Linked Data Adaptation and Practice in Libraries: an Indian Panorama

¹Biswanath Dutta, ²Nabina Khamaru and ³Harikrishnan J S

Documentation Research and Training Centre, Indian Statistical Institute, Bangalore Centre

{¹bisu, ²nkhamaru, ³harikrishnan}@drtc.isibang.ac.in

Abstract

Purpose - This paper aims to gauge library and information science (LIS) professionals' awareness, readiness, and preparedness to embrace linked data (LD) technologies, standards, techniques, and methodologies for integrating those within Indian libraries.

Design/ methodology/ approach - This survey study employs a precisely crafted methodology, comprising three main phases and ten steps, to delve into core research questions (RQs) by mapping them to detailed survey questions (SQs) section-wise.

Findings - The study reveals insights regarding the conception of LD technologies among LIS professionals, their inclination towards its adaptation for enhanced library operations, and their level of preparedness for its successful integration into library services. Despite the lack of technical knowledge and confidence among professionals, they seem interested in this new technological advancement and are ready to train themselves. So, it offers a positive outlook for the future adaptation of LD in libraries. Moreover, it adds value to the understanding of policymakers, government bodies, researchers, and practitioners, for further magnifying the quality and reach of library activities within society.

Originality/value – As per the authors' knowledge, this paper serves as the first research output in understanding LD adaptation in the Indian landscape. Though the National Library of Sweden showcases initiatives in implementing LD techniques in libraries, this study underscores a lack of similar endeavours within the Indian context.

Keywords Indian libraries, Linked data, Technological adaptation, Library and information science professional, Proficiency in LD technologies, Semantic Web, Survey, Methodology

1. Introduction

Present-day libraries are emerging in connectivity and outreach following technological advancements. They are transcending the traditional functionalities of organization, preservation, and accessibility with more challenging task resolution in the current era. After the Online Computer Library Center (OCLC) drives the exploration of linked data (LD) within the library community, "LD and Libraries" has become an evolving area of focus among many

organizations or institutes worldwide. LD (coined by Tim Berners-Lee) tends to receive divergent interpretations (Bizer *et al.*, 2009). OCLC overviews linked data (LD) as the "communities agreeing on the meaning of their data and sharing it in a massively networked information space" (OCLC, 2024a). It defines LD as "a way to organize and connect data on the web so it can be easily, automatically, and programmatically shared and used by various systems and services" (OCLC, 2024b). LD is "a set of design principles for sharing machine-readable interlinked data on the Web" (Ontotext, 2024a). In LD, information representation is performed in a standardized format, and relationships between different pieces of data are explicitly defined (encoded in 'triples' (subject-predicate-object) using RDF) to structure and interconnect data on the Web (Nagumothu *et al.*, 2021). Linking diverse datasets can enable meaningful discovery of resources, enforce collaboration and resource sharing among peers (decentralization), enhance adaptability to present changes, and resolve interoperability issues (Dutta, 2014).

Libraries have always been adaptive to emerging technologies to unravel the implicational benefits through actions and activities. The significance and applications of LD are being tested across domains such as biology, medicine, agriculture, e-commerce, cultural heritage, education, scientific research, and publications (Schmachtenberg *et al.*, 2014). The libraries are not less. OCLC Research advocated that libraries and their data must be visible on the internet as it is the place where the quest for information begins. Their studies acknowledge that "Linked data enables library staff to provide greater context for information and build rich connections across library resources, their communities, and beyond" (OCLC, 2024b).

LIBRIS, the Swedish Union Catalogue, is an implementation of LD technology, by exporting bibliographic data and relations from an integrated library system (ILS) into a minimal API for the exposure of library data to linked library data (LLD) (Malmsten, 2009). The National Library of Sweden became the first national library to fully transform into an LD library (Unterstrasser, 2023). Besides, the Library of Congress (https://id.loc.gov/), British Library (https://data.europa.eu/data/datasets/the-linked-open-british-national-

bibliography?locale=en), French National Library (https://data.bnf.fr/), German National Library

(https://www.dnb.de/EN/Professionell/Metadatendienste/Datenbezug/LDS/lds_node.html),

National Library of Spain (https://datos.bne.es/inicio.html), and Hungarian National Library (https://lists.w3.org/Archives/Public/public-lod/2010Apr/0155.html) have already paved the pathway for adapting LD in libraries.

Apart from the worldwide study, we searched for literature specifying the LD initiatives and activities in Indian libraries. Unfortunately, we have not come across any such evidence in the literature. We found a few articles specifying the LD research. For instance, Kumar *et al.* (2013) show the transformation of MARC21 data into RDF data for bibliographic records, Bhargav *et al.* (2016) develop an open library semantic bibliographic catalogue (BookLOD), Dutta (2017) describes the various LD applications and LD's relationships with ontologies, and Sharma *et al.* (2018) discusses the processing and storing of library data in LD format using Apache Spark. More discussion on literature is provided in section 2. The primary objective of the current study is to investigate the present state of LD adaptation and its practices in Indian libraries. It investigates LD awareness and popularity among Indian library professionals, their LD technology competency, implementation experiences, involvement in LD projects, and most importantly their willingness to transform the library data into LLD. Secondarily, the

study investigates the present usage of library tools and technologies in Indian libraries for delivering regular library services and carrying out routine technical operations, such as library housekeeping operations, library asset management, electronic resources management, and so forth. This would give hints about LIS professionals' progress in technology usage and their willingness to embrace advanced technologies like LD. Furthermore, the study aims to identify the availability of various kinds of data that the libraries generate, store, and process for various purposes and uses.

This would allow us to envision and explore the possible opportunities to develop and exploit LD technologies and applications beyond the bibliographic and authority data that most of the above-mentioned worldwide studies are based on. Thus the study not only brings out the present state of LD technology usage and implementation in Indian libraries but also guides the researchers, developers, and decision and policy makers to plan for successful LD implementation and design various applications.

To achieve the above-stated objectives, we formulated six research questions (RQs) as provided below. We used a closed questionnaire survey method to seek answers to these questions. The entire study method is provided in *Methodology* section 3. The questionnaire was distributed among the publicly funded libraries in India as detailed in step 2 of section 3. The mapping of these RQs are shown in Table I of Appendix A and further discussed in the *"Mapping SQs to RQs"* part of the *Results and Analysis* section 4.

RQ1: How aware/knowledgeable are the LIS professionals in India about LD technologies and techniques?

RQ2: How far are the library personnel familiar with and their level of understanding of LD technologies?

RQ3: What is the level of willingness of LIS professionals to adapt LD techniques?

RQ4: Are Indian libraries associated with any LD project and are the library professionals cognizant of such initiatives nationally or internationally?

RQ5: Do libraries and their staff feel ready/show interest in transforming their library data into LLD?

RQ6: Do the working professionals feel the need for proper training, through workshops, for LD implementation in libraries?

