Opening up repositories for the Semantic Web

Dr. Timo Borst
ZBW – Leibniz Information Centre for Economics

COAR Annual Meeting
Istanbul, Turkey, May 7-8 2013
At a glance

1. Repositories and data interoperability

2. Semantic Web approaches in current repository projects

3. EconStor-LOD: Publishing a repository’s content as Linked Data

Wrap-up: LOD and the issue of interoperability of repositories
A brief history on OAI-PMH and DublinCore...

• OAI-PMH = standard protocol for exchanging data between repositories and/or service providers

• Established in 2001

• Nowadays supported by almost any institutional repository (+ metadata store like e.g. VUFind/SOLR) because of out-of-the-box distribution by EPrints, DSpace & Co., standardization/certification (DINI) etc.

• Used by large service providers like OAIlster, BASE or Institutional Repository Search (IRS)
A brief history on OAI-PMH and DublinCore

Some pitfalls:

• Interoperability is supported only on a formal level: test of base URL, DC format

• Only predefined subsets / timestamps can be queried

• Continuous harvesting can be cumbersome...

• OAI-PMH built on top of HTTP (and OSI reference stack), but less reliable
Semantic Web principles and techniques : Brave new world for data interoperability?

• No further protocol (just HTTP, with some hacks...)
• Use of (commonly used) vocabularies for expressing classes (groups of things) and properties in a machine-readable way
• LOD induced interoperability = web based access to data + semantics
• Opportunities:
  • Linking to other external – especially non-repository located - data
  • Crosswalks between different repositories by using mappings expressed in SKOS-notation (during retrieving or indexing)
SemWeb approaches in repository projects

Performing triplification on the level of aggregators/service providers, e.g.

- **RKB-Explorer** (Univ. of Southampton)
  - Collecting data from over 100 different sources (metadata stores from publishers in Computer Science)
  - Converting them into RDF and providing them as Triple-Store/SPARQL
  - Normalizing URIs (e.g., by introducing “sameAs” relations for the same concepts or publications)
  - Providing a demo interface for browsing the mash-up

- **CORE project**
  - Converting harvested metadata into RDF and providing them as Triple-Store/SPARQL
SPARQL query:

```sparql
select ?x ?y ?w
where {?x <http://core.kmi.open.ac.uk/data/hasOAIIdentifier> "oai:eprints.bbk.ac.uk.oai2:128".
  FILTER (?x != ?y)
}
```

Example queries
SemWeb approaches in repository projects

On the level of local repositories / data providers:

• 'Heavyweight‘ approach:
  • Converting local databases into RDF-triplestores
  • Exposing RDF via
    • SPARQL endpoints
  • Embedded RDF (=RDFa) in a repository‘s web pages
  • RDF pages

• 'Lightweight‘ approach:
  • Using LOD datasets (e.g. for vocabularies, persons) for cross-repository indexing and retrieval
Lightweight integration: LOD as a means for cross-repository indexing and retrieval

• Basic idea: Publishing resp. using controlled vocabularies as LOD
• Input terms (for retrieval or indexing) are handled as labels to their concepts modelled as LOD
• Typical SKOS relations (broader, narrower, related, sameAs,...) are used for
  • synonyms and translations
  • mappings to other vocabularies
  • search term expansion
Using LOD for indexing (concepts)

Title
Estimating the shape of economic crises under heter

Please enter the name(s) of the author(s) as follows: surname, first names

Author(s)
Dovern, Jonas 132157322
Jannson, Nils 133933768

Please enter an abstract of the publication.

Abstract
During the ongoing financial crisis the analysis of similar historical crises has gained more and more attention among economic researchers and forecasters. Existing studies, however, do not tackle the

Please enter your keywords once more and select the corresponding term from the controlled vocabulary "Standard Thesaurus Wirtschaft".

STW-Terms
financial or
STW-Terms
Financial crisis

Please enter the JEL classification codes that best describe the publication.

