BM2LOD: PLATFORM FOR PUBLISHING BIBLIOGRAPHIC DATA AS LINKED OPEN DATA

Yusniel Hidalgo Delgado[1], Liudmila Reyes Álvarez[1], Amed Leiva Mederos[2], María del Mar Roldán García[3] and José F. Aldana Montes[3]

[1]*Universidad de las Ciencias Informáticas, Carretera a San Antonio de los Baños, Km. 2 1/2. Torrens, La Lisa. La Habana, Cuba*

[2]*Universidad Central "Marta Abreu" de las Villas, Carretera a Camajuaní, Km. 5½, Universidad Central de Las Villas, Santa Clara, Cuba*

[3]*Universidad de Málaga, Departamento de Lenguajes y Ciencias de la Computación, Málaga, España*

ABSTRACT

The development of Information and Communication Technologies has increased the number of scientific journals in digital format, speeding up dissemination and access to content. A large number of scientific journals have been included in the Open Access Initiative, giving full access to the complete texts of their scientific papers. In addition to complete text, some of these journals offer their bibliographic data via Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH). However, bibliographic data are dispersed, without being given an explicit relationship making their discovery and reuse for other information systems impossible. To address these issues, we propose a platform for publishing bibliographic data following the linked data principles. The proposed approach has been applied in a use case in the context of the DBJournal project for publishing bibliographic data from Open Access journals in Cuba.

KEYWORDS

Linked Data, Metadata Harvesting, OAI-PMH, Semantic Web

1. INTRODUCTION

The Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) (Lagoze et al., 2002) is a low-barrier mechanism for repository interoperability. It provides an application-independent interoperability framework based on metadata harvesting (Khan, 2013). There are two classes of participants in the OAI-PMH framework: (1) Data Providers are repositories that display structured metadata via OAI-PMH. (2) Service Providers use metadata harvested via the OAI-PMH as a basis for building value-added services. The OAI-PMH provides a simple technical option for data providers to make their metadata available to services, based on open standards such as HTTP, URI and XML. Currently, there are some open source digital library systems that act as data providers and can easily display metadata via OAI-PMH by default. Some of these tools are Dspace (Masár, 2012), EPrints (Team, 2012) and Fedora (Davis, 2009). On the other hand, several journals affiliated to the open access initiative have adopted the Open Journal System, a well-known journal management and publishing system that provides an OAI-PMH Endpoint. Although this protocol has been widely adopted, it has some drawbacks, including utilizing non-dereferenceable identifiers, limitations of selective access to metadata, and limitations in the representation format (Hakimjavadi et al., 2012):

Utilization of non-dereferenceable identifiers: According to version 2 of the OAI-PMH specification, the format of the unique identifier for all items in an OAI-PMH repository must correspond to that of the URI syntax. PMH is a web-based protocol, but it does not follow the World Wide Web Consortium (W3C) recommendation that states all URIs should be dereferenceable through the HTTP protocol and therefore, by requesting an URI, an HTTP client is not able to retrieve any records.

Limitations of selective metadata retrieval: Querying the repository to ask for all records on a specific author is a common need for some users; however, the OAI-PMH protocol is not able to do this. This is because the only selection criteria to retrieve a given record are the item identifier, the metadata format, the sets, and the record's creation date intervals.

Limitation in the representation format: There is no user interface for records, sets, and metadata formats defined in protocol specifications. However, most of the repositories, in their PMH base-URL, optionally provide an interface for human users. By default, all records in OAI-PMH are represented in XML, even though it provides very limited semantics, and even with this limited semantics, it is quite ambiguous. In other words, it does not have enough restrictions to successfully express semantics.