The main contribution of this study is it investigates the current state of LD awareness and knowledge among LIS professionals in India. It also studies the state of preparedness of LIS professionals in terms of their LD skills and readiness to embrace the technologies for further development of libraries. The other contributions of this study are: it presents a systematic approach that outlines the step-by-step processes used in the research and provides a questionnaire that can be adapted for similar studies with minimal modifications. Notably, to the best of our knowledge, this is the first study of its kind focused on Indian libraries.

The rest of the article is organized as follows: section 2 delves into the allied works already performed in the current study area. Section 3 outlines the detailed method followed for this study. Section 4 inspects the survey responses using descriptive statistical methods to deduce answers for the RQs and its discussion part provides analytical viewpoints on the received responses. Finally, section 5, summarises the present study providing future directions, and potential opportunities it can open for the LIS fraternity.

2. Literature Review

The section presents an overview of strategies, adaptation, implementation, practices, and experiments with LD technologies in library activities at a global level. Databases consulted for the literature search are Google Scholar, Scopus, and Semantic Scholar. The review of related works is divided into three distinct sections: benefits and potential, case studies of implementation, and awareness and perception studies related to LD in libraries.

Benefits and Potential

LD offers significant potential for libraries, in the evolving digital landscape. Hallo et al. (2016) examined that LD enables digital libraries to connect and share their data with a broader web ecosystem, thereby improving resource visibility and user experience. In the library environment, LD is valuable for expanding bibliographic and authority data (Park and Kim, 2014). For instance, the Bibliographic Framework (BIBFRAME) approach for harvesting and sharing bibliographic metadata using LD principles, allowed libraries to integrate disparate metadata sources and enrich their catalogues with external web resources (Tharani, 2015). Latif et al. (2016) discussed various benefits of LOD in libraries, which include distributed data management, improved data integration, and enhanced knowledge discovery. Singer (2009) argued that while LIS professionals have adopted MARC as a common language, the library records remain isolated strings divided by silos, which can be resolved using LD technologies. DeWeese and Segal (2015) summarised the importance of LD and highlighted some initiatives (e.g., W3C library LD incubator group) taken in the library domain for their practical implications. Despite challenges like data modelling and long-term preservation (Chen, 2017), LD benefits libraries in areas like web search, authority control, classification, data portability, and disambiguation (Byrne and Goddard, 2010). So, LD provides libraries with the potential to transform their catalogues into active knowledge repositories for improved resource discoverability and access.

Case Studies of Implementation

Several case studies have reported the implementation of LD technologies in library settings. Worldwide, some major organizations like the Europeana Library, the Library of Congress, the British Library, and the National Library of France implemented LD technologies focusing on vocabularies, ontologies, and data models (Hallo et al., 2016). However, the adaptation of LD in libraries is still in its early stages. It requires a conceptual shift from document-centric to data-centric models, by embracing RDF (Alemu et al., 2012). Deliot (2014) discussed the initiative taken by the British Library to publish their MARC21 bibliographic data as LOD through the RDF model. Schiavone et al. (2018) created a triplification pipeline and formed the CoBiS network to provide 65 libraries with an infrastructure for LOD publishing. Niu (2020) performed content analysis of literature and concluded that diffusion and adaptation of LD among individual libraries follows a decentralized process. Papadakis et al. (2015) identified the specific fields in MARC entry that can hold LD information for library authority records, using URIs. Few national libraries were inspected to find their acceptance of LD principles in publishing content for more accessibility. The PCC task group initiated the deployment of MARC encoders into LD identifiers following a survey, discussion, proposal, and pilot test which laid the groundwork for LLD transition (Shieh, 2020). Smith et al. (2017)

provided a roadmap for individual librarians to plan the transition of LD operation in libraries. Tamper *et al.* (2023) created a LOD and semantic service portal, "*BiographySampo*", utilizing bibliographical data from the National Library of Finland. Fallgren (2015) experimented with BIBFRAME to create a shareable vocabulary of new cataloguing data for the National Library of Medicine. These studies pave the way for the future exploitation of LD in libraries.

Awareness and Perception Studies

A series of publications by (Warraich and Rorissa, 2018, 2020, 2024) suggested that Pakistani LIS professionals are willing to adopt LD technologies in their university libraries, to link bibliography data for enhanced record navigation. McKenna et al. (2018) conducted a survey study addressing 185 information professionals, which suggested the design of a workflow for using LD tools in Libraries, Archives, and Museums to integrate and interlink datasets under the expertise and guidance of information professionals. In 2019, the same authors designed NAISC to expose library domain data for wider outreach and publicity, by interlinking the internal LD with the authoritative LD published by creators and data holders. The professionals rated NAISC as a useful and easy-to-use LD interlinking approach (McKenna et al., 2019). Desmeules et al. (2020) evaluated data models and ontology development in the context of LD projects and surveyed the roles performed by information professionals in modelling such project work. Pennington and Cagnazzo (2019) examined the inclination of information professionals toward adapting LD technologies in the national libraries of Scotland and European countries to discover library resources. Debattista et al. (2018) analyzed information professionals' experience with LD applications in digital libraries. LaPolla (2013) surveyed the perception and adaptive nature of library cataloguers and technical professionals regarding semantic web (SW) applications in academic libraries. It showed growing interest among professionals and highlighted some potential advantages.

The literature search for the present study reveals that no work has been conducted in the past to find the LD state of adaptation and practices in Indian libraries. So, with the need to assess the knowledge and perception of LIS professionals about LD technologies and understand their willingness for change, the current study adds value in reshaping library operations according to the latest technological improvements.

3. Methodology

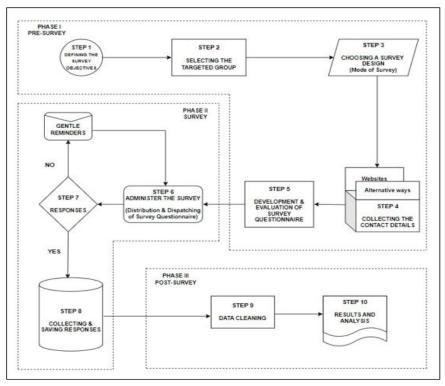
The authors have developed a formal and comprehensive methodology to provide a holistic view of the current study. Primarily, the survey is conducted in three phases: Phase 1- Pre-Survey, Phase 2- Survey, and Phase 3- Post-Survey. Each phase consists of distinct steps, culminating in a total of ten steps, as illustrated in Figure 1.

Phase I: Pre-survey: This phase defines the study objectives and answers the questions, such as, what is the purpose of the study? How do we select a targeted group to conduct a survey? How do we collect contact details of a particular population for the survey? etc. as detailed through steps 1 to 5.

Phase II: Survey: This is a vital part of the methodology, during which survey questionnaires are distributed to the targeted group and responses are collected in a convenient format.

