

Linking Data for Library Services

*Sneha Tripathi**

Sayaji Rao Gaekwad Library, Banaras Hindu University, Varanasi, Uttar Pradesh, India

Abstract

Since its inception, the web is changing and evolving continuously to make the web more structured and accessible. It has evolved through the “web of hypertexts” into “web of data” where the data gets the primary citizenship of the web. Linked Data is a data-publishing technique that enables related data to be connected and easily accessible using common Web technologies. The data published could be read and interpreted not only by humans but also by machines. Thus linked open data seeks to enhance the capacity of web from mere a data-storehouse to an intelligent, responsive web that compliments human inquest. The following study elaborates various aspects of linked open data and its implementation to enhance library services to the community.

Keywords: *Linked data, Linked Open Data, Libraries, Community Libraries*

***Author for Correspondence** E-mail: snehatripathi@gmail.com

INTRODUCTION

The Web technology has been changing and evolving continuously to make the web more structured and accessible. Initially, it was only a “web of hypertexts” where the web documents were interlinked through hypertexts given within the documents. It was meant only for humans to read and interpret. Gradually, it has evolved into “web of data” by publishing and interlinking open data sources on the Web based on well-established standards such as; RDF (Resource Description Framework) and URIs (Unique Resource Identifiers) [1,2]. The data published could be read and interpreted not only by humans but also by machines. Thus linked open data seeks to enhance the capacity of web from mere a data-storehouse to an intelligent, responsive web that compliments human inquest.

Linked Data is a data-publishing technique that enables related data to be connected and easily accessible using common Web technologies. Linked Open Data realizes this in an open manner, where everyone can access, re-use, enrich and share the data published. The web of linked data represents a collection of open and linked structured data on the Web. It calls upon two initiatives:

- (a) the Linked Data initiative and
- (b) the Linking Open Data initiative.

“Linked Data” was introduced by Tim Berners-Lee in 2006 in context to Semantic Web where he emphasized on establishing links between the data from various resources to enhance the current web through structured data. Linked Data uses RDF to establish typed links. The vision of the linked data is to interlink every piece of data that exists on web [3].

Linked Open Data

Linked Open Data (LOD) is data that is openly available on the Internet in a machine readable form, and linked to other data using RDF triples. However, there exists considerable difference between the two concepts, i.e., Linked Data and Open Data. A Linked data is one that generally complies with the four principles of Tim Berners-Lee explained in the following sections whereas, an Open data refers to the data which is available free for use and re-use. According to Miller, “A Linked Data may be open, and Open Data may be linked, but it is equally possible for Linked Data to carry licensing or other restrictions that prevent it being considered open or for Open Data to be made available in ways that do not easily enable linking [4]”.

Linking Open Data (LOD) is a W3C-SWEO (Semantic Web Education and Outreach) community project. It aims to

facilitate the emergence of a web of linked data, by means of publishing and interlinking open data on the Web in RDF.

The central dataset of LOD is DBpedia - an RDF extract of the Wikipedia. DBpedia is a sort of a hub in the LOD graph, which guarantees certain level of connectivity. It also provides easy entry points to find resources of interest using their Wikipedia names in DBpedia and through it in the LOD network.

The RDF Triples

The Linked Data has following basic tenets [5]:

- a. use the RDF data model to publish structured data on the Web
- b. use RDF links to interlink data from different data sources

An RDF data model is based on entity-relationship model wherein statement about resources is expressed in the form of following components:

- the *subject*, which is an RDF URI reference or a blank node
- the *predicate*, also known as property, which is an RDF URI reference
- the *object*, which is an RDF URI reference, a literal or a blank node

For example:

The http://en.wikipedia.org/wiki/S._R._Ranganathan webpage

an has the title "S. R. Ranganathan". The primary topic of the resource is a "Person" whose name is "S. R. Ranganathan" and who has worked for an organization identified by <http://drtc.isibang.ac.in/DRTC/>.

