

7-1-2022

Empowering linked data in cultural heritage institutions: A knowledge management perspective

Lei Zhang
San Jose State University, lei.zhang@sjsu.edu

Follow this and additional works at: https://scholarworks.sjsu.edu/faculty_rsca

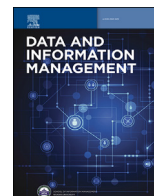
Recommended Citation

Lei Zhang. "Empowering linked data in cultural heritage institutions: A knowledge management perspective" *Data and Information Management* (2022). <https://doi.org/10.1016/j.dim.2022.100013>

This Article is brought to you for free and open access by SJSU ScholarWorks. It has been accepted for inclusion in Faculty Research, Scholarly, and Creative Activity by an authorized administrator of SJSU ScholarWorks. For more information, please contact scholarworks@sjsu.edu.

Contents lists available at [ScienceDirect](#)

Data and Information Management

journal homepage: www.journals.elsevier.com/data-and-information-management

Empowering linked data in cultural heritage institutions: A knowledge management perspective



Lei Zhang

School of Information, San José State University, San José, CA, 95192, USA

ARTICLE INFO

Keywords:

Linked data
Cultural heritage institutions
Knowledge management

ABSTRACT

This reported research explores the barriers and challenges in linked data implementation in cultural heritage institutions, i.e., libraries, archives, and museums. Various data were collected from different sources regarding the linked data use cases related to libraries, archives, and museums over the past decade and analyzed from multiple facets. The analysis revealed very few activities of effective knowledge management in the linked data implementation and suggested that the crucial role of knowledge management and innovation should deserve enough attention in linked data projects and services. The findings will add value to the literature on knowledge management in the context of linked data and the semantic web.

1. Introduction

The diagram of the Linked Open Data (LOD) cloud (<https://lod-cloud.net/>) contains information on datasets that have been published according to the linked data principles and clearly shows the growth of datasets. In May 2007, the LOD cloud contained only 12 datasets; by September 2011, the number increased to 295 datasets; and the latest, as of May 2021, showed a total of 1301 datasets.

The traditional cultural heritage institutions (i.e., libraries, archives, and museums) are information silos, with data and records long been stored in their own catalogs and databases. With the linked data principles and technologies, the valuable data that reside in libraries, archives, and museums can be made linkable, searchable, and useable on the Web. The goal of LOD is to promote connected and enriched data and the release of data for sharing in a networked information space. The development of linked data has brought forth a profound influence on the information professions and information professionals.

Classified as a subfield of information organization and retrieval by the Association for Library and Information Science Education (ALISE) Research Taxonomy, linked data exploit the natural synergy between data science and metadata. However, except for the technical issues, many observed challenges in the linked data initiatives are more related to data and information management. The perspective of knowledge management may shed new light on these challenges and offer insightful guidance.

The purpose of this research is to delineate a picture of the components and characteristics of linked data use cases in libraries, archives,

and museums to inform their future linked data policies and practices. The researcher has collected the linked data use cases related to libraries, archives, and museums over the past decade and analyzed a wide variety of sources from multiple facets. This research is to track the development of linked data, identify the key issues and trends that are common and unique across the settings, from standards and approaches to resources and services. As part of a larger project, the matter that is reported in this paper is focused on knowledge management in the context of linked data in cultural heritage institutions.

2. Methodology

In this research, a wide variety of sources have been examined including journal articles, conference proceedings, and technical reports regarding linked data use cases in libraries, archives, and museums, such as bibliographic data, authority data, controlled vocabularies, metadata element sets, etc. For the first phase, the relevant literature was collected through searching for the linked data use cases via the San José State University (SJSU) library databases and via the search engines following the preferred reporting items for systematic reviews and meta-analyses (PRISMA) guidelines.

First, an advanced search was conducted through SJSU Library OneSearch, with the subject containing library linked data, archive linked data, museum linked data, with the material type as articles, in English language, published between 01/01/2011 and 12/31/2021. In addition to the above criteria, the articles with a significant component addressing the barriers and challenges in linked data implementation

E-mail address: lei.zhang@sjsu.edu.

<https://doi.org/10.1016/j.dim.2022.100013>

Received 16 October 2021; Received in revised form 6 December 2021; Accepted 6 December 2021

2543-9251/© 2022 Wuhan University. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

other than technical details were considered as relevant to this reported research. From a total of 905 search results retrieved, a set of 22 full-text articles was identified by the researcher as relevant based on the article title, abstract, and conclusion, and an additional 8 articles were identified as relevant from the references of the first set of articles and retrieved through Google search.

