Linked Data Technologies in Libraries: An Appraisal

Nosheen Fatima Warraich*

Abstract

This study aims to explore the information professionals' insight about the potential benefits of Linked Data application in libraries. It also discussed the basic concepts of LD to create the awareness among information professionals.

Linked Data is an emerging set of standards and technologies that could potentially enable information environments such as libraries to publish and interlink their hidden data on the web for better global accessibility. However, it is also one of the least understood phenomena, especially with respect to its adoption and application by libraries.

This work is part of a larger study that was designed to investigate perception and application of Linked Data technologies in Pakistani libraries. This is a quantitative study based on a questionnaire survey. To meet the objective of the study, a survey of academic library information professionals was designed and circulated online to collect data from geographically scattered population.

This paper found that information professionals perceived that LD has the potential to interlink and publish library data on the Web and it will also assist patrons to discover relevant information through links to other repositories. They were of the view that LD will be useful for library patrons to find content on web and it will supplement the library's records from online sources. Respondents also opined that LD would improve users' overall search experiences with existing library catalogs.

It is the base line study on Linked Data application in libraries in Pakistani context as no literature available on it. This study provides practical suggestion for decision makers and opens the future research avenues for researchers.

1. Introduction

Linked Data (LD) is an emerging concept in service organization like libraries. The main principle of LD is to publish structured data on the Web, by using

_____

*Author is Associate Professor, Department of Information Management, University of the Punjab, Lahore – Pakistan.

Resource Description Framework (RDF) and interlinking these data items from different external data sources. Tim Berners-Lee, Internet pioneer and inventor of Linked Data term, (2006) said, "the original vision for the Internet entailed a semantically linked web of data, but this goal of Linked Data has only become technically feasible in the past decade". During the last one decade, an information society has been emerging and traditional services in all fields of life are shifting to electronic format i.e. e-learning, e-library, and e-govt. etc. The government has been investing in ICT that positively affect on academia, research and libraries. In this digital environment, academic library users have high expectations and information professionals have to design different services to make relevant information available. Linked Data is a potential technology to be used in these libraries for the better accessibility. This study is the continuation of the prevailing cutting-edge technologies in libraries.

Due to the availability of structured datasets on the web LD is used in many fields to share and exchange data. In medicine, LD technologies are used comprehensively to facilitate drug discovery that establish associations between medicines and its side effects. In media, there is examples of BBC News and The New York Times etc. NYT has published its vocabulary of almost 10,000 subject headings developed over last 150 years as Linked Data. It also has been working to expand its coverage by consuming these vocabularies and linking them with other external online resources (New York Times, n.d). BBC uses Linked Data to make content more "findable by search engines and more linkable through social media; to add additional context from supplemental resources in domains like music or sports; and to propagate linkages and editorial annotations beyond their original entry target to bring relevant information forward in additional contexts"(British Broadcasting Corporation, n.d).

Now the web has transformed the web of document to web of data to create a meaningful relationship among the related objects. Semantic Web and linked data are emerging concepts in information environment; and many studies focus on the LD initiatives in national libraries along with LD technology application in library catalog. The W3C's Library Linked Data Incubator Group offered assistance to many LD initiatives in Libraries.

Libraries have MARC-based bibliographic records that usually do not expedite its conversion to linked data compliant with RDF format. It is assumed that currently library data is unable to convert automatically due to non-standards URI tags of many objects (Greenberg & Robertson 2002, October 13–17). Gonzales (2014) stated, "the benefits of Linked Data to libraries and their users are potentially great, but so are the many challenges to its'

implementation". This point is significant for the present paper because it focuses on information professionals' perceived benefits to explore LD technology for compatible bibliographic data.

2. Literature Review

Linked Data Technology can make it possible to provide access to every piece of information on the web by extracted and converted library data into RDF.

Hooland and Verborgh (2014) in their book on "Linked Data for Libraries, Archives, and Museums" discuss concepts, method, and tools for Linked Data. This is a sort of how to do it manual that focuses on how to clean, enrich, and publish "library metadata" on the web.

