



3<sup>rd</sup> International Conference on Integrated Information (IC-ININFO)

## Exposing Bibliographic Information as Linked Open Data using Standards-based Mappings: Methodology and Results

Nikolaos Konstantinou<sup>a,\*</sup>, Nikos Houssos<sup>b</sup>, Anastasia Manta<sup>b</sup>

<sup>a</sup>*School of Electrical and Computer Engineering / National Technical University of Athens, Heron Polytechniou 9, 15780, Zografou, Greece*

<sup>b</sup>*National Documentation Centre / National Hellenic Research Foundation, Vas. Konstantinou 48, 11635, Athens, Greece*

---

### Abstract

The Linked Open Data (LOD) movement is constantly gaining worldwide acceptance. In this paper we describe how LOD is generated in the case of digital repositories that contain bibliographic information, adopting international standards. The available options and respective choices are presented and justified while we also provide a technical description, the methodology we followed, the possibilities and difficulties in the way, and the respective benefits and drawbacks. Detailed examples are provided regarding the implementation and query capabilities, and the paper concludes after a discussion over the results and the challenges associated with our approach, and our most important observations and future plans.

© 2013 The Authors. Published by Elsevier Ltd.

Selection and peer-review under responsibility of the 3<sup>rd</sup> International Conference on Integrated Information.

*Keywords:* Linked Open Data; Bibliographic information; Repositories; R2RML; Databases; RDF; SPARQL Endpoint; Ontology; Mapping

---

### 1. Introduction

The Linked Open Data (LOD) movement is gaining more and more acceptance worldwide, as a general shift toward openness is observed in many aspects of public information. Access to government data ([www.data.gov.uk](http://www.data.gov.uk)), financial data ([www.openspending.org](http://www.openspending.org)), news data ([www.guardian.co.uk/data](http://www.guardian.co.uk/data)) or cultural heritage data ([www.europeana.eu](http://www.europeana.eu)), are only some of the examples where openness is nowadays applicable. In the same time, general awareness is raised towards this direction as openness contributes to transparency. Regarding

---

\* Corresponding author. Tel.: +30-210-772-4483; fax: +30-210-772-4488.

*E-mail address:* [nkons@cn.ntua.gr](mailto:nkons@cn.ntua.gr)

publishing, new business models are invented that render this shift profitable for the publishers as well as for the target audience.

As to what contributes to this direction, we can observe that technologies and standards are mature nowadays in the Semantic Web domain, which lies in the base of things. The description languages, including RDF and OWL since long have been official W3C recommendations, query languages such as SPARQL also. The related technologies that serve as a basis for publishing and interlinking information on the web have reached the required maturity level. These building blocks have existed for more than a decade, while the technical knowhow that has developed in the whole Semantic Web ecosystem is nowadays bringing results outside the academia. Thus, the technical groundwork has been laid, in order to enable markets to appear.

As far as the digital libraries domain is concerned, rich experience has led to the creation of software systems that demonstrate flawless performance over maintaining bibliographic archives. These systems operate at a high level of accuracy and fulfill their purpose. One could argue, however, that there is no significant reason for evolution, while it is also uncertain to which direction this evolution should take place. In order, however, to keep up with an ever-changing environment, regarding data and knowledge description, adoption of newer technologies is a necessity. This is firstly in order for the solutions to remain competitive but also, because of the benefits new technologies entail. Regarding digital repositories, the LOD paradigm facilitates:

- *Integration*, typically materialized using OAI-PMH harvesting interfaces. The problem is that these interfaces, while often implemented and used in the digital libraries domain, do not ease integration with data models from other domains of discourse.
- *Expressiveness* in describing the information. OAI-PMH for instance, as far as information description is concerned, allows for a tree structure that extends to a depth-level of two. On the contrary, RDF allows for a graph-based description that can practically cover much more sufficiently the needs for information exchange in the digital libraries domain.
- *Query answering* can be realized using conventional systems, to the extent where standard keyword-based queries are evaluated efficiently and respective results can span several repositories, offering a basic solution. Contrarily, querying graphs using graph patterns allows for much more complex queries, enabling formation of more descriptive queries.

