

# An Overview of the Current State of Linked and Open Data in Cataloging

Irfan Ullah, Shah Khusro,  
Asim Ullah, and  
Muhammad Naeem

---

## ABSTRACT

*Linked Open Data (LOD) is a core Semantic Web technology that makes knowledge and information spaces of different knowledge domains manageable, reusable, shareable, exchangeable, and interoperable. The LOD approach achieves this through the provision of services for describing, indexing, organizing, and retrieving knowledge artifacts and making them available for quick consumption and publication. This is also aligned with the role and objective of traditional library cataloging. Owing to this link, major libraries of the world are transferring their bibliographic metadata to the LOD landscape. Some developments in this direction include the replacement of Anglo-American Cataloging Rules 2nd Edition by the Resource Description and Access (RDA) and the trend towards the wider adoption of BIBFRAME 2.0. An interesting and related development in this respect are the discussions among knowledge resources managers and library community on the possibility of enriching bibliographic metadata with socially curated or user-generated content. The popularity of Linked Open Data and its benefit to librarians and knowledge management professionals warrant a comprehensive survey of the subject. Although several reviews and survey articles on the application of Linked Data principles to cataloging have appeared in literature, a generic yet holistic review of the current state of Linked and Open Data in cataloging is missing. To fill the gap, the authors have collected recent literature (2014–18) on the current state of Linked Open Data in cataloging to identify research trends, challenges, and opportunities in this area and, in addition, to understand the potential of socially curated metadata in cataloging mainly in the realm of the Web of Data. To the best of the authors' knowledge, this review article is the first of its kind that holistically treats the subject of cataloging in the Linked and Open Data environment. Some of the findings of the review are: Linked and Open Data is becoming the mainstream trend in library cataloging especially in the major libraries and research projects of the world; with the emergence of Linked Open Vocabularies (LOV), the bibliographic metadata is becoming more meaningful and reusable; and, finally, enriching bibliographic metadata with user-generated content is gaining momentum. Conclusions drawn from the study include the need for a focus on the quality of catalogued knowledge and the reduction of the barriers to the publication and consumption of such knowledge, and the attention on the part of library community to the learning from the successful adoption of LOD in other application domains and contributing collaboratively to the global scale activity of cataloging.*

## INTRODUCTION

With the emergence of the Semantic Web and Linked Open Data (LOD), libraries have been able to make their bibliographic data publishable and consumable on the web, resulting in an increased understanding and utility both for humans and machines.<sup>1</sup> Additionally, the use of Linked Data principles of LOD has allowed connecting related data on the web.<sup>2</sup> Traditional catalogs as

---

**Irfan Ullah** ([cs.irfan@uop.edu.pk](mailto:cs.irfan@uop.edu.pk)) is doctoral candidate, **Shah Khusro** ([khusro@uop.edu.pk](mailto:khusro@uop.edu.pk)) is Professor, **Asim Ullah** ([asimullah@uop.edu.pk](mailto:asimullah@uop.edu.pk)) is doctoral student, and **Muhammad Naeem** ([mnaeem@uop.edu.pk](mailto:mnaeem@uop.edu.pk)) is Assistant Professor, at the Department of Computer Science, University of Peshawar.



---

collections of metadata about library content have served the same purpose for a long time.<sup>3</sup> It is, therefore, natural to establish a link between the two technologies and exploit the capabilities of LOD to enhance the power of cataloging services. In this regard, significant milestones have been achieved, which includes the use of Linked and Open Data principles for publishing and linking library catalogs, BIBFRAME, and Europeana Data Model (EDM).<sup>4</sup> However, the potential of Linked and Open Data for building more efficient libraries and the challenges involved in that direction are mostly unknown due to the lack of a holistic view of the relationship between cataloging and the LOD initiative and the advances made in both areas. Likewise, the possibility of enriching the bibliographic metadata with user-generated content such as ratings, tags, and reviews to facilitate the search for known-items as well as exploratory search has not received much attention.<sup>5</sup> Some studies of preliminary extent have, however, appeared in literature an overview of which is presented in the following paragraphs.

Several survey and review articles have contributed to different aspects of cataloging in the LOD environment. Hallo et al. investigated how Linked Data is used in digital libraries, how the major libraries of the world implemented it, and how they benefit from it by focusing on the selected ontologies and vocabularies.<sup>6</sup> They identified several specific challenges to applying Linked Data to digital libraries. More specifically, they reviewed the Linked Data applications in digital libraries by analyzing research publications regarding the major national libraries (obtaining five-stars by following Linked Data principles) and published from 2012 to 2016.<sup>7</sup> Tallerås examined statistically the quality of Linked Bibliographic Data published by the major libraries including Spain, France, the United Kingdom, and Germany.<sup>8</sup>

Yoose and Perkins presented a brief survey of LOD uses under different projects in different domains including libraries, archives, and museums.<sup>9</sup> By exploring the current advances in the Semantic Web, Robert identified the potential roles of libraries in publishing and consuming bibliographic data and institutional research output as Linked and Open Data on the web.<sup>10</sup> Gardašević presented a detailed overview of Semantic Web and Linked Open Data from the perspective of library data management and their applicability within the library domain to provide a more open and integrated catalog for improved search, resource discovery, and access.<sup>11</sup>

Thomas, Pierre-Yves, and Bernard presented a review of Linked Open Vocabularies (LOV), in which they analyzed the health of LOV from the requirements perspective of its stakeholders, its current progress, its uses in LOD applications, and proposed best practices and guidelines regarding the promotion of LOV ecosystem.<sup>12</sup> They uncovered the social and technical aspects of this ecosystem and identified the requirements for the long-term preservation of LOV data. Vandenbussche et al. highlighted the features, components, significance, and applications of LOV and identified the ways in which LOV supports ontology & vocabulary engineering in the publication, reuse and data quality of LOD.<sup>13</sup>

