Latif, Atif; Borst, Timo; Tochtermann, Klaus

Article

Exposing data from an Open Access repository for Economics as Linked Data

D-Lib Magazine: The Magazine of Digital Library Research

Suggested Citation: Latif, Atif; Borst, Timo; Tochtermann, Klaus (2014) : Exposing data from an Open Access repository for Economics as Linked Data, D-Lib Magazine: The Magazine of Digital Library Research, ISSN 1082-9873, Corporation for National Research Initiatives (CNRI), Reston, VA, Vol. 20, Iss. 9/10,
https://doi.org/10.1045/september2014-latif

This Version is available at:
http://hdl.handle.net/11108/146

D-Lib Magazine

September/October 2014
Volume 20, Number 9/10

Exposing Data From an Open Access Repository for Economics As Linked Data

Atif Latif, Timo Borst and Klaus Tochtermann
ZBW — German National Library of Economics
{a.latif, t.borst, k.tochtermann}@zbw.eu

doi:10.1045/september2014-latif

Abstract

This article describes an approach to publishing metadata on the Semantic Web from an Open Access repository to foster interoperability with distributed data. The article describes two significant movements which led to the application of Semantic Web principles to repository content: the development of Open Access repositories and software systems implementing them, and the Semantic Web movement, especially within libraries and related institutions. The article provides both conceptual and practical insight into the process of converting a legacy relational dataset to machine understandable semantic statements, a.k.a. 'triplification', and gives an overview of current software frameworks that can fulfill this goal. One important outcome of this effort and proof-of-concept is that a repository's content can be straightforwardly published as Linked Open Data. Another result is the ability to link to valuable external datasets, enabling a repository's data to become more contextualized and 'meaningful'. Finally, SPARQL querying shows the potential of triplifications to surpass the querying and use of local data. From both practical and software engineering points of view, the article describes an approach towards publishing a repository's content as Linked Open Data, which is of interest to librarians, repository managers and software developers who work for libraries.

1. Introduction and Motivation

In the last few years, repositories as systems for collecting, publishing, disseminating and archiving digital scientific content have become one of the most prominent types of digital library applications. Especially with respect to Open Access publishing, repositories nowadays serve as platforms for acquiring and disseminating scientific content, which before had been almost exclusively released by commercial publishers. But while the Open Access movement fostered the distribution of out-of-the-box repositories (or the other way round), it reproduced the traditional paradigm of paper-based scholarly communication and thereby fails to benefit from the effects of publishing content on the web as 'open'. Therefore, this traditional paradigm can be described this way: Scientific output in the shape of files is uploaded into a repository, described by means of bibliographic metadata, and crawled by popular search engines. While this bibliographic metadata is generally ignored by search engines (and thus, from the point of view of information fulfillment, redundant if not irrelevant) it is still vital for establishing data interoperability by means of exposing and harvesting data via OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting). The development of an international repository infrastructure and 'ecology' has yielded an impressive yet marginal proliferation of Open Access scientific work (in light of the overall scientific output). Both the Semantic Web and the Open Data movement claim to go beyond this by introducing the concept of machine readable and reusable data.

While the Open Data protagonists want (research) data to be exploited for the specific purpose of (re-)producing scientific results, the Semantic Web community wants this data to be reused for principally any purpose, and the most prominent one probably is understanding the contextualizing and interlinking information. Both agree that the art of open publishing must support not only free access to documents, but machine operations with data on the web. We agreed with these two opinions, and planned to open up our Open Access repository EconStor as Linked Data (machine readable) by publishing the metadata of scientific working papers into the Semantic Web and therefore supporting the publishing process and dissemination of current research results in Economics.

Below are the envisioned goals of opening up EconStor Open Access repository as Linked Open Data:

- Publishing scientific working papers into the Semantic Web and thereby supporting the publishing process and dissemination of current research results in Economics,
- Opening up content from typical repository systems like DSpace for the Semantic Web and integrating it into the mainstream of Linked Data by data publishing,
- Promoting Open Access in the context of other Linking Open Data (LOD) projects (to boost OA and appropriate licensing by

reusing the data in connection with other relevant LOD) thereby promoting the OA approach as such,

- Suggesting interlinking of (standard) elements like author or topic to other datasets,
- Increasing the visibility of datasets in Economics in the Linked Data cloud as well as disseminating contributions.

To achieve these goals, we do not necessarily have to dissolve a scientific paper into a set of triples modeling scientific assertions, but we do commit ourselves to bibliographic metadata as a source for connecting scientific output to real world entities like persons, organizations and subjects. In this sense, we expect repository data to be connected with external data, hence becoming 'assimilated into the web' (Lagoze *et al.*, 2005) and interoperable especially with data outside from repositories. In this work, we outline our guidelines, approach and techniques to publish our repository's metadata on the web, while at the same time connecting and interlinking it with external data from different sources.

This article is structured as follows: Section 2 introduces Linked Open Data basic concepts and highlights its potential with respect to legacy library data management tools. Section 3 provides details about both the EconStor Open Access repository infrastructure and the DSpace framework. Section 4 discusses the state of the art about conversion of a relational database to machine readable, 'RDFzied' data. Section 5, 6 and 7 explain the process of triplifying and present the design of the study for publishing and interlinking, followed by a discussion of benefits for querying LOD. Conclusions and future works are discussed in Section 8.

