
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 OAIster, 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

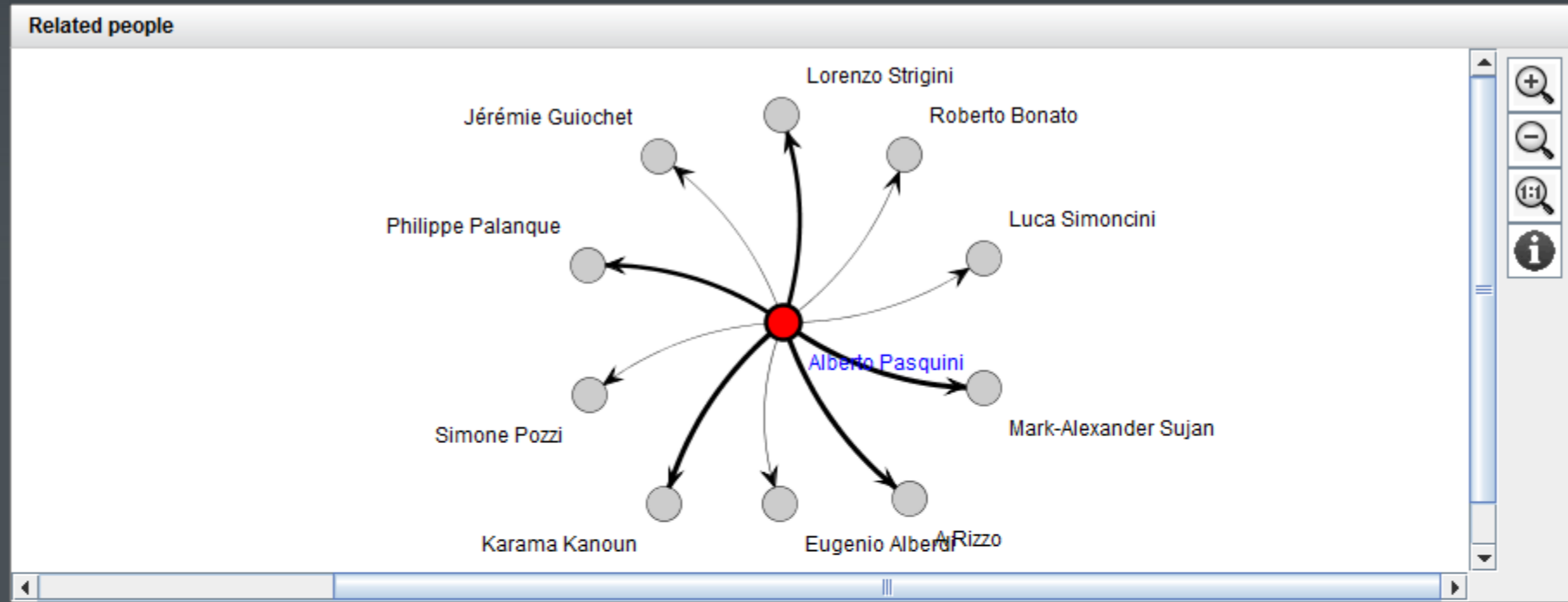
Semantic Web principles and techniques : Brave new world for data interoperability?

- 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



Details

Name
[Alberto Pasquini](#)
[A PASQUINI](#)
[A Pasquini](#)

Address
[DeepBlue, Piazza Buenos Aires 20, 00198, Roma, Italy](#)

Tel
[+39-6-30486189](#)
[+63-04-86189](#)
[+39 06 4544 3773](#)

Email
alberto.pasquini@dblue.it

Raw Linked Data:
 Resolvable URI

- People** ▾
- [Mark-Alexander Sujan](#) ?
 - [A Rizzo](#) ?
 - [Karama Kanoun](#) ?
 - [Lorenzo Strigini](#) ?
 - [Philippe Palanque](#) ?
 - [Roberto Bonato](#) ?
 - [Eugenio Alberdi](#) ?
 - [Jérémie Guiochet](#) ?
 - [Luca Simoncini](#) ?

- Organisations** ▾
- [Deep Blue Srl](#) ?
 - [Deep Blue](#) ?
 - [ENTE PER LE NUOVE TECNOLOGIE, L'ENERGIA E L'AMBIENTE'](#) ?

- Publications** ▾
- [Safety assessment of experimental Air Traffic management procedures](#) ?
 - [Contradictions and critical issues during system evolution](#) ?
 - [Analysis of Incidents Involving Interactive Systems](#) ?
 - [Dependability Evaluation](#) ?
 - [Quantitative Evaluation and Operative Usage of Interactive Systems](#) ?
 - [Dependability Evaluation: Model and Method Based on Activity Theory.](#) ?
 - [Computer Safety, Reliability and](#) ?