Tosaka and Park performed a detailed literature review of RDA (2005–11) and identified its fundamental differences from AACR2, its relationship with the metadata standards, and its impact on metadata encoding standards, users, practitioners, and the training required.<sup>14</sup> Sprochi presented the current progress in RDA, FRBR (Functional Requirements for Bibliographic Records), and BIBFRAME to predict the future of library metadata, the skills and knowledge required to handle it, and the directions in which the library community is heading.<sup>15</sup> Gonzales identified the limitations of MARC21 and the benefits of and challenges in adopting the BIBFRAME

---

framework.<sup>16</sup> Taniguchi assessed BIBFRAME 2.0 for the exchange and sharing of metadata created in different ways for different bibliographic resources.<sup>17</sup> He discussed BIBFRAME 1.0 from RDA point of view.<sup>18</sup> He examined BIBFRAME 2.0 from the perspective of RDA to uncover issues in its mapping to BIBFRAME including RDA expressions in BIBFRAME, mapping RDA elements to BIBFRAME properties, and converting MARC21 metadata records to BIBFRAME metadata.<sup>19</sup> Fayyaz, Ullah, and Khusro reported on the current state of LOD and identified several prominent issues, challenges, and research opportunities.<sup>20</sup> Ullah, Khusro, and Ullah reviewed and evaluated different approaches for bibliographic classification of digital collections.<sup>21</sup>

By looking at the above survey and review articles, one may observe that these articles target a specific aspect of cataloging from the perspective of LOD. The holistic analysis and a complete picture of the current state of cataloging in transiting to LOD ecosystem are missing. This paper adds to the body of knowledge by filling this gap in the literature. More specifically, it attempts to answer the following research questions (RQs):

RQ01: How Linked Open Data (LOD) and Vocabularies (LOV) are transforming the digital landscape of library catalogs?

RQ02: What are the prominent/major issues, challenges, and research opportunities in publishing and consuming bibliographic metadata as Linked and Open Data?

RQ03: What is the possible impact of extending bibliographic metadata with the user-generated content and making it visible on the LOD cloud?

The first section of this paper answers RQ01 by discussing the potential role of LOD and LOV in making library catalogs visible and reusable on the web. The second section answers RQ02 by identifying some of the prominent issues, challenges, and research opportunities in publishing, linking, and consuming library catalogs as Linked Data. It also identifies specific issues in RDA and BIBFRAME from LOD perspective and highlights the quality of LOD-based cataloging. The third section answers RQ03 by reviewing the state-of-the-art literature on the socially curated metadata and its role in cataloging. The last section concludes the paper followed by references cited in this article.

## **THE ROLE OF LINKED OPEN DATA AND VOCABULARIES IN CATALOGING**

The catalogers, librarians, and information science professionals have always been busy defining the set of rules, guidelines, and standards to record the metadata about knowledge artifacts accurately, precisely, and efficiently. The AACR2 are among the widely used rules and guidelines for cataloging. However, it has several issues with the nature of authorship, the relationships between bibliographic metadata, the categorization of format-specific resources, and the description of new data types.<sup>22</sup> In an attempt to produce its revised version, AACR3, the cataloging community noticed that a new framework should be developed with the name of RDA.<sup>23</sup> Based on FRBR conceptual models, RDA is a “flexible and extendible bibliographic framework” that supports data sharing and interoperability and is compatible with MARC21 and AACR2.<sup>24</sup> According to the RDA Toolkit, RDA describes digital and non-digital resources by taking advantage of the flexibilities and efficiencies of modern information storage and retrieval technologies while at the same time is backward-compatible with legacy technologies used in conventional resource discovery and access applications.<sup>25</sup> It is aligned with the IFLA’s



---

(International Federation of Library Associations and Institutions) conceptual models of authority and bibliographic metadata (FRBR, FRAD [Functional Requirements for Authority Data], FRSAD [Functional Requirements for Subject Authority Data]).<sup>26</sup> RDA accommodates all types of content and media in digital environments with improved bibliographic control in the realm of Linked and Open Data; however, its responsiveness to user requirements needs further research.<sup>27</sup>

The discussion of the cataloging rules and guidelines stays incomplete without the metadata encoding standards and formats that give practical shape to these rules in the form of library catalogs. The most common encoding formats include Dublin Core (DC) and MARC21. Dublin Core (<http://lov.okfn.org/dataset/lov/vocabs/dce>) is a [general-purpose metadata encoding scheme and] vocabulary of fifteen properties with “broad, generic, and usable terms” for resource description in natural language. It is advantageous as it presents relatively low barriers to repository construction; however, it lacks in standards to index subjects consistently as well as to offer a uniform semantic basis necessary for an enhanced search experience.<sup>28</sup> The lack of uniform semantic basis is due to the individual interpretations and exploitations of DC metadata by the libraries, which in turn originated from its different and independent implementations at the element level.<sup>29</sup>

MARC21 is the most common machine process-able metadata encoding format for bibliographic metadata. It can be mapped to several formats including DC, MARC/XML (<http://www.loc.gov/standards/marcxml/>), MODS (<http://www.loc.gov/standards/mods>), MADS (<http://www.loc.gov/standards/mads>), and other metadata standards.<sup>30</sup> However, MARC21 has several limitations such as only library software and librarians understand it, it is semantically inexpressive and isolated from the web structure, and it lacks in expressive semantic connections to relate different data elements in a single catalog record.<sup>31</sup> Besides its limitations, MARC metadata encoding format is vital for resource discovery especially within the library environment, and therefore, ways must be found to make visible the library collections outside the libraries and available through the major web search engines.<sup>32</sup> One such effort is from the Library of Congress (<http://catalog.loc.gov/>) that introduced a new bibliographic metadata framework, BIBFRAME 2.0, which will eventually replace MARC21 and allow Semantic Web and Linked Open Data to interlink bibliographic metadata from different libraries. Other metadata encoding schema and frameworks include Schema.org, EDM, and the International Community for Documentation (CIDOC)’s Conceptual Reference Model (CIDOC-CRM).<sup>33</sup>