---

2. Linked Open Data and Digital Library Uptake

The recent paradigm shift to openness has seen new levels of acceptance of openness on a global scale. This trend has been observed in many types of public data, e.g. government data, financial data, heritage data, research data and scholarly communication data (Open Access). This shift to openness is working as a catalyst to the success and acceptability of Linked Open Data projects worldwide. The 'Linked Open Data' effort (Bizer *et al.*, 2007) was conceived in 2007 as a community project to disengage practices of wall-gardening data and motivating people to publish their datasets as open structured data. It's an effort for creating a global data space (Heath & Bizer, 2011), where related information is better connected. This will make information more reusable and discoverable; as well as lead toward unrestrictive data usage for better data interlinking, querying and intelligent application development. It is based on the four Linked Data principles stated by Tim Berners-Lee (Berners-Lee, 2006) which are:

1. Use URIs as names for things.
2. Use HTTP URIs so that people (and machines) can look up those names (see also (Sauermann *et. al.*, 2008)).
3. When someone looks up a URI, provide useful information.
4. Incorporate other URI's so that more data can be exposed.

Basically, these rules provide a set of guidelines for publishing Linked Data. Firstly, this emphasizes identifying real and abstract concepts within the datasets and then assigning identified resources unique URIs in Resource Description Framework (RDF) (RDF, 2004) format which are further dereferenceable to present more meaningful information. One of the benefits of this new paradigm is that linked data applications can use any number of datasets for searching and can be connected with them at run time. As of January 2014, it was estimated there were more than 925 data sets consisting of over 60 billion RDF triples which are interlinked by around 673 million RDF links in the Linked Data cloud. (See http://stats.lod2.eu/).

Also in the last few years, the Digital Library community has observed a rise in interest in use of semantic technologies and Linked Data in the context of traditional bibliographic tasks. There have been several initiatives for introducing new flexible data models in place of old data models to incorporate more features: e.g., the Library of Congress initiative to introduce a new Bibliographic Framework (Library of Congress, 2011), as well as the recommendation of W3C's Library Linked Data Incubator Group (Baker *et al.*, 2011), have successfully highlighted the benefits in this paradigm for the library community. The recent uptake of Linked Data by various national libraries to serve their bibliographic metadata as Linked Data (Biblioteca Nacional De España, 2012) (British Library, 2012) (Malmsten, 2008) (Deutsche National Bibliothek, 2012) (German National Library of Economics, 2013) are showcasing the level of awareness of the library community with respect to recent advances in the Linked Data movement. The ever increasing interest of local and international libraries participating in the recently concluded and successful Semantic Web in Libraries SWIB'13 conference series is a clear indicator of the acceptability and penetration of Linked Data technologies in the library world.

Currently, there are many stable institutional repository software packages and frameworks such as DSpace, Eprints, OPUS, Digital Commons and Fedora Commons, which provide faultless managing of bibliographic databases. This is mainly due to the evolution of these systems during the last decade as a result of rich experiences and expertise which librarians as well as software developers shared over this period of time. Now, these systems provide highly accurate results and seamlessly fulfill the 'Open Access repositories' purpose. In this situation, one can argue why is there a need to shift from traditional repository systems to new technologies? On one hand, this hesitation makes sense, while on the other hand it is also very important to keep up with the ever-changing paradigm of knowledge and data description. Hence adopting new technologies is an important resource for libraries, letting them becoming more competitive by providing better solutions to ever-changing user needs. Importantly, LOD with its flexible RDF data offers some significant advantages over old bibliographic standards which have been in use by the bibliographic community for decades. Some of these important advantages are:

- Better conceptualization and diversity in bibliographic terms, taxonomies and vocabularies by the use of state of the art ontologies, with every concept dereferenced via a unique identifier (URI). This practice will lead to better representation within the system, discoverability, linking and reusability.
- Better management, processing and visualization of content by using an already available wide range of semantic tools.
- Integration of bibliographic items residing internally or externally in the repository will be easier, as all concepts are modeled with a single Resource Description Framework (RDF), bringing consistency to structured data representation and leading to interoperability at various levels. The same applies to the external bibliographic datasets, which are also modeled with the same RDF framework, hence promoting more interoperability as compared to the OAI-PMH standard.
- Better query capacity to answer simple and complex requests due to the graph model implemented by RDF. Queries of different interconnected datasets can lead to the discovery of hidden patterns and relationships.
- More and better granularity of metadata, since every metadata field is now expressed and referenced as a web resource.

These benefits of LOD do compensate for the obvious shortcoming of having to invest labor in exploring additional resources for creating, managing and migrating data from one traditional paradigm to another (LOD). Taken as an indicator, many prominent digital libraries have recently opened their bibliographic repositories as Linked Data (for example, The Open University, datahub and AIMS (Agricultural Information Management Standards)).

In the next section we introduce EconStor as an Open Access repository and its bibliographic framework.

---

3. EconStor — Open Access Repository of Economics

EconStor is the Open Access server of the German National Library of Economics — Leibniz Information Centre for Economics. EconStor provides a platform for publishing grey literature in Economics. Any research institution willing to publish and permanently preserve relevant publications in Economics under the terms of Open Access can participate on the basis of a license agreement. EconStor offers a full depositing service for working paper series and e-journal articles. This includes the uploading of PDF files as well as the processing of all corresponding title data such as abstracts, keywords and JEL codes. EconStor currently provides access to working papers from approximately 100 institutions as well as full text access to more than 5 million papers. Among these are the six top-rated German economic research institutes, including Kiel Institute for the World Economy and DIW Berlin, and the Deutsche Bundesbank.