Gathering responses may take time, and it's not uncommon for the initial attempt to yield an insufficient number of responses. This process is elaborated in steps 6 to 8.

Phase III: Post-survey: This is the concluding phase. It initiates the process of data cleaning followed by results and analysis. The survey findings relate to the research questions by data analysis and interpretation using statistical methods as detailed in steps 9 and 10.



Source: Figure by authors

Figure 1. Schematic representation of the methodology

Step 1: Defining the survey objectives

This preliminary planning phase outlines the essential requirements and clear-cut survey workflow. The primary objectives focus on ascertaining the prevalence of LD technology within Indian libraries and assessing the level of awareness and willingness among library professionals to adopt LD tools, standards, techniques, and methodologies.

Step 2: Selecting the targeted group

Identifying the specific demographic for the survey is essential. The selected audience for this survey comprises professionals working in various libraries affiliated with central and state government-sponsored institutions, including publicly funded arts and science, engineering and technical, management, medical, law, science research, and development institutes, as well as state central libraries and public libraries in India.

The list contains, for example, the Central Universities, Indian Institutes of Technology (IITs), National Institutes of Technology (NITs), Indian Institutes of Information Technology (IIITs), Indian Institutes of Management (IIMs), All India Institute of Medical Sciences (AIIMS),

National Law Schools, various national research institutes, such as Indian Space Research Organisation (ISRO), Defence Research and Development Organisation (DRDO), Indian Institute of Space Science and Technology (IIST), Indian Institute of Science Education and Research (IISER), and Indian Institute of Science (IISc). A total of 231 distinct libraries affiliated with publicly funded higher educational institutes were contacted.

Step 3: Choosing a survey design

This step focuses on selecting the most effective survey mode for reaching the targeted group and distributing the SQs. The priority is to identify the fastest and most convenient method for generating and delivering questions. While multiple methods are available, including phoneins, letters, in person interactions, and email, we have chosen an online survey using Google Forms. This platform offers customizable templates with options like checkboxes, dropdown lists, and fields for both objective and descriptive responses, making it ideal for capturing feedback. Additionally, Google Forms automatically records and organizes responses in spreadsheet, simplifying data analysis.

Step 4: Collecting the contact details

This step involves collecting contact details of the targeted respondent group, as defined in step 2. We gathered this information through various sources, including institute websites, referrals, directories, and association registries. Additionally, we obtained email addresses of librarians from esteemed institutions like IITs and IIMs through professional connections. To maximize response rates, we collected the email IDs of all library professionals within each targeted institution, ensuring a higher likelihood of receiving at least one reply per library.

Step 5: Development and evaluation of survey questionnaire

The first step in developing the questionnaire involves reviewing existing literature to identify relevant studies for reuse, which saves time and helps create a more comprehensive questionnaire. However, multiple evaluations are necessary to refine the SQs. We found few studies closely aligned with our research, with the most notable being Warraich and Rorissa's study (discussed in Section 2). While these studies provide useful references, their reliance on open-ended questions limits their reusability, prompting us to develop a new set of SQs for our study.

We conducted a three-round evaluation of the questionnaire, including a pseudo-survey with five participants, to assess its quality. The goal was to refine the SQs for simplicity, clarity, relevance, and ease of response. Initially, we created a pool of 47 questions. After reviewing the survey's objectives, research questions, and the need for time efficiency, we focused on using straightforward language and reducing lengthy, paragraph-style questions. We also prioritized concise, single-line, and multiple-choice questions. This process led to a final set of 30 questions (see Appendix A), which were then integrated into Google Forms.

Questionnaire in Brief: The 30-question questionnaire is designed to collect data on various aspects, including demographic information, Library Management Systems (LMS) and other software used in libraries, types of data generated and stored, the understanding of LD technologies among LIS professionals, knowledge of LD tools and initiatives, opportunities for experimenting with LD in the library sphere, LD projects in institute libraries, and the

willingness of professionals to engage in LD initiatives for linking and sharing library data. Each question is aligned with the RQs, which are mapped to the survey questions in Appendix A and discussed in Section 4.

Step 6: Administer the survey

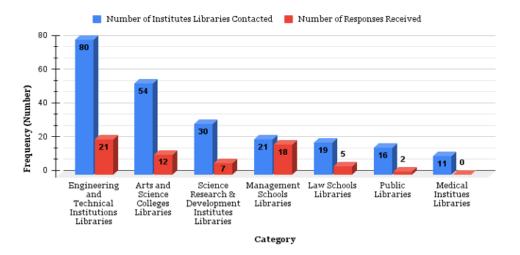
This step begins the survey process based on the design decisions made earlier. It involves distributing the survey questions (SQs) via email or directly engaging with the target respondents. In this study, the Google Form link and a cover letter were sent to all email addresses collected in Step 4.

Step 7: Responses

The next phase involves awaiting responses, with two possible outcomes: receiving responses or not. To ensure reliable results, it is crucial to gather a sufficient number of responses, which we aimed to achieve by sending multiple gentle reminders. The survey was launched on August 8, 2023, with a three-month deadline due to the study's schedule and funding constraints. To maximize responses within this timeframe, reminders were sent on August 16, 2023; August 28, 2023; September 5, 2023; September 21, 2023; and October 7, 2023.

The survey received a total of 68 responses, but three were found to be repetitive. Following the decision to retain only one response per library, the responses were prioritized by professional rank, starting with the librarian and moving to scientific/technical assistant professionals if the former was absent. This approach was based on the assumption that senior professionals have greater administrative authority to make decisions about new initiatives, while lower-ranked staff typically lack decision-making authority. After eliminating duplicates, 65 unique responses were retained.

When comparing category-wise responses, management institutes contributed the most (18 out of 21), followed by law schools and technical institutes (see Figure 2). Notably, only two out of the sixteen public libraries contacted responded. This lack of response may suggest that these libraries find the survey irrelevant, possibly due to their limited access to the latest technologies, particularly LD and related facilities, which may make them feel disconnected or uninterested in participating.



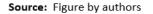
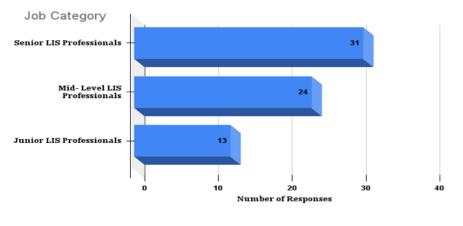


Figure 2. Institute categories and number of responses

The responses came from LIS professionals with varied job titles and expertise levels. To gain more structured insight, we categorized the participants into three groups based on their job titles (see Figure 3).



