



ELK
Asia Pacific Journals

www.elkjournals.com

OVERVIEW OF THE LAND ESCAPE OF LINKED OPEN DATA IN DIGITAL LIBRARY PERSPECTIVE

GIZEALEW ALAZIE DAGNAW

(Information Science, faculty of Informatics/
University of Gondar, Gondar, Ethiopia,
gizeinstra@gmail.com

SISAY EBABYE TSIGIE

(Information Science, faculty of Informatics/
University of Gondar, Gondar, Ethiopia,
Sis.ebe2007@gmail.com

ABSTRACT

This study aims to explore about the adoption of Linked Data technologies along with their level of interest to adopt these technologies in different libraries. This review examines the quality of linked bibliographic data published by the national libraries of Spain, France, the United Kingdom, and Germany. The examination is mainly based on a statistical study of the vocabulary usage and interlinking practices in the published data sets. The review finds that the national libraries successfully adapt established Linked Data principles, but issues at the data level can limit the fitness of use. In addition, the review reveals that these four libraries have chosen widely different solutions to all the aspects examined. 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 us 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 review about the current uses of Linked Data in digital libraries, including the most important implementations around the world, is presented. The study focuses on selected vocabularies and ontologies, benefits and problems encountered in implementing Linked Data in digital libraries. In addition, it also identifies and discusses specific challenges that digital libraries face, 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 Linked Data Incubator Group in 2018, and scientific publications from Google Scholar, Scopus, ACM and Springer from the last 5 years. The selected libraries for the study are the National Library of France, the Europeana Library, and the Library of Congress of the USA, the British Library and the National Library of Spain. In this paper, we outline the best practices found in each experience and identify gaps and future trends.

Key word:- linked open data, digital library, open data

Introduction

Open data is information that is provided by public entities to be accessed and reused. Publishing open

data is an excellent way to improve an organization's transparency and provide insight into the value of the organization. Libraries are uniquely

positioned to assist their patrons in the use of open data by making them aware of it and helping them access and use it[1]. This article provides a short history of open data and explores ways that library trailblazers are already using open data as well as contributing open data for others to use. Getting started with open data involves identifying key open data resources such as data.gov, identifying library information that would be beneficial to publish, and creating programs that provide digital literacy training and create opportunities for patrons to engage with open data in new and creative ways[2].

Linked Open Data (LOD) is a core Semantic Web technology that makes knowledge and information spaces of different knowledge domains manageable, reusable, shareable, exchangeable, and interoperable. The LOD approach achieves this through the provision of services for describing, indexing, organizing, and retrieving knowledge artifacts and making them available for quick consumption and publication. This is also aligned with the role and objective of traditional library cataloging. Owing to this link, major libraries of the world are transferring their bibliographic metadata to the LOD landscape[3]. Some developments in this direction include the replacement of Anglo-American Cataloging Rules 2nd Edition by the Resource Description and Access (RDA) and the trend towards the wider adoption of BIBFRAME 2.0. An interesting and related development in this respect is the discussions among knowledge resources managers and library community on the possibility of enriching bibliographic metadata with socially curate or user-generated content. The popularity of Linked Open Data and its benefit to librarians and knowledge management professionals warrant a comprehensive survey of the subject. Although several reviews and survey articles on the application of Linked Data principles to cataloging have appeared in literature, a generic yet holistic review of the current state of Linked and Open Data in cataloging is missing. To fill the gap, the authors have collected recent literature (2014–18) on the current state of Linked Open Data in cataloging to identify research trends, challenges, and opportunities in this area and, in addition, to understand the potential of socially curate metadata in cataloging mainly in the realm of

the Web of Data[4]. To the best of the authors' knowledge, this review article is the first of its kind that holistically treats the subject of cataloging in the Linked and Open Data environment. Some of the findings of the review are: Linked and Open Data is becoming the mainstream trend in library cataloging especially in the major libraries and research projects of the world; with the emergence of Linked Open Vocabularies (LOV), the bibliographic metadata is becoming more meaningful and reusable; and, finally, enriching bibliographic metadata with user-generated content is gaining momentum. Conclusions drawn from different study include the need for a focus on the quality of catalogued knowledge and the reduction of the barriers to the publication and consumption of such knowledge, and the attention on the part of library community to the learning from the successful adoption of LOD in other application domains and contributing collaboratively to the global scale activity of cataloging [3].