Therefore, we can safely deduce that the benefits (query expressiveness, inherent semantics, and integration with third party sources) outweigh the disadvantages (resources investment in creating and maintaining the data). Because of these, and many other reasons as well, we can observe that more and more libraries open their data. For instance, the National Libraries in Spain (Biblioteca Nacional De Espana, 2012), Germany (Deutsche National Bibliothek, 2012) and Great Britain (British Library, 2012) are exposing free data services according to the LOD paradigm.

Now, as it is commonly understood, Linked Data can originate from a variety of sources through numerous methodologies and therefore there is no global solution for its production (Villazon-Terrazas, Vila-Suero, Garijo, Vilches-Blazquez, Poveda-Villalon, Mora, Corcho, & Gomez-Perez, 2012). Thus, the approach presented in this paper serves as a description of a methodology that was followed in order to achieve certain results under certain conditions. In particular, we present the methodology behind exposing as LOD the bibliographic repository contents maintained at the National Documentation Centre (EKT), part of the National Hellenic Research Foundation (NHRF).

The EKT has rich information to maintain, often several decades old. For instance, its archives are constantly enriched with scanned dissertations from the pre-personal computing era. It is of crucial importance that the systems used to store this information demonstrate excellent ability in preserving information and thus only mature systems tested through time can be considered. The challenge under these circumstances is to offer state-of-the-art services around the bibliographic archives while not deviating from the main *raison-d'être*, which is digital preservation, and all that in a modern and sustainable manner.

This paper is structured as follows: Section 2 presents a number of related implementations that outline the technological environment in which the implementation is called to operate, Section 3 describes the design and implementation of the system and the respective methodology with which it is operated, Section 4 discusses our choices while Section 5 concludes the paper with our most important observations and future plans.

## 2. Related Work

Publishing LOD originating from relational databases is a domain where much work has been conducted and several approaches exist in the bibliography (Sahoo, Halb, Hellmann, Idehen, Thibodeau, Auer, Sequeda & Ezzat, 2009), (Konstantinou, Spanos, & Mitrou, 2008). Typically, the goal is to describe the relational database contents using an ontology in a way that allows queries that are posed to the RDF schema to be answered with data originating from the relational database.

In order to publish RDF while maintaining its relational database origin accessible to other software systems that may require it present for their operation, there are in general two approaches. Both of them have as a prerequisite the presence of a mapping definition that defines the way we can obtain RDF triples from the data that resides in the source database.

The first category of approaches involves synchronous transformations in the queries over the result, from queries over the RDF using SPARQL to queries over the relational database using SPARQL. These systems include D2RQ, a solution introduced in (Bizer & Seaborne, 2004). D2RQ translates the mapping definitions into SQL in order to expose the relational database contents as triples and additionally translates SPARQL queries into SQL. D2RQ evolved its own mapping language (Cyganiak, Bizer, Garbers, Maresch, & Becker, 2012). Essentially, SPARQL queries on the result are translated to nested SQL queries on the database (Cyganiak, 2005). The Virtuoso universal server (Erling & Mikhailov, 2007) is another solution that supports RDF Views over its relational database contents making it possible to publish a virtual RDF graph generated on top of a relational database.

The second category of approaches involves a mapping file as well and dictates that using the mapping file, we can obtain a snapshot of the relational database contents and export it as an RDF graph. These include the authors' previous work, introduced in (Konstantinou, Spanos, Houssos, & Mitrou, 2013), an approach that appears to perform much faster in RDF-izing relational database contents compared to real-time SPARQL-to-SQL translators. This option is being supported by D2RQ as well alongside its main function as an RDF server. Triplify (Auer, Dietzold, Lehmann, Hellmann, & Aumueller, 2009) can also serve towards this goal. The Virtuoso server can also be used to this end, as it allows hosting of multiple RDF graphs, plus offering the means to query and navigate them.

A third possibility to be noted is the one that operates natively on a semantic web triplestore. Tools that enable this approach include the popular Jena framework (Carroll, Dickinson, Dollin, Reynolds, Seaborne, & Wilkinson, 2004), the Sesame framework ([www.openrdf.org](http://www.openrdf.org)), or even native RDF support that can be found in modern RDBMS's such as Oracle. Apache Stanbol is a framework for semantic content management ([stanbol.apache.org](http://stanbol.apache.org)) with a set of tools that extend semantically the standard set of CMS functionality.

