

Current state of Linked Data in Digital Libraries

Hallo Maria

National Polytechnic School (ECUADOR)

Luján-Mora Sergio

University of Alicante (SPAIN)

Maté Alejandro

University of Alicante (SPAIN)

Juan Trujillo

University of Alicante (SPAIN)

Abstract

The Semantic Web encourages institutions, including libraries, to collect, link and share their data across the Web in order to ease its processing by machines to get better queries and results. Linked Data technologies enable to connect related data on the Web using the principles outlined by Tim Berners-Lee in 2006.

Digital libraries have great potential to exchange and disseminate data linked to external resources using Linked Data. In this paper, a study about the current uses of Linked Data in digital libraries including the most important implementations in the world is presented. The study focuses on selected vocabularies and ontologies, benefits and problems encountered in implementing Linked Data on digital libraries. Besides, it also identifies and discusses specific challenges that digital libraries presents offering suggestions for ways in which libraries can contribute to the Semantic Web.

The study uses an adapted methodology for literature review, to find data available to answer research questions. It is based on the information found in the library websites recommended by W3C Library Incubator Group in 2011, and scientific publications from Google Scholar, Scopus, ACM, and Springer from the last 5 years. The selected libraries for the study are National Library of France, Europeana Library, Library of Congress, British Library, and National Library of Spain. In this paper, we outline the best practices found in each experience and identify gaps and future trends.

Keywords:

Linked Data; Digital Libraries; Semantic Web.

I. Introduction

Digital libraries allow online access to devices that contain digital knowledge. Libraries have traditionally worked on publishing data, however to solve more complex queries it is necessary the connection to external data sources. Moreover, it is necessary to share knowledge on the Web which involves improving the method of administration of current libraries including new artifacts as ontologies, semantic content descriptions, data links and new forms of collaboration using: social networks, specialized communities, wikis, collaborative games and mashups (applications that use and combine data, presentation and functionality from one or more sources). Furthermore, applications can be developed on metadata to build

Corresponding author:

María Hallo, National Polytechnic School, Isabel la Católica El I-253, Quito, Ecuador PO-Box 17-01-2759

Email: maria.hallo@epn.edu.ec

The last version of this article was published at: <http://jls.sagepub.com/content/early/2015/07/18/0165551515594729.abstract>

services on top of them such as adding maps to locate resources, putting together resources with similar issues, making recommendations, allowing users to create semantic annotations, etc.

Linked Data is a publication technique using standard web technologies to connect related data and make them available on the Web by following principles recommended by Tim Berners-Lee [1].

Data integration and the building of new services may be favored with the support of Linked Data technology using unique identifiers (URIs) for resources (places, people, events, etc.) and RDF (Resource Description Framework) to express relations between resources. The use of URIs allows different descriptions for the same resource building relationships from external data sources to libraries. Moreover, it is possible to add data in a collaborative way by providing information about a resource that would be integrated into the overall graph.

In 2010, Tim Berners-Lee proposed a 5-star model to encourage people to publish in a Linked Open Data environment. In Table I, a summary of the 5-star model is presented.

Table I. 5-star Linked Open Data Model.

Stars	Description
1-star (*)	Available on the Web but with open license to be Open Data.
2-star (**)	Available as machine readable structured data.
3-star (***)	As non-proprietary format.
4-star (****)	All the above plus, use open standards from W3C (RDF and SPARQL) to identify things, so that people can point at your stuff.
5-star (*****)	All the above, plus link your data to others people data to provide context.

In recent years, digital libraries have given more importance to Linked Data in several ways: creating metadata models such as Europeana Data Model [2] or the bibliographic framework of the Library of Congress [3] and publishing Linked Data and library catalogs or integrating authority files of national libraries of several continents using services like the Virtual International Authority File (VIAF). Other Linked Data has been obtained by extracting information from Wikipedia on the Dbpedia¹ or from projects such as LinkedGeoData² that exports data generated by users from those contained in OpenStreetMap³. Some software systems such as Evergreen⁴, Hydra⁵ and Omeka⁶ have started producing Linked Data automatically. Several global digital libraries such as the Library of France, the Library of the Congress in USA, the British Library, the National Library of Spain, among others, are implementing this new technology. It is expected in the future to obtain relationships with the community to improve data quality, annotate, recommendations, links, etc. [4].