In parallel, several institutions and governments around the world have published their data following the linked data principles. Linked data refers to a set of best practices for publishing and interlinking structured data on the web (Berners-Lee, 2006). Over the last few years, the W3C has played an important role in the development of standards such as Resource Description Framework (RDF) and SPARQL Protocol and RDF Query Language (SPARQL). The first is a data model that describes structured data in a machine readable format based on a directed graph model. SPARQL is the standardized query language for RDF data models. The linked data principles proposed by Tim Berners-Lee are: (1) Use URIs to denote things; (2) Use HTTP URIs so that these things can be referred to and looked up ("dereferenced") by people and user agents. (3) Provide useful information about the thing when its URI is dereferenced, leveraging standards such as RDF, SPARQL; (4) Include links to other related things (using their URIs) when publishing data on the Web. The first principle proposes the use of URIs to identify not only Web documents and digital content, but also to serve as a -reference for real-world objects and abstract concepts. The second principle proposes the use of HTTP URI to identify objects and abstractions, enabling these URI to be dereferenceable over HTTP protocol and to provide a description of the object or concept identified. The third principle proposes the use of a single data model to publish structured data on the web. The RDF data model is a good choice. It uses classes and properties from ontologies to express semantics of data in a machine readable way. Finally, the fourth principle proposes the use of hyperlinks to connect Web documents.

In recent years, researchers have developed several tools for converting bibliographic metadata into linked data (Malmsten, 2008) (Coppens et al., 2009) (Haslhofer and Schandl, 2010) but these tools do not use the OAI-PMH protocol to produce such metadata. Although the OAI2LOD tool (Malmsten, 2008) allows for metadata using OAI-PMH, this is still only possible from a single data provider and, metadata is stored in the memory, causing scalability problems. Other tools only allow an RDF graph from a relational database to be generated, so they do not include a prior step of re-collection and pre-processing of metadata (Bizer and Seaborne, 2004). In order to solve the aforementioned problems, we propose a platform for publishing bibliographic data following the linked data principles. The proposed approach has been applied in a use case in the context of the DBJournal project for publishing bibliographic data from Open Access journals in Cuba. The paper is structured as follows: Section 2 describes the BM2LOD platform architecture and its components. Section 3 describes the methodological guideline followed for publishing bibliographic metadata as linked open data using the proposed platform. Section 4 mentions some of the relevant related work found in the literature and finally, Section 5 summarizes the key points presented in this paper and explains our plans for future work.

2. BM2LOD PLATFORM

Publishing structured data as Linked Open Data involves making complex technical decisions. Over the last few years, several tools with specific goals have been developed by Linked Open Data practitioners. The BM2LOD (Bibliographic Metadata to Linked Open Data) platform reuses some of these tools following a pipeline approach. BM2LOD extends the life cycle of a common Linked Open Data project for publishing structured data following the linked data principles. Given the lack of a platform that performs the processes of collection, pre-processing, generation, bounding and publication of bibliographic metadata harvested via the OAI-PMH protocol, we took on the task of integrating, in a single platform, a set of tools, some already existing in the scientific community of bound data. BM2LOD has two main contributions: (1) the formalization of the workflow described above using an iterative and incremental approach (2) a platform that will support the aforementioned workflow. The platform consists of several tools:

Metharto (Hidalgo-Delgado et al., 2013): Metharto is an ongoing effort still in the development stage, intended to cover metadata extraction and preprocessing tasks. It is able to extract metadata from data providers that disseminate their metadata through OAI-PMH and can perform some interesting tasks, such as: selective and bash metadata retrieval, connection over proxy server and support for HTTP and HTTPS

connections. Some preprocessing tasks have been implemented, such as author name cleaning and disambiguation.

D2RQ (Bizer and Seaborne, 2004): D2RQ is a mapping language and platform for treating non-RDF relational databases as virtual RDF graphs. Its aim is to display relational databases on the Semantic Web in order to provide access via SPARQL queries and Linked Data interface. Several RDF and OWL vocabularies can be reused in the mapping process. We use the dump-rdf shell script to generate the RDF graph. This script is executed periodically with a scheduled task.

Silk (Jentzsch et al., 2010): The Silk Link Discovery Framework provides tools to generate links between data items based on user-provided link specifications. The Silk Link Discovery Framework consists of two main components: (1) a console application used to interlink two datasets, and (2) an HTTP server, which receives an incoming RDF stream and creates links between data items. Both applications provide a flexible configuration language, the Silk Link Specification Language (Silk-LSL), to specify the conditions that the data items must fulfill so as to be interlinked.