De Boer et al (2013) also discussed the Amsterdam Museum Linked Open Data set. "The dataset is a five-star Linked Data representation and comprises the entire collection of the Amsterdam Museum consisting of more than 70,000 object descriptions. Furthermore, the institution's thesaurus and person authority files used in the object metadata are included in the Linked Dataset".

Miller (2011) in his article 'Linked Data and Libraries' emphasized on the role of libraries in implementing LD "not only can libraries contribute to linked data efforts, but they also are positioned to lead them. This is a huge opportunity for libraries, because they are already familiar with how to collect, curate, access, preserve, and deliver data to users". Singer (2009) pointed on interest of librarians to adopt LD "it is critical for libraries to become of the Web … if librarians have the will, Linked Data may just have the way". Salo, (2010) talked about the potential of LD in libraries with service context and mentioned "Linked data deserves special mention here, not so much for its technical details as for the mindset of building data and metadata with the express intent of easy sharing and remixing."

Lapolla (2013) study '*Perceptions of Librarians Regarding Semantic Web and Linked Data Technologies*' is relevant to the current study. The author explored the potential of Semantic Web technology application to the library catalog and librarians' levels of understanding to the key concepts and attitudes regarding Linked Data. Its target population is library cataloguer from United States. This study found that librarians participating in the study tended to believe that Semantic Web technologies are important to the future. However, it is found that there are several barriers to implement LD in libraries including lack of best practices, economic and social issues.

Relevant literature of LD adoption and application in the libraries of developing countries is rare. Overall literature establishes a positive note on the role and potential of Linked Data in libraries and describes few studies on its implementation. However, there is no empirical study conducted to know the level of interest of information professionals to adopt Linked Data technologies in libraries in developing country context.

3. Linked Data Key Terminology

Linked Data is a set of principles and best practices for sharing data and building relationships between data on the web so that machines can *understand and interpret* the data. It is a technology that creates link between systems or entities for self-describing interrelations of data accessible on the web. Basically, it is a paradigm shift from publishing data in human readable HTML documents to machine-readable documents. LD is a publishing paradigm for making data and not just human-readable documents fully accessible and inter-linkable anywhere on the Internet.

3.1 Linked Data

Linked Data refers to "data published on the Web in such a way that it is machine-readable, its meaning is explicitly defined, it is linked to other external data sets, and can in turn be linked to from external data sets. Linked Data relies on documents containing data in RDF (Resource Description Framework) format" (Coyle, 2012).
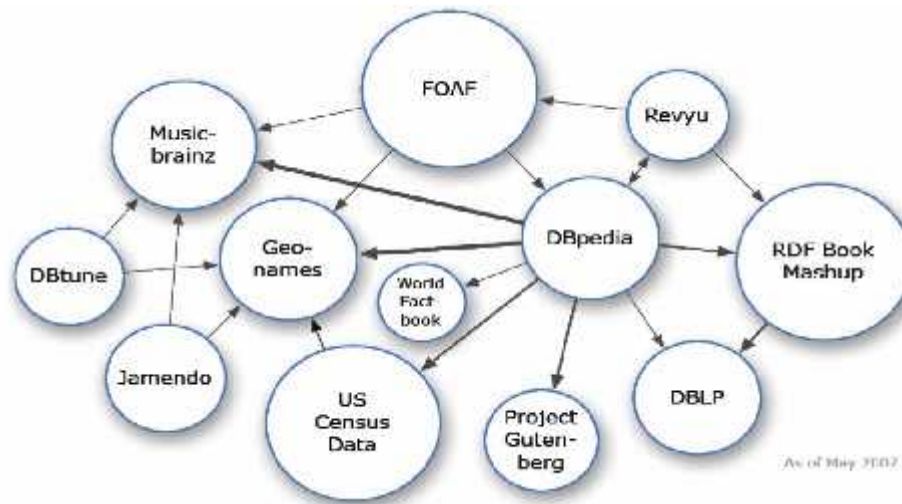

Fig. 2 Linked Data Structure

Now the web has transformed the web of document to web of data to create a meaningful relationship among the related objects. Linked Data uses RDF to link arbitrary things that create the Web of Data or we may say it as a web of things.