Today, the bibliographic metadata records are available on the web in several forms including MARC21, Online Public Access Catalogs (OPACs), and bibliographic descriptions from online catalogs (e.g., Library of Congress), online cooperative catalogs (e.g., OCLC’s WorldCat [<https://www.oclc.org/en/worldcat.html> program]), social collaborative cataloging applications (e.g., LibraryThing [<https://www.librarything.com>]), digital libraries (e.g., IEEE Xplore digital library [<https://ieeexplore.ieee.org/Xplore/home.jsp>]), ACM digital library(<https://dl.acm.org>), book search engines such as Google Books, and commercial databases including e.g., Amazon.com. Most of these cataloging web applications use either MARC or other legacy standards as metadata encoding and representation schemes. However, the majority of these applications are either considering or transiting to the emerging cataloging rules, frameworks, and encoding schemes so that the bibliographic descriptions of their holdings could be made visible and reusable as Linked and Open Data on the web for the broader interests of libraries, publishers, and end-users.

---

The presence of high-quality reusable vocabularies makes the consumption of Linked Data more meaningful, which is made possible by Linked Open Vocabularies (LOV) that bring value-added extensions to the Web of Data.<sup>34</sup> The following two subsections attempt to answer the RQ01 by highlighting how LOD and LOV are transforming the current digital landscape of cataloging.

### ***Linked and Open Data***

The Semantic Web and Linked Open Data have enabled libraries to publish and make visible their bibliographic data on the web, which increases the understanding and consumption of this metadata both for humans and machines.<sup>35</sup> LOD connects and relates bibliographic metadata on the web using Linked Data principles.<sup>36</sup> Publishing, linking, and consuming bibliographic metadata as Linked and Open Data brings several benefits. These include improvements in data visibility, linkage with different online services, interoperability through universal LOD platform, and the credibility due to user annotations.<sup>37</sup> Other benefits include: the semantic modeling of entities related to bibliographic resources; ease in transforming topics into SKOS; ease in the usage of linked library data in other services; better data visualization according to user requirements; linking and querying linked data from multiple sources; and improved usability of library linked data in other domains and knowledge areas.<sup>38</sup> Different users including scientists, students, citizens and other stakeholders of library data can benefit from adopting LOD in libraries.<sup>39</sup>

Linked Data has the potential to make bibliographic metadata visible, reusable, shareable, and exchangeable on the web with greater semantic interoperability among the consuming applications. Several major projects including BIBFRAME, LODLAM (Linked Open Data in Libraries Archives and Museums [<http://lodlam.net>]), and LD4L (Linked Data for Libraries [<https://www.ld4l.org>]) are in progress, which advocates for this potential.<sup>40</sup> Similarly, Library Linked Data (LLD) is LOD-based bibliographic datasets, available in MODS and MARC21 and could be used in making search systems more sophisticated and may also be used in LOV datasets to integrate applications requiring library and subjects domain datasets.<sup>41</sup>

Bianchini and Guerrini report on the current changes in the library and cataloging domains from Ranganathan's point of view of trinity (library, books, staff), which states that changes in one element of this trinity undoubtedly affect the others.<sup>42</sup> They found several factors including readers, collections, and services influence this trinity and emphasize for a change:

- Readers moved to the web from libraries and wanted to save their time but want many capabilities including searching and navigating the full-text of resources by following links. They want resources connected to similar and related resources. They want concepts interlinked to perform an exploratory search, find serendipitous results to fulfill their information needs.
- Collections encompass several changes from their production to dissemination, from search and navigation to the representation and presentation of content. The ways the users access them and catalogers describe them are changing. Their management is moving beyond the boundaries of their corresponding libraries to the open and broader landscape of Open Access context and exposure to LOD environment.
- Services are moving from bibliographic data silos to the Semantic Web. This affects moving the bibliographic model to a more connected and linked data model and environment of Semantic Web. The data is moving from bibliographic database management systems to large LOD graph, where millions of MARC records are reused and converted to new



---

encoding formats that are backward compatible with MARC21, RDA, and others and provide opportunities to be exploited fully by the Linked and Open Data environment.

Thinking along this direction, new cataloging rules and guidelines, such as RDA, are making us a part of the growing global activity of cataloging. Therefore, catalogers should take keen interest in and avail themselves of the opportunities that lie in Linked and Open Data for cataloging. Otherwise, they (as a service) might be forgotten or removed from the trinity, i.e., from collections and readers.<sup>43</sup>

Several major libraries have been actively working to make their bibliographic metadata visible and re-usable on the web. The Library of Congress through its Linked Data Service (<http://id.loc.gov>) enables humans and machines to access its authority data programmatically.<sup>44</sup> It exposes and interconnects data on the web through dereferenceable Uniform Resource Identifiers (URIs).<sup>45</sup> Its scope includes providing access to the commonly found LOC standards and vocabularies (controlled vocabularies and data values) for the list of authorities and controlled vocabularies that LOC currently supports.<sup>46</sup> According to the LOC, the Linked Data Service brings several benefits to the users including: accessing data at no cost; providing granular access to individual data values; downloading controlled vocabularies and their data values in numerous formats; enabling linking to LOC data values within the user metadata using Linked Data principles; providing a simple RESTful API, clear license and usage policy for each vocabulary; accessing data across LOC divisions through a unified endpoint; and visualizing relationships between concepts and values.<sup>47</sup> However, to fully exploit the potentials of LOD, LOC is mainly focusing on its BIBFRAME initiative.<sup>48</sup>