OpenLink Virtuoso (Haslhofer et al., 2011): The OpenLink Virtuoso Universal Server is a hybrid storage solution for a range of data models, including relational data, RDF and XML, and free text documents. Through its unified storage, it can serve as an integration point for data from different, heterogeneous sources. Virtuoso is offered as an open-source version; for commercial purposes several licensed models are available. It can be deployed as a server on major platforms.

Pubby (Cyganiak and Bizer, 2008): Pubby is a linked data frontend that rewrites URI requests into SPARQL DESCRIBE queries against the underlying RDF store. Besides RDF, Pubby also provides a simple HTML view over the data store and takes care of handling 303 redirects and content negotiation between the two representations.

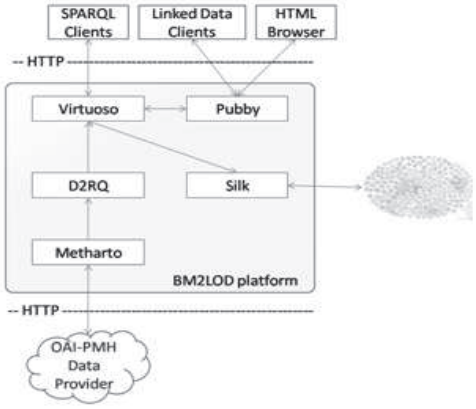

Figure 1. BM2LOD Platform architecture

Figure 1 shows the general architecture of BM2LOD platform. The workflow begins with the metadata extraction step. At this stage, the Metharto tool is used for the harvesting, cleaning and storage in a relational database of existing metadata fields in each of the data providers. Then, a D2RQ tool generates a mapping file using the D2RQ mapping language. In this file, tables and fields are mapped, from relational schema with classes and properties from previously identified ontologies, respectively. Then, an RDF graph is generated by a scheduled task using the dump-rdf script. The RDF graph generated is loaded into the Virtuoso RDF triples store, which provides a SPARQL endpoint to query the data. Then, the Silk tool is configured to generate the links type *owl: sameAs* between our graph and the DBLP graph. Finally, we use the Pubby tool to publish the RDF graph in the web, making it accessible to both humans and computers.

3. USE CASE: CUBAN JOURNALS

In Cuba, several academic journals disseminate their bibliographic data through OAI-PMH. In order to measure the performance and test the features of the proposed platform, we carry out a use case choosing a sample of these journals in the context of the DBJournal project. This project is an initiative, the aim of which is to enrich the Web of Data with bibliographic metadata harvested from Cuban journals (see Table 1).

Table 1. Sample of Cuban journals used in the case study.

Journal name	OAI-PMH endpoint
ACIMED	http://www.acimed.sld.cu/index.php/acimed/oai
Revista Cubana de Ingeniería	http://rci.cujae.edu.cu/index.php/rci/oai
Medisur	http://www.medisur.sld.cu/index.php/medisur/oai
Revista Cubana de Ciencias Informáticas	http://rcci.uci.cu/index.php/rcci/oai
Revista Finlay	http://www.revfinlay.sld.cu/index.php/finlay/oai
Serie Científica UCI	http://publicaciones.uci.cu/index.php/SC/oai
Revista Cubana de Cardiología	http://www.revcardiologia.sld.cu/index.php/revcardiologia/oai
Revista Cubana de Oftalmología	http://www.revoftalmologia.sld.cu/index.php/oftalmologia/oai
Revista de CM de P. del Río	http://publicaciones.pri.sld.cu/index.php/publicaciones/oai

Our approach automatizes six steps: *metadata extraction*, *preprocessing*, *data modeling*, *transformation*, *linking* and *publication*.

3.1 Step 1: Metadata Extraction

The goal at this stage is to extract, clean and store bibliographic metadata from data providers. The input of this stage is one or more OAI-PMH endpoints for specific journals and the output is a relational database with the bibliographic metadata harvested. This relational database stores metadata related to four main entities: journals, records, sets and authors. For this purpose the Metharto tool has been used.