The collected literature relevant to this reported research was then analyzed with Ranganathan's five facets (PMEST):

Personality: entity in question.

Matter: materials, substances, properties, etc.

Energy: operations, processes, activities, etc.

Space: geographic areas and features.

Time: periods, dates, seasons, etc.

These five facets can also be described as Who, What, How, Where, and When. Facet analysis was originated from Ranganathan's five fundamental facets, and these five facets have been expanded as thirteen categories by the Classification Research Group and developed as Aitchison et al.'s four facets and their sub-facets: entities, things, and objects; actions or activities; space, place, location, or environment; time. Facets have been widely used in the development of controlled vocabularies such as taxonomies and thesauri, the organization system and the navigation system of websites, and other types of information structures, to reveal the various aspects, attributes, or dimensions of a topic.

3. Findings

From the collected literature, almost all deals with the linked data use cases in the developed countries in Europe and North America, except one focusing on the developing country. Sixty-seven percent of the collected literature was published in the past five years.

The most interesting findings in this reported research were revealed from two facets: the Matter facet, the materials, substances, properties of knowledge management, and the Energy facet, the operations, processes, activities that take place concerning knowledge management.

3.1. Common issues

As early as 2011, the Library Linked Data Incubator Group reported the following observations: fewer bibliographic datasets had been published as linked data than value vocabularies and element sets, the quality of and support for available data varied greatly, linking across datasets had begun but required further effort and coordination, and pointed out the issues of rights that hindered their release as LOD. These were echoed by Mitchell's two reports (2013, 2016) that addressed the themes of metadata quality; open data and business models — the value of data openness; literacy, training, and cross-community engagement; aggregation, provenance, and trust — the importance of standards and approaches for defining and maintaining standards; and implementation, interoperability, and scale — approaches to system development.

The more recent 2018 OCLC international linked data survey results showed the status of linked data projects or services, with 75% of the implementations in production and 40% of them in production for more than four years (Smith-Yoshimura, 2018). The barriers to publishing linked data were listed in the following order: steep learning curve for staff, inconsistency in legacy data, selecting appropriate ontologies to represent our data, lack of resources, little documentation or advice on how to build the systems, establishing the links, lack of tools, immature software, ascertaining who owns the data, among which "lack of resources" was added as an option (a popular "other" response in 2015) and tied with little documentation. For the barriers to consuming linked data in the 2018 survey, the biggest difference from the 2015 survey was the increasing number of responses pointing to unstable endpoints and service reliability.

Studies on linked data projects and communities worldwide show the challenges to implementing linked data including technological issues related to ontology, linked data tools, existing datasets, lack of financial

resources and skilled human resources, low level of awareness among the community, lack of open licenses, and non-availability of standards and best practices (Ali & Warraich, 2018; LaPolla, 2013; Raza et al., 2019; Wahid et al., 2018). With data open and linked to be more visible, integrated, and reused by other services, it arouses more attention to the ownership rights, data quality control, and social justice (Hilario et al., 2014; Ruddock, 2011; Schreur & Carlson, 2020; Ullah et al., 2018; Wenz, 2013).

3.2. Use cases

Most respondents in all three OCLC linked data surveys in 2014, 2015, 2018 (OCLC Research, 2021) were from the library domain, with research libraries and national libraries occupying the top slot, while museums are second to last ("others"). Exploring two groups of linked data use cases at both ends of the spectrum — national libraries and museums — may reveal the key issues and trends of knowledge management that are common and unique across the settings.

The national libraries that are the linked data pioneers include the British Library's British National Bibliography and the National Library of France's data.bnf.fr. Other national libraries joined the linked data implementation more recently such as the National Library of Finland and the National Library of Hungary.