RDF Syntax

```
1: <?xml version="1.0"?>
2: <rdf:RDFxmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
3: xmlns:dc="http://purl.org/dc/elements/1.1/"
4: xmlns:foaf="http://xmlns.com/foaf/0.1/">
5: <rdf:Descriptionrdf:about="http://en.wikipedia.org/wiki/S._R._Ranganathan">
6: <dc:title>Ranganathan, S. R.</dc:title>
7: <foaf:primaryTopic>
8: <foaf:Person>
9: <foaf:name>Ranganathan, S. R.</foaf:name>
10: <foaf:workplaceHomepagerdf:resource="http://drtc.isibang.ac.in/DRTC/">
11: </foaf:Person>
12: </foaf:primaryTopic>
13: </rdf:Description>
14: </rdf:RDF>
```

In a traditional web environment, the hyperlinks connect two web documents. However, in a Linked data environment, the hyperlink connects two digital objects described in RDF format. Hence, the name RDF links is used for it (Table 1).

Table 1: RDF Triples.

Number	Subject	Predicate	Object
1	http://en.wikipedia.org/wiki/S._R._Ranganathan	http://purl.org/dc/elements/1.1/title	"Ranganathan SR"
2	genid:A451236	http://www.w3.org/1999/02/22-rdf-syntax-ns#type	http://xmlns.com/foaf/0.1/Person
3	http://en.wikipedia.org/wiki/S._R._Ranganathan	http://xmlns.com/foaf/0.1/primaryTopic	genid:A451236
4	genid:A451236	http://xmlns.com/foaf/0.1/name	"Ranganathan SR"
5	genid:A451236	http://xmlns.com/foaf/0.1/workplaceHomepage	http://drtc.isibang.ac.in/DRTC/

Deploying Linked Open Data: The 5 Star System

Tim Berners-Lee, the inventor of the Web and Linked Data initiator, suggested a 5 star deployment scheme for Linked Open Data. Each star depicts the level of compliance of the web data to be linked to other data [6]. A single star represents a simple web document or a scanned image available on web whereas,

a five star ranking is given to the web documents which strictly adheres to the principles of Linked data as explained in the following text, in an open manner, under Open Data License.

Underpinning the evolution of web into global data space is a set of principles given by Tim Berners-Lee in his web architecture note.

These principles are known as Linked Data principles which are elaborated below [7]:

1. *To use URIs as names for things*

This principle mandates for extending the scope of the web from online resources to encompass any object or concept in the real world.

2. *To use HTTP URIs so that people can look up those names*

The HTTP protocol is the Web's universal access mechanism. In the classic Web, HTTP URIs are used to combine globally unique identification with a simple, well-understood retrieval mechanism. Thus, the second Linked Data principle advocates the use of HTTP URIs to identify objects and abstract concepts, enabling these URIs to be *dereferenced*, (i.e., looked up) over the HTTP protocol into a description of the identified object or concept.

3. *When someone looks up a URI, provide useful information, using the standards (RDF*, SPARQL)*

This principle advocates use of a single data model for publishing structured data over web for interoperability. Linked open data uses Resource Description Framework (RDF), a simple graph-based data model, based on subject-object-predicate relationship model has been designed for use in the context of the Web.

4. *Include links to other URIs, so that they can discover more things*

This principle advocates the use of hyperlinks to connect between web objects such as; web documents, an image, a music file or any type of thing. These hyperlinks carry additional information to identify the relationship existing between the interlinked objects. For example, a hyperlink of the type *author of* may be set between a researcher and a research article or a book. Thus, these hyperlinks of traditional web are called as RDF links in the context of Linked Data.

Libraries and Linked Data

The Linked data is emerging as a powerful discover tool. It has also redefined the way information can be represented on the web. Libraries can make a great deal out of it. Some of the benefits include [8–10]:

- Increased discoverability: Linked data allows description of resources in collaboration with other libraries. This

provides diverse description for same resource which ultimately improves data discoverability.