BIBFRAME is not only a replacement for the current MARC21 metadata encoding format it is a new way of thinking how the available large amount of bibliographic metadata could be shared, reused, and made available as Linked and Open Data.<sup>49</sup> The BIBFRAME 2.0 (<https://www.loc.gov/bibframe/docs/bibframe2-model.html>) model organizes information into *work* (the details of the about the work information), *instance* (work on specific subject quantity in numbers), *item* (format: print or electronic), and *nature* (copy/original work). BIBFRAME 2.0 elaborates the roles of the persons in the specific work as *agents*, and the subject of the work as *subjects* and *events*.<sup>50</sup>

According to Taniguchi, BIBFRAME 2.0 takes the bibliographic metadata standards to the Linked and Open Data with model and vocabulary that makes the cataloging more useful both inside and outside the library community.<sup>51</sup> To achieve this goal, it needs to fulfill two primary requirements. These include (1) accepting and representing metadata created with RDA by replacing the MARC21, and therefore, working as creating, exchanging, and sharing RDA metadata; (2) accepting and accommodating descriptive metadata for bibliographic resources created by libraries, cultural heritage communities, and users for the wide exchange and sharing. BIBFRAME 2.0 should comply with the Linked Data principles including the use of RDF and URIs.

In addition to the Library of Congress, OCLC through its Linked Data Research has also been actively involved in research on transforming and publishing its bibliographic metadata as Linked Data.<sup>52</sup> Under this program, OCLC aims to provide a technical platform for the management and publication of its RDF datasets at a commercial scale. It models the key bibliographic entities including *work* and *person* and populates them with legacy and MARC-based metadata. It extends

---

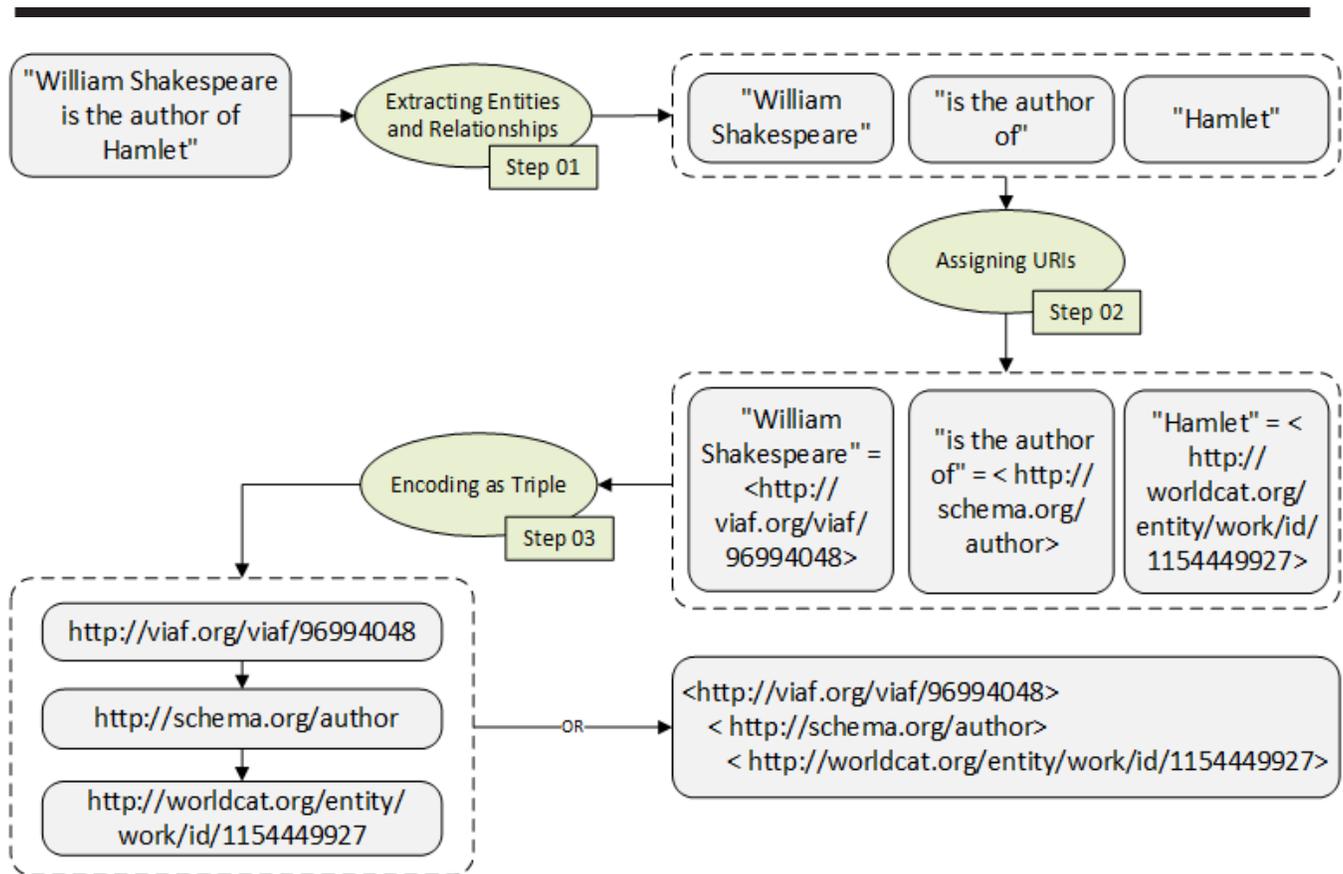
models to efficiently describe the contents of digital collections, art objects, and institutional repositories, which are not very well-described in MARC. It improves the bibliographic description of works and their translations. It manages the transition from MARC and other legacy encoding formats to Linked Data and develops prototypes for native consumption of Linked Data to improve resource description and discovery. Finally, it organizes teaching and training events.<sup>53</sup> Since 2012, OCLC has been publishing bibliographic data as Linked Data with three major LOD datasets including *OCLC Persons*, *WorldCat works*, and *WorldCat.org*.<sup>54</sup> Inspired from Google Research, currently, they have been working on *Knowledge Vault pipeline process* to harvest, extract, normalize, weigh, and synthesize knowledge from bibliographic records, authority files, and the web to generate Linked Data triples to improve the exploration and discovery experience of end-users.<sup>55</sup>