A primary challenge is the lack of knowledge and expertise, the lack of a clear definition and purpose, and the lack of a clear strategy at the beginning (Cagnazzo, 2017; Rasmussen Pennington & Cagnazzo, 2019). As the early linked data adopter, the British Library could not have existing examples to follow suit and had to do most of the work from the scratch. Similarly, the National Library of France lacked the practical and technical knowledge of linked data requirements that determined the adoption of this technology. However, the skills developed through the experience from previous projects were helpful to the accomplishment of data.bnf.fr. The observed challenges struggling with so many ways of publishing linked data, having difficulty publishing the linked data without training and procedural documentation, lacking resources, staff, and technology tools to complete the projects promptly are all because most institutions do not have a written policy to cover all aspects of linked data implementation and use. The National Library of Germany did not consider they had many problems because they had the buy-in from senior management since the beginning, built knowledge networks as part of their long-term strategy, and refined the linked data service as part of their strategic priorities. However, the issues that remain unresolved across European national libraries include difficulties with obtaining management buy-in which requires tangible proof for investment and licensing constraints that make it hard to track by whom and how their data is being used. An investigation of European and US national libraries noted that several issues related to technical, administrative, and social aspects must be solved, such as the development of new tools and of policies to publish and manage data (Hallo et al., 2016).

The Library of Congress released their high-level BIBFRAME primer document in 2012 and the updated BIBFRAME 2.0 model in 2016. The goal of BIBFRAME is to replace MARC, the legacy standard for bibliographic description, with a linked data model. The BIBFRAME, though still in the experimental stage and in the process of development, has been used by some other institutions to various extents, including the National Library of Sweden, the first national library to fully transition to BIBFRAME in 2018. However, "there is currently no universal model for how to represent bibliographic metadata as Linked Data, even though many attempts for such a model have been made" (Suominen & Hyvönen, 2017, p. 1). A study of four national libraries, such as the national libraries of Spain, France, Britain, and Germany, showed they all applied different vocabularies for data representation, largely linked to different external sources, and chose different bibliographic models for their structures (Tallerås, 2017). The data quality achieved varied with the National Library of Spain, the National Library of France, the British National Bibliography, and the Miguel de Cervantes Virtual Library, as

evaluated with the metrics of accuracy, trustworthiness, consistency, relevancy, completeness, timeliness, ease of understanding, interoperability, accessibility, licensing, and interlinking (Candela et al., 2020). For the libraries worldwide including national libraries, there was a great deal of diversity that existed from metadata elements to ontologies, schemas, and data models in describing the same materials with the same general concepts (Park & Kipp, 2019).

The members of the Canadian Linked Data Initiative were working collaboratively to investigate possibilities and start the transition by planning digital projects, evaluating tools, promoting linked data education and training, and enhancing legacy metadata (van Ballegooye et al., 2017), while the Pakistan university librarians found a general lack of awareness of the basic concepts of linked data and the best practices for the emerging linked data technology (Warraich & Rorissa, 2018).

Linked data is gaining prominence in the archives and museums communities as well as in libraries. The challenges these institutions encountered are similar: which records are appropriate to share for open data. Though licensing is also an issue for other types of institutions, it is even more noticeable for the museums with the holding of artworks and other sensitive data. That means not all the museum data could be included in their LOD projects. The Smithsonian American Art Museum, together with 13 other art museums, joined the American Art Collaborative (AAC) LOD Initiative (Fink, 2018) but concluded that a smaller case was best for testing out wider LOD usage. Though the AAC eventually wanted each member institution to control their own data, they had hired university students who were not experts and produced work inconsistent with others. They also needed more vocabularies and ontologies to fully describe the diverse datasets that were in their collections but had limited resources and staff with expertise and lacked ready-to-use software. Unlike AAC, the Rijksmuseum released its high-quality images instead of sending only the data (not the images) for conversion to LOD (Dijkshoorn et al., 2018).

4. Discussion

4.1. A knowledge management perspective

Only a handful of literature from the search results was found relevant to the barriers and challenges in linked data implementation other than technical details. Furthermore, the above findings regarding the linked data in cultural heritage institutions failed to suggest there were operations, processes, and activities that took place for effective knowledge management; they were more related to the materials, substances, and properties of knowledge management, but we only found a lack of resources including staff, tools, guidance, and licenses in addition to the technical challenges.

The institutions implementing linked data as well as the linked data research community are more concerned with the technologies to publish and consume linked data but neglected the knowledge processing cycle for knowledge being transformed, disseminated, and applied (Dalkir, 2017). For example, early adopters such as the British Library and the National Library of France may offer valuable knowledge to other cultural heritage institutions undertaking linked data projects or services. However, not all the linked data projects are fully visible to the public yet. There is a need for data sharing and collaboration between institutions to fill the knowledge gap.