Adoption of such approaches enables the implementation of systems relying purely on triplestores such as the museum collection available at [www.clarosnet.org](http://www.clarosnet.org) that uses the CIDOC-CRM vocabulary ([cidoc-crm.org](http://cidoc-crm.org)) to describe cultural heritage concepts and relations among them. The discussion of such approaches is out of the scope of this paper. It remains however, a possibility, after exposing a snapshot of the relational database as an RDF graph to host it and enable the development of a custom application over it.

## 3. System Description

### 3.1. System Architecture

The EKT hosts numerous repositories of contents with nationwide interest. The repositories are being maintained and updated by a team of software engineers, librarians, and domain experts, creating a living organism around this information. In order to enable the provision of LOD in a sustainable manner, first we took into account the fact that existing systems had to remain intact, in order not to discard established practices and knowhow.

The importance of the data, however rigorous backup approaches may be, leaves no room with experimentation. Therefore it has been chosen not to replace existing practices with newer technologies, before running them alongside for a period of time.

As depicted in Figure 1, an RDF quadstore lies at the core of the system that exposes our data via [data.ekt.gr](http://data.ekt.gr). Two basic services over the data are exposed, comprising a SPARQL endpoint and a faceted browser. The LOD repository is operating independently of the sources, which continue uninterruptedly their operation. Currently, the repositories exposed as LOD are (a) [www.phdtheses.ekt.gr](http://www.phdtheses.ekt.gr), a repository containing more than 28,500 dissertations and (b) [www.helios-eie.ekt.gr](http://www.helios-eie.ekt.gr) containing more than 5,500 publications by researchers affiliated with the NHRF.

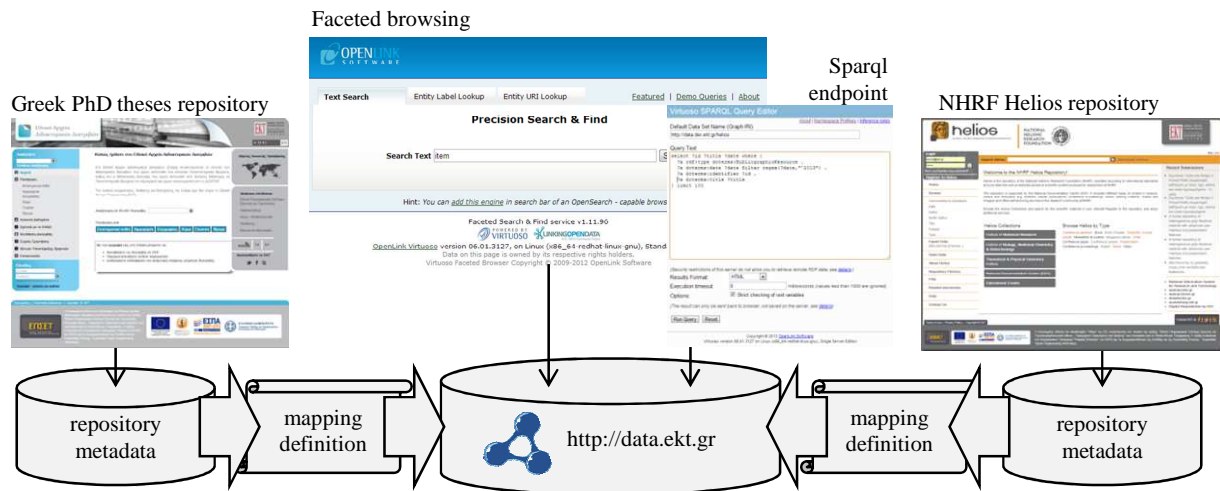


Fig. 1. Architecture overview.

### 3.2. Methodology and Examples

In order to preserve existing practices, the asynchronous approach was adopted: data is exported periodically in RDF from the repositories, according to a mapping definition. There is a mapping file for each repository in order to deal with the distinct features each implementation demonstrates. Then, after the data is exported in RDF, they are imported into the Virtuoso quadstore.

In order to render this methodology sustainable, our approach relies on the recently standardized R2RML language (Das, Sundara, & Cyganiak, 2012) which allows vendor-independent RDF extraction. Based on a mapping file for each repository, and using the tool introduced in (Konstantinou, Spanos, Houssos, & Mitrou, 2013), we were able to extract the database contents in an RDF graph. An example of how the mappings are formed, using R2RML, follows next.