WorldCat.org publishes its bibliographic metadata as Linked Data by extracting a rich set of entities including persons, works, places, events, concepts, and organizations to make possible several web services and functionalities for resource discovery and access.<sup>56</sup> It uses Schema.org (<http://schema.org>) as the base ontology, which can be extended with different ontologies and vocabularies to model WorldCat bibliographic data to be published and consumed as Linked Data.<sup>57</sup> Tennant presents a simple example of how this works. Suppose we want to represent the fact “*William Shakespeare is the author of Hamlet*” as Linked Data.<sup>58</sup> To do this, the important entities should be extracted along with their semantics (relationships) and represented in a format that is both machine-processable and human-readable. Using Schema.org, Virtual International Authority File (VIAF.org), and WorldCat.org, the sentence can be represented as a Linked Data triple, as shown in figure 1 based on Tennant.<sup>59</sup>

The Digital Bibliography & Library Project (DBLP) is an online Computer Science bibliography that provides bibliographic information about major publications in Computer Science with the goal of providing free access to high-quality bibliographic metadata and links to the electronic version of these publications.<sup>60</sup> As of October 2018, it has indexed more than 4.3 million publications from more than 2.1 million authors and has indexed more than 40,000 journal volumes, 38,000 conference/workshop proceedings, and more than 80,000 monographs.<sup>61</sup> Its dataset is available on LOD that allows for faceted search and faceted navigation to the matching publications. It uses GrowBag graphs to create topic facets and uses DBLP++ datasets (an enhanced version of DBLP) and additional data extracted from the related webpages on the web.<sup>62</sup> A MySQL database stores the DBLP++ dataset that is accessible through several ways including (1) getting the database dump; (2) using its web services; (3) using D2R server to access it in RDF; and (4) getting the RDF dump available in N3 serialization.<sup>63</sup>

The above discussions on LOC, OCLC, and DBLP make it clear that LOD can potentially transform the cataloging landscape of libraries by making bibliographic metadata visible and reusable on the web. However, this potential can only be exploited to its fullest if relevant vocabularies are provided to make the Linked Data more meaningful. LOV fulfills this demand for relevant and standard vocabularies, discussed in the next subsection.





**Figure 1.** An Example of Publishing a Sample Fact as Linked Data (Based on Tennant<sup>64</sup>).

### **Linked Open Vocabularies**

Linked Open Vocabularies (LOV) are a “high-quality *catalog* of reusable vocabularies to describe Linked and Open Data.”<sup>65</sup> They assist publishers in choosing the appropriate vocabulary to efficiently describe the semantics (classes, properties, and data types) of the data to be published as Linked and Open Data.<sup>66</sup> LOV interconnect vocabularies, version control, the property type of values to be matched with a query to increase the score of the terms, and offers a range of data access methods including APIs, SPARQL endpoint, and data dump. The aim is to make the reuse of well-documented vocabularies possible in the LOD environment.<sup>67</sup> The LOV portal brings value-added extensions to the Web of Data, which is evident from its adoption in several state-of-the-art applications.<sup>68</sup> The presence of vocabulary makes the corresponding Linked Data meaningful, if the original vocabulary vanishes from the web, linked data applications that rely on it no longer function because they cannot validate against the authoritative source. LOV systems prevent vocabularies from becoming unavailable by providing redundant or back-up locations for these vocabularies.<sup>69</sup> The LOV catalog meets almost all types of search criteria including search using metadata, ontology, APIs, RDF dump, and SPARQL endpoint enabling it to provide a range of services regarding the reuse of RDF vocabularies.<sup>70</sup>

Linked Data should be accompanied by its meaning to achieve its benefits, which is possible using vocabularies especially RDF vocabularies that are also published as Linked Data and linked with each other forming an LOV ecosystem.<sup>71</sup> Such an ecosystem defines the health and usability of Linked Data by making its meaningful interpretation possible.<sup>72</sup> For an ontology or vocabulary to be included into the LOV catalog, it must be of an appropriate size with low-level and normalized

---

constraints and represented in RDFS or Web Ontology Language (OWL); it must allow creating instances and support documentation by permitting comments, labels, definitions, and descriptions to support end users.<sup>73</sup> The ontology must have additional characteristics such as those described in Semantic Web languages like OWL, published on the web with no limitations on its reuse, and support for content negotiation using searchable content and namespace URIs.<sup>74</sup> The LOV catalog offers four core functionalities that make it more attractive for libraries. The *Aggregate* accesses vocabularies through dump file or (a SPARQL) endpoint. The *Search* finds classes/properties in a vocabulary or ontology. The *Stat* displays descriptive statistics of LOV vocabularies. Finally, *Suggest* enables the registry of new vocabularies.<sup>75</sup>