Source: Figure by authors

Figure 3. Responses according to job category

The first category consists of "Senior LIS Professionals" including roles like - 'Librarians, Director of Library, Deputy Librarians, Professor-in-charge cum librarian, and Chief library officer'. The second category, named "Mid-level LIS professionals" encompasses roles such as - 'Assistant librarian, Assistant manager, and Senior Library and Information officer'. The third category, "Junior LIS professionals" includes - 'Junior Library officers, Library assistants, Scientific officers, Professional assistants, Junior technicians', and similar positions. As most of the responses are from senior and medium-level categories, we can be sure that our survey was conducted among the most experienced and highly qualified professionals who are currently heading and managing esteemed libraries in India.

Step 8: Collecting and saving responses

Store the responses in a convenient format for future data manipulation and cleaning, preferably .csv, .tsv, or .xlsx. In this study, the responses were saved in a Google spreadsheet.

Step 9: Data cleaning

Data cleaning involves correcting or eliminating inaccurate, corrupted, improperly formatted, duplicated, or incomplete data. With large sets of categorical data, cleaning is often necessary. In our survey, objective answers made some data easier to handle, while descriptive answers required corrections, especially in fields like institute category, experience, and job titles. This ensures smooth statistical analysis in subsequent steps.

Step 10: Results and analysis

This step addresses the RQs outlined in the Introduction. Data analysis is essential for drawing insights, making decisions, and reaching conclusions. It involves methods like diagnostic analysis, cohort analysis, regression analysis, time series analysis, and statistical analysis. In this study, we used both diagnostic and descriptive statistical methods. Statistical analysis

identifies overall data patterns, while diagnostic analysis explores the causes behind specific data behaviours ("cause-and-effect relationships") (Taherdoost, 2022). Further details are provided in Section 4.

4. Results and Analysis

Mapping SQs to RQs

The questionnaire starts with demographic and general information (SQ 1-6) to gather details such as the participant's name, contact information, job title, experience, etc. SQ 7-12 focus on the current use of LMS and other software for library management, the types of data generated and stored, and how this data is utilized. In the course of our study, we aimed to comprehend the possibilities of transforming library data into LLD, extending beyond bibliographic records. So, the identification of key data and software, commonly employed in the library domain, serves to illuminate the challenges associated and the future course of actions to be undertaken for implementing LD in libraries.

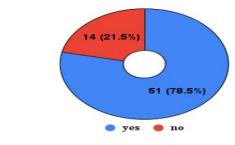
The subsequent section of the survey questionnaire, from SQ 13 ("Do you know about LD?") to SQ 18 ("Do you know any utility software tools for LD creation and publication?") addresses the core of the survey's main objective. It focuses on assessing the level of understanding of LIS professionals regarding LD technologies, standards, tools, initiatives, and programs. The results will provide insights into their awareness of LD and help evaluate the current state of LD applications in Indian libraries, directing to RQ 1-2.

Questions SQ 19-26 aim to assess whether library professionals have implemented LD technology and their approach to using it in library operations, addressing RQ 3-4. Questions SQ 27-30 focus on gauging their interest in learning about LD through workshops and their views on its application in libraries, addressing RQ 5-6.

Results

Survey results from SQ 7-12 show that most institute libraries use library automation software, with only one Art & Science library still in the automation process. Over 60% of libraries use Koha as their LMS. In the management category, Koha, Libsys, and Virtua are most common, while Koha, NewGenlib, and Libsys dominate in Technical and Science & Research libraries. Law and Arts & Science libraries prefer Koha, E-granthalaya, and SOUL, while public libraries reported using Koha and Libsys.

Moreover, it's important to highlight that libraries are integrating additional software to enhance various operations and extend their services. A diverse range of tools, including DSpace, ERP, Zotero, Mendeley, Grammarly, Drupal, LS discovery, and VuFind, is being utilized for these purposes.

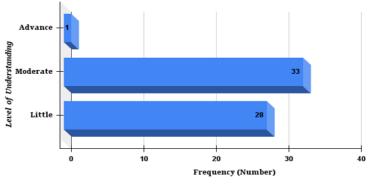


Source: Figure by authors

Figure 4. LD knowledge among LIS professionals

All 65 respondents answered SQ 13. Figure 4 shows that the majority of library professionals (78%) have either heard of or are familiar with the term LD, with only a small percentage (21%) answering 'No' to the same question. All percentage values in the study are rounded to the nearest integer.

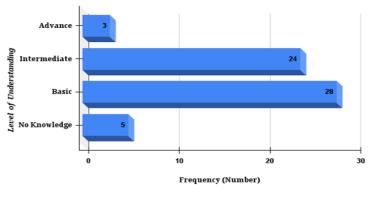
While assessing the level of understanding of LD among professionals, it is noteworthy that only one respondent possesses 'advanced' knowledge and working experience in the LD field. Additionally, (53%) of individuals have a 'moderate' level of knowledge, while (45%) have very 'little' knowledge as depicted in Figure 5.



Source: Figure by authors

Figure 5. Level of understanding of LD among LIS professionals

SQ 15 assesses the level of understanding of LIS professionals concerning the most prevalent and basic building blocks of LD technology, i.e., HTTP Uniform Resource Identifier (URI), Resource Description Framework (RDF), Ontology, and SPARQL, as outlined by Tim Berners-Lee (Bizer *et al.*, 2009). The responses to these subparts are exhibited in Figures 6-9.



Source: Figure by authors

Figure 6. Level of understanding of HTTP URI among LIS professionals

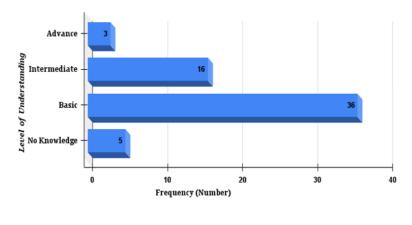




Figure 7. Level of understanding of RDF among LIS professionals

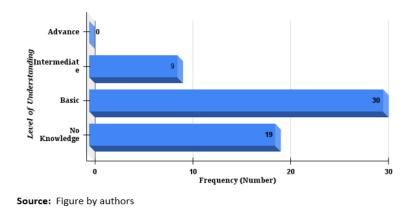
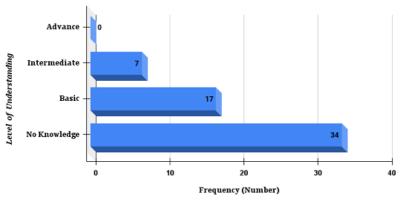


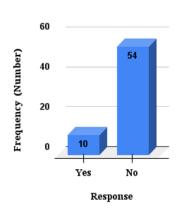
Figure 8. Level of understanding of Ontology among LIS professionals



Source: Figure by authors

Figure 9. Level of understanding of SPARQL among LIS professionals

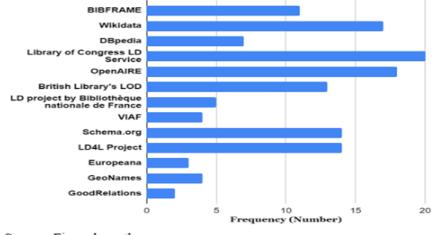
From the above Figures 6-9, we can conclude that most of the professionals have knowledge of URI, RDF, and Ontology. However, it is noteworthy that only 42% of professionals are aware of SPARQL. The rest lack a basic understanding of SPARQL, as depicted in Figure 9. Through our web search, we discovered that numerous workshops on LD have been conducted by various national and international organizations over time. However, it is learned from our survey that only 16% of LIS professionals have attended LD-related workshops, while the majority, 84%, have not attended any, as depicted in Figure 10.