In the example in Table 1, on the left we can see the triples mapping definition and on the right the logical table that retrieves the data from the input database. The items are formed obtaining URI's according to their respective handle id assigned to them by DSpace upon completion of the submission workflow. They are registered as instances of the `dcterms:BibliographicResource` class, and they are assigned a property, `dcterms:abstract`, whose value is the value of the repository metadata field `dc.description.abstract`. In a similar manner, values can be extracted from the DSpace repository and offer full descriptions. Consider for instance the record in Table 2 below. Using the transformation, based on R2RML, the record on the left of the table is transformed into RDF as presented on the right.

The data generated in this manner, are then concentrated in a central quadstore. In order to distinguish among the various datasets, a separate namespace was reserved for the contents of each repository. Note that we chose to not install a triplestore next to each repository, because this would require substantially more effort and resources.

Table 1: Example R2RML Mapping excerpt

Triples Mapping definition
<pre>@prefix map: &lt;#&gt;. @prefix rr: &lt;http://www.w3.org/ns/r2rml#&gt;. @prefix dcterms: &lt;http://purl.org/dc/terms/&gt;. map:items   rr:logicalTable &lt;#item-view&gt;;   rr:subjectMap [     rr:template '&lt;http://data.ekt.gr/helios/item/{ "handle" }&gt;';     rr:class dcterms:BibliographicResource;   ]. map:dc-description-abstract   rr:logicalTable &lt;#dc-description-abstract- view&gt;;   rr:subjectMap [ rr:template '&lt;http://data.ekt.gr/helios/item/{ "handle" }&gt;';   ];   rr:predicateObjectMap [     rr:predicate dcterms:abstract;     rr:objectMap [ rr:column "text_value" ];   ].</pre>

Logical Table View
<pre>&lt;#dc-description-abstract-view&gt; rr:sqlQuery ""   SELECT h.handle AS handle, mv.text_value AS text_value   FROM handle AS h, item AS i, metadatavalue AS mv, metadataschemaregistry AS msr, metadatafieldregistry AS mfr WHERE   i.in_archive=TRUE AND   h.resource_id=i.item_id AND   h.resource_type_id=2 AND   msr.metadata_schema_id=mfr.metadata_schema_id AND   mfr.metadata_field_id=mv.metadata_field_id AND   mv.text_value is not null AND   i.item_id=mv.item_id AND msr.namespace = '&lt;http://dublincore.org/documents/dcmi-terms/' AND   mfr.element='description' AND   mfr.qualifier='abstract' "".</pre>

As far as the vocabularies are concerned, the core vocabulary is based on the Dublin Core, as it is geared towards describing digital objects (Nilsson, Powell, Johnston, & Naeve, 2008). Lexvo, (<http://lexvo.org/>), part of the LOD cloud, serves as an exhaustive set of human languages URIs integrated in a global approach of terminology (De Melo & Weikum, 2008). It was used to map the `dc.language` field and connect our repository to the LOD cloud. A second external link has been realized using the Library of Congress Linked Data Services (<http://id.loc.gov/>). Information stored in our repositories in the `dc.subject` field has been mapped to the Library’s Linked Data Service resource URIs, thus to URIs such as `<http://id.loc.gov/authorities/classification/N.html>`.

Table 2: Example DSpace record

DSpace field	Values
dc.creator	Kollia, Zoe Sarantopoulou, Evangelia Cefalas, Alciviadis Constantinos Kobe, S. Samardzija, Z.
dc.date	2004
dc.format.extent	379-382
dc.identifier.uri	<a href="http://hdl.handle.net/10442/7055">http://hdl.handle.net/10442/7055</a>
dc.language	eng
dc.publisher	Springer
dc.title	Nanometric size control and treatment of historic paper manuscript and prints with laser light at 157 nm
dc.type	Article
dc.subject	Printmaking and Engraving