3.2 Step 2: Preprocessing

The quality of bibliographic metadata is a crucial point that significantly affects the visibility and discovery of the resources described. The goal of this step is to clean up and normalize some fields of metadata, thereby considerably improving their quality. Some problems found in this step are: (1) disambiguation of authors' names and, (2) duplicate detection of authors' affiliations. These problems arise due to a lack of authority control by the journals. In the first step, we implement a small algorithm for automatic cleaning by removing stopwords such as honorific titles and special characters.

Although the algorithm (Algorithm 1) is not able to build authority control files, it can provide a method to ensure uniqueness of the authors in the data set identifying each one with an URI in the space of the web as proposed in the linked data principles. To do this, the AUTHORIS tool (Amed Leiva-Mederos et al., 2013), takes into account the obsolescence of MAhine-Readable Cataloging (MARC is a specific format in RDF) and its replacement by newer standards such as Functional Requirements for Bibliographic Records (FRBR) and Resource Description and Access (RDA). Among its many features the ability to process data of authority following the linked data principles is especially noteworthy. For data transformation tasks learning rules are used.

Algorithm 1: Cleaning author's name removing stopwords

Input: 1) List *S* with stopwords and 2) List *N* with names of authors
Output: 1) List *R* with names cleaned

```
1: Define list R
2: for all n_i ϵ N do
3:     if n_i contain ";" then
4:         k ← split n_i for ";"
5:         for all s_j ϵ S do
6:             if k contain s_j then
7:             r1 ←k.replace(s_j, " ")
8:             r1.trim()
9:             add r1 to list R
10:        end for
11: end for
12: return list R
```

3.3 Step 3: Data Modeling

After the aforementioned tasks, in which the metadata was stored and cleaned, we determine the domain ontologies used for modeling bibliographic metadata, in particular, research papers from scientific journals. The most important recommendation in this context is to reuse available vocabularies as much as possible. Currently there are some useful repositories for finding available vocabularies, such as Swoogle and LOV. We use some of these systems for searching classes and properties related to bibliographic fields. We use the following ontologies for modeling journals, authors, sets and records:

FaBiO (Silvio Peroni and David Shotton, 2012) is the FRBR-aligned Bibliographic Ontology for recording and publishing bibliographic records of scholarly endeavors on the Semantic Web.

The Bibliographic Ontology (BiBo) (Giasson and D'Arcus, 2009) describes all things bibliographic, on the semantic Web by means of RDF. This ontology can be used as a citation ontology, as a document classification ontology, or simply as a way to describe any kind of document in RDF. It has been inspired by many existing document description metadata formats, and can be used as a common basis for converting other bibliographic data sources.

Dublin Core (Coppens et al., 2009, Board, 2012) is a broadly accepted descriptive schema. The power of this schema is its simplicity and generality. It consists of fifteen fields, e.g., creator, subject, coverage, description, and date. It can answer the basic questions: Who, What, Where, and When. All the fields in Dublin Core are optional and repeatable.

FOAF (Friend of a Friend) (Brickley and Miller, 2010) is a project that publishes OWL definitions of a set of classes and properties providing a structured vocabulary for describing people and their interconnections in RDF format. Some popular classes from this vocabulary are *foaf:Person* and *foaf:Agent*. The first, only represents people and is a sub-class of *foaf:Agent*. In the second, agents could be considered as people, groups or organizations.

3.4 Step 4: Transformation

In this step, the D2RQ tool has been used to perform the transformation of the relational database in an RDF graph. In the latest version of the platform, this transformation is performed periodically, by executing a set of scheduled scripts in bash language. These scripts are capable of generating the RDF graph and loading it into the Virtuoso RDF store, through the interface that it provides for just this purpose.

The D2RQ platform provides an additional tool for querying and publishing virtual RDF graph called the D2R Server. However, the D2R Server is quite slow, because SPARQL queries are translated to SQL queries. Then, SQL queries are executed on the relational database system. For this reason, we decided to create and load an RDF graph into a Virtuoso as an external RDF store. Figure 2 presents a view of the Linked Data frontend in the BM2LOD tool.