JEL-Classification
Using LOD for indexing (persons)

Title

Author(s)

Siebert, Horst

Author names

Siebert, Horst, PND: 120273152

Abstract

Siebert, Horst (1930 - 2009; Wirtschaftswissenschaftler, Dt. Wirtschaftswissenschaftler, Hein-Nördorf-Professor für Europäische Integration und Wirtschaftspolitik an der Johns Hopkins Univ., weiter tätig in Neuwied, MünsterWestfalen, Mannheim, Konstanz, Kiel. Professor i.R. für Volkswirtschaftslehre, internationale Wirtschaftsbeziehungen, Umwelt und natürliche Ressourcen; bis 2003 Präsident des Instituts für Weltwirtschaft (IWF), Professor an der Christian-Albrechts Universität zu Kiel, Lehrstuhl für Theoretische Volkswirtschaftslehre bis 2003, Institut für Weltwirtschaft <kiel>

Siebert, Horst, PND: 120273152

Siebert, Horst, PND: 120272814

Siebert, Horst (1939 - ; Pädagoge, Professor für Erwachsenenbildung)
Using LOD for retrieving (search term expansion)

<table>
<thead>
<tr>
<th>Datum</th>
<th>Titel</th>
<th>Autoren</th>
</tr>
</thead>
<tbody>
<tr>
<td>2010</td>
<td>Industrial ecology in policy making: What is achievable and what is not?</td>
<td>Pothen, Frank</td>
</tr>
<tr>
<td>1998</td>
<td>Kurz kommentiert: Rentenversicherung - Kombilohn - Autobahnaut - EZB - Umweltbilanz</td>
<td></td>
</tr>
<tr>
<td>2008</td>
<td>Pflichtpfand: Wirkungsloses Instrument</td>
<td>Wacker-Theodo</td>
</tr>
<tr>
<td>2008</td>
<td>Zuwanderung: Chance vertan</td>
<td>Koppel, Oliver</td>
</tr>
<tr>
<td>2008</td>
<td>Mitarbeiterkapitalbeteiligung: Kumulation von Risiken</td>
<td>Ragnitz, Joachir</td>
</tr>
</tbody>
</table>
EconStor-LOD: Publishing a repository’s content as Linked Open Data

About EconStor

• One of the top (subject) repositories for OA documents in Economics
• > 50,000 OA documents (mainly Working Papers)
• RePEc data provider
• DINI certificate
• Based on DSpace (1.4.2), customization work mainly done on
  • OAI-PMH crosswalks (oai_dc, didl, oai_wgl, oai_xmdp)
  • PDF optimization (automatically generated cover page)
  • Optimization of indexing by Google Scholar
  • Usage statistics compliant with COUNTER code of practice
EconStor-LOD: Publishing a repository’s content as Linked Open Data

Goals of EconStor-LOD:
• Disseminating OA content (metadata) into the Semantic Web
• Better retrieving of information
• (Automatic) Linking to other datasets
• Gaining experience on providing a local LOD infrastructure

Means:
• Regular PostgreSQL database dump
• Converted via D2R-server
• Use of links to other datasets
This is a database published with D2R Server. It can be accessed using:

1. your plain old web browser
2. Semantic Web browsers
3. SPARQL clients.

1. HTML View

You can use the navigation links at the top of this page to explore the database.

2. RDF View

You can also explore this database with Semantic Web browsers like Disco or Marbles. To start browsing, open this entry point URL in your Semantic Web browser:

http://linkeddata.econstor.eu/beta/all

3. SPARQL Endpoint

SPARQL clients can query the database at this SPARQL endpoint:

http://linkeddata.econstor.eu/beta/sparql

The database can also be explored using this AJAX-based SPARQL Explorer.
**Trade dynamics in the market for federal funds**