The [Library Linked Data Incubator Group \(2011\)](#) addressed the stakeholders involved in the linked data initiatives including the library leaders, the library standards bodies, the data and systems designers, and the librarians and archivists. The role of knowledge management deserves the utmost attention to ensure a smooth workflow of linked data projects and services. As stated by [Vasudevan et al. \(2006\)](#), knowledge management in an enterprise often involves:

- identifying, selecting, and cataloging information resources that are pertinent to the enterprise's needs;

- identifying information flow patterns among individuals and among groups (e.g., finding out who asks what questions, learning what information is obtained from which sources, determining what types of information are not easily available or accessible); and
- designing and developing user-friendly systems for accessing the enterprise's knowledge base.

In the case of library linked data, it involves the diffusion of related innovations of software tools, linked data vocabularies, and datasets between libraries, within libraries, and between librarians, and is facilitated by the leading libraries, professional organizations, vendors, and external funders (Niu, 2020). Many challenges to implementing linked data in cultural heritage institutions are essentially the issues of knowledge management and knowledge innovation within organizations and across organizations. Lack of strategies in managing staff, tools, guidance, and licenses have resulted in so many issues that surfaced regarding quality, reliability, control, security, and ethics of data and records in the linked data format.

4.2. Looking ahead

It calls for more cooperation, collaboration, sharing, and investigation among the stakeholders, the continuing education and institutional support for professional development, such as training librarians, introducing linked data courses, and organizing seminars and workshops, coordinating with vendors, and increasing the role of cultural heritage institutions and professional associations in motivating linked data implementation (Park & Tosaka, 2017; Raza et al., 2019; Wahid et al., 2018).

With the development of linked data over the years, more rules and regulations are emerging to guide the linked data implementation. There is empirical evidence of the benefits of using best practices for publishing linked data, especially standard practices, integrability, and uniformity (Feitosa et al., 2018). Five of the ten steps in the working group report *Best Practices for Publishing Linked Data* (Hyland et al., 2014) are focused on the policy and social-good issues rather than the technical issues: prepare stakeholders, select a dataset, model the data, specify an appropriate license, the role of "good URIs" for linked data, standard vocabularies, convert data to linked data, provide machine access to data, announce to the public, the social contract of a linked data publisher. Management of that knowledge will increase the power of linked data.

5. Conclusion

This paper reports the barriers and challenges in linked data implementation in libraries, archives, and museums from the perspective of knowledge management, which suggests that the crucial role of knowledge management and knowledge innovation should deserve enough attention to empower the linked data projects and services. Due to the space limit, the matter reported in this paper was primarily based on the linked data use cases in the national libraries and the museums from the collected literature in phase one. The next phase study will further explore the limitations and implications of LOD with a complete set of use cases and in a broader context.

References

- Ali, I., & Warraich, N. F. (2018). Linked data initiatives in libraries and information centres: A systematic review. *The Electronic Library*, 36(5), 925–937.
- van Ballegooye, M., Borie, J., & Senior, A. (2017). The Canadian linked data initiative: Charting a path to a linked data future. *The Serials Librarian*, 72(1–4), 207–213.
- Cagnazzo, L. F. (2017). *Linked Data: Implementation, use, and perceptions across European national libraries*. Thesis. University of Strathclyde. ResearchGate.
- Candela, G., Escobar, P., Carrasco, R. C., & Marco-Such, M. (2020). Evaluating the quality of linked open data in digital libraries. *Journal of Information Science*. <https://doi.org/10.1177/0165551520930951>
- Dalkir, K. (2017). Knowledge management. In J. D. McDonald, & M. Levine-Clark (Eds.), *Encyclopedia of library and information sciences* (4th ed., pp. 2640–2648). Boca Raton: CRC Press.