The following types of documents are published in EconStor:

- Working papers
- Articles from journals and edited volumes (as post prints)
- Conference proceedings
- Dissertations, habilitations and theses
- Research reports and expert opinions
- Books and Festschriften (celebratory (piece of) writing)
- Complete issues of journals

The repository is based on a DSpace system and maintained by a team of domain experts, librarians and software engineers. The DSpace data model provides a basis on which bibliographic metadata can be stored. As depicted in Figure 1 below, the EconStor Open Access repository follows the conventional DSpace hierarchical scheme, i.e. Communities include a set of collections, which in turn contain a set of Publications (=items). A Community is the highest level of hierarchy in repository content, which can be an institute, a research center, a research department or a lab. A Community can also have Sub-communities, and inside every Community there can be more than one Collection. A Collection is a sub-hierarchy of Community which contains information about an item, i.e. a publication and its author. Interestingly, a Collection can belong to one or more Communities, also known as a Shared Collection, which is normally formed when more than one Community contributed. Similarly, one Publication can also belong to more than one Collection.

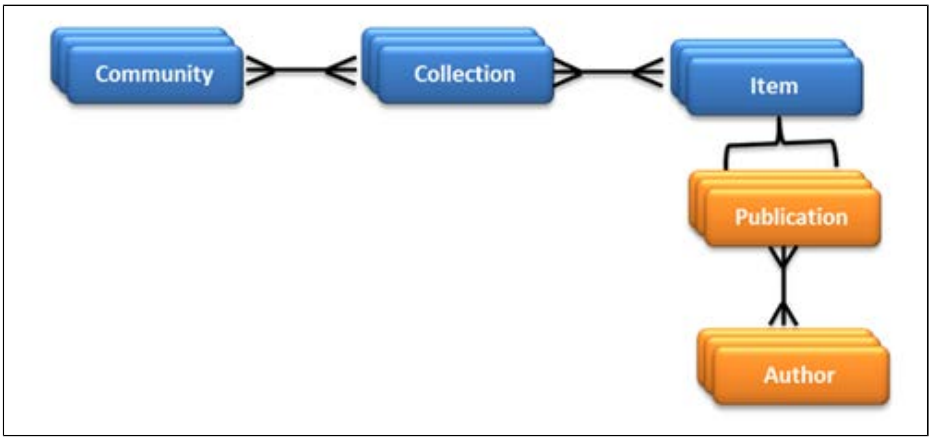

Figure 1: DSpace hierarchical scheme in EconStor

With the maturity of Linked Open Data projects, it must be granted that there are many approaches in use for opening data as Linked Data, and there is no set of global or fixed approaches for Linked Data production (Villazon-Terrazas *et al.*, 2012). One has to adopt certain methodologies and techniques for getting a certain kind of results under certain conditions. This article will describe the methodology we used for exposing the bibliographic content of EconStor.

4. Related Work

A lot of work has been invested in the publishing of relational data as Linked Data, and several approaches have been tested and mentioned in various research studies (Sahoo *et al.*, 2009) (Latif *et al.*, 2013) (Konstantinou *et al.*, 2008). At present, the bulk of data available as Linked Data in the Linked Data cloud is either extracted or replicated from relational databases. The main objective is to describe the relational database content with the help of RDF schema and ontologies in a way that can provide SPARQL query capabilities and further on, can respond with results originating from the relational database. After surveying the related literature, we found that three categories of systems exist which use different approaches in converting Linked Data from a relational database while keeping the database structure intact and accessible to other software operations. In two systems, a mapping file is mandatory which works as a connector for obtaining RDF triples from the data originally residing in the relational database, while the third system natively operates on a Semantic Web triple store where data is replicated into a dedicated RDF triple store. Details about these systems are given below.

The first category of systems work on synchronous transformation of a relational database to RDF which implies real time transformation of queries over the result. The D2RQ platform (Bizer & Seaborne, 2004) is one of the most popular systems. D2RQ provides its own mapping language (Cyganiak *et al.*, 2012), a declarative language for describing the relationship between ontology and a relational data model. With the help of a mapping file it creates a virtual and read-only RDF graph over the relational database. In the backend, the D2R server and D2RQ engine first translate all the mapping definitions provided in the mapping file to SQL for producing and exposing RDF triples over relational database content, which leads to SPARQL query and browsing capabilities in HTML. Additionally, it transforms the SPARQL queries back to SQL for producing results from the relational database. The Open Link Virtuoso Universal server (Erling & Mikhailov, 2007) is another system providing RDF data over relational database content. It also makes use of a declarative mapping file which comprises quad map patterns that specify how the column values are mapped to RDF triples.

The second category also makes use of a mapping file, but works with snapshots and views over relational database contents which are further exposed as an RDF Graph. Triplify (Auer *et al.*, 2009) is one example. It works on the creation of a mapping file, where the important SQL queries that fetch information from the underlying relational database are defined. This mapping file generates a database view which is then used by Triplify to convert the view into RDF. A recent study (Konstantinou *et al.*, 2013) presented a similar approach which claimed to perform much faster in producing RDF data from relational database contents when compared to real-time SPARQL-to-SQL translators.