Linked Data

Linked open data is expressed using a simple three-part data structure: the "RDF Triple," as defined by a W3C standard, Resource Description Framework (RDF). An individual RDF Triple makes a simple statement of the form "A is related to B," as in "Book X was authored by Person Y" or "Book X has the title *Being There*." Some elements of RDF statements are string values, such as "Being There." For all other elements, RDF leverages the globally managed address space of uniform resource identifiers (URIs) on the World Wide Web – for example, the ubiquitous <http://> URL – as a source of identifiers to denote not just specific web pages, but also non-digital entities such as people, books, and concepts. URIs makes people, books, and concepts globally citable. RDF triples use "vocabularies" of properties (to express relationships such as "was authored by") and classes (to say that Resource X is a "book"). Vocabularies bring meaning to data. RDF vocabularies are themselves expressed as linked data – i.e. they are defined using RDF triples and published on the web, using URIs as identifiers for their terms. In today's web environment, RDF vocabularies are created by a wide range of people and institutions, from individual researchers to

national libraries and for-profit corporations, for a wide range of descriptive requirements [2].

The Semantic Web isn't just about putting data on the web. It is about making links, so that a person or machine can explore the web of data. With linked data, when you have some of it, you can find other, related, data[5].

Like the web of hypertext, the web of data is constructed with documents on the web. However, unlike the web of hypertext, where links are relationships anchors in hypertext documents written in HTML, for data they links between arbitrary things described by RDF. The URIs identifies any kind of object or concept. But for HTML or RDF, the same expectations apply to make the web grow[6]:

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

The Evolution of the Idea of Bibliographic Transition into a Post MARC Future

idea of replacing MARC with a linked-data metadata structure, covering the period from 2002 through the 2012 release of the draft of the proposed bibliographic framework, BIBFRAME. Works proposing the replacement of MARC or exploring linked data in a library context are examined. In particular, key documents leading to the creation of the Library of Congress Bibliographic Framework Transition Initiative are examined, along with some of the critical responses they received, to better understand the chain of ideas shaping BIBFRAME[7].

Literature review

Open Data: What it is and why we should be concerned

Open Data refers to data collected and shared with others to use as they wish, without restrictions on copyright or usage. Traditional examples of open data include government-collected data (e.g., weather reports, crime incident reports, postal or Zip codes) as well as some academic sources (e.g., open-access journals, raw polling and survey data, scientific experiment results). Businesses are also realizing the benefits of sharing data and making it available for use[8].

***RDA: Resource description & access*—>a survey of the current state of the art**

Resource Description & Access (RDA) is intended to provide a flexible and extensible framework that can accommodate all types of content and media within rapidly evolving digital environments while also maintaining compatibility with the Anglo American Cataloguing Rules, 2nd edition (AACR2)[9]. The cataloging community is grappling with practical issues in navigating the transition from AACR2 to RDA; there is a definite need to evaluate major subject areas and broader themes in information organization under the new RDA paradigm. This article aims to accomplish this task through a thorough and critical review of the emerging RDA literature published from 2011 to 2018[10]. The review mostly concerns key areas of difference between RDA and AACR2, the relationship of the new cataloging code to metadata standards, the impact on encoding standards such as Machine-Readable Cataloging (MARC), end user considerations, and practitioners' views on RDA implementation and training. Future research will require more in depth studies of RDA's expected benefits and the manner in which the new cataloging code will improve resource retrieval and bibliographic control for users and catalogers alike over AACR2. The question as to how the cataloging community can best move forward to the post-AACR2/MARC environment must be addressed carefully so as to chart the future of bibliographic control in the evolving environment of information production, management, and use[11].

Quality of Linked Bibliographic Data: The Models, Vocabularies, and Links of Data Sets Published by Four National Libraries

Berners-Lee (2006) introduced principles for Linked Data, large quantities of bibliographic descriptions have been published on the Web, resulting in linked bibliographic data (LBD). Linked Data principles are intended to facilitate a Semantic Web of data, enabling a variety of novel applications. A satisfactory level of output quality is essential to realize this vision. The library community continuously discusses issues concerning involved operations, such as data modeling, transformation, and interlinking. Less effort, however, has been devoted to systematic examination of the actual output, particularly the organization of data and various aspects of data quality. This article examines bibliographic metadata published as Linked Data by four European national libraries[7]:

How do prominent agents (and experts) in the library community organize and represent bibliographic collections of metadata when they publish these collections as Linked Data on the Web? How do these Linked Data sets conform to established measurements of Linked Data quality for vocabulary usage and interlinking? To answer these questions, concrete dimensions of Linked Data quality are analyzed statistically. A qualitative close reading of selected corpus samples supplements the statistical data. The first section of this article presents background information on LBD data and quality dimensions, clarifying the scope of the study. The following sections summarize previous research and present the corpus data and methodological considerations[12].