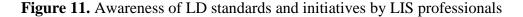
Source: Figure by authors

Figure 10. LD related workshops attended by LIS professionals

As shown in Figure 11, responses for SQ 17 shows LIS professionals' awareness of different LD standards and initiatives. A larger part of the respondents are familiar with LD initiatives such as Wikidata¹, BIBFRAME², schema.org³, the British Library's LOD⁴, the LD project by the Bibliothèque nationale de France, the LOC LD initiative, and LD4L⁵ (see Figure 11). We received only six 'No or Nil' responses for question SQ 17.



Source: Figure by authors



From SQ 18, 'Do you know any utility software tools for LD creation and publication?', as referred to in Figure 12, it was found that some handful of tools mentioned were Protege, Fuseki, OpenRefine, RD4R, XML, XML Schema, and RDF. One respondent mentioned Winzip and antivirus programs, which are in no way related to the question. Out of the 22

¹ https://www.wikidata.org/wiki/Wikidata:Introduction

² https://www.loc.gov/bibframe/

³ https://schema.org/

⁴ https://old.datahub.io/es/dataset/bluk-bnb

⁵ https://wiki.lyrasis.org/display/ld4lGW

attempted answers, 18 respondents replied in negation. This shows professionals' lack of exposure to and minimal knowledge of utility software tools for LD creation and publication.

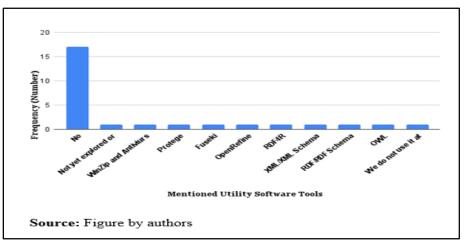
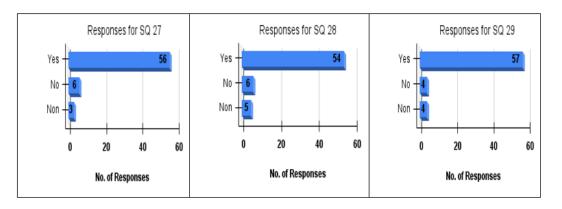


Figure 12. Utility software tools for LD creation and publication as mentioned by LIS professionals

While analysing responses to SQ 19-26, as noted, only one library having a professional occupying the role of Senior Library and Information Assistant, was associated with an LD initiative in his library. However, this respondent did not furnish extensive details about the undertaken project. The majority of the LIS professionals who participated in the study were not engaged in any LD initiatives. Consequently, they have not sought to utilize LD technology to enhance library resource accessibility or improve decision-making processes. This critical insight can be derived from the considerable number of 'No' responses to SQ 19-22. Additionally, responses to SQ 25 and 26 contribute to this understanding. Some individuals attempted to respond to SQ 23 and 24 using common sense and hypothetical assumptions, even though they had not been involved in any LD projects. Most of them, in response to SQ 24, identified the following main challenges encountered while implementing LD in their library. The numeric figure written in parentheses indicates the number of responses.

- 1. Data integration and interoperability issues (44)
- 2. Data quality and consistency (40)
- 3. Lack of training and trained LD personnel (50)
- 4. Lack of availability of software tools (42)
- 5. Lack of motivation among the staff members (17)
- 6. Lack of infrastructural support (46)
- 7. Shortage of staff (46)

SQ 27-29 received an overwhelmingly positive response, as shown in Figure 13. Only a few professionals expressed reluctance to participate in LD initiatives or receive training on LD and related technologies. Specifically, six negative responses were recorded for SQ 28, as respondents stated they were not willing to provide access to library data for LD experimentation even if contacted.



Source: Figure by authors

Figure 13. LD related workshops attended by LIS professionals

The majority of positive responses collected, along with the descriptive answers to SQ 30, indicate a strong interest in experimenting with LD technologies and exploring ways to enhance current library services. This highlights the potential for increased engagement and collaboration to implement LD in libraries and develop an LD framework. Furthermore, our survey results suggest a clear need for a national-level workshop on LD, with considerable interest among LIS professionals in attending such an event.

Non-Response Bias

We observed a predominant occurrence of null or non-responses in SQ 13-18, as shown in Figure 14. This phenomenon, where some individuals are reluctant to answer each question, introduces bias in the survey analysis, known as non-response bias (Berg, 2005).

Survey Question Number	SQ 13	SQ 14	(SQ SQ 15.1 to	15 15.4)		SQ 16	SQ 17	SQ 18
(SQ No.)			SQ 15.1	SQ 15.2	SQ 15.3	SQ 15.4			
Null									
Response / Non-	0	3	5	4	7	7	1	10	15
Response Count									

Figure 14. Counting non-responses

Discussion

A conclusive view of the study can be drawn by thorough analysis of the response data received from the survey. This offers insights into the current state of LD awareness and adaptation

within esteemed libraries in India. Incorporating LD in library operations could enhance service discovery (Chen and Paik, 2013), navigation of related resources (Rosati and Mayernik, 2013), and improved decision-making (Adams *et al.*, 2021). Day-to-day library activities produce various types of data, such as acquisition data, bibliographic data, serials data, cataloguing data, circulation data, reservation and holding data, security and access data, vendor data, etc. Library data conversion into LLD has great potential if utilized fully, and libraries need to experiment with this possibility (Dutta and Bhuvaneshwari, 2023).

While analysing RQ 1-2, along with the responses to survey questions up to SQ 19 (excluding the demographic part, see Table I), it becomes evident that almost all professionals are aware of LD or have heard about it. However, they lack detailed technical proficiency in LD technologies. As observed, for example, SPARQL, an essential technology for retrieving and managing data stored in RDF triplestores (Ontotext, 2024b) is unknown to most of them. The reasons for insufficient knowledge of LD technologies at the individual level could be many, for instance, lack of resources in terms of finance, manpower (trained professionals), training programs, and lack of applications for ready use in libraries.

The result shows that most institutional libraries in India have adopted library automation software, with Koha being the most widely used LMS. This aligns with previous studies (Kumar V & Jasimudeen, 2012; Sonker & Jayakanth, 2003), which highlight Koha's popularity due to its free, open-source nature and robust features. Notably, 21% of libraries still use older versions (V.16 to V.20) of Koha, reflecting a group of LIS professionals' reluctance to adopt newer technologies or their limited tech-savviness. This may present a challenge for implementing advanced technologies like LD in libraries. However, as Koha is open-source and adaptable, it could potentially integrate with other tools for LD implementation. Future research is needed to explore this potential, suggesting a promising future for LD in libraries.