The third category works natively on Semantic Web triple stores, and data can be migrated from a relational database by use of R2RML (Das *et al.*, 2012). These systems only provide the storage and retrieval features for triples and do not support any mapping definitions. Some of the tools which are used for this purpose are 4store, Arc2, Jena (Carroll *et al.*, 2004) and Sesame framework. All these tools work in the same basic way, but differ in terms of performance, scalability and stability features.

5. System Design

Currently, the EconStor Open Access repository serves more than 70,000 full text documents along with their bibliographic metadata. As a basic consideration, it was crucial to retain the features of the current operational repository, e.g. download count, citations analysis and easy-to-use paper submission function. Therefore it was decided to create a separate system using semantic technologies

alongside of the operational repository for better sustainment of Linked Data. This approach gave us both the time and the freedom to test and evaluate the impact of newly incorporated technologies.

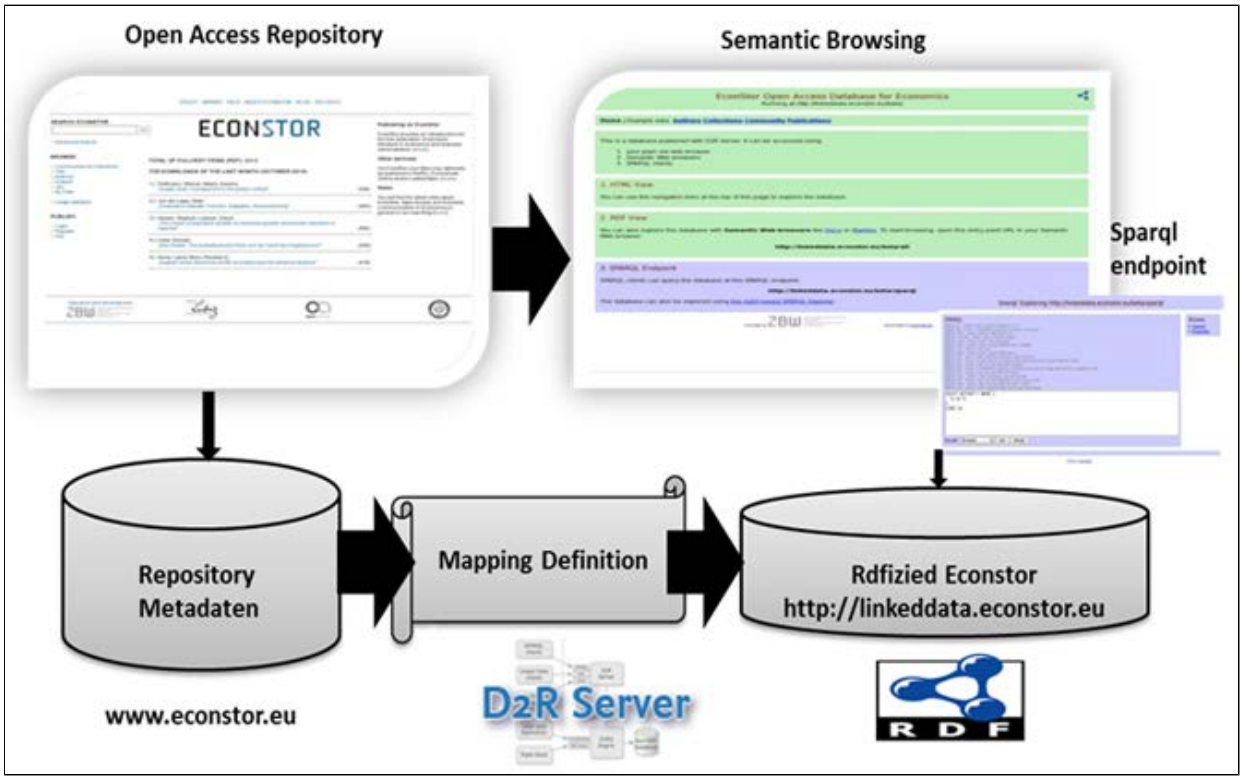

Figure 2: System Architecture

As illustrated in Figure 2, Open Access repository metadata in the form of a relational database was acquired and then transformed into Linked Data by using the mapping definition of D2R Server. The interface is accessible at http://linkeddata.econstor.eu. Two services are provided for faceted browsing and querying via SPARQL. To maintain the data currency and to include new bibliographic items which appeared after triplification of data, a regular update is planned.

5.1 Data Modeling and Assignment of URIs

The first step in the modeling process was to identify the important resources along with their structure and properties. Complying with the DSpace hierarchical structure, we selected Communities, Collections and Items (Publications and Authors) as the main resources to be exposed as Linked Data.

The second step in the modeling was to look for vocabularies as alternative naming for the defined resources fields. Therein lies a certain challenge in exposing these fields by choosing appropriate vocabulary concepts to describe them. The general recommendation by the LOD community to reuse already available ontologies is well backed by the major work conducted by LOD, which has led to the creation of widespread and diversified vocabularies and concepts which cover all of the concepts described here. So rather than create a new vocabulary, we reused established ones for better interoperability and discovery of our content. Some of the popular vocabularies used for modeling EconStor repository content are shown in the graph structure of Figure 3.