Linked data Berners-Lee (2006) first described Linked Data identifying four principles to help support bottom-up adoption of the Semantic Web: Use Uniform Resource Identifiers (URIs) as names for things. Use HTTP URIs so people can look up those names. When someone looks up a URI, provide useful information, using the standards (Resource Description Framework (RDF), SPARQL protocol and RDF query language (SPARQL)). Include links to other URIs, so that

users can discover more things. To further “encourage people along the road to good Linked data,” Berners-Lee (2006) later added a rating system of five stars reflecting these principles. The principles have since evolved into comprehensive collections of best practice recommendations, both as general guidelines (see, e.g., Heath & Bizer, 2011; Hyland, Atemezing, & Villazón-Terrazas, 2014) and as guidelines targeting data providers in specific domains (e.g., van Hooland & Verborgh, 2014). The emphasis on standards and transparency indicates a lingua franca approach to solving hetero genetic conflicts across domains and datasets[13].

Linked Open Vocabularies (LOV)

One of the major barriers to the deployment of Linked Data is the difficulty that data publishers have in determining which vocabularies to use to describe the semantics of data. This systematic report describes Linked Open Vocabularies (LOV), a high-quality catalogue of reusable vocabularies for the description of data on the Web. The LOV initiative gathers and makes visible indicators such as the interconnections between vocabularies and each vocabulary's version history, along with past and current editor (individual or organization)[14]. The report details the various components of the system along with some innovations, such as the introduction of a property-level boost in the vocabulary search scoring that takes into account the property's type (e.g., dc:comment) associated with a matching literal value. By providing an extensive range of data access methods (full-text search, SPARQL endpoint, API, data dump or UI), the project aims at facilitating the reuse of well-documented vocabularies in the Linked Data ecosystem. The adoption of LOV by many applications and methods shows the importance of such a set of vocabularies and related features for ontology design and the publication of data on the Web[4].

Where Are We Headed? Resource Description and Access, Bibliographic Framework, and the Functional Requirements for Bibliographic Records Library Reference Model

The technical services world is in a state of chaotic transformation at this moment, and undoubtedly, librarians are feeling the growing pains. Three major initiatives in the cataloging field are driving the revolution, and their adoption will mean big changes in the way that catalogers and metadata specialists approach their work. Given the many projects, models, and papers being disseminated and discussed within the library community, the question becomes, what will library cataloging and metadata creation look like in the next decade? Where are we headed? Also, what knowledge and skills will be needed to function in this increasingly digital and mechanized world? The metadata and data management initiatives that are currently being developed, both in and outside the library realm, are numerous. The most transformative for those providing metadata and cataloging services in libraries. The first is the International Federation of Library Associations and Institutions (IFLA)'s Functional Requirements for Bibliographic Records (FRBR) Library Reference Model (FRBR-LRM; Riva, Le Boeuf, & Žumer, 2016) which seeks to harmonize the various FRBR models into one[12].

Linked Data Technologies and What Libraries Have Accomplished So Far

What is Linked Data? According to David Wood, the co-chair of the W3C's (World Wide Web Consortium) RDF Working Group which lays the foundation for Linked Data and the Semantic Web, Linked Data is a set of techniques to represent and connect structured data on the web. Linked Data makes the World Wide Web into a global database that we call the Web of Data[15]. Linked Data technologies, with its broader concept, Semantic Web, have gained rapid momentum and popularity on the World Wide Web. The Linked Data technologies hold the potential to evolve the current Web of document into the Web of Data. Imagine that in the future Internet world, not only web documents but all data are Wang & Yang /

International Journal of Librarianship, connected. More importantly, these connected data are not only accessible to human but to machine also. In other words, all devices that are connected on the Internet can access and process those linked data and thereby make smart decisions automatically. This will greatly enhance the way we access information and make informed decisions[16].

W3C Library Linked Data Incubator Group Final Report published

The W3C Library Linked Data Incubator Group has published its final report at <http://www.w3.org/2005/Incubator/lld/XGR-lld-20111025/>, examining how Semantic Web standards and Linked Data principles can be used to make the information assets of libraries more widely visible and re-usable on the Web. The report is supplemented by descriptions of use cases and by an inventory of datasets, value vocabularies, and metadata element sets currently available as Linked Data[17].