Resulting RDF snippet in turtle syntax
<pre>&lt;http://data.ekt.gr/helios/item/10442/7055&gt;   a dcterms:BibliographicResource ;   dcterms:creator "Kobe, S." , &lt;http://data.ekt.gr/person/48&gt; , &lt;http://data.ekt.gr/person/14&gt; , "Samardzija, Z." , &lt;http://data.ekt.gr/person/112&gt; ;   dcterms:date "2004" ;   dcterms:extent "379-382" ;   dcterms:identifier "http://hdl.handle.net/10442/7055" ;   dcterms:language &lt;http://www.lexvo.org/page/iso639-3/eng&gt; ;   dcterms:publisher "Springer" ;   dcterms:title "Nanometric size control and treatment of historic paper manuscript and prints with laser light at 157 nm" ;   dcterms:type "Article"   dc:subject &lt;http://id.loc.gov/authorities/classification/NE1- NE978&gt;.</pre>

A second RDF graph was created based on the MADS-based, authority file which the EKT preserves internally for its researchers. The authority file keeps a record for every researcher working in the organization, containing alternative authoritative labels, personal data (email, telephone, homepage) and data concerning the specific department related to the person. This information was published as an RDF graph under the namespace `http://data.ekt.gr/persons/{researcher_id}`, based on the XML authority file and the Bibliotransformation engine transformation tool (Stamatis, Konstantinou, Manta, Paschou, & Houssos, 2012). In this way, the researchers can be uniquely identified, the authority files can be shared and reused, and the NHRF authors can accept inbound links and references, allowing the graph to serve as an island of persistence.

Also, we have to note that the digital repository implementations include, as it is often the case, custom fields that served specific in-house purposes. The challenge in exposing these fields' values lies in choosing the appropriate ontology concepts to describe them, based on existing ones. We tried to avoid introducing “yet another” ontology/vocabulary since the work conducted in the LOD domain has led to the creation of a large number of widespread ontologies and concepts covering to a large extent the majority of the concepts to be described. As a result, custom fields that exist in our implementation, besides the ones belonging to Dublin Core, have yet to be exposed after being mapped to established vocabularies.

#### 4. Discussion

It is important, first, to acknowledge that exposing data as LOD entails several benefits as well as drawbacks. Among the key benefits we can consider the following:

- *Semantic annotation.* The data that is exposed, in contrast to the initial data presentation available to the end user/consumer (i.e. repository browsing) is semantically annotated. This means that the data itself conveys its semantics in the sense that it is unambiguously understood and interpreted. Thus, it can be consumed by software clients as well as by humans, containing at all times the same meaning.
- *Query simplification.* Information representation as a graph, allows queries much more complex than the ones allowed over the current web application user interfaces. For instance, in the repository user interface, end users can simply browse the data (e.g. by author, title, type, etc.) or use the advanced search capabilities that allow slightly more complex queries in comparison to the basic approach. The creation, however, of new relations among concepts of the RDF graph can simplify the queries, since querying the resulting graph using SPARQL will be of significantly reduced complexity compared to querying the database using SQL. For instance, a query over the resulting graph, of the form “Find all item ids and abstracts”, in SPARQL would be similar to the following:

```

SELECT ?id ?abstract
FROM <http://data.ekt.gr/helios>
FROM <http://data.ekt.gr/phdtheses>
WHERE {
  ?a rdf:type dcterms:BibliographicResource .
  ?a dcterms:identifier ?id .
  ?a dcterms:abstract ?abstract }

```

This SPARQL query is much simpler (it contains only 3 triple patterns) than the SQL query that retrieves only the `dc.description.abstract` field values (contains 9 JOIN conditions), as presented in Section 3.2, on the right of Table 1. Thus, once deriving an RDF graph from the data, there are far more possibilities in the ways in which we can pose queries, in unprecedented ways.

- *Increased discoverability.* Exposing the data as RDF graphs opens new possibilities in discovering it. Firstly, the data can be accompanied by a description over their contents (e.g. using the VoID vocabulary). Also, links can be realized towards instances in other parts of the LOD cloud, and also, inbound links are welcome as well, forming an interconnected island in the LOD cloud. Therefore, the dataset can be linked to other datasets, thus increasing its discoverability.
- *Reduced effort* required for schema modifications. Creation of new relations is allowed without altering the database schema. New classes and properties can be defined in the RDF schema that will apply to subsets of the repository data, without the need to modify neither the relational database schema nor its contents.