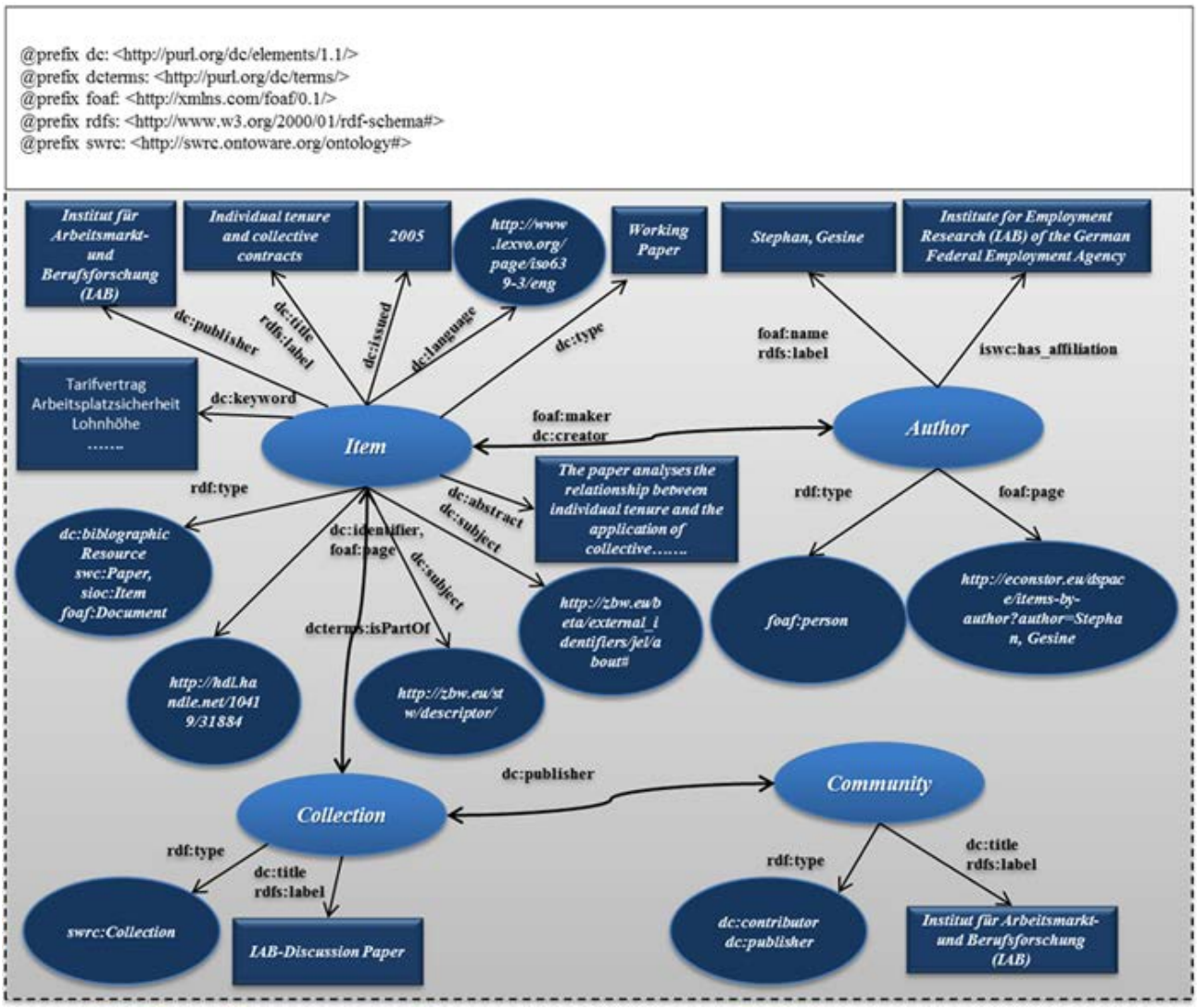

Figure 3: Modeling (See larger version of Figure 3)

The core vocabulary is based on the Dublin Core metadata scheme which is primarily dedicated to describing digital objects (Nilsson *et al*., 2008) and already an integral part of the DSpace legacy metadata. Additionally, Friend of a Friend (FOAF), SWRC (Semantic Web for Research Communities) and Semantic Web Conference Ontology (SWC) were also used. While producing the mapping file, an emphasis was laid on adding rich descriptions for the concepts to achieve better interoperability between ontologies which will lead to better visibility and discovery of related data.

The third step after vocabulary concept assignment was to design a unique and persistent URI for dereferencing the resources. It is also very important to design URIs in a way that keeps data provenance information intact. For the design of URIs we chose the unique handle ID scheme which is commonly used for identifying items in a DSpace system. The subdomain of the EconStor namespace '#econstor.eu' was selected to preserve information about the origin and owner of the resources. Semantified resources in the repository are expressed and addressed according to the following URI pattern:

http://linkeddata.econstor.eu/beta/(resource or page)/(resource type)/handle id

'resource type' comprises four resource types (community, collection, publication and author), which are identified by their handle id. The '(resource or page)' is subject to a content negotiation process, meaning that only information resources like pages are rendered as HTML, whereas any other resources are redirected to get their RDF description. For example, this URI

http://linkeddata.econstor.eu/beta/resource/authors/9085848

identifies and describes author 'Dorothea Kübler' whereas this URI

http://linkeddata.econstor.eu/beta/resource/publications/24800

describes her publication. A similar pattern is provided for communities and collection.

5.2 Interlinking with External Resources

Because interlinking with external linked datasets is a crucial step in publishing Linked Data, interlinking with the following three external datasets was realized in this study.

