Economists Online as a building block of a VRE solution

OAI6 Conference, Geneva
18 June, 2009

Benoit PAUWELS - Université Libre de Bruxelles
Agenda

• NEEO project
  - Facts
  - Infrastructure
  - Economists Online portal

• EO Data model
  - extensibility

• Enrichment of publications in EO
  - metadata
  - datasets

• EO and OAI-ORE
Acknowledgements

- Thomas Place - Universiteit van Tilburg
- Peter van Huisstede - Erasmus Universiteit Rotterdam
- Fred Vos - Universiteit van Tilburg
NEEO project - facts

- **Key objectives**
  - Improve usability and global visibility of economics research
  - Provide easier and open access to high-quality multilingual academic output of leading economics institutes and their researchers
  - Via a sustainable portal with aggregated and enhanced metadata enabling an infrastructure for new services

- **Background**
  - 18 partners (LSE, Oxford, Tilburg, Leuven, UCL, ...)
  - Initiated by Nereus - 23 academic institutions/libraries with strengths in economics
  - eContentPlus, €2m, 30 months (Sep 2007 - March 2010)
Meresco Metadata Harvester

Objects

HTTP

OAI-PMH

Logs

General infrastructure

Metadata exchange format

DIDL / MODS
NEEO specs

Usage metadata exchange format

SWUP
OFI Comm Profile

Harvester

Crawler

Metadata

Meresco

Lucene

SRU

RePEc

OAI-PMH

RSS

EO portal
Homemade - FOSS

Exporter engine
Homemade - FOSS

Other portals
EO portal - main features

• EO Gateway
  - harvesting 20 institutional repositories + all RePEc archives
  - +/- 800,000 bibliographic references | +/- 650,000 publications available online | +/- 250,000 open access
  - Lucene indexing; Meresco FOSS (http://www.cq2.nl)

• EO portal
  - search & find, facets, sorting, links to OA full-text, download statistics, permalink, OpenURL
  - export: APA citations, RIS; complete publication lists for registered authors (PDF | RTF) ; Coins/Zotero
  - full text searching
  - every search = 1 RSS feed
  - MLIA: search query English, find also French publications
  - public launch 28 January 2010

• Open standards:
  - Information exchange: OAI-PMH, SRU, RSS, OpenURL
  - Metadata formats: MPEG-21/DIDL, MODS, DDI, COINS, ReDIF, RDF/FOAF, SWUP
Welcome to Economists Online

Economists Online showcases some of the world's leading institutions, their scholars and their academic publications and datasets.

Economists Online content is being combined with RePEc archives to provide a new information service to the economist.

At the end of April 2009, over 700,000 bibliographic references were recorded in the system, with many links open access full text.

This is increasing day by day.

Economists Online is run by the Nereus consortium. co-funded by the European Union

### Institutions

<table>
<thead>
<tr>
<th>Institution</th>
<th>Homepage</th>
<th>IR</th>
<th>Scholars</th>
<th>More info</th>
</tr>
</thead>
<tbody>
<tr>
<td>London School of Economics and Political Science</td>
<td><a href="http://www.lse.ac.uk">Home</a></td>
<td><a href="http://www.lse.ac.uk">IR</a></td>
<td><a href="http://www.lse.ac.uk">Scholars</a></td>
<td><a href="http://www.lse.ac.uk">More info</a></td>
</tr>
<tr>
<td>Tiburg University</td>
<td><a href="http://www.tiburg.nl">Home</a></td>
<td><a href="http://www.tiburg.nl">IR</a></td>
<td><a href="http://www.tiburg.nl">Scholars</a></td>
<td><a href="http://www.tiburg.nl">More info</a></td>
</tr>
<tr>
<td>Katholieke Universiteit Leuven</td>
<td><a href="http://www.kuleuven.be">Home</a></td>
<td><a href="http://www.kuleuven.be">IR</a></td>
<td><a href="http://www.kuleuven.be">Scholars</a></td>
<td><a href="http://www.kuleuven.be">More info</a></td>
</tr>
<tr>
<td>Maastricht University</td>
<td><a href="http://www.maastrichtuniversity.nl">Home</a></td>
<td><a href="http://www.maastrichtuniversity.nl">IR</a></td>
<td><a href="http://www.maastrichtuniversity.nl">Scholars</a></td>
<td><a href="http://www.maastrichtuniversity.nl">More info</a></td>
</tr>
<tr>
<td>Université libre de Bruxelles</td>
<td><a href="http://www.ulb.ac.be">Home</a></td>
<td><a href="http://www.ulb.ac.be">IR</a></td>
<td><a href="http://www.ulb.ac.be">Scholars</a></td>
<td><a href="http://www.ulb.ac.be">More info</a></td>
</tr>
<tr>
<td>London School of Economics and Political Science</td>
<td><a href="http://www.lse.ac.uk">Home</a></td>
<td><a href="http://www.lse.ac.uk">IR</a></td>
<td><a href="http://www.lse.ac.uk">Scholars</a></td>
<td><a href="http://www.lse.ac.uk">More info</a></td>
</tr>
<tr>
<td>Tiburg University</td>
<td><a href="http://www.tiburg.nl">Home</a></td>
<td><a href="http://www.tiburg.nl">IR</a></td>
<td><a href="http://www.tiburg.nl">Scholars</a></td>
<td><a href="http://www.tiburg.nl">More info</a></td>
</tr>
<tr>
<td>Katholieke Universiteit Leuven</td>
<td><a href="http://www.kuleuven.be">Home</a></td>
<td><a href="http://www.kuleuven.be">IR</a></td>
<td><a href="http://www.kuleuven.be">Scholars</a></td>
<td><a href="http://www.kuleuven.be">More info</a></td>
</tr>
<tr>
<td>Maastricht University</td>
<td><a href="http://www.maastrichtuniversity.nl">Home</a></td>
<td><a href="http://www.maastrichtuniversity.nl">IR</a></td>
<td><a href="http://www.maastrichtuniversity.nl">Scholars</a></td>
<td><a href="http://www.maastrichtuniversity.nl">More info</a></td>
</tr>
<tr>
<td>Université libre de Bruxelles</td>
<td><a href="http://www.ulb.ac.be">Home</a></td>
<td><a href="http://www.ulb.ac.be">IR</a></td>
<td><a href="http://www.ulb.ac.be">Scholars</a></td>
<td><a href="http://www.ulb.ac.be">More info</a></td>
</tr>
</tbody>
</table>