Adoption of linked data technologies among university librarians in Pakistan: Challenges and prospects

This study aims to explore librarians' perspective about the adoption of Linked Data technologies along with their level of interest to adopt these technologies in Pakistani university libraries. It also identifies their perceived hindrances that obstruct them to adopt and effectively implement Linked Data technologies in libraries. This is a quantitative study based on survey research design. To meet the objective of the study, a self-constructed and validated questionnaire was used to collect data. Findings show that university librarians believe in the effective adoption of Linked Data technology in libraries. They perceive that Linked Data technology can enhance navigation between the traditional online tools to access library resources and Linked Data will soon be the standard for creating metadata and records for information resources management in libraries. They are eager to attend events about Linked Data application in their libraries and willing to explore ways to incorporate Linked Data standards in bibliographic records

management as well. Barriers do exist such as the general lack of awareness of basic Linked Data concepts and best practices for this emerging technology. To foster the research and to set Linked Data best practices in libraries and other cultural heritage institutions, there is a potential to invest in this area in terms of financial and social capital [2].

Report Finding

For the past few years libraries have been working attentively towards Linked Data and the Semantic Web. Due to the complexity and vast scope of Linked Data, many people have a hard time to understand its technical details and its potential for the library community. This paper aims to help librarians better understand some important concepts by explaining the basic Linked Data technologies that consist of Resource Description Framework (RDF), the ontology, and the query language[9]. It also includes an overview of the achievements by libraries around the world in their efforts to turn library data into Linked Data including those by Library of Congress, OCLC, and some other national libraries. Some of the challenges and setbacks that libraries have encountered are analyzed and discussed. In spite of the difficulties, there is no way to turn back[10], [9].

For many in the library field, linked open data (LOD) is both a common and an enigmatic phrase. Linked data has been the topic of many articles, books, conference presentations, and workshops in recent years. The topic, however, is one that many are still working to understand. idea of replacing MARC with a linked-data metadata structure, covering the period from 2002 through the 2012 release of the draft of the proposed bibliographic framework, BIBFRAME. Works proposing the replacement of MARC or exploring linked data in a library context are examined[12]. Library cataloging and metadata are undergoing a major revolution. The International Federation of Library Associations and Institution's Functional Requirements for Bibliographic Records Library Reference Model, the ongoing development of Resource Description and Access, and the Library of Congress' Bibliographic Framework initiative are the three programs currently under way which will, in all probability, profoundly affect how library bibliographic data is recorded, stored, and retrieved.

These initiatives will also allow library holdings, for the first time, to be visible on the Web and discoverable for users, and, therefore, sharable with other cultural resource communities. There are several definitions of linked data in library literature. It's variously described as a set of best practices required for publishing and connecting data that is structured in a way that machines can use it and also as the actual relationships between data objects in a much more granular fashion than the current relationships or links between documents on the web that we have now that is made readable by both machines and humans[18].

Discussion

Ontology and Linked Data (LD) are the two prominent web technologies that have emerged in the recent past. Both of them are at the center of Semantic Web and its applications[1]. Linked Open Data (LOD) is a core Semantic Web technology that makes knowledge and information spaces of different knowledge domains manageable, reusable, shareable, exchangeable, and interoperable. The LOD approach achieves this through the provision of services for describing, indexing, organizing, and retrieving knowledge artifacts and making them available for quick consumption and publication. This is also aligned with the role and objective of traditional library cataloging. 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 us 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 around the world, is presented. The study focuses on selected vocabularies and ontologies, benefits and problems encountered in implementing Linked Data in digital libraries. In addition, it also identifies and discusses specific challenges that digital libraries face, 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[2]. It is based

on the information found in the library websites recommended by W3C Library Linked Data Incubator Group in 2011, and scientific publications from Google Scholar, Scopus, ACM and Springer from the last few years. The selected libraries for the study are the National Library of France, the European Library, and the Library of Congress of the USA, the British Library and the National Library of Spain. In this paper, we outline the best practices found in each experience and identify gaps and future trends[3].

Conclusion

This overview describes recent developments in the provision of data services from the perspective of a research library. Some of the issues concerning re-dissemination and communication in a community of advanced users may shed light on the overall relationship between data providers and users. As the open data movement matures, it is critical that we do not create greater digital inequities by assuming that everyone and every organization will have the same time, skills and resources to invest in learning how to find and use this data. This study aims to explore about the adoption of Linked Data technologies along with their level of interest to adopt these technologies in deferent libraries[19]. It also identifies their perceived hindrances that obstruct them to adopt and effectively implement Linked Data technologies in libraries. They perceive that Linked Data technology can enhance navigation between the traditional online tools to access library resources and Linked Data will soon be the standard for creating metadata and records for information resources management in libraries. They are eager to attend events about Linked Data application in their libraries and willing to explore ways to incorporate Linked Data standards in bibliographic records management as well. Barriers do exist such as the general lack of awareness of basic Linked Data concepts and best practices for this emerging technology[7]. It makes sense to integrate open data training and knowledge in these important community institutions. By empowering librarians with the knowledge to find and use this data, and creating tools for libraries to teach everyone in their community to find and use this data. Linked data is not yet widely used in any field, including in