1. *The STW — Thesaurus for Economics*. This thesaurus provides vocabulary on subjects in Economics and is maintained at the ZBW. The approximately 6,000 standardized subject headings and 18,000 entry terms support individual keywords and also work as a gateway to the technical terms used in law, sociology and politics, as well as geographic names. The STW is already part of the LOD Cloud (Neubert, 2009) (Borst & Neubert, 2009) and has been actively linked to from other vocabularies such as Agrovoc and TheSoz. Keywords associated with publications which are stored in our published repository as the dc:subject field have been mapped to their corresponding STW linked data descriptor, e.g. <http://zbw.eu/stw/descriptor/20018-4>.
2. *Lexvo*. This external link connects the information stored in our repository in the dc:language field with Lexvo, which is part of the LOD Cloud and serves other resources like human languages, words, characters, and other human language-related entities, as URIs for interlinking.
3. *Journal of Economic Literature (JEL) Classification System*. This dataset contains the Journal of Economic Literature (JEL) Classification System, which is created and maintained by the American Economic Association. The JEL dataset is mirrored at the ZBW to mint preliminary identifiers for Semantic Web applications and to publish the translations from the original (English) version to German, French and Spanish, which were created by André Davids, K.U. Leuven. Information stored in our repository about JEL classification as dc:subject is linked to the JEL Classification for Linked Open Data resource URI.

After completing these steps, a D2RQ mapping file was created to establish a connection between the EconStor relational database schema and RDF vocabularies. Consider for instance the record in Table 1 showing the transformation result produced by the mapping file. The data in the table on the left side is transformed into RDF and presented on the right side.

Item Field	**Values**	**Resulting RDF Snippet**
		<http://linkeddata.econstor.eu/beta/resource/publications/20165> a dcterms:BibliographicResource, foaf:Document, swc:Paper, sioc:Item; rdfs:label "Do You Need a Job to Find a Job?";
Title:	Do You Need a Job to Find a Job?	dc:title "Do You Need a Job to Find a Job?";
Authors:	Cobb-Clark, Deborah A.; Frijters, Paul; Kalb, Guyonne	dc:creator, foaf:maker <http://linkeddata.econstor.eu/beta/resource/authors/9082740>, <http://linkeddata.econstor.eu/beta/resource/authors/9082739>, <http://linkeddata.econstor.eu/beta/resource/authors/9082738>;
Identifier:	http://hdl.handle.net/10419/20464	dcterms:identifier "http://hdl.handle.net/10419/20464";
Issue Date:	2004	dc:issued "2004";
Series/Report no.:	IZA Discussion paper series 1211	dcterms:isPartOf <http://linkeddata.econstor.eu/beta/resource/collections/25>;
Language:	en	dc:language <http://www.lexvo.org/page/iso639-3/eng>;
Abstract:	"This paper investigates whether job offers arrive more frequently for those in employment than for those in unemployment. To this end, we take..."	dcterms:abstract "This paper investigates whether job offers arrive more frequently for those in employment than for those in unemployment. To this end, we take...";
Subjects:	job-offer arrival rates; reservation wages; wage-offer distribution; directed search	dc:subject <http://zbw.eu/stw/descriptor/19037-3>, <http://zbw.eu/stw/descriptor/17935-2>, <http://zbw.eu/stw/descriptor/11266-0>, <http://zbw.eu/stw/descriptor/18140-1>, <http://zbw.eu/stw/descriptor/24623-2>, <http://zbw.eu/stw/descriptor/11299-6>;
JEL:	J64	dc:subject

	C14 C41	<http://zbw.eu/beta/external_identifiers/jel/about#J64>, <http://zbw.eu/beta/external_identifiers/jel/about#C14>, <http://zbw.eu/beta/external_identifiers/jel/about#C41>;
Document Type:	Working Paper	dc:type "Working Paper";
Appears in Collections:	IZA Discussion Papers, Forschungsinstitut zur Zukunft der Arbeit (IZA)	dc:publisher "Forschungsinstitut zur Zukunft der Arbeit (IZA)";

6. SPARQL Query Example

The conversion of the EconStor Open Access repository into Linked Data results in a connected RDF graph. This transformation allows complex requests in the form of SPARQL queries, which had been lacking in traditional system search. For instance, in traditional web-based information systems end users can simply browse through data, i.e. authors, publications, communities and collections, or use the advanced search option for filtering results. Due to the unstructured and untyped relationships within concepts in traditional systems, query options are limited and users have to divide the search task into subtasks for achieving appropriate results. However, by creating new explicit relationships among concepts in this repository, the RDF graph has simplified the query structure. Querying EconStor data's RDF graph using SPARQL queries has significantly reduced complexity when compared to SQL querying. For example, examine the results of the query "Find me all author names of publications which have been published after 2011 and are of document type 'conference paper'". For SQL queries, the result requires more than ten joins, whereas in a SPARQL query the answer can be retrieved by means of six simple query patterns.

```
SELECT ?authorName ?abstract WHERE {
   ?publication rdf:type swc:Paper.
     ?publication dc:type ?type.
    ?publication dc:issued ?date.
?publication dcterms:abstract ?abstract.
    ?publication dc:creator ?author.
     ?author foaf:name ?authorName.
FILTER regex(?type,"Conference Paper").
  FILTER(?date > "2010"^^xsd:gYear).
                  }
```

By producing the RDF graph from relational data, many more possibilities are available for making queries which can help repository managers as well as end users find complex and even hidden relationships in quick succession. One example of such a query would be: "Show me all articles in 'debt theory', which have been published by European research institutes". Normally this information could not be obtained without other databases or integrating the data into the database of the local repository. By linking to the external authority record of a research institution exposed as Linked Open Data, the information could be obtained via querying the geolocation of the institution (see for instance the authority files published by the German National Library). Another example would be: "Show me all articles in 'service sector' or related topics". Normally, such queries are supported outside of a local repository by means of external dictionaries or thesauri, which often lack a programming interface for querying and integrating the data. Published as a linked dataset, a repository's data becomes easier to extend by linking its descriptive keywords to the concept URIs of an externally published thesaurus, e.g. the 'Standard Thesaurus for Economics'.