Despite awareness among LIS professionals about major LD resources (e.g., DBpedia (Holze, 2022), Wikipedia, GeoNames⁶), initiatives (e.g., BIBFRAME, VIAF⁷), and projects (e.g., LD4L project), Indian libraries are not part of any such efforts. This may be due to the lack of a governing body for library activities and insufficient administrative support for new initiatives. Directing to RQ4, these challenges highlight the need for a collaborative LIS community to provide guidance and foster encouragement.

The statistics clearly demonstrate a strong demand for an introductory program to explore the potential of LD technology in the library field, especially among aspiring LIS professionals. There is a notable need for workshops and training programs with practical demonstrations tailored to the LIS domain. This is further supported by the overwhelmingly positive responses to SQ 27-30, reflecting participants' enthusiasm to join the LD community and their interest in exploring ways to enhance library services, addressing RQ 5-6.

In summary, the survey, which engaged libraries and professionals from leading institutions (see Figure 2), reveals that these organizations have yet to experiment with LD. While LIS professionals in India possess basic knowledge of LD, their use and experimentation with LD tools and concepts remain limited. Several factors may contribute to this. It raises questions about whether this reluctance stems from professional attitudes or waning enthusiasm for adopting new technologies in libraries, or if it reflects the government's indifference toward

⁶ https://www.geonames.org/

⁷ https://viaf.org/

advancing library technology. Alternatively, it may be linked to the less-explored technological content in the current LIS curriculum and practices in Indian universities.

It is recommended that higher authorities facilitate workshops and offer opportunities for hands-on experience with advanced technologies, such as LD, to keep pace with global developments. Such initiatives would enhance library services across various areas and ensure that libraries and library professionals remain at the forefront of the evolving landscape of librarianship.

5. Conclusion

This study examines the LIS professionals' awareness, proficiency level, preparedness, and eagerness for LD implementation in Indian libraries. It provides a picturesque example of professionals' know-how in LD technologies and techniques. So, it paves the way for adapting LD technologies and its successful implementation within library services. A closed questionnaire survey method was designed to commence the study. In the Indian scenario, this study is the first of its kind. It deciphers the professionals' interest and zeal to learn more about LD techniques for their future implementation and concludes by revealing a tapestry of opportunities and possibilities it brings for LIS professionals. Additionally, it guides policymakers and administration about the required technological advancement for upgrading library activities.

The study testimonies that library professionals are yet to be exposed to LD technologies and their potentiality. Thereby, it points out the need for training the LIS fraternity to transform library data into LLD. This will help to enrich and integrate the local catalogue of networked libraries, improve discovery by enabling effective search, and create a resilient and flexible data management system.

As a next step, the study aims to examine the adaptation, implementation, and use of LD technology in Indian libraries, focusing on internal factors (e.g., behavioural intention, perceived ease of use, and perceived usefulness) and external factors (e.g., user training, participation, and the implementation process), as outlined in the Technology Acceptance Model (Davis, 1985; Venkatesh & Davis, 2000). Additionally, the study seeks to develop guidelines and toolkits for effective LD operations in libraries.

Appendix A.

Table I: Shows mapping of survey questions (SQs) to research questions (RQs) (under each section) without the various options against each SQ to avoid lengthiness.

SQ No.	SQ	RQ No.
Section Section	on 1 Questions related to Demographic & General Information	
~	Your name (Please try to begin with any of the suitable prefix given: Mr/Mrs/Ms) Job Title	
SQ 3	Organization & Library name	
SQ 4	Years of Experience as a Library professional	
SQ 5	Email	
SQ 6	Mobile number	
Secti	on 2 Practice of library technology & data utility in libraries	
	Which library management system (ILS/LMS) does your library currently use? If none from the list then specify your own (below). What is the version of your LMS?	
SQ 9	How long has your library been using the current LMS?	
SQ 10 SQ	What types of data are routinely generated and stored in your library? (Select all that apply) If any other data, please mention below. How does your library use the above-mentioned data for administrative or library	
11	usage enhancement? (Select all that apply) If any other usage, please mention below.	
SQ 12	Besides LMS do you use any other software tools for your library activities (e.g. report generation, data visualization, digital library software, ERMS, CMS, discovery tools)? If yes, mention them.	
Secti	• •	
SQ 13 SQ	Do you know about Linked Data (LD)? How do you rate your level of understanding of LD?	
14 SQ 15	Please indicate your level of understanding for each of the following LD technologies: HTTP Uniform Resource Identifier (URI)	RQ 1 & RQ 2
	Resource Description Framework (RDF) Ontology SPARQL	
SQ 16	Have you received any formal training or attended workshops related to LD technology in libraries?	
SQ 17	Are you familiar with the following LD standards and initiatives?	
SQ 18	Do you know any utility software tools for LD creation and publication? If yes, mention the tool name.	
Secti	Existing LD technology practices in libraries and involvement	
SQ	What LD software tool(s) or technologies have been used in your library?	
19 SQ 20	Has your library implemented/undertaken any LD initiative?	RQ 3 & RQ 4
SQ 21	Is your library associated with any LD initiative nationally or internationally?	(· · · · · (·
SQ 22	Have you been involved in any LD initiative?	

SQ	How has the adaptation of LD technology improved the accessibility of library				
23	resources and in the decision-making process?				
SQ	What were the main challenges encountered while implementing LD in your library?				
24					
SQ	Is there any application built utilizing the created LD in your library?				
25					
SQ	If the generated LD is published or shared on the web, please provide the URI of the				
26	dataset.				
Section 5 Willingness of LIS professionals for transformations of library data to LLD					
SQ	Is your library willing to participate in a LD initiative?				
27					
SQ	Would your library provide access to library data, if you are contacted?				
28		RQ 5 & RQ 6			
SQ	Would your library staff participate in receiving training on LD tools and				
29	technologies?				
SQ	Is there anything else you would like to share in the given context?				
30					

References

- Adams, C., Allen, J. and Flack, F. (2021), "Sharing Linked Data for Health Research: Towards Better Decision Making: Synopsis by the authors", *International Journal of Population Data Science*, Vol. 6 No. 2.
- Alemu, G., Stevens, B., Ross, P. and Chandler, J. (2012), "Linked Data for libraries: Benefits of a conceptual shift from library-specific record structures to RDF-based data models", *New Library World*, Vol. 113 No. 11, pp. 549–570, doi: 10.1108/03074801211282920.

Berg, N. (2005), "Non-Response Bias", SSRN Scholarly Paper, Rochester, NY.