- Multiple usability of the data: RDF data stored in one data sink can easily be used as such by different services within an organization and beyond.
- Interoperability and re-usability: Web standards facilitate re-use by reducing the need for conversion processes and post-processing.
- Establishing “Cloud” culture among libraries: Linked Data is a cost effective way to represent the data over web where the small institutions can make themselves more visible and connected without major infrastructural cost. This approach would eventually lead to the establishment of "cloud-based" culture among these smaller institutions or individuals.
- Data Preservability: Linked Data focuses on attaching semantics or meaning to the data unlike traditional web where the focus was majorly on structuring and formatting of data. Thus, Linked Data retains its meaning across changes of format.
- Less Dependency on Vendors: Today's library technology is specific to library data formats and provided by an Integrated Library System industry specific to libraries, libraries will be able to use mainstream solutions for managing Linked Data. Adoption of mainstream Linked Data technology will give libraries a wider choice in vendors, and the use of standard Linked Data formats will allow libraries to recruit from and interact with a larger pool of developers.
- Mobilization of Organizational Synergy: Linked Open Data best practices leads to standardization of the data produced by the different services. This will bring in intra- and inter-organizational synergy resulting in improvement or deriving additional services from the existing ones.
- Interdisciplinary Research and Citation Management: Automation of citations retrieval from Linked Data will facilitate linking of information thereby supplementing Interdisciplinary research. Management of citations will be easier in compare to traditional web as each

resource will be identified through a unique URI.

- Granularity: With Linked Data, different kinds of data about the same asset can be produced in a decentralized way by different actors, which then can be aggregated to a common platform to describe a resource with higher granularity.
- Multilinguality: Multilingual functionality for data and user services is well supported by linked data, such as the labeling of concepts identified by a language-agnostic URIs.

ISSUES

1. Library Standards: According to Breeding, the biggest obstacle to transform the library data which is mostly the bibliographic data; is the fact that it is encoded with standards like MARC which squeezes data into the smallest possible package [11]. It is difficult to derive semantic relationship out of this encoded data. Further, library cataloguing rules like AACR2 does not support automated computer processing as most of the information rests in the mnemonic form. Even its current form, i.e., RDA fails to fully comply with the linked data. Realizing this, Library of Congress has taken initiative through BIBFRAME to transform their existing data into Linked data.
2. Data Integrity: It is important that the vocabulary terms used are consistent, authorized and recorded in a uniform manner.
3. Integration of data existing in local databases: Much of the data also exists in local databases whose web interfaces are not fairly well connected with other data resources existing on the Web.
4. Duplication: Though the efforts towards Linked open data came up to prevent duplication nevertheless a number of datasets exist representing the same concept. Hence, a system is to be evolved where the communities are informed with existing vocabularies and terms to avoid such duplication of effort.
5. Manpower Development: Linked data is an opportunity for libraries to make their data more discoverable as their data will

be enriched with metadata which provides additional and new ways to explore the data. However, this will require improved workflow and specialized manpower for technical services. As such the technology is on continuous evolve and there are still scope of improvement and simplification. Thus, the management needs to continuously work on their existing policies and procedures to keep themselves abreast with the technology. Decline in library budgets will be challenging for effective planning.

6. Privacy: It is one of the major concerns for many interested in linked data. The more data are linked, the more it is possible that personal information will be exposed. Librarians, with their long tradition of protecting the privacy of patrons, will have to take an active role in linked data development to ensure rights are protected [12].
7. Dependency on Proprietary Software: Digital divide is a known fact which leads to dependency on proprietary systems by less privileged ones. These proprietary systems mostly come with their own mechanism of data representation providing little flexibility to their end users to encode the data.