However, libraries are facing several problems to implement Linked Data services. In 2010, Hannemann and Kett [5] classified these problems in technical (e.g. poorly documented development tools), conceptual (little experience with Linked Data models for libraries, variables schemes, e.g. Marc21 vs RDF) and legal (publishing rights and licensing of Linked Data). Another problem they also exposed was the lack of experience in defining Linked Data services.

The purpose of this study is to review Linked Data applications in digital libraries (LDDL), analyzing papers and other publications of the last five years related to the world's major national digital libraries.

2. Methodology

This study was carried out based partially on the guidelines proposed by Kitchenham et al (2009) for the main stages of a Systematic Literature Review (SLR) [6], combined with Mapping Review (MP) Guides [7], extracting information from a group of selected publications and national libraries websites that expose Linked Data. The use of SLR provides a systematic approach to answering a research question by finding relevant research outcomes from primary empirical studies. MP is an open form of SRL, intended to map out the research that has been undertaken related to a topic. Such a study tries to identify gaps in the set of primary studies, where new primary studies are required.

The stages followed in our research were:

- definition of research questions,
- search process,
- data collection,

- data synthesis, and,
- discussion.

The definition of research questions outlines the problems to be addressed by the review. In the MP, the research questions address a wider scope of the study.

Search process includes the identification of primary studies that may contain relevant research results. This stage could be a target manual search of selected journals and conferences or an automated search. According to Kitchenham [6] automated searches find more studies than manual restricted searches but they may be of poor quality. Target manual searches intend to omit low quality papers. In both cases, the selected papers are examined for quality.

Data collection stage focuses on the design of data extraction procedures to record the information needed to address the review questions such as: the source of data, the topic area, the authors and institutions working in the related research topics.

Data synthesis requires summarising the results of the included primary studies: data are tabulated in relation to the review questions showing the features and findings of the studies. Data from the studies could be presented narratively and/or statistically.

Discussion stage includes analysis of results, main findings related to the research questions, their meanings, magnitude of the effects, and applicability of the findings.

3. Results

The proposed methodology in section 2, was applied to find the current state of the use of Linked Data on Digital Libraries.

3.1. Definition of research questions.

In the first stage we decided to look for the current state of development of LDDL taking the most important applications in the world, the research questions are:

- (1) What are the benefits of LDDL.
- (2) What are the vocabularies and ontologies used in LDDL.
- (3) What are the problems reported for LDDL.
- (4) What are the future trends in research on LDDL.

3.2. Search Process

Based on the study by the Library Linked Data Incubator Group (2011), five national digital libraries presenting advances in the use of Linked Data were chosen. In this study, publications reported in Google Scholar, Scopus, ACM, Springer and information from the websites of libraries were reviewed. The queries included the terms: linked data combined with the name of the studied libraries. Few publications for each specific library were found and cited in the corresponding sections.

3.3. Data Collection

Data extracted from each study were focused on the research questions: benefits of using LDDL, vocabularies used, current issues and future trends.

3.4. Data Synthesis

A summary of the studies started by Library Linked Data Incubator Group is presented in section 3.4.1 and a table of features found for each digital library is developed in Table II. Following we present narratively the main features of each project:

3.4.1. Library Linked Data Incubator Group

The mission of the W3C Library Linked Data Incubator Group was stated in 2010: “to help increase global interoperability of library data on the Web, by bringing together people involved in Semantic Web activities—focusing on Linked Data—in the library community and beyond, building on existing initiatives, and identifying collaboration tracks for the future” [4]. In its final 2011 report, the W3C group summarize: the benefits of Linked Data, the current situation of libraries and data sets

published in the Semantic Web, the problems of copyright and the recommendations for using Semantic Web standards and Linked Data principles to ensure more visibility and reuse of resources such as bibliographic data and concept schemes.

