

Building integration environment based on OAI-PMH protocol

**Novytskyi Oleksandr
Institute of Software Systems NAS Ukraine
Alex@zu.edu.ua**

Roadmap

- What is OAI-PMH?
- Requirements for infrastructure
- Step by step building integration environment
- Additional services for end-users
- Future directions for integration
- Some issues
- Questions

What is OAI-PMH?

OAI-PMH (*Open Archives Initiative Protocol for Metadata Harvesting*) is a protocol developed by the Open Archives Initiative. It is used to harvest (or collect) the metadata descriptions of the records in an archive so that services can be built using metadata from many archives. An implementation of OAI-PMH must support representations metadata in Dublin Core, but may also support additional representations.

What is OAI-PMH?

Protocol Requests and Responses

Requests	Responses
GetRecord	Metadata item
Identify	Information about repository, schema, protocol version and other information
ListIdentifiers	Returns record header it contains: <ul style="list-style-type: none">· unique identifier -- the unique identifier of an item in a repository;· datestamp -- the date of creation, modification or deletion of the item for the purpose of selective harvesting;· zero or more setSpec elements -- the set elements belong to item for the purpose of selective harvesting.
ListMetadataFormats	is used to retrieve the metadata formats available in a repository
ListRecords	is used to harvest records from a repository
ListSets	is used to retrieve the set structure of a repository, useful for selective harvesting.

Requirements for infrastructure

- Data provider
 - Eprints (<http://www.eprints.org>)
 - Dspace (<http://www.dspace.org>)
 - Fedora (<http://fedora-commons.org>)
 - Other software applications which support uploading data in to OAI-PMH
- Service provider
 - Open Harvester Systems (<http://pkp.sfu.ca/?q=harvester>)
 - Omeka (<http://omeka.org>)

Step by step for building integration environment

1. Select hosting for installation Service provider
 - aggregations many big repositories need a lot of RAM (~3 GB) for PHP execution
 - Crontab
 - Long time execution script



- **Needs dedicated server or VPS**

1. Installation and configuration Service provider
 1. Installation Service provider
 2. Customization user interface
 3. Connection digital libraries
 4. Configuration Reading Tools

Step by step guide for building integration environment

PKP Open Harvester Systems

- Easy to install
- Long term support
- Easy to upgrade
- Many plugins
- Supports ETD-MS, MARC, Dublin Core, MODS metadata formats
- Good performance

search in public archives of Ukraine oai.org.ua

The project <http://oai.org.ua> started in 2008 and its main goal is to integrate open archives/repositories of Ukraine.

Currently, it integrates 35 open archives of Ukraine. And it contains more than 126,000 records.

The system was developed and is maintained by Institute of Software Systems of the NAS of Ukraine.

Addition services for end-users

PKP Harvester offers a range of tools that you can use in scientific discovery.

For example, the Reading Tools. This tool allows to carry out a quick search of related resources.

Reading Tools

Record Details

Інтеграція анотацій в Semantic Web. Модель інформаційного середовища

Zhytomyr State University Library

[VIEW ARCHIVE INFO](#)

FIELD	VALUE
Title	Інтеграція анотацій в Semantic Web. Модель інформаційного середовища
Creator	Новицький, О.В.
Subject	QA75 Electronic computers. Computer science
Publisher	Інститут програмних систем НАН України
Date	2009-09-29
Type	Monograph NonPeerReviewed
Format	application/pdf other
Identifier	http://eprints.zu.edu.ua/2935/1/%D0%97%D0%B2%D1%96%D1%82.pdf http://eprints.zu.edu.ua/2935/2/%D0%97%D0%B2%D1%96%D1%82.docx Новицький, О.В. (2009) Інтеграція анотацій в Semantic Web. Модель інформаційного середовища. Technical Report. Інститут програмних систем НАН України. (Unpublished)
Relation	http://eprints.zu.edu.ua/2935/

READING TOOLS

For this record

- [Look up terms](#)

Find related items among

- [Author's work](#)
- [Government policy](#)
- [Book searches](#)
- [Book reviews](#)
- [Dissertations](#)
- [Online forums](#)
- [Quotations](#)
- [Pay-per-view](#)
- [Resources](#)
- [Media reports](#)
- [Web search](#)

USER

You are logged in as...

admin

- [My Profile](#)
- [Log Out](#)

SEARCH

LANGUAGE

English ▾

Reading Tools

Record Details

Інтеграція анотацій в Semantic Web. Модель інформації

Zhytomyr State University Library

[VIEW ARCHIVE INFO](#)

FIELD	VALUE
Title	Інтеграція анотацій в Semantic Web. Модел
Creator	Новицький, О.В.
Subject	QA75 Electronic computers. Computer scienc
Publisher	Інститут програмних систем НАН України
Date	2009-09-29
Type	Monograph NonPeerReviewed
Format	application/pdf other
Identifier	http://eprints.zu.edu.ua/2935/1/ http://eprints.zu.edu.ua/2935/2/ Новицький, О.В. (2009) Інтеграція анотації програмних систем НАН України. (Unpublis
Relation	http://eprints.zu.edu.ua/2935/

READING TOOLS

For this record

- [Look up terms](#)

Author's work

"Інтеграція анотацій в Semantic Web. Модель інформаційного середовища"

For items with more than one author, leave all names checked to search for similarly co-authored work, or search one at a time by checking only a single name. To learn more about the resource searched, click on the information symbol. These resources have been selected because of their relevance and their open (free) access to all or part of their contents.

Новицький, О.В.

Search 1. Google Scholar [ABOUT](#)

Search 2. Наукова електронна бібліотека періодичних видань НАН України [ABOUT](#)

Search 3. Intute: Repository Search [ABOUT](#)

Search 4. OAIster (Open Archives Initiative research databases) [ABOUT](#)

Search 5. Public Knowledge Project Open Archives Harvester [ABOUT](#)

Close

Future directions for **INFORMATION** integration

- Protocol OAI-PMH provides a structured data integration, however leaves aside the semantics.

We plan to use the model of semantic description for each resource by using RDFa.

This will allow search engines to search for more information.

Future directions

INFORMATION

integration

- Support distributed full text search in the archives.
 - We have already developed for EPrints a set of web services that enable searching for metadata fields as well as full-text search. In the near future we plan to introduce this extension to our system <http://oai.org.ua>.
- The integration of archives based on ontologies
This direction is being studied by us, and we plan the use of ontologies to integrate heterogeneous resources that harvester collects.

Some issues

- Some users install DSpace without supporting OAI-PMH
- Some EPrints installations do not have unique OAI Repository Identifier
- Repositories can change their URL. It gives additional difficulties in supporting the list of repositories (keeping it up to date).

Questions

Thank you for your attention!