Libraries and Linked Data

Linked Data lies at the heart of what Semantic Web is all about: large scale integration of, and reasoning on, data on the Web [13]. Linked data provides enormous opportunities as far as user services are concerned. The Linked data will lead to standardization and interlinking of library data thus enabling an environment of interoperability and unification of segregated, individual technical activities by the libraries. This will eventually help them to combat the ever increasing budget cuts and changing priorities of their library users for more précised and personalized search results. For example, a researcher can be facilitated with precise information enriched with additional related data. This will also suggest the researcher about the possible interdisciplinary research. Libraries across the world have realized the potential of Linked data to improve their library services. OCLC has taken several initiatives in this regard and has offered its various resources as linked

data. Some of them are [14]:

- VIAF (Virtual International Authority File)
- FAST (Faceted Application of Subject Terminology)
- Dewey Decimal Classification System

BIBFRAME (The Bibliographic Framework) is an initiative of Library of Congress which aims to define a bibliographic framework based on linked data principles for library data [15]. BIBFRAME has been designed to replace the MARC standards and to make bibliographic data more useful both within and outside the library community. OpenAGRIS is a web application that aggregates information from different Web sources such as; DBPedia, FAO fisheries dataset, AGRIS serials dataset and so on to expand the AGRIS, an information system for agricultural sciences, thus, providing description of each concept as much as possible [16]. It uses Agrovoc for Vocabulary control. The datasets of OpenAGRIS is now merged to AGRIS and exist only as AGRIS. LIBRIS, the Swedish National Union Catalogue made a switch to Linked data from MARC21 from 2008 onwards [17]. They make use of a mix of ontologies to describe their resources.

The Poda Project by Oslo Public Library was initiated in 2008 to rejuvenate their online catalogue by adopting Linked data technology so as to make it more open with a flexible API (Application Program Interface). A flexible API empowers its users to write their own program using semantic web technology [18]. Open Library is a project started in 2006 by nonprofit Internet Archive (digital library at California) funded by California State Library and Kahle/Austin Foundation [19]. It aims to build web page for every book ever published. The majority of the information is collected from Library of Congress and Amazon.com. The Google Knowledge Graph, launched in May 2012 is a linked based system that describes how a set of objects are connected or interlinked [20]. It helps to understand facts about people, places and things and the relationship that exists between these entities. This knowledge graph was initially based on Freebase, a large collaborative knowledge base developed by Metaweb Technologies

(now merged to Google). The knowledge base is continuously augmenting its database through Wikipedia and CIA World Factbook and other web based resources. Some popular Open Vocabularies available as Linked data are:

1. Geonames Geographic Data
2. BBC Wildlife Vocabulary
3. New York Times People, organizations and subjects
4. Library of Congress Vocabularies (LCSH, subject authorities, classification, others)
5. FAST in RDF
6. RDA Vocabularies and Elements
7. Dewey Summaries Dewey Decimal System in RDF
8. Rameau subject headings Subject headings from the National Library of France
9. Deutsche national Bibliothek subjects Subject headings from the National Library of Germany
10. AGROVOC (FAO thesaurus)

The amount of linking varies, and may change over time as these systems and the available resources grow.

CONCLUSION

Karen Coyle, a librarian and consultant on digital libraries, says: "If library records were open access on the web, it would be possible to create bibliographies that go beyond the holdings of any one library. Open records could create a virtual library of books published in that geographical region, which would allow scholars to study the literature and history of that area in a way that isn't possible today with our separate, physical libraries" [21]. Definitely, Linked data is an innovative way to locate, disseminate and combine information and enhance the capability of current web to perform more astutely for their users. Librarians can be helpful in the effective categorization of information as they are the ones responding to the variant and ambiguous queries of their users in this physical world. However, they need to be trained with the technology so that they can establish comfort to use these tools. Data duplication and privacy are some of the important issues to be looked in while the handing of the datasets.