According to the final report of the W3C Library Linked Data Incubator Group, “few libraries have published their data as Linked Data with great variation in support and quality”. However, there has been interest in Linked Data projects in national libraries of Sweden, Hungary, Germany, France, United States of America, United Kingdom and international organizations such as the Food and Agriculture Organization of the United Nations among others. The recommendations cited in the study include defining an organization that assigns URIs to local resources to avoid duplication and to use Linked Data design guidelines. It is also recommended to include activities designed to preserve Linked Data which mostly are in the cloud.

The final report of the group also suggests that the success of Linked Data in any domain depends on the ability to identify, reuse or connect data and data models available. Moreover, the same report highlights the complexity of the vocabularies available and the lack of similarity with the vocabulary and Linked Data sets used in research communities in the Semantic Web. In addition, the W3C study [4] presents an inventory of vocabularies and ontologies mainly for classification systems, authority files, themes, thesauri and other controlled vocabularies. Some examples of vocabularies and ontologies published in RDF widely used in digital libraries are: SKOS (Simple Knowledge Organization System), Marc21, FRBR (Functional Requirements for Bibliographic Records), FOAF (Friend of a friend), VoiD (Vocabulary of InterLinked Datasets), CITO (Citation Type Ontology), BIBO (Bibliographic Ontology), Dublin Core, FOAF (Friend of a Friend) and NeoGeo to describe geographic data. Finally, a large number of national libraries are using Linked Data. In the following, we present the experiences of a group of them chosen by the importance of the LDDL and the size of the data sets.

3.4.2. National Library of France (NLF)

The National Library of France (data.bnf.fr) brings together information from different catalogs using Semantic Web technologies created from data in various formats such as INTERMARC Bibliographic format for the main catalog, XML-EAD (Encoded Archival Description) for inventory files and Dublin Core for the digital library [8]. These data are automatically retrieved, shaped, enriched and published in RDF. The NLF provides URIs for resources built based on ARK (Archival Resource Key) identifiers and designed to access information objects. Among the motivations for publishing Linked Data are difficulties of cataloging a growing number of resources without increasing the staff, but if metadata are published as Linked Data it can be disseminated more widely and enriched from external sources also increasing its reuse [9]. Furthermore, data from relational systems, not accessed from search engines on the Web, could be more visible.

Most of the Linked Data in the National Library of France are based on the Functional Requirements for Bibliographic Records (FRBR) model. The website generates web pages providing standardized information, references and links on authors, works and themes. The web pages provide filtering capabilities by type, export links to other online services and open data recovery. In addition, the vocabulary used is Rameau and was transformed to a version of Simple Knowledge Organization System (SKOS) vocabulary exposing as Linked Data about 160,000 concepts [10].

3.4.3. Europeana

Europeana (data.europeana.eu) is an European digital library of free access, offering multilingual publications and linked metadata from multiple European institutions. This library provides information on millions of items digitized from museums, libraries, archives and multimedia collections [11].

The data model EDM (Europeana Data Model) is based on the principles and practice of the Semantic Web and Linked Data. The model is built using RDF standards and established vocabularies as: OAI-ORE (Open Archives Object Reuse and Exchange), SKOS, Dublin Core. EDM acts as a common top-level ontology that enables interoperability while maintaining original data models .

Other ontologies that are of special interest to EDM are Dublin Core (DC) and Friend-of-a-Friend (FOAF). Dublin Core provides a vocabulary for describing the essential features of cultural objects (creators, relationships with other resources, subject indexing, etc.) in an appropriate manner for the Semantic Web[12]. FOAF is an ontology describing persons, their activities and their relations to other people and objects.

RDF has been used as the meta-model of EDM, and any object of interest in the space of Europeana (either an object of cultural heritage, a person, a place, a concept, etc.) is considered a resource, identified by a URI. EDM provides a set of constructors (classes and properties) that can be used by the metadata providers. The model captures a description of the digital environment of the object sent to Europeana adding descriptive information resources that are part of that environment. Furthermore, EDM includes descriptive and contextual properties that capture the features of a resource and relate to others in this context. For descriptive metadata, EDM provides builders that can be represented focused on objects or events metadata [13]. A pilot project has built links to several datasets like: people in DBpedia, Geonames gazetteer, GEMET thesaurus. The concepts are described using SKOS [14].