3.2 Linked Data Principles

Berners-Lee (2006) outlined a set of 'rules' for publishing data on the Web. These are also called 'Linked Data principles' through which all published data becomes part of a single global data space. Following LD principals provide a simple approach to connect and publish data on the web by adhering to its architecture and standards:

- "Use URIs as names for things
- Use HTTP URIs so that people can look up those names
- When someone looks up a URI, provide useful information, using the standards (RDF, SPARQL)
- Include links to other URIs, so that they can discover more things"

3.3 Resource Description Framework (RDF)

RDF provides a way to describe anything you encounter in the world: people, places, concepts, words, documents, etc. RDF provides a way to describe relationships: how one thing relates to another thing. RDF was adopted as a W3C recommendation in 1999.

Following is the Components of RDF:

Statement is a combination of *a resource*, *a property*, and *a property value.* It usually referred to as Triples.

- Resource a Subject
- Property a Predicate
- Property Value an Object

We may provide the example of RDF triples as follows:

"The trees have the colour green ". In this case a subject represent "the trees", a predicate represent " have the colour ", and an object represent "green". Therefore, RDF has a resource, a property, and a property value in object-oriented design: a resource (trees), a property (color), and a property value (green).

3.4 Five-star Linked Data

Tim Berners-Lee, the inventor of the Web and the Linked Data, suggested a 5 star scheme for Linked Data. The 5 Star Linked Data system is cumulative. Each additional star supposes that data fulfill the standards of the prior stage(s). The details of 5 stars are given below:

"☆ Data is available on the Web, in whatever format.

☆☆ Available as machine-readable structured data, (i.e., not a scanned image)

☆☆☆ Available in a non-proprietary format, (i.e, CSV, not Microsoft Excel)

☆☆☆☆ Published using open standards from the W3C (RDF and SPARQL).

☆☆☆☆☆ Link your data to other data to provide context"

3.5. SKOS (Simple Knowledge Organization System)

Simple Knowledge Organization System (SKOS), a data standard for sharing and linking knowledge organization systems i.e. such as thesauri, taxonomies, classification schemes and subject heading systems (for example Library of Congress subject heading), via the Web. SKOS share a similar structure, and are used in similar applications.

The SKOS offers a standard and economical migration path for transferring and sharing existing knowledge organization systems to the Semantic Web. SKOS may be used with formal knowledge representation languages such as the Web Ontology language (OWL).

4. Research Objectives

Linked Data is an emerging set of standards and technologies that could potentially enable information environments. However, it is also one of the least understood phenomena, especially with respect to its adoption and application by libraries. Following are the objectives of this study:

- To present a discussion of Linked Data key terminology to make information professionals familiar with this concept
- To explore the potential benefits to implement LD technologies in information environment as perceived by professionals

5. Research Design

This work is part of a larger study that was designed to investigate perception and application of Linked Data technologies in Pakistani libraries. To meet the objective of the study, an online questionnaire survey was used to collect data from information professionals working in higher education institutes in Pakistan.

There is no sampling frame available to do the random sample. There are 177 public and private sector universities in Pakistan and almost 350 professionals are working in theses libraries & information centers. Social media networks, emails, and phone calls were used to contact the potential population (N=350) and get their response. The questionnaire link was shared on social media websites along with professional list-serves.

6. Data Analysis

6.1 Demographic information

During the data cleaning process, it was found that 50 responses were partially filled among the 136 received questionnaires. The fully completed 86 responses were used for analysis. The response rate is almost 24.57%.

Data shows that more than half of the respondents 45(52.33%) have designation '*librarian*' as it is the most common title among information professionals in Pakistan. Librarian is considered the initial professional rank in organizational hierarchy after assistant librarian. There are 13(15.12%) early career professionals whose designation is '*assistant librarian'*. This is the first professional grade in many public sector institutes, then librarian, senior librarian; deputy librarian and chief librarian are in the hierarchy. Most of the organizations promote them on next level on the basis of length of service. Six (6.98%) have the title of 'senior librarian' followed by only one (1.16%) has 'deputy chief librarian'. Eight respondents (9.30%) have been working as 'chief librarian', in-charge or director of libraries.