- Bhargav, H.S., Akalwadi, G., Kishan, K. and Mahesh, K. (2016), "Curating a Semantic Bibliographic Catalog", *Proc. IEEE Int. Conf. Semantic Comput., ICSC*, presented at the Proceedings 2016 IEEE 10th International Conference on Semantic Computing, ICSC 2016, Institute of Electrical and Electronics Engineers Inc., pp. 172–173, doi: 10.1109/ICSC.2016.41.
- Bizer, C., Heath, T. and Berners-Lee, T. (2009), "Linked Data The Story So Far", *International Journal on Semantic Web and Information Systems (IJSWIS)*, IGI Global, Vol. 5 No. 3, pp. 1–22, doi: 10.4018/jswis.2009081901.

- Byrne, G. and Goddard, L. (2010), "The strongest link: Libraries and linked data", *D-Lib Magazine*, Vol. 16 No. 11–12, doi: 10.1045/november2010-byrne.
- Chaurasia, A. and Singh, A.P. (2024), "Navigating the Narrative: Unveiling the Journey of Public Libraries in India Through Challenges and Opportunities", *Public Library Quarterly*, Routledge, Vol. 43 No. 4, pp. 524–542, doi: 10.1080/01616846.2023.2299501.
- Chen, W. and Paik, I. (2013), "Improving efficiency of service discovery using Linked databased service publication", *Information Systems Frontiers*, Vol. 15 No. 4, pp. 613–625, doi: 10.1007/s10796-012-9381-x.
- Chen, Y.-N. (2017), "A review of practices for transforming library legacy records into linked open data", in Koutsomiha D., Garoufallou E., Siatri R., and Virkus S. (Eds.), *Commun. Comput. Info. Sci.*, Vol. 755, presented at the Communications in Computer and Information Science, Springer Verlag, pp. 123–133, doi: 10.1007/978-3-319-70863-8_12.
- Davis, F.D. (1985), A Technology Acceptance Model for Empirically Testing New End-User Information Systems: Theory and Results, Thesis, Massachusetts Institute of Technology.
- Debattista, J., McKenna, L. and Brennan, R. (2018), "Understanding Information Professionals: A Survey on the Quality of Linked Data Sources for Digital Libraries", in Panetto, H., Debruyne, C., Proper, H.A., Ardagna, C.A., Roman, D. and Meersman, R. (Eds.), On the Move to Meaningful Internet Systems. OTM 2018 Conferences, Springer International Publishing, Cham, pp. 537–545, doi: 10.1007/978-3-030-02671-4_32.

- Deliot, C. (2014), "Publishing the British national bibliography as linked open data", *Catalogue & Index*, Vol. 174, pp. 13–18.
- Desmeules, R.E., Turp, C. and Senior, A. (2020), "Exploring Methods for Linked Data Model Evaluation in Practice", *Journal of Library Metadata*, Routledge, Vol. 20 No. 1, pp. 65–89, doi: 10.1080/19386389.2020.1742434.
- DeWeese, K.P. and Segal, D. (2015), "Linked Data and Library Initiatives", in DeWeese, K.P. and Segal, D. (Eds.), *Libraries and the Semantic Web*, Springer International Publishing, Cham, pp. 25–38, doi: 10.1007/978-3-031-02038-4_3.
- Dutta, B. (2014), "Symbiosis between Ontology and Linked Data", *Librarian*, Vol. 21, pp. 15–24.
- Dutta, B. (2017), "Examining the interrelatedness between ontologies and Linked Data", *Library Hi Tech*, Emerald Group Publishing Limited, Bradford, United Kingdom, Vol. 35 No. 2, pp. 312–331, doi: 10.1108/LHT-10-2016-0107.
- Dutta, Biswanath and Bhuvaneshwari, V. (2023), "Towards standardizing the library circulation metadata", in Garoufallou, E. & Vlachidis, A. (Ed.s.), (MTSR 2023): 17th Metadata and Semantics Research Conference, University of Milano-Bicocca, Italy, October 25-27, 2023, CCIS, Springer Nature, Cham (Switzerland), Vol. 2048, pp. 117-131, doi: <u>https://doi.org/10.1007/978-3-031-65990-4_10</u>.
- Fallgren, N. (2015), "Experimentation with BIBFRAME at the National Library of Medicine", *GitHub, Last Updated June*, Vol. 24.
- Hallo, M., Luján-Mora, S., Maté, A. and Trujillo, J. (2016), "Current state of Linked Data in digital libraries", *Journal of Information Science*, SAGE Publications Ltd, Vol. 42 No. 2, pp. 117–127, doi: 10.1177/0165551515594729.

- Kumar, S., Ujjal, M. and Utpal, B. (2013), "Exposing MARC 21 Format for Bibliographic Data As Linked Data With Provenance", *Journal of Library Metadata*, Routledge, Vol. 13 No. 2–3, pp. 212–229, doi: 10.1080/19386389.2013.826076.
- Kumar V, V. and Jasimudeen, S. (2012), "Adoption and user perceptions of Koha library management system in India", ALIS Vol.59(4) [December 2012], NISCAIR-CSIR, India.
- LaPolla, F. (2013), "Perceptions of Librarians Regarding Semantic Web and Linked Data Technologies", *Journal of Library Metadata*, Routledge, Vol. 13 No. 2–3, pp. 114–140, doi: 10.1080/19386389.2013.826063.
- Latif, A., Scherp, A. and Tochtermann, K. (2016), "LOD for Library Science: Benefits of Applying Linked Open Data in the Digital Library Setting: Retrospects and Research Topics", *KI - Kunstliche Intelligenz*, Springer Science and Business Media Deutschland GmbH, Vol. 30 No. 2, pp. 149–157, doi: 10.1007/s13218-015-0420-x.

Malmsten, M. (2009). "Exposing Library Data as Linked Data", paper presented at the IFLA satellite preconference, 19–20 August, Florence, available at: https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=2e0791d88a65cb2517e28 4c2bfca02b7c6660f30;(accessed 2October 2024).

- McKenna, L., Debruyne, C. and O'Sullivan, D. (2018), "Understanding the Position of Information Professionals with regards to Linked Data: A Survey of Libraries, Archives and Museums", *Proc. ACM IEEE Joint Conf. Digit. Libr.*, presented at the Proceedings of the ACM/IEEE Joint Conference on Digital Libraries, Institute of Electrical and Electronics Engineers Inc., pp. 7–16, doi: 10.1145/3197026.3197041.
- McKenna, L., Debruyne, C. and O'Sullivan, D. (2019), "NAISC: An authoritative linked data interlinking approach for the library domain", in Bonn M., Wu D., Downie S.J., and

Martaus A. (Eds.), *Proc. ACM IEEE Joint Conf. Digit. Libr.*, Vol. 2019-June, presented at the Proceedings of the ACM/IEEE Joint Conference on Digital Libraries, Institute of Electrical and Electronics Engineers Inc., pp. 11–20, doi: 10.1109/JCDL.2019.00012.