It is expected that the new knowledge generated with LDDL will play an important role in the usability of Europeana in areas such as improving functionality and content search queries semantically related. The principal problems reported for the adoption of Open Linked Data are: lack of metadata expressed in EDM, lack of links to other sources and lack of agreements to provide data [15].

In the future, it is expected to solve the cited problems and to have a greater involvement of the community by adding new data sets, defining specifications of links in each institution and improving data quality. It is also expected an increase in the visibility of data integrating initiatives such as Schema.org.

3.4.4. Library of Congress USA

The Linked Data Service defined by the Library of Congress (www.loc.gov) began in 2009, exposing around 260,000 authority records. Currently it provides access to standards and vocabularies promulgated by the Library of Congress such as: classification files, themes, geographic areas, countries, languages, preservation vocabularies, event types, types of software, etc. [16].

Since 2011, the Library of Congress began a process of change to implement a new bibliographic environment facilitating the interconnection of network resources. The BIBFRAME (Bibliographic Framework) model of Linked Data was implemented like a starting point for integration on the Web of data [17]. The BIBFRAME model contains classes: *item* (creative work), *body*, *authority*, *annotation*. The origin of the relationship *item / instance* may reflect FRBR relationships in terms of graphs rather than hierarchical relationships. In the context of the entity-relationship model FRBR recognizes entities, attributes and relationships between entities like web resources. The BIBFRAME model is defined in RDF, identifying all entities (resources), attributes and relationships between entities (properties), this allows annotations such as mapping to other vocabularies. In addition, there are defined connection between BIBFRAME and vocabularies such as Dublin Core, FOAF, SKOS and others.

The screenshot displays the Library of Congress website's 'LC Linked Data Service' page. At the top, there is a navigation bar with the Library of Congress logo, buttons for 'ASK A LIBRARIAN', 'DIGITAL COLLECTIONS', and 'LIBRARY CATALOGS', and a search box labeled 'Search Loc.gov' with a 'GO' button. Below this, a breadcrumb trail reads 'The Library of Congress > Linked Data Service'. The main content area is titled 'LC Linked Data Service Authorities and Vocabularies'. On the left, there is a sidebar with a 'LIBRARY OF CONGRESS LINKED DATA SERVICE' header and a list of links: 'Linked Data Service', 'About', 'Main Dataset Descriptions', 'Preservation Dataset Descriptions', 'Small Datasets Descriptions', 'Search', 'Download', 'Technical Center', and 'Contact Us'. The main content area features a decorative header with a repeating pattern of 'rel' and 'http://www.loc.gov' followed by a search box. The search box contains the text 'Enter Keyword or Phrase' and a dropdown menu with options: 'All', 'LC Subject Headings', 'LC Name Authority File', 'LC Classification', and 'LC Children's Subject Headings'. To the right of the search box, there is a note: '**Please Note: LC Classification entries are not included in general search results. You must explicitly select LC Classification in order to search the schema. This is temporary while the impact of adding LOC to the current system is better understood.' Below the search box are 'Search' and 'Reset' buttons.

Figure 1. Library of Congress website

To support the developed model there are tools to migrate from Marc21 to BIBFRAME model and interfaces for publication like Exhibit. This model could be integrated with an event model with annotation capabilities. Moreover, projects like

Reference extract can be used to extract references to answer questions in libraries that hold this service (this allows the user to verify the quality of the sources of the information provided). In addition, the Library of Congress uses the services of linkable authority files and virtual international authority VIAF (Virtual International Authority File). In the future, the development of tools for migration from Marc21 to BIBFRAME and its use expected to increase improving the services associated with the model. Figure 1 shows the Library of Congress website.

3.4.5. British Library

The British Library (www.bl.uk) has millions of items from various countries in print and digital format. The British Library is developing a version of linked open data from books and serial publications among others. Unlike previous initiatives they do not transform their collections of MARC (MACHINE-Readable Cataloging) record to RDF / XML but they model things of interest such as people, places and events related to a book [18]. Figure 2 shows the British Library website.