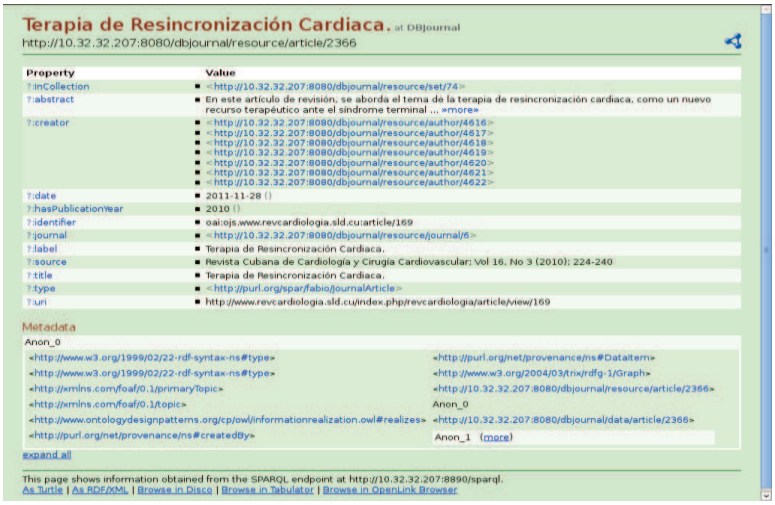

Figure 2. Linked data frontend

3.5 Step 5: Linking

The fourth Linked Data principle is to set RDF links to point out other data sources on the Web. External RDF links connect data islands in a global and interconnected data space, enabling other applications to discover additional data sources in a follow-your-nose fashion (Heath and Bizer, 2011).

To satisfy this principle, we used the Silk framework as specified in the sections above. To do this, a DBLP dataset was downloaded and loaded into the Virtuoso RDF store (See Figure 3). Once we had both datasets available via a SPARQL endpoint, the tool Silk was configured to generate the links of type *owl: sameAs*, comparing the names of the authors in both datasets. Silk provides several algorithms to establish a measure of similarity between strings, in this case the name of the authors. In our proposal we use the Levenshtein distance (Levenshtein, 1966). A new set of RDF triples are generated from this process and mixed with the RDF graph originally created and uploaded into the RDF store.

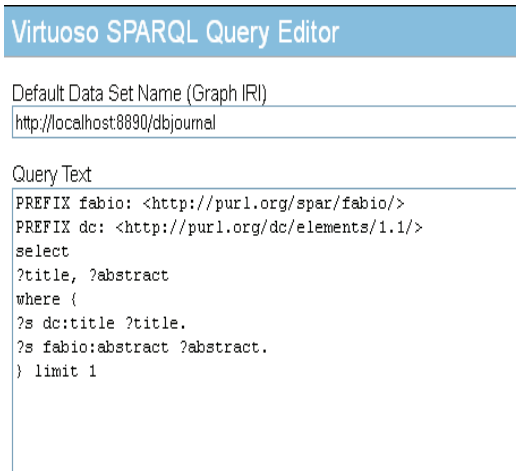

Figure 3. Query executed in the Virtuoso SPARQL Endpoint

3.6 Step 6: Publication

The purpose of this step is to publish the RDF graph in the Web space, making it accessible to both humans and computers. This stage can be considered as a technical solution to the third linked data principle, which states that it is necessity to provide useful information about the published data using standards such as RDF and SPARQL. In our approach, we use the Pubby linked data interface in front of the SPARQL endpoint. Pubby rewrites URI dereferencing requests into SPARQL DESCRIBE queries, comparing them against the underlying RDF store, and handles 303 redirects and content negotiation.

4. RELATED WORK

In this paper, we have presented our own experience in the transformation of bibliographic data into Linked Open Data. In recent years, several institutions and research projects have published high quality bibliographic data following linked data principles. We have therefore found some relevant tools and approaches in the literature. One of the earliest initiatives for publishing bibliographic data as linked data was proposed by the Swedish Union Catalogue (Malmsten, 2008). In this approach, the authors present an RDF Server Wrapper for transforming MARC 21 (an application-specific format, in RDF prior to OAI-PMH) records into RDF graphs. This approach uses the eXtensible Stylesheet Language Transformations (XSLT) to carry out the transformation. The main difference between our approach and theirs is the data format used for the data provider and the quality of the obtained metadata. In our approach, we use an unqualified Dublin Core format harvested from open access journals. In some cases, the quality of the metadata harvested is not the best, and therefore, we need to clean it. For this, we propose a preprocessing task for data cleaning.