Radio and Hanrath uncovered the concerns regarding transitioning to LOV including how pre-existing terms could be mapped while considering the potential semantic loss.<sup>76</sup> They describe this transition in the light of a case study at the University of Kansas institutional repository, which adopted OCLC's FAST vocabulary and analyzed the outcomes and impact of exposing their data as Linked Data. To them, a vocabulary that is universal in scope and detail can become "bloated" and may result in an aggregated list of uncontrolled terms. However, such a diverse system may be capable of accurately describing the contents of an institutional repository. In this regard, adopting Linked Data vocabulary may serve to increase the overall quality of data by ensuring consistency with greater exposure of the resources when published as LOD. However, such a transition to a Linked Data vocabulary is not that simple and gets complicated when the process involves reconciling the legacy metadata especially when dealing with the issues of under or misrepresentation.<sup>77</sup>

Publishers, commercial entities, and data providers such as universities are taking keen interest and consortial participation, and therefore the library community must contribute to, benefit from, and consider this inevitable opportunity seriously.<sup>78</sup> Considering, the core role of libraries in connecting people to the information, they should come forward to make available their descriptive metadata collections as Linked and Open Data for the benefit of the scholarly community on the web. It is time to move from strings (descriptive bibliographic records) to things (data items) that are connected in a more meaningful manner for the consumption of both machines and humans.<sup>79</sup>

Besides the numerous benefits of the LOV, there are some well-documented [and well-supported] vocabularies that are "not published or no longer available."<sup>80</sup> While focusing on the mappings between Schema.org and LOV, Nogales et al. argue that the LOV portal is limited as "some of the vocabularies are not available here."<sup>81</sup> In other words, the LOV portal is growing, but currently, it is at the infant stage, where much work is needed to bring all or at least the missing well-documented and well-supported vocabularies. This way the true benefits of LOV could be exploited to the fullest when such vocabularies are linked and made available for the consumption and reuse of the broader audience and applications of the Web of Data.

## CHALLENGES, ISSUES, AND RESEARCH OPPORTUNITIES

To answer the RQ02, this section attempts to identify some of the prominent/key challenges and issues regarding publishing and consuming bibliographic metadata as Linked and Open Data. The sheer scale and diversity of cataloging frameworks, metadata encoding schemes, and standards make it difficult to approach cataloging effectively and efficiently. The quality of the cataloging data is another dimension that needs proper attention.



---

### ***The Multiplicity of Cataloging Rules and Standards***

The importance and critical role of standards in cataloging are clear to everyone. With standards, it becomes possible to identify authors uniquely; link users to the intended and the required resources; assess the value and usage of the services a library or information system provides; operate efficiently different transactions regarding bibliographic metadata, link content, preserve metadata, and generate reports; and enable the transfer of notifications, data, and events across machines.<sup>82</sup> The success of these standards is because of the community-based efforts and their utility for a person/organization and ease of adoption.<sup>83</sup> However, we are living in a “jungle of standards” with massive scale and complexity.<sup>84</sup> We are facing a flood of standards, schemas, protocols, and formats to deal with bibliographic metadata.<sup>85</sup> It is necessary to come up with some uniform and widely accepted standard, schema, protocol, and format, which will make possible the uniformity between bibliographic records and make way for records de-duplication on the web. Also, because of the exponential growth of the digital landscape of document collections and the emerging yet widely adopted Linked Data environment, it becomes necessary for librarians to be part of this global scale activity of making their bibliographic data available as Linked and Open Data.<sup>86</sup> Therefore, all these standards need reenvisioning and reconsideration when libraries transit from the current implementations to a more complex LOD-based environment.<sup>87</sup>

RDA is easy to use, user-centric, and retrieval-supportive with a precise vocabulary.<sup>88</sup> However, it has lengthier descriptions with a lot of technical terms, is time-consuming, needs re-training, and suffers from the generation gap.<sup>89</sup> RDA is transitioning from AACR2 to produce metadata for knowledge artifacts, and it will be adaptive to the emerging data structures of Linked Data.<sup>90</sup> Although librarians could potentially play a vital role in making RDA successful, it is challenging to bring them on the same page with publishers and vendors.<sup>91</sup>

While studying BIBFRAME 2.0 from RDA point of view, Taniguchi observed that:

- BIBFRAME has no class correspondence with RDA, especially making a distinction between *Work* and *Expression* is challenging.
- Some RDA elements have no corresponding properties in BIBFRAME, and therefore, cannot be expressed in BIBFRAME. In other cases, BIBFRAME properties cannot be converted back to RDA elements due to the many-to-one and many-to-many mappings between them.
- The availability of multiple MARC21-to-BIBFRAME tools results in the variety of BIBFRAME metadata, which makes its matching and merging in the later stages challenging.<sup>92</sup>

To understand whether BIBFRAME 2.0 is suitable as a metadata schema, Taniguchi examined it closely for domain constraint of properties and developed four additional methods for implementing such constraints, i.e., defining properties in BIBFRAME.<sup>93</sup> In these methods, method 1 is the strictest one for defining such properties, method 2 from BIBFRAME, and the remaining gradually loosen. Method 1 defines the domain of individual properties as work or instance only, which is according to the method in RDA. Method 2 defines properties using multiclass structure (work-instance-item) for descriptive metadata. Method 3 introduces a new class *BibRes* to accommodate *Work* and *Instance* properties. Method 4 uses two classes *BibRes* and *Work* for representing a bibliographic resource. Method 5 leaves the domain of any property unspecified and uses *rdf:type* to represent whether a resource belongs to the *Work* or *Instance*. He observed that:

- The multi-class structure used in BIBFRAME (method 2) questions the consistency between this structure and the domain definition of the properties.
- If the quality of the metadata is concerned especially matching among converted metadata from different source metadata, then method 1 works better than method 2.
- If metadata conversion from different sources is required, then method 4 or 5 should be applied.<sup>94</sup>