Following the practices adopted by various institutions on the Web, the British Library have modeled the “things of interest” reusing the most descriptive existing schemes adding their own terms in the cases where there were no published terms. The following vocabularies are used to describe the resources of the library: BIBO (Bibliographic Ontology), Bio (Biographic Information), Dublin Core, ISB (International Standard Bibliographic Description), Org (Ontology Organizational), SKOS, RDF Schema, OWL, FOAF, EVENT (Event ontology), WGS84 Geo Positioning and draft of RDA (Resource Description and Access). The data is connected to other sources of Linked Open Data particularly VIAF, LCSH (Library of Congress Subject Headings), Lexvo, GeoNames (for country of publication), MARC (for country code and language), and RDF Book Mashup (application to makes information about books, their authors, reviews, and online bookstores available on the Semantic Web), Dewey.info (a prototype for linked Dewey Decimal Classification data) [19].

The screenshot shows the website for 'THE BRITISH NATIONAL BIBLIOGRAPHY as Linked Open Data'. The page features a search bar at the top right with a 'SEARCH' button. Below the search bar is a navigation menu with links for Home, Getting Started, Documentation, FAQs, and Contact Us. The main content area begins with a heading 'Welcome to bnb.data.bl.uk' followed by a paragraph explaining the platform's access to the British National Bibliography via SPARQL services. A second search bar with a 'SEARCH' button is provided for plain text queries. The page concludes with a paragraph describing the history and scope of the British National Bibliography (BNB), noting its focus on the United Kingdom and the Republic of Ireland since the 1950s, and its inclusion of metadata for various types of publications.

Figure 2. British National Library website

3.4.6. National Library of Spain (NLS).

The project Linked Data on the NLS (datos.bne.es) began in 2011 with the support of the Polytechnic University of Madrid and it is published using the IFLA (International Federation of Library Associations) models and the vocabularies: FRBR,

FRAD, FRISAD, ISBD (International Standard for Bibliographic Description). The objectives of this project are: to expose the bibliographic and authority data as Linked Data for improving user experience navigating with internal and external interrelated data and moreover, and to achieve multilingualism. In the project RDF is used to add elements of several vocabularies such as: DC, RDA and to MADS / RDF. Metadata are built in FRBR and RDF from Marc21 using the software Marimba. Authority files, monographs, sound recordings were migrated establishing links to Open library, VIAF, LIBRIS, SUDOC and DBpedia [20].

Several observations were made to the model namely: field names are numeric, no comprehensible codes based on IFLA standards. This is correct internationally speaking but makes the initial mapping to consume more time. In addition, there are not reports about the reuse of published data outside the library environment. In the process, mapping problems were identified, not all basic relations of FRBR could be extracted from MARC records [21].

In the future, it is expected to work with better tools for visualization, mapping refinements, links to selected databases, connections to new data sources, better cataloging procedures. Other advances include the development of the list of topics (subjects) in SKOS, the development of NLS for schools with teachers recommending annotated content and MARC21 records converted with the ontology of Learning Object Model (LOM) enriched with other sources. It is also expected to migrate resources with specific vocabularies increasingly commons [22].

3.5. Discussion

In this section we discuss the answers to the research questions that were presented in section 2.1. A comparison of the features of the selected libraries is shown in Table II.

3.5.1. Benefits of LDDL.

The first research question was “What are the benefits of LDDL?”. The benefits of LDDL found in the libraries are summarized in the following:

- The visibility of the data is improved.
- It is possible to establish links to other online services.
- The transformation of topics in SKOS is facilitated.
- Open data recovery is improved.
- The interoperability is enabled without affecting the data source models.
- It is possible to query linked metadata from multiple institutions.
- It allows modelling things of interest related to a bibliographic resource such as people, places, events, themes.
- The end user resources annotations improve their credibility.

In 2004, W3C recommended to libraries publish their data warehouses using semantic web technologies to increase its digital impact and social utility [23]. When the library data becomes more richly linked and open the user will have improved capabilities for discovering and using data browsing in a global information graph. First, the library linked data, using RDF as a format unique, can help to improve and link the knowledge in several domains, improving the networked science. In addition, the library data could be reused for new services in the library and outside helping to improve the impact of using library resources. Finally, different kind of users will be beneficiated of this technology: librarians, scientists, students, all citizens. Despite of the benefits cited, no data have been submitted to quantify them.