Thirteen respondents (15.12%) respondents tick 'others' and mentioned their designations as Archive officer, Metadata librarian, Assistant Director, Regional library coordinator and academic designations i.e. lecturer, assistant and associate professor.

Data shows that a significant number of respondents, 58(67.44%), had masters' degree followed by 25(29.7%) had MPhil degree. MPhil is a research degree equivalent to 18 years of education. It is a 30-credit program

containing 24 credit hours coursework followed by 6-credit research thesis. Only 3 (3.49%) respondents have completed their doctorate after MPhil and have been working on academic positions.

6.2 Potential Benefits to Adopt Linked Data Technology in Libraries as perceived by Respondents

Respondents were asked to mention the potential benefits to adopt to adopt LD technologies in libraries. There were seven statements in this construct to understand information professionals' perceived benefits on LD initiatives in libraries. The Cronbach's Alpha value of these statements is 0.877 that shows a good reliability of items.

Figure 1 shows that respondents are interested in 'LD has the potential to interlink and publish library data on the Web' with highest mean value 5.59. It is considered good mean value on the 7-point scale. However the Standard Deviation (1.508) that is also highest among these seven statements. It shows that respondents' opinion on publish library data on the Web have deviation from each other.
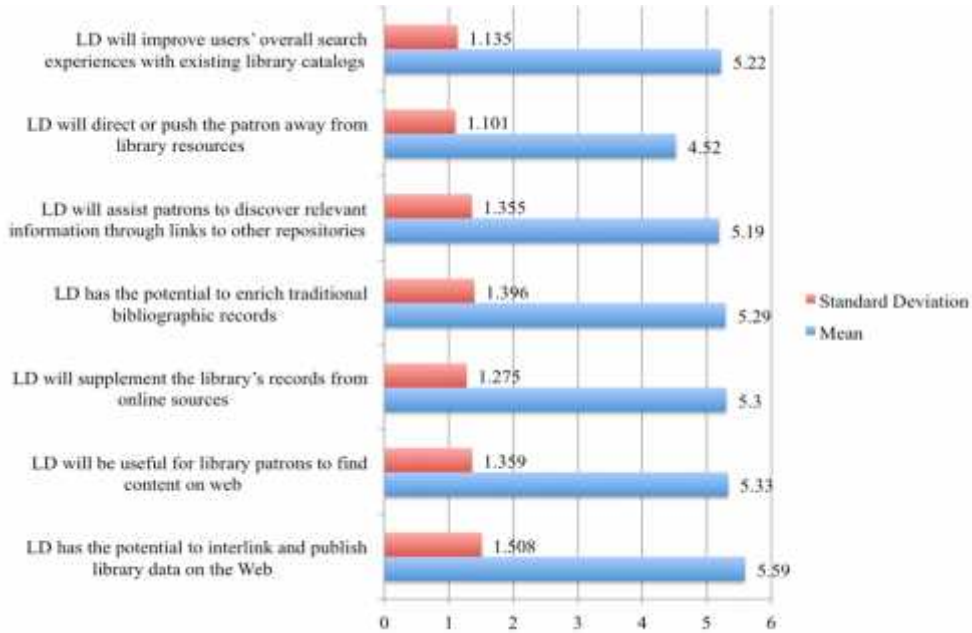

Fig. 2Potential Benefits to Adopt LD Technologies in Libraries

Scale: 1=Strongly Disagree, 2=Disagree, 3=Slightly Disagree, 4=Neutral (Neither Disagree nor Agree), 5=Slightly Agree, 6=Agree, 7=Strongly Agree, Don't Know/Not sure/Not Applicable

Respondents were also thought that 'LD will be useful for library patrons to find content on web' with the second highest mean (5.33) and SD (1.359). The SD (1.359) value indicates that respondents' opinion is not numerically very different with each other.

The statement '*LD will supplement the library's records from online sources'* has third highest mean value (5.30) with third lowest SD (1.275). Respondents' opinions on ' *LD will direct or push the patron away from library resources'* are more consistent with lowest SD (1.101) and lowest mean (4.52). It shows that respondents thought that LD will not push the user from library resources and they will use library resources in future.