Taniguchi concludes that BIBFRAME's domain constraint policy is unsuitable for descriptive metadata schema to exchange and share bibliographic resources, and therefore, should be reconsidered.<sup>95</sup>

According to Sprochi, bibliographic metadata is passing through a significant transformation.<sup>96</sup> FRBR, RDA, and BIBFRAME are among the three major and currently running programs that will affect the recording, storage, retrieval, reuse and sharing of bibliographic metadata. IFLA focuses on reconciling FRBR, FRAD, and FRSAD models into one model namely FRBR-Library Reference Model (RFBR-LRM [<https://www.ifla.org/node/10280>]), published in May 2016.<sup>97</sup> Sprochi further adds that it is generally expected that by adopting this new model, RDA will be changed and revised significantly. BIBFRAME will also get substantial modifications to become compatible with FRBR-LRM and the resulting RDA rules.<sup>98</sup> These initiatives, on the one hand, makes possible their visibility on the web, but on the other hand, introduces several changes and challenges for the library and information science community.<sup>99</sup> To cope with the challenges of making bibliographic data visible, available, reusable, and shareable on the web, Sprochi argues that:<sup>100</sup>

- The library and information science community must think of the bibliographic records in terms of data that is both human-readable and machine-understandable, which can be processed across different applications and databases with no format restrictions. Also, this data must support interoperability among vendors, publishers, users, and libraries and therefore, should be thought of beyond the notion that “only library create quality metadata (as quoted in Coyle (2007)” and cited by Sprochi<sup>101</sup>).
- A shared understanding of Semantic Web, LOD, data formats, and other related technologies is necessary for the library and information science community for more meaningful and fruitful conversations with software developers, Information & Library Science (ILS) designers, and IT & Linked Data professionals. At least some basic knowledge about these technologies will enable the library community to take active participation in publishing, storing, visualizing, linking, and consuming bibliographic metadata as Linked and Open Data.
- The library community must show a strong commitment to more ILS vendors to “post-MARC” standards such as BIBFRAME or any other standard that is supportive of the LOD environment. This way we will be in a better position to exploit Linked Data and Semantic Web to their fullest. The library community must be ready to adopt LOD in cataloging.

Transitioning from MARC to Linked Data needs collaborative efforts and requires addressing several challenges. These challenges include:



- 
- committing to a single standard by getting all units in the library, so that the *Big Data problem* resulting from using multiple metadata standards by different institutions could be mitigated;
  - bringing individual experts, libraries, university, and governments to work together and organize conferences, seminars, and workshops to bring Linked Data into the mainstream;
  - translating the BIBFRAME vocabulary into other languages;
  - involving different users and experts in the area; and
  - obtaining funding from the public sector and other agencies to continue the journey towards Linked Data.<sup>102</sup>

In the current scenario of metadata practices, the interoperability for the exchange of metadata varies across different formats.<sup>103</sup> The Semantic Web and LOD support different library models such as FRBRoo, EDM, and BIBFRAME. These conceptual models and frameworks suffer from the interoperability issue, which makes data integration difficult. Currently, several options are available for encoding bibliographic data to RDF (and to LOD), which further complicates the interoperability and introduces inconsistency.<sup>104</sup>

Existing descriptive cataloging methodologies and the bibliographic ontology descriptions in cataloging and metadata standards set the stage for redesigning and developing better ways of improved information retrieval and interoperability.<sup>105</sup> Besides the massive heaps of information on the web, the library community (especially digital libraries) has devised standards for metadata and bibliographic description to meet the interoperability requirements for this part of the data on the web.<sup>106</sup> Semantic Web technologies could be exploited to make information presentation, storage, and retrieval more user-friendly for digital libraries.<sup>107</sup> To achieve such interoperability among resources, Castro proposed an architecture for semantic bibliographic description.<sup>108</sup> Gardašević emphasizes on employing information system engineers and developers to understand resource description, discovery, and access process in libraries and then extend these practices by applying Linked Data principles.<sup>109</sup> This way bibliographic metadata will be more visible, reusable and shareable on the web. Godby, Wang, and Mixer stress collaborative efforts to establish a single and universal platform for cataloging rules, encoding schema, and model to a higher level of maturity, which requires initiatives such as RDA, BIBFRAME, LD4L, and BIBLOW (<https://bibflow.library.ucdavis.edu/about>).<sup>110</sup>

The massive volume of metadata (available in MARC and other legacy formats) makes data migration to BIBFRAME challenging.<sup>111</sup> Although BIBFRAME challenges the conventional ground of cataloging, which aims to record tangible knowledge containers, it is still in the infant stage at both theoretical and practical levels.<sup>112</sup> For BIBFRAME to be more efficient, enhanced, and enriched, it needs the attention of librarians and information science experts who will use it to encode their bibliographic metadata.<sup>113</sup> Gonzales suggests that librarians must be willing to share metadata and upgrade metadata encoding standards to BIBFRAME; they should train, learn, and upgrade their systems to efficiently use BIBFRAME encoding scheme and research new ways of bringing interoperability between BIBFRAME and other legacy metadata standards; and they should ensure the data security of patrons and mitigate the legal and copyright issues in making visible their resources as Linked and Open Data.<sup>114</sup> Also, LOV must be exploited from the cataloging perspective by finding out ways to create a single, flexible, adaptable, and representative vocabulary. Such a vocabulary will bring the cataloging data from different

---

libraries of the world and make it accessible and consumable as a single Library Linked Data to get free from the jungle of metadata vocabularies [and standards].