7. Discussion

Normally, a Linked Dataset is made available to the public just by linking it to the LOD cloud and by providing a machine readable interface. "Public" in this sense is not a human agent, but rather other applications consuming or linking to the data. These applications (including content aggregators like Semantic Search Engines) still have to be built upon the data. Although we do not yet provide a human readable interface, we do provide a proof-of-concept for publishing an OA repository's data into the Semantic Web. One way to build a human readable interface would be by linking author names to other sources for further biographical information (the German National Authority file (GND), or DBPedia), a task which is on our agenda. It is still important, however, to identify and discuss the important benefits of exposing our Open Access repository as Linked Data, which include:

- **Removing Data Silos:** This practice will remove data silos, turning inaccessible and unconnected data into a single connected giant global graph. There have been few connections between distributed data in the field of economics to date. Hence, the publishing of EconStor metadata will act as a stepping stone in introducing quality Linked Data for the field. Moreover, this

practice provides a manageable way to push repositories' content (as it is provided by many libraries and institutions) into the Semantic Web.

- **Semantic Annotation:** The data which is exposed will be semantically annotated in contrast to the former modeling as relational data. Linked data will be self-describable and unambiguous, and it can be used by software clients as well as by humans.
- **Discoverability and Reusability:** Exposing data as RDF will increase the options for its discoverability. With semantic annotation, it will be subject to inbound and outbound linkage. External datasets can link to it, hence increasing discoverability. Additionally, reusability of data will be improved, e.g. third parties can use this published data in their system or can link to it.
- **Visibility and Provenance:** Interlinking by third parties and related libraries from social sciences will bring in more value and provide better visibility for both the data and its providers. This will also increase the visibility of authors and lead to improved dissemination of their publications. Additionally, the description of dc:license and dc:publisher accompanied by data will help to retain the provenance information about the data.
- **Applications Development:** The publishing of library data as Linked Data will create the potential for development of mash up-applications which can curate data from different relevant linked data stores and expose hidden relations between library content and generate knowledge along the linked data chain.

---

8. Conclusion and Future Work

In this article, we presented an approach for generating and publishing Linked Open Data from EconStor, one of the major Economics Open Access repositories. To meet the obligation to keep current repository system services intact while addressing the requirement for better and more sustainable Linked Data, this approach showcased a manageable way to push a repository's content into the Semantic Web alongside its legacy system. The approach to converting relational repository data to RDF was proposed after choosing one of the available RDB2RDF software tools. Keeping in mind the promise of Linked Open Data with respect to discoverability, reusability and interoperability among already available semantic bibliographic databases, well established vocabularies were used for modeling the concepts within the repository. Linking to external linked datasets was performed in order to contribute to the Linked Open Data cloud.

We conclude that exposing data from Open Access repositories will pay off greatly, and will help to bring more value into repositories. We assume that this study will be a step forward in promoting Open Access in the context of LOD. Moreover, it will push other related Open Access repositories to produce their data as Linked Data so that semantic information related to Economics literature can be accessed from a single connected global graph.

In the future we plan to enrich our dataset by interconnecting it with other related social, agricultural and medical repositories, i.e. OpenAgris, TheSoz, ZBMED, etc. We are also planning to query the German National Authority File (DNB) and the Virtual Internet Authority File (VIAF) for related information about EconStor authors to construct a real time profile system (Latif *et al.*, 2010) which will be useful for scholarly communication.

References

[1] Auer, S, Dietzold, S., Lehmann, J., Hellmann, S. and Aumueller, D. (2009), "Triplify — Light-Weight Linked Data Publication from Relational Databases", 18th international conference on World wide web (WWW '09), pp. 621-630.

[2] Baker, T., Bermes, E., Coyle, K., Dunsire, G., Isaac, A., Murray, P., Michael, P., Schneider, J., Singer, R., Summers, E., William, W., Young, J., and Zeng, M. (2011), "Library Linked Data Incubator Group Final Report", W3C Incubator Group Report.

[3] Berners-Lee, T. (2006), "Linked Data — Design Issues".

[4] Biblioteca Nacional De España (2012), "Linked Data at Spanish National Library".

[5] Bizer, C. and Seaborne, A. (2004), "D2RQ — treating non-RDF databases as virtual RDF graphs". 3rd International Semantic Web Conference (ISWC2004)

[6] British Library (2012), "Linked Data at the British Library".

[7] Borst, T. and Neubert, J. (2009), "W3C Case Study: Publishing STW Thesaurus for Economics as Linked Open Data. Case Study".