3.5.2. Vocabularies and ontologies used in LDDL.

The second research question was “What are the vocabularies and ontologies used in LDDL?”.

In the analyzed libraries the following vocabularies and ontologies are used: Dublin Core, BIBO, BIO, FOAF, FRBR, FRAD, FRISAD, IFLA, ISBD, INTERMARC, MADS / RDF, XML-EAD, OAI_ORE, RDF, RDF Schema ORG, OWL, RDA, SKOS, WGS84, events Ontology. The cited vocabularies in RDF are represented using elements of RDF Schema (RDFS) and Web Ontology Language (OWL) helping to the interoperability and reusing of existing vocabularies where possible but there is difficult to match the terms without a supporting tool. Moreover, each library is developing his own tool for the RDF generation such as Marimba in the Spain National Library. Furthermore, the National Library of France has adapted the FRBR model linking information about documents, persons, organizations and subjects. Finally, the RDF data sets must to have the corresponding metadata to ensure adequate provenance identification using ontologies such as PROV-O (Provenance Ontology).

Table II. Comparison of features related to Linked Data in the selected libraries.

Library	Benefits	Vocabularies and ontologies	Problems	Future
National Library of France.	Increase data visibility. Links to other online services. Publication of topics in SKOS. Open data recovery.	FRBR, SKOS, InterMarc, XML-EAD, Dublin Core, RDF	Difficulty of cataloging resources.	Community participation in cataloging and quality control (Citizen Science).
Europeana	Enables interoperability without affecting the source data models. Provides queries to multiple linked metadata from European institutions.	RDF, OAI_ORE, SKOS, FOAF, Dublin Core	Lack of agreements to provide data, difficulty of migrating data to EDM.	Greater involvement of the community with new datasets, specifying links, improving the quality of data. Improve visibility integrating to initiatives such as Schema.org.
Library of Congress USA	Interconnection to other data sources. Model people, places and events related to a book reusing existing schemes.	BIBO, BIO DUBLIN CORE, ISBD, ORG, SKOS, RDF SCHEMA, OWL, FOAF, WGS84, RDA.	Need to develop tools to transform metadata to Linked Data. Lack of experts in different areas for the transformations.	The user could check the quality of the information sources. It's needed to develop browsers on BIBFRAME. Improve Marc21 to BIBFRAME migration tools.
British Library	Data linked to other data sources. Shape things of interest related to a bibliographic resource such as people, places, events, themes. Improved visibility.	BIBO, BIO, Dublin Core, ISBD, ORG SKOS, RDF Schema, OWL, FOAF, Ontología de eventos. WGS84.	Lack of applications consuming Linked Data.	Identify linking needs. Improve data linking tools. Obtain data use feedback.
National Library of Spain	Linking to other data sources. List of topics in SKOS. NSL scholar with annotations and recommendations.	FRBR, FRAD, FRSAD, IFLA, ISBD. RDF, DC, RDA, MADS/RDF, Dublin Core	Mapping problems: not all the basic relationships of FRBR could be extracted from MARC records.	Develop tools for visualization. Mapping refinement. To get new links. To obtain new data sources.

3.5.3. Problems

The third research question was “What are the problems reported in publications?”.

In the analyzed publications the following problems are reported:

- difficulty of cataloging resources,
- so much vocabularies for the same metadata,
- lack of agreements to provide data, difficulty of migrating data to new models,
- need to develop tools for Linked Data transformation,
- lack of experts in different areas for the transformations,
- lack of applications consuming Linked Data,
- mapping problems: for example not all basic relations of FRBR could be extracted from MARC records,
- need for more useful links of datasets,
- ownership definition and control,
- quality control of the datasets, and,
- lack of indicators about the use of Linked Data.

In this study it is observed that library community and Semantic Web community have different vocabularies for the same metadata causing a complication in the mapping process. In addition, the objectives are not well defined for the use of the RDF data sets, the potential users should be linked to this process to define priorities of data sets and applications to consume linked data. Another complex problem cited is the rights ownership, the policies and rules vary from country to country making difficult to publish the data sets on an Open Linked Data environment.