REFERENCES

1. Bizer, Chris, *et al.* Linked Data on the Web (LDOW2008), 2008. Available at: <http://www2008.org/papers/pdf/p1265-bizer.pdf> (Accessed on 2.12.2015)
2. Hassanzadeh Oktie, Introduction to Semantic Web Technologies & Linked Data, 2011. Available at: <http://www.cs.toronto.edu/~oktie/slides/we-b-of-data-intro.pdf> (Accessed on 3.12.2015)
3. Campbell Lorna M, Sheila MacNeill. The Semantic Web, Linked and Open Data. Available at: http://wiki.cetis.ac.uk/images/1/1a/The_Semantic_Web.pdf (Accessed on 3.12.2015)
4. Miller Paul, Linked Data and Government, 2010. Available at: <http://www.epsiplatform.eu/sites/default/files/ePSPlatform%20Topic%20Report%20No.%207%20%20Linked%20Data%20and%20Gov.pdf> (Accessed on 4.12.2015)
5. Bizer Chris, Cyganiak R, Heath T. How to Publish Linked Data on the Web, 2007. Available at: <http://wifo5-03.informatik.uni-mannheim.de/bizer/pub/LinkedDataTutorial/#intro> (Accessed on 4.12.2015)
6. Berners-Lee Tim, Linked Data, 2009. Retrieved from: <http://www.w3.org/DesignIssues/LinkedData.html> (Accessed on 4.12.2015)
7. Heath Tom, Talis Bizer Christian, Linked Data: Evolving the Web into a Global Data Space. (n.d.), 2011. Retrieved from: <http://linkeddatabook.com/editions/1.0/#htoc8> (Accessed on 6.12.2015)
8. W3C, Benefits of the Linked Data Approach, 2011. Retrieved from: <http://www.w3.org/2005/Incubator/lld/wiki/Benefits> (Accessed on 6.12.2015)
9. Ostrowski F, Pohl A, Towards an Editable, Versionized LOD Service for Library Data. *Liber Quarterly*, 22(4), 2013. Retrieved from: <http://liber.library.uu.nl/index.php/lq/article/view/URN%3ANBN%3ANL%3AUI%3A10-1-114290/8669> (Accessed on 1.1.2016)
10. Byrne Gillian, Goddard Lisa, The Strongest Link: Libraries and Linked Data. *D-Lib Magazine*, 16(11/12), 2010. Retrieved from: <http://www.dlib.org/dlib/november10/byrne/11byrne.html> (Accessed on 3.1.2016)
11. Breeding Marshall, Linked Data: The Next Big Wave or Another Tech Fad? 2013. Available at: <http://www.infotoday.com/cilmag/apr13/Breeding--Linked-Data--The%20Next-Big-Wave-or-Another-Tech-Fad.shtml> (Accessed on 3.2.2016)
12. Byrne Gillian, Goddard Lisa, Op. cit, 2010. (Accessed on 4.1.2016)
13. W3C. Linked Data. Available at: <https://www.w3.org/standards/semanticweb/data> (Accessed on 4.1.2016)
14. Breeding Marshall, Op. cit, 2013. (Accessed on 6.1.2016)
15. BIBFRAME. Available at: <https://www.loc.gov/bibframe> (Accessed on 10.1.2016)
16. Celli Fabrizio, *et al.* The OpenAGRI(s)mashup: Using linked data and web services to augment AGRIS repository content, 2012. Available at: http://eprints.rclis.org/17736/1/Celli-The_OpenAGRI_%28s%29mashup-282_b%5B1%5D.pdf (Accessed on 11.1.2016)
17. Martin, LIBRIS available as Linked Data, 2008. Available at: <http://librisbloggen.kb.se/2008/12/03/libris-available-as-linked-data/> (Accessed on 12.1.2016)
18. Titangos Hui-Lan H. (ed). *Local Community in the Era of Social Media Technologies: A Global Approach*. Chandos Publishing. 2013, 129p.
19. Open Library. Available at: <https://openlibrary.org/about> (Accessed on 12.1.2016)
20. Singhal Amit, Introducing the Knowledge Graph: things, not strings, 2012. Available at: <https://googleblog.blogspot.in/2012/05/introducing-knowledge-graph-things-not.html> (Accessed on 12.1.2016)
21. Grossman Wendy M, *Why you can't find a library book in your search engine*. The Guardian (London), 2009. (Accessed on 15.1.2016).

Cite this Article

Sneha Tripathi. Linking Data for Library Services. *Journal of Advancements in Library Sciences*. 2016; 3(1): 29–34p.