However, respondents opined that *'LD has the potential to enrich traditional bibliographic records'* with second lowest mean (5.19) and SD (1.355). Data shows that traditional bibliographic records will have better accessibility by adopting LD technologies. Overall, respondents are positive and optimistic about the adoption of LD technologies in libraries and know about its potential benefits.

The statement '*LD will assist patrons to discover relevant information through links to other repositories'* has second lowest mean (5.19) with SD=1.355. It shows that respondents perceived that LD would be very helpful to library users to explore other information sources from repositories those are designed and developed to provide free access to them.

It is interesting to note that respondents are enthusiastic regarding LD and highlighting its benefits. They identify that *'LD will improve users' overall search experiences with existing library catalogs'* (Fig. 2). It will consequently improve the library services and enhance users' satisfaction with library.

The findings of this survey indicate that information professionals are very keen to get the benefits from LD technologies to improve library sources and services that will subsequently enhance the user' satisfaction to library. Survey respondents are certainly curious to avail the potential benefits of LD technologies by interlinking and publishing library data on the Web. One point is worth noting that that it is their self-perception and no method is used to verify these findings experimentally.

7. Conclusion

This study describes the basic notion of Linked Data technology in information environment. The findings show the potential benefits to learn about and implement LD technologies in libraries as perceived by information professionals. Respondents are eager to avail the benefits of LD technologies for better service quality to patterns in their institutes. To foster the research and to set LD best practices in libraries and other cultural heritage institutes, there is a potential to invest in this area in terms of financial and social capital. On the basis of the finding of the study it is recommended that training programs are substantial for information professionals to get them updated with Linked Data application.

For future research, it is needed to map the Linked data constructs qualitatively by conducting interview with these professionals to find out more about LD adoption in libraries. There is a potential to explore the ethical, social and technological issues regarding LD application in libraries and cultural heritage. This is a multidisciplinary concept that is being adopted in different sphere of life. LD also needs to be explored in the context of E-Government initiatives in Pakistan

References

Berners-Lee, T., &Swick, R. (2006).*Semantic Web Development.* Massachusetts Inst of Tech Cambridge.

British Broadcasting Corporation (n.d). Retrieved from: http:// www. Cmswire .com/cms/information-management/bbcs-adoption-of-semantic-web-technologies-an-interview-017981.php?pageNum=2

Coyle, K. (2012). Chapter 2: Semantic Web and Linked Data. Library Technology Reports 48(4): 10-14.

De Boer, V., Wielemaker, J., van Gent, J., Oosterbroek, M., Hildebrand, M., Isaac, A., ... & Schreiber, G. (2013). Amsterdam museum linked open data. Semantic Web, 4(3), 237-243.

Gonzales, B. M. (2014). Linking libraries to the web: linked data and the future of the bibliographic record. *Information Technology and Libraries (Online)*, *33*(4), 10.

Greenberg, J., & Robertson, W. (2002, October 13–17). Semantic Web construction: An inquiry of authors' views on collaborative metadata generation. In *Proceeds of the International Conference on Dublin Core and Metadata for e-Communications,* Florence, Italy (pp. 45–52). Retrieved from http://dcpapers.dublincore.org/pubs/article/view/693/689

Greenberg, J., White, H. C., Carrier, S., & Scherle, R. (2009). A metadata best practice for a scientific data repository. Journal of Library Metadata 9(3-4): 194-212.

Hannemann, J. and J. Kett (2010). Linked data for libraries. Proc of the world library and information congress of the Int'l Federation of Library Associations and Institutions (IFLA).

Hooland, S., & Verborgh, R. (2014). Linked Data for Libraries, Archives and Museums: How to clean, link and publish your metadata. Facet.

LaPolla, F. (2013). Perceptions of Librarians Regarding Semantic Web and Linked Data Technologies. Journal of Library Metadata 13(2-3): 114-140.

Miller, E. and M. Westfall (2011). Linked data and libraries. The Serials Librarian 60(1-4): 17-22.

New York Times (n.d). Retrieved from: http://data.nytimes.com

Singer, R. (2009). Linked library data now!.*Journal of Electronic Resources Librarianship*, *21*(2), 114-126.