<table>
<thead>
<tr>
<th>Property</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>dcterms.abstract</td>
<td>We develop a model of the market for federal funds that explicitly accounts for its two distinctive features: banks have to search for a suitable counterparty and the rate depends on the size of the loan and the repayment. The theory is used to answer a number of positive and normative questions: What are the determinants of the market able to achieve an efficient reallocation of funds? We also use the model for theoretical and quantitative analyses of policy issues facing the Federal Reserve.</td>
</tr>
<tr>
<td>dc:creator</td>
<td><a href="http://linkeddata.econstor.eu/beta/resource/authors/9133509">http://linkeddata.econstor.eu/beta/resource/authors/9133509</a></td>
</tr>
<tr>
<td>dc:creator</td>
<td><a href="http://linkeddata.econstor.eu/beta/resource/authors/9133510">http://linkeddata.econstor.eu/beta/resource/authors/9133510</a></td>
</tr>
<tr>
<td>dcterms.isPartOf</td>
<td><a href="http://linkeddata.econstor.eu/beta/resource/collections/493">http://linkeddata.econstor.eu/beta/resource/collections/493</a></td>
</tr>
<tr>
<td>dc:issued</td>
<td>2012 (xsd:Year)</td>
</tr>
<tr>
<td>dc:keyword</td>
<td>bargaining (xsd:string)</td>
</tr>
<tr>
<td>dc:keyword</td>
<td>fed funds market (xsd:string)</td>
</tr>
<tr>
<td>dc:keyword</td>
<td>over-the-counter (xsd:string)</td>
</tr>
<tr>
<td>dc:keyword</td>
<td>ssr (xsd:string)</td>
</tr>
<tr>
<td>rdfs:label</td>
<td>Trade dynamics in the market for federal funds (xsd:string)</td>
</tr>
<tr>
<td>dc:language</td>
<td>en (xsd:string)</td>
</tr>
<tr>
<td>foaf:maker</td>
<td><a href="http://linkeddata.econstor.eu/beta/resource/authors/9133509">http://linkeddata.econstor.eu/beta/resource/authors/9133509</a></td>
</tr>
<tr>
<td>foaf:maker</td>
<td><a href="http://linkeddata.econstor.eu/beta/resource/authors/9133510">http://linkeddata.econstor.eu/beta/resource/authors/9133510</a></td>
</tr>
<tr>
<td>foaf:page</td>
<td><a href="http://hdl.handle.net/10419/69774">http://hdl.handle.net/10419/69774</a></td>
</tr>
<tr>
<td>dc:publisher</td>
<td>Federal Reserve Bank of New York (xsd:string)</td>
</tr>
<tr>
<td>dc:subject</td>
<td><a href="http://zbw.eu/beta/external_identifiers/jel/about#G78">http://zbw.eu/beta/external_identifiers/jel/about#G78</a></td>
</tr>
<tr>
<td>dc:subject</td>
<td><a href="http://zbw.eu/beta/external_identifiers/jel/about#G83">http://zbw.eu/beta/external_identifiers/jel/about#G83</a></td>
</tr>
<tr>
<td>dc:subject</td>
<td><a href="http://zbw.eu/beta/external_identifiers/jel/about#44">http://zbw.eu/beta/external_identifiers/jel/about#44</a></td>
</tr>
<tr>
<td>dc:subject</td>
<td><a href="http://zbw.eu/beta/external_identifiers/jel/about#G1">http://zbw.eu/beta/external_identifiers/jel/about#G1</a></td>
</tr>
<tr>
<td>dc:title</td>
<td>Trade dynamics in the market for federal funds (xsd:string)</td>
</tr>
<tr>
<td>dc:type</td>
<td>Working Paper (xsd:string)</td>
</tr>
<tr>
<td>rdf:type</td>
<td>ssw:Paper</td>
</tr>
<tr>
<td>rdf:type</td>
<td>sioc:Item</td>
</tr>
<tr>
<td>rdf:type</td>
<td>foaf:Document</td>
</tr>
</tbody>
</table>
Wrap-up: LOD and the issue of interoperability of repositories

- OAI-PMH has proven successful for exchanging data mainly between repositories and service providers supporting the protocol.
- Interoperability in terms of SemWeb and LOD means: at least from the point of view of infrastructure, it should not matter into which repository a publication is uploaded (albeit dedicated repositories are still important for branding, Corporate Identity and acquisition).
- Current repository packages still lack of sufficient support for RDF publishing – one way would be developing and maintaining RDF(a) core packages.
- Lightweight services built upon LOD (e.g. for classifications or persons) can already foster interoperability on a semantical level.
Timo Borst
t.borst@zbw.eu