Another similar approach is proposed by (Haslhofer and Schandl, 2008) and (Haslhofer and Schandl, 2010). The authors propose OAI2LOD, a server wrapper that displays metadata of OAI-PMH compliant data sources as Linked Data on the Web and provides a SPARQL query interface for this metadata. OAI2LOD is able to generate links *owl:sameAs* between two server instances. In order to do this, the server compares the values of a set of manually selected attributes according to their lexical similarity using the Levensthein string distance. Finally, this tool is based on D2RQ Server (Bizer and Seaborne, 2004) and transforms from RDF/XML via XSLT. The OAI2LOD server has an important limitation related to the ability to manage several data providers in a single instance server. Along the same lines, in (Coppens et al., 2009) the authors enhance the OAI2LOD Server with import data taken from different OAI-PMH repositories. Similarly, our approach is able to manage any amount of data providers and store the metadata harvester in a relational database, ensuring that our solution is more scalable over time. Finally, the Academic Bibliographic Database of the University of Economics in Praga consists of bibliographic records of publications concerning journal articles, conference papers, lecture notes and monograph chapters created by the academic staff of the university. In (Hladka et al., 2012) the authors discuss their experiences in the process of transformation of MARC records to RDF graphs. During this process several challenges have been addressed, such as the quality of bibliographic data from the data provider and the ambiguity of authors' names. Some of these challenges have also been addressed in our approach.

5. CONCLUSION AND FUTURE WORK

In this paper, we have proposed the BM2LOD platform for publishing bibliographic metadata following the linked data principles. BM2LOD is able to extract metadata from an OAI-PMH endpoint and automate the most common tasks for publishing it as linked data. Several challenges have been addressed in the preprocessing task, such as disambiguation of authors' names and the affiliations of authors.

The BM2LOD platform is able to integrate metadata from different data providers, using a simple method of integration based on a workflow, where the output of the previous stage is used for subsequent stages. With the publication of bibliographic metadata, following the linked data principles, we provide a mechanism of interoperability that enables the discovery and reuse of this metadata by other computer systems.

We have identified some problems which we hope to resolve in future versions of the tool. One of the most important is to find an automatic mechanism for incremental updates in the RDF graph. Currently, the stable graph in the RDF store is deleted. So, the new graph is loaded into the RDF store. This obviously means that when the database grows a lot, it requires more computational resources. We will also work on improving the algorithms used in the disambiguation of the names of the authors, as well as on obtaining an efficient mechanism for incremental updating RDF graphs from relational databases.

REFERENCES

AMED LEIVA-MEDEROS, JOSÉ A. SENSO, SANDOR DOMÍNGUEZ-VELASCO & PEDRO HÍPOLA 2013. AUTHORIS: a tool for authority control in the Semantic Web. *Library Hi Tech,* Vol 31**,** No 3, pp 536-553.

BERNERS-LEE, T. 2006. *Linked Data - Design Issues* [Online]. W3C. Available: http://www.w3.org/DesignIssues/LinkedData.html [Accessed October 17 2013].

BIZER, C. & SEABORNE, A. 2004. D2RQ – Treating Non-RDF Databases as Virtual RDF Graphs. *Proceedings of the 3rd International Semantic Web Conference.* Japan: Springer.

BOARD, D. U. 2012. *DCMI Metadata Terms* [Online]. DCMI. Available: http://dublincore.org/documents/2012/06/14/dcmi-terms/?v=elements [Accessed October 13 2013].

BRICKLEY, D. & MILLER, L. 2010. *FOAF Vocabulary Specification* [Online]. Available: http://xmlns.com/foaf/spec/ [Accessed October 13 2013].

COPPENS, S., MANNENS, E. & WALLE, R. V. D. 2009. Disseminating heritage records as linked open data. *International Journal of Virtual Reality,* Vol 8**,** No 3, pp 39–44.

