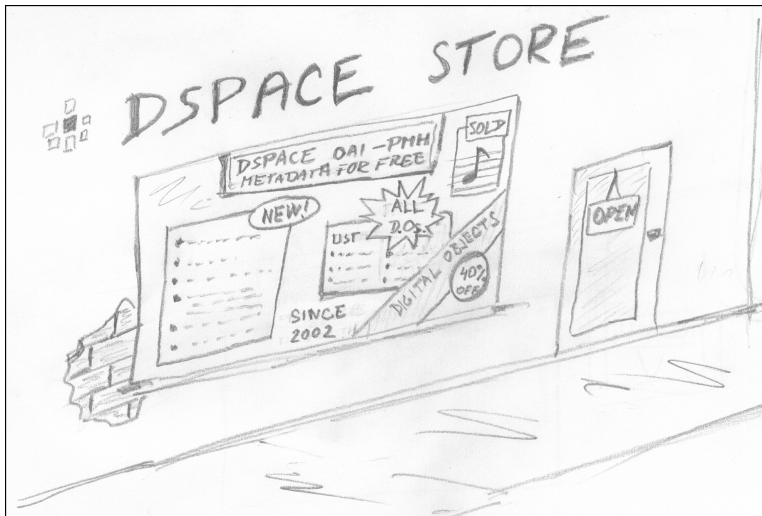
Protokol OAI-PMH



The Open Archives Initiative Protocol for Metadata Harvesting

`http://www.openarchives.org/OAI/
openarchivesprotocol.html`

Protokol OAI-PMH II



Protokol OAI-PMH III

- sdílení metadat (nezávisle na formátu)
 - musí podporovat nekvalifikovaný Dublin Core
- bezstavový protokol (HTTP + XML)
- poskytovatel (server) vs. klient (harvester)
- jednoznačné identifikátory
- „sklizení“ metadat (sets a datestamp)
- sklizení dat? (Ano, ale. . .)

Protokol OAI-PMH IV

OAI server v DSpace:

`http://DSPACE_URL/dspace-oai/request?<dotaz>`

Příklady:

`http://dspace.muni.cz/dspace-oai/request?`

`http://oai.dml.cz/request?`

Protokol OAI-PMH V

<dotaz>

Informace:

verb=Identify

verb=ListMetadataFormats

Co je v repozitáři:

verb=ListIdentifiers&metadataPrefix=oai_dc

A získání metadat:

verb=GetRecord&identifier=oai:dspace.muni.cz:

ics_muni_cz%2F47&metadataPrefix=oai_dc

Konfigurace, metadata

Podporované formáty metadat:

- základní Dublin Core (`oai_dc`)
- kvalifikovaný Dublin Core (`qdc`)
- MODS (`mods`)

Podporované zapouzdřovací formáty:

- METS (`mets`)
- MPEG-21 DIDL (`didl`)

Kromě `oai_dc` je nutno všechno explicitně zapnout.

Konfigurace QDC a MODS

Příklad (výchozí nastavení):

```
verb=GetRecord&identifier=oai:dspace.muni.cz:  
ics_muni_cz%2F47&metadataPrefix=mods
```

Příklad pro modifikaci qdc:

```
dc.description.volume =  
<dc:description qualifier="volume" />
```

Výsledek:

```
verb=GetRecord&identifier=oai:dml.cz:  
10338.dmlcz%2F102493&metadataPrefix=qdc
```

Změna může znamenat porušení XML schéma!
A v příkladu se porušení děje!

DIDL (dc) a METS (mods)

DIDL obsahuje i data zakódovaná v **base64** (výhoda?).

Konfigurace MODS ovlivňuje i MODS v METS.

Zapouzdření jiných metadatových formátů - zásah do kódu DSpace.

XSLT crosswalks

DSpace vnitřně pracuje s formátem DIM (DSpace Intermediate Metadata). Je možné využít XSL šablony pro převod z DIM do libovolného formátu.

Tyto šablony pracují na bázi self-named pluginů – není tedy nutný zásah do kódu DSpace.

Výhoda – lze si nadefinovat vlastní XML schéma (řeší problém proušení schéma u refefinice `qdc` a `mods`).

Soubory

- soubory lze „přímo vystavit“ přes mets a didl
- je možné je získat i nepřímo z metadat (extrakcí URL z webu)
- ORE?

Interoperabilita

- harvestery – chceme jich co nejvíce
- mnoho formátů – kdo je umí harvestovat
- oai_dc umí všichni!
- oai_dc je moc jednoduché :-)

Jak z toho ven?

Interoperabilita II

Koncept DML - GDZ

Koncept používaný v **Göttinger Digitalisierungszentrum**
<http://gdz.sub.uni-goettingen.de/>.

Klasický DC:

```
<dc:identifier>1214-8148</dc:identifier>
```

GDZ unqualified DC:

```
<dc:identifier>ISSN:1214-8148</dc:identifier>
```

Řešení v DML-CZ

Pravděpodobně na bázi GDZ. Předpokládám konfigurovatelný „plugin“, který bude volně k dispozici. Sledujte web

<http://dSPACE.cz/>

Děkuji za pozornost.