- Dijkshoorn, C., Jongma, L., Aroyo, L., van Ossenbruggen, J., Schreiber, G., ter Weele, W., & Wielemaker, J. (2018). The Rijksmuseum collection as linked data. *Semantic Web*, 9(2), 221–230.
- Feitosa, D., Dermeval, D., Ávila, T., Bittencourt, I. I., Lóscio, B. F., & Isotani, S. (2018). A systematic review on the use of best practices for publishing linked data. *Online Information Review*, 42(1), 107–123.
- Fink, E. E. (2018). In *American art collaborative (AAC) linked open data (LOD) initiative: Overview and recommendations for good practices*. <https://s3.amazonaws.com/assets.saam.media/files/documents/2020-07/OverviewandRecommendationsAccessible.pdf>.
- Hallo, M., Luján-Mora, S., Maté, A., & Trujillo, J. (2016). Current state of Linked Data in digital libraries. *Journal of Information Science*, 42(2), 117–127.
- Hilario, A. B. R., Fernández, T. F., & Campo, D. M. (2014). From bibliographic records to data: Changes in the library Environment with the application of linked open data technologies. *Information Resources Management Journal*, 27(3), 28–41.
- Hyland, B., Atemezing, G., & Villazón-Terrazas, B. (2014). In *Best practices for publishing linked data*. <https://www.w3.org/TR/ld-bp/>.
- LaPolla, F. (2013). Perceptions of librarians regarding semantic web and linked data technologies. *Journal of Library Metadata*, 13(2–3), 114–140.
- Mitchell, E. T. (2013). Library linked data: Research and adoption. *Library Technology Reports*, 49(5).
- Mitchell, E. T. (2016). Library linked data: Early activity & development. *Library Technology Reports*, 52(1).
- Niu, J. (2020). Diffusion and adoption of Linked Data among libraries. *Proceedings of the Association for Information Science and Technology*, 57(1), 1–9.
- OCLC Research. (2021). In *International linked data survey*. <https://www.oclc.org/research/areas/data-science/linkedata/linked-data-survey.html>.
- Park, H., & Kipp, M. (2019). Library linked data models: Library data in the semantic web. *Cataloging & Classification Quarterly*, 57(5), 261–277.
- Park, J. R., & Tosaka, Y. (2017). Emerging information standards and technologies: Cataloging and metadata professionals' perspectives. *Library Hi Tech News*, 34(4), 22–26.
- Rasmussen Pennington, D., & Cagnazzo, L. (2019). Connecting the silos: Implementations and perceptions of linked data across European libraries. *Journal of Documentation*, 75(3), 643–666.
- Raza, Z., Mahmood, K., & Warraich, N. F. (2019). Application of linked data technologies in digital libraries: A review of literature. *Library Hi Tech News*, 36(3), 9–12.
- Ruddock, B. (2011). Linked data and the LOCAH project. *Business Information Review*, 28(2), 105–111.
- Schreur, P. E., & Carlson, A. J. (2020). Bridging the worlds of MARC and linked data: Transition, transformation, accountability. *The Serials Librarian*, 78(1–4), 48–56.
- Smith-Yoshimura, K. (2018). Analysis of 2018 international linked data survey for implementers. *The Code4Lib Journal*, 42. <https://journal.code4lib.org/articles/13867>.
- Suominen, O., & Hyvönen, N. (2017). From MARC silos to linked data silos? *O-Bib*, 4(2), 1–13. <https://doi.org/10.5282/o-bib/2017H2S1-13>
- Tallerås, K. (2017). Quality of linked bibliographic data: The models, vocabularies, and links of data sets published by four national libraries. *Journal of Library Metadata*, 17(2), 126–155.
- Ullah, I., Khusro, S., Ullah, A., & Naeem, M. (2018). An overview of the current state of linked and open data in cataloging. *Information Technology and Libraries*, 37(4), 47–80.
- Vasudevan, M. C., Mohan, M., & Kapoor, A. (2006). Information system for knowledge management in the specialized division of a hospital. In K. S. Raghavan, & K. N. Prasad (Eds.), *Knowledge organization, information systems and other essays: Professor A. Neelameghnan Festschrift* (pp. 247–261). Chennai: Ranganathan Centre for Information Studies.
- W3C Library Linked Data Incubator Group. (2011). In *Library linked data incubator group final report*. <https://www.w3.org/2005/Incubator/ld/XGR-ld-20111025/>.
- Wahid, N., Warraich, N. F., & Tahira, M. (2018). Mapping the cataloguing practices in information environment: A review of linked data challenges. *Information and Learning Science*, 119(9/10), 586–596.
- Warraich, N. F., & Rorissa, A. (2018). Adoption of linked data technologies among university librarians in Pakistan: Challenges and prospects. *Malaysian Journal of Library & Information Science*, 23(3), 1–13.
- Wenz, R. (2013). Linked open data for new library services: The example of data.bnf.fr. *JLIS.it: Italian Journal of Library and Information Science*, 4(1), 403–415.