Despite the presented problems continued migration from new data sets in national libraries to RDF is worldwide observed.

3.5.4. Future trends

The fourth research question was “What are the future trends in research on LDDL?”.

In the future, it is expected:

- community participation in cataloging and quality control of published data,
- increased participation of the community with new datasets, specifying links to external sources,
- improved visibility through integration to initiatives such as Schema.org,
- quality of information sources checked by the user,
- new models browsers and data migration tools,
- data links needs to be defined,
- better tools for mapping the data links,
- feedback about data usage,
- tools for visualization, mapping refinement and data analysis,
- policies for managing RDF vocabularies and URIs,
- library data standards compatible with Linked Data, and,
- a better discussion about rights of Open Linked Data.

In short, it is generally expected a greater diffusion of this technology with the contribution of the librarians, community and the development of standards and applications to exploit Linked Data but several cited issues must be solved related to technical, administrative and social aspects such us: the development of new tools, the definition of policies to publish and manage the Web of data among others.

4. Conclusions

A growing number of national libraries worldwide are using architectures to get and publish Linked Data. Mainly, authority and bibliographic records are published using RDF data model. The vocabularies and ontologies used vary from implementations but it is possible to establish and publish equivalence relations keeping the source data models.

Benefits of the publication and use of these Linked Data from digital libraries are reported, such as improved data visibility, data linked with external resources, easy resource annotation process and reuse of data, but there are not enough reports that quantify them. Most of the reviewed libraries are achieving the 5-star in the corresponding open Linked Data model, publishing open bibliographic data with open license, in machine readable and non-proprietary format, using open standards from W3C (RDF and SPARQL) and linking the data to external sources.

There are problems to be solved for using Linked Data technologies in digital libraries such as the need of support tools, mechanisms for data quality control, better querying interfaces, the lack of technical staff with knowledge of these new technologies and difficulty of defining open Linked Data rights among others. Furthermore, few data, more of them presented like statistics, about the use of Linked Data in library evaluation were found.

It is expected to have a better number of datasets providers, achieving citizen participation in enriching data through data annotation processes and to develop more applications to consume Linked Data. It is also expected that the online library catalogues be enriched with recommendations and search rankings based on the popularity of an item and the activity of the user. Moreover, the preservation of Linked Data sets should be considered in the librarian tasks. In addition, the dissemination of best-practice design patterns will help to increase library participation in Semantic Web. In conclusion, libraries should adopt the Web of data considering the 5-star model, by publishing the data sets in the new non-proprietary open formats, linking to other resources, bringing new services and using the community to enrich the data and quality to improve the knowledge discovery and management.

Notes

1. <http://dbpedia.org>.

2. <http://linkedgoedata.org>.
3. <http://www.opnastreetmap.org>.
4. <http://evergreen-ils.org/>.
5. <http://projecthydra.org/>.
6. <http://omeka.org/>.

Acknowledgements

This work was supported by CEDIA through the project: Integration platform of bibliographic resources with Semantic Web, by the project GEODAS-BI (TIN2012-37493-C03-03) supported by the Ministry of Economy and Competitiveness of Spain (MINECO) and by the Prometeo Project from the Secretary of Higher Education, Science, Technology and Innovation (SENESCYT) of the Ecuadorian Government.