- Nagumothu, D., Eklund, P.W., Ofoghi, B. and Bouadjenek, M.R. (2021), "Linked data triples enhance document relevance classification", *Applied Sciences (Switzerland)*, MDPI AG, Vol. 11 No. 14, doi: 10.3390/app11146636.
- Niu, J. (2020), "Diffusion and adoption of Linked Data among libraries", Proceedings of the Association for Information Science and Technology, John Wiley and Sons Inc, Vol. 57 No. 1, doi: 10.1002/pra2.216.

OCLC. (2024a). "Linked Data Overview", OCLC, Dublin, OH, available at: https://www.oclc.org/research/areas/data-science/linkeddata/linked-data-overview.html; (accessed 9 September 2024).

OCLC. (2024b). "Linked Data: The Future of Library Cataloging", OCLC, Dublin, OH, available at: https://doi.org/10.25333/71w4-vq71; (accessed 1 October 2024).

Ontotext. (2024a). "What Are Linked Data and Linked Open Data?", Ontotext, available at: https://www.ontotext.com/knowledgehub/fundamentals/linked-data-linked-open-data/(accessed 1 October 2024).

Ontotext. (2024b). "What Is an RDF Triplestore?", Ontotext, available at: https://www.ontotext.com/knowledgehub/fundamentals/what-is-rdf-triplestore/ (accessed 9 September 2024).

Papadakis, I., Kyprianos, K. and Stefanidakis, M. (2015), "Linked data URIs and libraries: The story so far", *D-Lib Magazine*, Corporation for National Research Initiatives, Vol. 21 No. 5–6, p. 1, doi: 10.1045/may2015-papadakis.

- Park, Z. and Kim, H. (2014), "Organizing and Sharing Information Using Linked Data", New Directions in Information Organization, Vol. 7, Emerald Group Publishing Limited, pp. 61–87, doi: 10.1108/S1876-0562(2013)0000007008.
- Pennington, D. R. and Cagnazzo, L. (2019), "Connecting the silos: Implementations and perceptions of linked data across European libraries", *Journal of Documentation*, Emerald Publishing Limited, Vol. 75 No. 3, pp. 643–666, doi: 10.1108/JD-07-2018-0117.
- Rosati, A. and Mayernik, M. (2023). "Facilitating Data Discovery by Connecting Related Resources", available at: https://www.semantic-webjournal.net/system/files/swj484.pdf (accessed 2 October 2024).
- Schiavone, L., Morando, F., Allavena, D. and Bevilacqua, G. (2018), "Library data integration: The CoBiS linked open data project and portal", in Serra G. and Tasso C. (Eds.), *Commun. Comput. Info. Sci.*, Vol. 806, presented at the Communications in Computer and Information Science, Springer Verlag, pp. 15–22, doi: 10.1007/978-3-319-73165-0_2.
- Schmachtenberg, M., Bizer, C. and Paulheim, H. (2014), "Adoption of the Linked Data Best Practices in Different Topical Domains", in Mika, P., Tudorache, T., Bernstein, A., Welty, C., Knoblock, C., Vrandečić, D., Groth, P., et al. (Eds.), *The Semantic Web – ISWC 2014*, Springer International Publishing, Cham, pp. 245–260, doi: 10.1007/978-3-319-11964-9_16.
- Sharma, K., Marjit, U. and Biswas, U. (2018), "Efficiently processing and storing library linked data using apache spark and parquet", *Information Technology and Libraries*, American Library Association, Vol. 37 No. 3, pp. 29–49, doi: 10.6017/ital.v37i3.10177.

- Shieh, J. (2020), "PCC's Work on URIs in MARC", *Cataloging & Classification Quarterly*, Routledge, Vol. 58 No. 3–4, pp. 418–427, doi: 10.1080/01639374.2019.1705951.
- Singer, R. (2009), "Linked Library Data Now!", Journal of Electronic Resources Librarianship, Routledge, Vol. 21 No. 2, pp. 114–126, doi: 10.1080/19411260903035809.
- Smith, MacKenzie, Stahmer, Carl G., Li, Xiaoli, and Gonzalez, Gloria. (2017), "BIBFLOW: A Roadmap for Library Linked Data Transition", University Library, University of California, Davis and Zepheira, Inc., UC, available at: https://library.ucdavis.edu/wpcontent/uploads/2022/07/bibflow_roadmap_revised_3_14_2017.pdf; (accessed 2 October 2024).
- Sonker, S.K. and Jayakanth, F. (2003), "Koha An Open Source Integrated Library Automation System", *Journal of Information and Knowledge (Formerly SRELS Journal of Information Management)*, Vol. 40 No. 2, pp. 135–142.
- Taherdoost, H. (2022), "Different Types of Data Analysis; Data Analysis Methods and Techniques in Research Projects", SSRN Scholarly Paper, Rochester, NY, 1 August.
- Tamper, M., Leskinen, P., Hyvönen, E., Valjus, R. and Keravuori, K. (2023), "Analyzing biography collections historiographically as Linked Data: Case National Biography of Finland", *Semantic Web*, IOS Press, Vol. 14 No. 2, pp. 385–419, doi: 10.3233/SW-222887.
- Tharani, K. (2015), "Linked data in libraries: A case study of harvesting and sharing bibliographic metadata with BIBFRAME", *Information Technology and Libraries*, American Library Association, Vol. 34 No. 1, pp. 5–19, doi: 10.6017/ital.v34i1.5664.

Unterstrasser, J. (2023), "Linked Data and Libraries : How the Switch to Linked Data Has Affected Work Practices at the National Library of Sweden", available at: https://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-506075 (accessed 02 October 2024).

- Venkatesh, V. and Davis, F.D. (2000), "A Theoretical Extension of the Technology Acceptance Model: Four Longitudinal Field Studies", *Management Science*, INFORMS, Vol. 46 No. 2, pp. 186–204, doi: 10.1287/mnsc.46.2.186.11926.
- Warraich, N.F. and Rorissa, A. (2018), "Adoption of linked data technologies among university librarians in Pakistan: Challenges and prospects", *Malaysian Journal of Library and Information Science*, Vol. 23 No. 3, pp. 1–13, doi: 10.22452/mjlis.vol23no3.1.
- Warraich, N.F. and Rorissa, A. (2020), "Application of linked data technologies in libraries: Pakistani information professionals' attitudes and perceptions", *The Electronic Library*, Emerald Publishing Limited, Vol. 38 No. 5/6, pp. 1035–1051, doi: 10.1108/EL-01-2020-0002.
- Warraich, N.F. and Rorissa, A. (2024), "Impediments to and readiness for Linked Data application in libraries: Pakistani information professionals' perspective", *Journal of Librarianship and Information Science*, SAGE Publications Ltd, Vol. 56 No. 2, pp. 341–352, doi: 10.1177/09610006221141690.