EO Data model

- Publication is described as a compound object

- Representation as an MPEG21/DIDL document according to NEEO application profile

- Aggregation of 3 types of components
  - descriptiveMetadata (MODS)
  - objectFile
  - humanStartPage
EO Data model

• Compound object
  • persistent identifier
  • modified date

• Each component
  • persistent identifier
  • modified date
  • type: URI
    • info:eu-repo/semantics/descriptiveMetadata
    • info:eu-repo/semantics/objectFile
    • info:eu-repo/semantics/humanStartPage
EO Data model

• “descriptiveMetadata” component
  • representation ‘by-value’
  • MODS v.3.2; NEEO application profile; Digital Author Identifier

• “objectFile” component
  • representation ‘by-ref’; object file at some network location
  • additional properties: version, accessRights, ...

• Extensibility
  • additional components: enrichment
  • vocabularies:
    • typing of components
    • properties of components
  • by-value / by-ref
Enrichment in EO

• Types of enrichment in EO
  - Automatically generated JEL codes
  - Automatically created reference lists
  - Generate text version of PDF object file

• Enrichment process in EO
  1. ES gets records to be enriched from EO, over SRU
     • Based on date of request for enrichment of certain type and version
     • Based on flag set in EO record
  2. ES creates enrichment record(s)
  3. ES makes enrichment records available to EO, over OAI-PMH
  4. EO harvests enrichment records from ES and integrates into original record
  5. EO reuses enrichment information in its services:
     - Full text searching
     - Index & present JEL and references through portal
Enrichment in EO

- Status of implementation
  - Prototype on the way
    - Text: by-ref
    - JEL / references: by-value (XML/MODS)
  - Unclear on vocabulary for typing of extra components
  - Full specs in project deliverable
Enriched publication / by-ref

IR / ES

- XML - MODS
- PDF
- HTML
- TXT
- XML - MODS

EO

DIDL[1]

Item[1]

Descriptor/Identifier (persistent identifier)
Descriptor/modified

Item[1..∞] (of type descriptiveMetadata)

Item[0..∞] (of type objectFile)

Item[0..1] (of type humanStartPage)

Item[0..∞] (of type text)

Item[0..∞] (of type enrichedMetadata)
Other types of enrichment

- (Peer-) review of a publication
  - third party reviews publication and makes resulting text available as ...
  - compound object (metadata, object files, ...)
  - extra component in original compound object

- Datasets
  - compound object of
    - metadata : DDI
    - object files: SPSS, XLS, ...
    - software components
    - ...
  - extra component in original compound object
  - NEEO
    - prototype for enrichment of publications with datasets
    - Hosted Dataverse repository / DDI metadata
    - represent as DDI or DIDL ?
Enriched publication

IR / ES

XML - MODS
PDF
HTML
TXT
XML - MODS

EO

DIDL[1]

Item[1]

Descriptor/Identifier (persistent identifier)
Descriptor/modified

Item[0..∞] (of type descriptiveMetadata)

Item[0..∞] (of type objectFile)

Item[0..1] (of type humanStartPage)

Item[0..∞] (of type text)

Item[0..∞] (of type enrichedMetadata)

Item[0..∞] (of type review)

Item[0..∞] (of type dataset)

Review

Descriptor/Identifier (persistent identifier)
Descriptor/modified

Item[1..∞] (of type descriptiveMetadata)

Item[0..∞] (of type objectFile)

Dataset

Descriptor/Identifier (persistent identifier)
Descriptor/modified

Item[1..∞] (of type descriptiveMetadata)

Item[0..∞] (of type objectFile)
EO and OAI-ORE

- **DIDL**
  - Item
  - Item
  - ... 
  - Item
  - Aggregation
  - Compound Object
  - Aggregated Resources

- **Enriched publication**
  - aggregation of resources
  - aggregation of aggregations

- **Mapping DIDL to ORE ResourceMap (Atom)**
  - XSLT transformation: DIDL 2 Atom

- All NEEO compatible repositories are ORE-aware
- All EO publications and their components are URI addressable
- All EO publications can be used in ORE-based applications
Some references

- NEEO technical guidelines

- NEEO DIDL application profile
  - [http://drcwww.uvt.nl/~place/neeo/didl%20application%20profile.doc](http://drcwww.uvt.nl/~place/neeo/didl%20application%20profile.doc)

- NEEO MODS application profile
  - [http://drcwww.uvt.nl/~place/neeo/Use%20of%20MODS%20for%20institutional%20repositories-version%201.1.doc](http://drcwww.uvt.nl/~place/neeo/Use%20of%20MODS%20for%20institutional%20repositories-version%201.1.doc)

- NEEO technical guidelines for the exchange of usage metadata / SWUP ContextObject

- Dataverse
  - [http://thedata.org/](http://thedata.org/)

- NEEO project web site (deliverables, contacts, ...)