- *Synthesis* by means of integration, fusion, mashups. The end user/developer can perform searches spanning various repositories, from a single SPARQL endpoint. Also, it is possible to download parts or the whole data in order to combine it to other data and process it according to his/her needs.
- *Inference* is enabled. With reasoning support, new implicit facts can be inferred, based on the existing ones that are explicitly derived from the relational database. These new facts can then be added to the graph, thus augmenting the knowledge base.
- *Reusability*. Third parties can reuse the data in their systems, either by including the information in their datasets, or by providing reference to the published resources.

On the other hand, technical difficulties encountered in the course of things are expected, as they may be caused by a series of factors, as the following:

- *Multidisciplinarity*. As a multidisciplinary task, involving both Computer and Library science, annotation alone is a task not to be underestimated, since it requires contributions from both scientific domains and close collaboration of the implementation team.
- *Technology barrier*. Tools are not as mature yet as to guide the inexperienced user, or warn for possible incorrect uses. For instance, when performing the mapping (i.e. writing the R2RML file), the whole procedure takes place manually, in a plain text editor. There is no way of assuring the validity of the mappings themselves, only after producing the resulting graph. This is, therefore, a step that requires the presence of an expert.
- *Result is prone to errors*. Even after producing the resulting graph, and in the case when it is syntactically correct, there is no validator service to perform an automatic check of whether the concepts and properties involved are used as intended. Errors or bad practices, even after being published can go unnoticed.
- *Concept mismatch*. It is not always possible to extract an RDF description from the stored values in the repository. Although RDF allows in general for more complex descriptions compared to the metadata model, identical mappings may not always be found and compromises have to be made.
- *Exceptions to the general rule*. When curating metadata, it is expected that not all records can be simultaneously updated: Typically, changes made automatically will apply to the majority of the data, while the remaining portion will require manual intervention. The same case applies to the database-to-RDF mapping: most of the fields will be mapped while contents of other fields will have to be subject to programmatic and/or manual corrections. Therefore, one can expect from the mapping file to cover the majority, but not the whole of the metadata fields, which may require post-transformation manual interventions.

It must also be noted that, as far as it concerns the choice between synchronous or asynchronous SPARQL-to-SQL translation in the digital repositories domain, the second option seems more viable. In the digital repositories domain, the choice of exposing data periodically comes at a low cost, because the information in this domain does not change as frequently as it would in other domains such as e.g., in sensor or social network data. Therefore, we did not need to carry the burden of synchronous translation that would, certainly, provide truly up-to-date information but in the same time impose an extra burden, both computational and regarding system sustainability.

## 5. Conclusions and Future Work

In this paper we presented an approach for generating and publishing Linked Data from a large bibliographic dataset. Our approach involved interaction by a multidisciplinary team of professionals, consideration of sustainability among the most important desired features of the result, and a nationwide responsibility. We presented our choices after considering the available options with respect to the software tools, the vocabularies, while keeping in mind that a viable methodology had to be adopted, that would ensure the datasets' high value and, most importantly, their viability.

We can conclude that LOD generation requires some initial resource investment in terms of man-hours which may make the whole attempt seem unprofitable. Nonetheless, the investment pays off in numerous ways, confirming that "a little semantics goes a long way" for many reasons as discussed in the previous Section.

Future plans include enriching the data, utilizing more constructs and defining new relationships. This can take place by offering alternatives to the already published metadata fields. For instance, exposing author names

publication as resources to the property `dcterms:creator` is an alternative to exposing them as resources to the property `mrel:cre`.

Among our intentions is also to generate denser links to other datasets. This can be materialized using the Silk framework (Volz, Bizer, Gaedke, & Kobilarov, 2009) and with special care to the librarian science concerns.

Apart from the bibliographic datasets that are hosted currently, our future plans also include publishing more datasets for which the EKT is responsible at a national level. These datasets include Greek researcher authority files, publications coauthored by Greek researchers, universities, research institutions and respective administrative information and indexes, etc. This could lead to the creation of a single point of reference with reusable information that in addition to its increased value could also save money as future needs by potentially interested parties will be covered.