libraries. There are some barriers to libraries deciding to use it, including the relative complexity of the technology; the aversion to risk many libraries have; and economic, political, and system limitations. However, some fairly recent forays into linked data outside of libraries have shown that the technology can be useful. Both Google's Knowledge Graph and Face book's Open Graph Protocol use linked data to, respectively, find related information in searches and connect people based on interests and common contacts. If the library uses linked data to present its materials on the web, a search for a book could bring up a similar graph, with information about the book and its availability in the library at the moment of searching. Links to the author and publisher and other information about the book would also be made available through the graph interface[20].

Reference

- [1] T. Bourke, "Integrating Data Collections with Open Data Resources: a communication-centred approach," no. June, pp. 1–4, 2017.
- [2] B. Yoose and J. Perkins, "The Linked Open Data Landscape in Libraries and Beyond," *J. Libr. Metadata*, vol. 13, no. 2–3, pp. 197–211, 2013.
- [3] Y. Wang and S. Q. Yang, "Linked Data Technologies and What Libraries Have Accomplished So Far," *Int. J. Librariansh.*, vol. 3, no. 1, p. 3, 2018.
- [4] N. F. Warraich and A. Rorissa, "Adoption of linked data technologies among university librarians in Pakistan: Challenges and prospects," *Malaysian J. Libr. Inf. Sci.*, vol. 23, no. 3, pp. 1–13, 2018.
- [5] J. Manyika, M. Chui, P. Groves, D. Farrell, S. Van Kuiken, and E. A. Doshi, "Open Data: Unlocking Innovation and Performance with Liquid Information," *McKinsey*, no. October, p. 24, 2013.
- [6] C. Ebikeme, S. Hodson, G. Boulton, H. Hackmann, A. Stevance, and L. Spini, "Open Data in a Big Data World: challenges and opportunities for sustainable development," *Br. GSDR*, pp. 1–3, 2016.
- [7] K. Tallerås, "Quality of Linked Bibliographic

- Data: The Models, Vocabularies, and Links of Data Sets Published by Four National Libraries,” *J. Libr. Metadata*, vol. 17, no. 2, pp. 126–155, 2017.
- [8] “Supporting sustainable development with open data | Open Data Institute 2015,” pp. 1–29, 2015.
- [9] S. Mahaboob, P. Kanakam, D. Suryanarayana, S. Gunnam, and S. S., “Semantic Information Retrieval: An Ontology and RDF-based Model,” *Int. J. Comput. Appl.*, vol. 156, no. 9, pp. 34–38, 2016.
- [10] F. Gandon, R. Krummenacher, S.-K. Han, and I. Toma, *The Resource Description Framework and its Schema. Handbook of Semantic Web Technologies*, no. November 2015. 2011.
- [11] J. Mynarz, “Linked Open Data for Public Sector Information,” 2012.
- [12] A. Kroeger, “The Road to BIBFRAME: The Evolution of the Idea of Bibliographic Transition into a Post-MARC Future,” *Cat. Classif. Q.*, vol. 51, no. 8, pp. 873–890, 2013.
- [13] S. Chattopadhyay, “Digital Reference Service using E-Resources : a Study at St . Xavier â€™s College Central Library , Kolkata Digital Reference Service using E-Resources : a Study at St . Xavier ’ s College Central Library , Kolkata .,” no. February, 2018.
- [14] F. Bauer and M. Kaltenböck, *Linked Open Data: The Essentials. edition mono/monochrom.* 2011.
- [15] R. Shaw, R. Troncy, and L. Hardman, “LODE: Linking open descriptions of events,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 5926 LNCS, no. Section 2, pp. 153–167, 2009.
- [16] J. Houghton and N. Gruen, “Open Research Data,” no. November, 2014.
- [17] Group on Earth Observations, “The Value of Open Data Sharing,” p. 42, 2015.
- [18] F. Matsumura, I. Kobayashi, F. Kato, T. Kamura, I. Ohmukai, and H. Takeda, “Producing and consuming linked open data on art with a local community,” *CEUR Workshop Proc.*, vol. 905, 2012.
- [19] D. D. Shonrock, “Impact of Digital Technology on Library Collections and Resource Sharing (review),” *portal Libr. Acad.*, vol. 3, no. 4, pp. 689–690, 2007.
- [20] E. E. Abdelaziz and S. M. Kaffas, “Proposal for Implementing Linked Open Data on Libraries Catalogue,” pp. 1–9, 2018.