Courses & Materials ▾

No results found

SPARQL query:

```
select ?x ?y ?w
where {?x <http://core.kmi.open.ac.uk/data/hasOAIIdentifier>
"oai:eprints.bbk.ac.uk.oai2:128".
    ?x <http://purl.org/ontology/similarity/element> ?z.
    ?y <http://purl.org/ontology/similarity/element> ?z.
    ?z <http://purl.org/ontology/similarity/weight> ?w.
    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

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

Author(s)

Dovern, Jonas 132157322

REMOVE THIS ENTRY

Jannsen, Nils 133933768

ADD MORE

i 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

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

STW-Terms

financial cr

ADD MORE

STW-Terms

Financial crisis

conomic Literature Classification System).


i Please enter the JEL classification codes that best describe

JEL-Classification


> CLOSE

ADD MORE


Using LOD for indexing (persons)

 Please enter the publication title.

Title

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

Author(s) [ADD MORE](#)


 Please enter an abstract of the

Abstract


Author names

[Siebert, Horst](#), PND: 120273152

Siebert, Horst (1938 - 2009; Wirtschaftswissenschaftler; Dt. Wirtschaftswissenschaftler. Hein-Nixdorf-Professor für Europäische Integration und Wirtschaftspolitik an der Johns Hopkins Univ., weiter tätig in Neuwied, Münster/Westfalen, Mannheim, Konstanz, Kiel. Professor i.R. für Volkswirtschaftslehre, internationale Wirtschaftsbeziehungen, Umwelt und natürliche Ressourcen; bis 2003 Präsident des Instituts für Weltwirtschaft (IfW), Professor an der Christian-Albrechts Universität zu Kiel, Lehrstuhl für Theoretische Volkswirtschaftslehre bis 2003; Institut für Weltwirtschaft <Kiel>)

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

Author(s) [ADD MORE](#)

 Please enter an abstract of the

Abstract

Author names

[Siebert, Horst](#), PND: 120273152

[Siebert, Horst](#), PND: 120272814

[Siebert, Horst \(1939 - ; Pädagoge; Professor für Erwachsenenbildung\)](#)

Using LOD for retrieving (search term expansion)

[EconStor](#) >

Suchergebnisse

Suchen:

EconStor gesamt

nach

EconStor-LOD: Publishing a repository's content as Linked Open Data

ECONSTOR

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

EconStor Open Access Database for Economics

Running at <http://linkeddata.econstor.eu/beta/>

Home | Example data: [Authors](#) [Collections](#) [Community](#) [Publications](#)

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](#).

Provided by:  Leibniz-Informationszentrum
Wirtschaft
Leibniz Information Centre
for Economics

Generated by [D2R Server](#)

Trade dynamics in the market for federal funds

Resource URI: <http://linkeddata.econstor.eu/beta/resource/publications/60831>

[Home](#) | [Example Publications](#)

Property	Value
dcterms:abstract	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 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's ability to achieve an efficient reallocation of funds? We also use the model for theoretical and quantitative analyses of policy issues facing the market.
dc:creator	< http://linkeddata.econstor.eu/beta/resource/authors/9133509 >
dc:creator	< http://linkeddata.econstor.eu/beta/resource/authors/9133510 >
dcterms:isPartOf	< http://linkeddata.econstor.eu/beta/resource/collections/493 >
dc:issued	2012 (xsd:gYear)
dc:keyword	bargaining (xsd:string)
dc:keyword	fed funds market (xsd:string)
dc:keyword	over-the-counter (xsd:string)
dc:keyword	search (xsd:string)
rdfs:label	Trade dynamics in the market for federal funds (xsd:string)
dc:language	en (xsd:string)
foaf:maker	< http://linkeddata.econstor.eu/beta/resource/authors/9133509 >
foaf:maker	< http://linkeddata.econstor.eu/beta/resource/authors/9133510 >
foaf:page	< http://hdl.handle.net/10419/60774 >
dc:publisher	Federal Reserve Bank of New York (xsd:string)
dc:subject	< http://zbw.eu/beta/external_identifiers/jel/about#C78 >
dc:subject	< http://zbw.eu/beta/external_identifiers/jel/about#D83 >
dc:subject	< http://zbw.eu/beta/external_identifiers/jel/about#E44 >
dc:subject	< http://zbw.eu/beta/external_identifiers/jel/about#G1 >
dc:title	Trade dynamics in the market for federal funds (xsd:string)
dc:type	Working Paper (xsd:string)
rdf:type	swc:Paper
rdf:type	sioc:Item
rdf:type	foaf:Document

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