## Acknowledgments

Part of this work was funded by the project "National Information System for Research & Technology/Social Networks-User Generated Content" – [www.epset.gr](http://www.epset.gr)

## References

- Biblioteca Nacional De Espana. (2012). Linked Data at Spanish National Library. Available at <http://www.bne.es/en/Catalogos/DatosEnlazados/index.html> (accessed January 2013).
- Deutsche National Bibliothek. (2012). The Linked Data Service of the German National Library. Available at [http://www.dnb.de/EN/Service/DigitaleDienste/LinkedData/linkedata\\_node.html](http://www.dnb.de/EN/Service/DigitaleDienste/LinkedData/linkedata_node.html) (accessed January 2013).
- British Library. (2012). Linked Data at the British Library. Available at <http://www.bl.uk/bibliographic/datafree.html> (accessed December 2012).
- Villazon-Terrazas, B., Vila-Suero, D., Garijo, D., Vilches-Blazquez, L.M., Poveda-Villalon, M., Mora, J., Corcho, O., Gomez-Perez, A. (2012). Publishing Linked Data - There is no One-Size-Fits-All Formula. In Proceedings of the European Data Forum 2012.
- Sahoo, S., Halb, W., Hellmann, S., Idehen, K., Thibodeau, T., Auer, S., Sequeda, J., Ezzat, A. (2009). A Survey of Current Approaches for Mapping of Relational Databases to RDF. Available at [http://www.w3.org/2005/Incubator/rdb2rdf/RDB2RDF\\_SurveyReport.pdf](http://www.w3.org/2005/Incubator/rdb2rdf/RDB2RDF_SurveyReport.pdf) (accessed 1 February 2013).
- Konstantinou, N., Spanos, D.E., Mitrou, N. (2008). Ontology and Database Mapping: A Survey of Current Implementations and Future Directions. In *J. Web Engineering*, 7 (1), 1–24.
- Bizer, C., Seaborne, A. (2004). D2RQ—treating non-RDF databases as virtual RDF graphs. In Proceedings of the 3rd International Semantic Web Conference (ISWC2004), (Hiroshima, Japan, November 2004).
- Cygniak, R., Bizer, C., Garbers, J., Maresch, O., Becker, C. (2012). The D2RQ Mapping Language. Available at <http://d2rq.org/d2rq-language> (accessed April 2013)
- R. Cygniak, R. (2005). A relational algebra for SPARQL. HP Technical Report, HPL-2005-170.
- Erling, O., Mikhailov, I. (2007). RDF support in the Virtuoso DBMS. In Proceedings of the 1st Conference on Social Semantic Web (Leipzig, Germany, September 2007), 59–68.
- N. Konstantinou, D.E. Spanos, N. Houssos, N. Mitrou. (2013). Exposing Scholarly Information as Linked Open Data: RDFizing DSpace contents. In *The Electronic Library Journal*, to appear.
- Auer, S., Dietzold, S., Lehmann, J., Hellmann, S., Aumüller, D. (2009). Triplify -- Light-Weight Linked Data Publication from Relational Databases. In Proceedings of the 18th international conference on World wide web (WWW '09), (New York, NY, USA) 621-630
- Carroll, J., Dickinson, I., Dollin, C., Reynolds, D., Seaborne, A., Wilkinson, K. (2004). Jena: implementing the semantic web recommendations. In Proceedings of the 13th World Wide Web Conference (New York City).
- Das, S., Sundara, S., Cygniak, R. (2012). R2RML: RDB to RDF Mapping Language. W3C Recommendation. Available at <http://www.w3.org/TR/r2rml/> (accessed 2 February 2013).
- Nilsson, M., Powell, A., Johnston, P., Naeve, A. (2008). Expressing Dublin Core metadata using the Resource Description Framework (RDF), DCMI Recommendation. Available at: <http://dublincore.org/documents/dc-rdf/> (accessed August 2012)
- De Melo, G., Weikum, G. (2008). Language as a Foundation of the Semantic Web. In Proceedings of the 7th International Semantic Web Conference (Karlsruhe, Germany).
- Stamatis, K., Konstantinou, N., Manta, A., Paschou, C., Houssos, N. (2012). Biblio-transformationengine: An open source framework and use cases in the digital libraries domain, 7th International Conference on Open Repositories, Edinburgh.
- Volz, J. Bizer, C., Gaedke, M., Kobilarov, G. (2009). Silk – A Link Discovery Framework for the Web of Data. In Proceedings of the 2nd Workshop about Linked Data on the Web (LDOW 2009) (Madrid, Spain, April 20).