References

- [1] Berners-Lee T. Design Note: Linked Data, <http://www.w3.org/DesignIssues/LinkedData.html> (2006, accessed December 2014).
- [2] Doerr M, Gradmann S, Hennicke S, Isaac A, Meghini C and Van de Sompel H. The Europeana Data Model (EDM). In: Proc. of the World Library and Information Congress: 76th IFLA general conference and assembly, Gothenburg, Sweden 10-15 August 2010, pp. 10-15.
- [3] Kroeger A. The road to Bibframe: the evolution of the idea of bibliographic transition into a post MARC future. *Cataloging & classification quarterly*, 2103; 51(8) 873-890.
- [4] Library Linked Data Incubator Group. W3C Incubator Group Report 25 October 2011. <http://www.w3.org/2005/Incubator/ld/XGR-ld-20111025>. (2011, accessed February 2015).
- [5] Hannemann J and Kett J. Linked Data for libraries. In: Proc. of the World Library and Information Congress: 76th IFLA general conference and assembly, Gothenburg, Sweden 10-15 August 2010, pp. 1-11.
- [6] Kitchenham B, Pearl O, Budgen D, Turner M, Bailey J., and Linkman S. Systematic literature reviews in software engineering. *Information and software technology* 2009; 51(1): 7-15.
- [7] Petersen, K., Feldt, R., Mujtaba, S., and Mattsson, M. Systematic mapping studies in software engineering. In 12th International Conference on Evaluation and Assessment in Software Engineering 2008; 17(1).
- [8] Agnès S, Wenz R, Vincent M, Adrien D. Publishing bibliographic records on the Web of data: opportunities for the BnF (French National Library). In: *The Semantic Web: Semantics and big data*. 2013; 563-577.
- [9] Wenz R.. Linked Data and libraries. *Catalogue & index* 2010; 160: 2-5.
- [10] Wenz R. Linked open data for new library services: the example of data. bnf. fr. *Italian Journal of Library & Information Science*, 2013; 4(1).
- [11] Haslhofer B and Isaac A. Data. europeana. eu: The Europeana Linked Open Data Pilot. In: *International Conference on Dublin Core and Metadata Applications*, Netherlands, 21 -23 September 2011; 94-104.
- [12] Borst T, Fingerle B, Neubert J, & Seiler. A How do Libraries Find their Way onto the Semantic Web?. *Liber Quarterly*, 2010. 19(3/4), 336-343.
- [13] Aloia N, Concordia C, Meghini C and Trupiano L. EuropeanaLabs: An Infrastructure to Support the Development of Europeana. In: *Bridging Between Cultural Heritage Institutions*, Springer Berlin. 2014; 53-58.
- [14] Europeana. Europeana Data Model Primer. Technical report. August 2010. <http://version1.europeana.eu/web/europeana-project/technicaldocuments/>. (2010, accessed January 2014).
- [15] Hanshofer B, Elaheh M, Gay M, and Rainer S. Augmenting Europeana content with linked data resources. In: Proc. of the 6th International Conference on Semantic Systems. ACM. 2010; 40.
- [16] Gheen T. Library of congress launches beta release of Linked Data classification. In: *Custodia Legis*, <http://blogs.loc.gov/law/2012/07/library-of-congress-launches-beta-release-of-linked-data-classification/> (2012, accessed January 2015).
- [17] Miller E, Ogbuji U, Mueller V and MacDougall K. Bibliographic Framework as a Web of Data: Linked Data Model and Supporting Services. Technical report, Library of Congress; Washington, DC, USA, 2012.
- [18] Bizer C, Cyganiak R and Gauß T. The RDF Book Mashup: From Web APIs to a Web of Data. In: Proc. of the 3rd International Workshop on Scripting for the Semantic Web (SFSW2007), June 6, 2007, Austria, online [CEUR-WS.org/vol-248/paper4.pdf](http://www.ceur-ws.org/vol-248/paper4.pdf).
- [19] Deliot C. Publishing the British National Bibliography as Linked Open Data. *Catalog & Index*, 2014. 174, 13-18. http://www.bl.uk/bibliographic/pdfs/publishing_bnb_as_lod.pdf. (2014, accessed January 2015).
- [20] Vila Suero D and Escolano Rodríguez E. Linked Data at the Spanish National Library and the Application of IFLA RDFS Models. *IFLA ScatNews*, 2011; 35: 5-6.
- [21] Vila Suero D, Villazón Terrazas B, Gómez-Pérez A. Datos.bne.es: A library linked dataset. *Semantic Web*, 2013; 4(3): 307-313.
- [22] Pattuelli M. Personal name vocabularies as linked open data: A case study of jazz artist names. *Journal of Information Science*, 2012; 38, 6: 558-565.
- [23] Miller E. Digital libraries and semantic web. W3C.2004. <http://www.w3.org/2001/09/06-ecdl/Overview.html>.(2004, accessed June 2015).