[8] Carroll, J., Dickinson, I., Dollin, C., Reynolds, D., Seaborne, A. and Wilkinson, K. (2004). "Jena: implementing the Semantic Web recommendations", 13th World Wide Web Conference. http://doi.org/10.1145/1013367.1013381

[9] Cyganiak, R., Bizer, C., Garbers, J., Maresch, O. and Becker, C. (2012), "The D2RQ Mapping Language".

[10] Das, S., Sundara, S. and Cyganiak, R. (2012), "R2RML: RDB to RDF Mapping Language". W3C Recommendation.

[11] Deutsche National Bibliothek (2012) "The Linked Data Service of the German National Library".

[12] Erling, O. and Mikhailov, I. (2007), "RDF support in the Virtuoso DBMS". 1st Conference on Social Semantic Web, pp. 59-68. http://doi.org/10.1007/978-3-642-02184-8_2

[13] German National Library of Economics (2013), "Linked Data EconStor Open Access Database for Economics at German National Library of Economics".

[14] Heath, T. and Bizer, C. (2011), Linked Data: Evolving the Web into a Global Data Space, Morgan & Claypool.

[15] Konstantinou, N., Spanos, D.E. and Mitrou, N. (2008). "Ontology and Database Mapping: A Survey of Current Implementations and Future Directions", *Journal of Web Engineering*, Vol. 7 No. 1, pp. 1-24.

[16] Konstantinou, N., Spanos, D.E. and Mitrou, N. (2013). "Transient and persistent RDF views over relational databases in the context of digital repositories". In *Metadata and Semantics Research (MTSR 2013)*, Thessaloniki, Greece, pp. 342-354.

[17] Lagoze, C., Krafft, D. B., Payette, S., and Jesuroga, S. (2005). "What is a digital library anymore, anyway". *D-Lib Magazine*, Vol. 11 No. 11. http://doi.org/10.1045/november2005-lagoze

[18] Latif, A., Afzal, M . T. and Maurer, H. A. (2013), "Weaving Scholarly Legacy Data into Web of Data", J. UCS, Vol. 18 No. 16, pp. 2301-2318. http://doi.org/10.3217/jucs-018-16-2301

[19] Latif, A., Saeed, A. U., Höfler, P., Tochtermann, K. and Afzal, M. T. (2010), "Harvesting Pertinent Resources from Linked Open Data" JDIM, Vol. 8 No. 3, pp. 205-.

[20] Library of Congress (2011), "A Bibliographic Framework for the Digital Age".

[21] Linking Open Data (2007), "W3C Community Project Linking Open Data".

[22] Malmsten, M. (2008), "Making a Library Catalogue Part of the Semantic Web", International Conference on Dublin Core and Metadata Applications.

[23] Neubert, J. (2009), "Bringing the" thesaurus for economics" on to the web of linked data", WWW Workshop on Linked Data on the Web (LDOW 2009)

[24] Nilsson, M., Powell, A., Johnston, P. and Naeve, A. (2008). "Expressing Dublin Core metadata using the Resource Description Framework (RDF)", DCMI Recommendation.

[25] RDF, "Resource Description Framework".

[26] Sahoo, S., Halb, W., Hellmann, S., Idehen, K., Thibodeau, T., Auer, S., Sequeda, J. and Ezzat, A. (2009). "A Survey of Current Approaches for Mapping of Relational Databases to RDF".

[27] Sauermann, L., Cyganiak, R., Ayers, D. and Völkel, M., "Cool URIs for the Semantic Web", W3C Interest Group Note.

[28] Villazon-Terrazas, B., Vila-Suero, D., Garijo, D., Vilches-Blazquez, L.M., Poveda-Villalon, M., Mora, J., Corcho, O. and Gomez-Perez, A. (2012), "Publishing Linked Data — There is no One-Size-Fits-All Formula". European Data Forum 2012.

About the Authors

Atif Latif received his PhD degree in Computer Science with a focus on Linked Open Data research from Graz University of Technology, Austria in 2011. Dr. Latif was affiliated with the Institute of Knowledge Management and Know-Center, Austria's COMET Competence Center for Knowledge Management. His main research areas are Linked Open Data, Knowledge Discovery and Digital Libraries. Since 2012, he has worked at the Leibniz Information Centre for Economics (ZBW) where he investigates solutions for applying Semantic and Linked Data technologies in digital library settings.

Timo Borst has an academic background in Computer Science and Political Science. For several years, he worked in industrial software projects on electronic publishing and web information systems. Dr. Borst's main research interests are in the field of Information Retrieval, Repositories, Semantic Web and Web Science from the perspective of collaborative software development. Since 2007, he is Head of the department for

Information Systems and Publishing Technologies (IIPT) at Leibniz Information Centre for Economics. He is a lecturer for Digital Libraries at Hamburg University of Applied Sciences (HAW).

Klaus Tochtermann is a professor of Computer Media at the Christian-Albrechts University of Kiel (Germany). In addition to his professorial duties, Professor Tochtermann is the Director of the ZBW — Leibniz Information Centre for Economics. With approximately 270 employees maintaining more than 4.2 million documents related to economics, the ZBW is the world's largest library for economics. From October 2000 to June 2010, he was the director of Austria's first industry-based research institute on knowledge management, Know-Center. He was head of the Institute for Knowledge Management at Graz University of Technology from 2004 to 2010. Klaus Tochtermann studied Computer Science and earned his PhD in 1995. In 1996, he spent his post-doc at Texas A&M University in College Station, Texas.

(This article was edited on September 23, 2014, to add a citation that was erroneously omitted from the References section.)