CYGANIAK, R. & BIZER, C. 2008. *Pubby – A Linked Data Frontend for SPARQL Endpoints* [Online]. Available: http://wifo5-03.informatik.uni-mannheim.de/pubby/ [Accessed October 13 2013].

DAVIS, D. 2009. *Basic OAI-PMH Provider* [Online]. Fedora Commons. Available: https://wiki.duraspace.org/display/FEDORA36/Basic+OAI-PMH+Provider [Accessed October 17 2013].

GIASSON, F. & D'ARCUS, B. 2009. *Bibliographic Ontology Specification* [Online]. Structured Dynamics LLC. Available: http://bibliontology.com/specification [Accessed October 13 2013].

HAKIMJAVADI, H., MASREK, M. N. & ALAM, S. 2012. SW-MIS: A Semantic Web Based Model for Integration of Institutional Repositories Metadata Records. *Science Series Data Report,* Vol 4, No 11, pp 57–66.

HASLHOFER, B., ROOCHI, E. M., SCHANDL, B. & ZANDER, S. 2011. Europeana RDF Store Report. Viena, Austria: University of Viena.

HASLHOFER, B. & SCHANDL, B. 2008. The OAI2LOD Server: Exposing OAI-PMH Metadata as Linked Data. *In: International Workshop on Linked Data on the Web*, 2008 Beijing, China.

HASLHOFER, B. & SCHANDL, B. 2010. Interweaving OAI-PMH data sources with the linked data cloud. *International Journal of Metadata, Semantics and Ontologies,* Vol 5, No 1, pp 17–31.

HEATH, T. & BIZER, C. 2011. *Linked data: Evolving the web into a global data space*, Morgan & Claypool.

HIDALGO-DELGADO, Y., RODRÍGUEZ-PUENTE, R., ORTIZ-MUÑOZ, E. & ALONSO-SIERRA, L. E. 2013. Herramienta para la recolección de metadatos bibliográficos mediante el protocolo OAI-PMH. *In: II Conferencia Internacional de Ciencias Computacionales e Informáticas*, La Habana, Cuba.

HLADKA, J., MYNARZ, J. & SKLENAK, V. 2012. Experience with transformation of bibliographic data into linked data. *Journal of Systems Integration,* Vol 3, No 1, pp 54-62.

JENTZSCH, A., ISELE, R. & BIZER, C. 2010. Silk – Generating RDF Links while publishing or consuming Linked Data. *In: International Semantic Web Conference Posters&Demos*, 2010 Shanghai, China.

KHAN, N. A. 2013. Emerging Trends in OAI-PMH Application. *In:* GULATI, T. A. A. P. A. (ed.) *Design, Development, and Management of Resources for Digital Library Services.* IGI Global.

LAGOZE, C., SOMPEL, H. V. D., NELSON, M. & WARNER, S. 2002. *Open Archives Initiative-Protocol for Metadata Harvesting-v. 2.0* [Online]. Available: http://www.openarchives.org/OAI/openarchivesprotocol.html [Accessed October 16 2013].

LEVENSHTEIN, V. I. 1966. Binary codes capable of correcting deletions, insertions and reversals. *In: Soviet physics doklady*, Vol 10. pp 707.

MALMSTEN, M. 2008. Making a Library Catalogue Part of the Semantic Web. *Intl Conf. on Dublin Core and Metadata Applications.* Korea: Dublin Core Metadata Initiative.

MASÁR, I. 2012. *Use the OAI-PMH interface* [Online]. DuraSpace. Available: https://wiki.duraspace.org/display/DSPACE/Use+the+OAI-PMH+interface [Accessed October 17 2013].

SILVIO PERONI & DAVID SHOTTON 2012. FaBiO and CiTO: ontologies for describing bibliographic resources and citations. *Web Semantics: Science, Services and Agents on the World Wide Web,* Vol 17, No December 2012, pp 33-43.

TEAM, E. 2012. *Synchronize your repository via OAI-PMH* [Online]. Available: http://wiki.eprints.org/w/Synchronize_your_repository_via_OAI-PMH [Accessed October 17 2013].