### ***Publishing and Consuming Linked Bibliographic Metadata***

According to the findings of one survey, there are several primary motives for publishing an institution's [meta]data as Linked Data. These include (in the order from most frequent/ essential to a lesser one):<sup>115</sup>

- making data visible on the web;
- experimenting and finding the potentials of publishing datasets as Linked Data;
- exposing local datasets to understand the nature of Linked Data;
- exploring the benefits of Linked Data for Search Engine Optimization (SEO);
- consuming and reusing Linked Data in future projects;
- increasing the data reusability and interoperability;
- testing Schema.org and BIBFRAME;
- meeting the requirements of the project; and
- making available the "stable, integrated, and normalized data about research activities of an institution."<sup>116</sup>

They also identified several reasons from the participants regarding the consumption of such data. These include (in the order from most frequent/essential to a lesser one):<sup>117</sup>

- improving the user experience;
- extending local data with other datasets;
- effectively managing the internal metadata;
- improving the accuracy and scope of search results;
- trying to improve SEO for local resources;
- understanding the effect of data aggregation from multiple datasets; and
- experimenting and finding the potentials of consuming Linked Datasets.

Publishing and consuming bibliographic data on the LOD cloud brings numerous applications. Kalou et al. developed a semantic mashup by combining Semantic Web technologies, RESTful services, and content management services (CMS) to generate personalized book recommendations and publish them as Linked Data.<sup>118</sup> It allows for the expressive reasoning and efficient management of ontologies and has potential applications in the library, cataloging services, and ranking book records and reviews. This application exemplifies how we can use the commercially [and socially] curated metadata with bibliographic descriptions from improved user experience in digital libraries using Linked Data principles. However, publishing and consuming bibliographic metadata as Linked and Open Data is not that simple and need addressing several prominent challenges and issues, which are identified in the following subsections along with some opportunities for further research.

### ***Publishing Linked Bibliographic Metadata***

The University of Illinois Library worked on publishing MARC21 records of 30,000 digitized books as Linked Library Data by adding links, transforming them to LOD-friendly semantics (MODS) and deploying them as RDF with the objective to be used by a wider community.<sup>119</sup> To them, using Semantic Web technologies, a book can be linked to related resources and multiple possible



---

contexts, which is an opportunity for libraries to build innovative user-centered services for the dissemination and uses of bibliographic metadata.<sup>120</sup> In this regard, the challenge is to utilize the existing book-related bibliographic maximally and descriptive metadata in a manner that parallels with the services (both inside the library and outside) as well as exploit to the fullest the full-text search and Semantic Web technologies, standards, and LOD services.<sup>121</sup>

While publishing the National Bibliographic Information as free open Linked Data, IFLA identifies several issues including:<sup>122</sup>

- dealing with the negative financial impact on the revenue generated from traditional metadata services;
- the inability to offer consistent services due to the complexity of copyright and licensing frameworks;
- the confusion in understanding the difference between “open” and “free” terms;
- remodeling library data as Library Linked Data;
- the limited persistence and sustainability of Linked Data resources;
- the steep learning curve in understanding and applying Linked Data practices to library data;
- making choices between sites to link to; and
- creating persistent URIs for library data objects.

From the analysis of the relevant literature, Hallo identified several issues in publishing bibliographic metadata as Linked and Open Data. These include difficulties in cataloging and migrating data to new conceptual models; the multiplicity of vocabularies for the same metadata; the lack of agreements to share data; the lack of experts and tools for transforming data; the lack of applications and indicators for its consumption; mapping issues; providing useful links of datasets; defining and controlling data ownership; and ensuring dataset quality.<sup>123</sup> Libraries should adopt to Linked Data five-stars model by adopting emerging non-proprietary formats to publish its data; link to external resources and services; participate actively in enriching; and improving the quality of metadata to improve knowledge management and discovery.<sup>124</sup> The cataloging has a bright future with more dataset providers by involving citizens and end-users in metadata enrichment and annotation; making ranking and recommendation as part of library cataloging services; and the increased participation of the library community to the body of Semantic Web and Linked Data.<sup>125</sup>

Publishing Linked Data poses several issues. These include data cleanup issues especially when dealing with legacy data; technical issues such as data ownership; the software maturity to keep Linked Data up-to-date; managing its colossal volume; and providing IT support for data entry, annotation, and modeling; developing representative and widely applicable LOVs; and handling the steep learning curve to understand and apply Linked Data principles.<sup>126</sup> Bull and Quimby stress understanding how the library community is transiting their cataloging methods, systems, standards, and integrations to the LOD for making them visible on the web and how they keep backward compatibility with legacy bibliographic metadata.<sup>127</sup>

It is necessary for the LOD data model to maintain the underlying semantics of the existing models, schemas, and standards, yet innovate and renew old traditions, where the quality of the conversion solely depends on the ability of this new model to cope with heterogeneity conflicts,

---

maintain granularity and semantic attributes and consequently prevent loss of data and semantics.<sup>128</sup> The new model should be semantically expressive enough to support meaningful and precise linking to other datasets. By thinking alternatively, these challenges are the significant research opportunities that will enable us to be part of Linked and Open Data community in a more profound manner.

#### *Consuming Linked Bibliographic Metadata*

Consuming Linked Data resources can be a daunting task and may involve resolving/mitigating several challenges. These challenges include:<sup>129</sup>

- dealing with the bulky or non-available RDF dumps, no authority control within RDF dumps, and data format variations;
- identifying terms' specificity levels during concept matching;
- the limited reusability of Library Linked Data due to lack of contextual data;
- harmonizing classes and objects at the institution level;
- excessive handcrafting due to few off-the-shelf visualization tools;
- manual mapping of vocabularies;
- matching, aligning, and disambiguating library and Linked Data;
- the limited representation of several essential resources as Linked Data due to non-availability of URIs;
- the lack of sufficient representative semantics for bibliographic data;
- the time-consuming nature of Linked Data to understand its structure for reuse;
- the ambiguity of terms across languages; and
- the non-stability of endpoints and outdated datasets.

Syndication is required to make library data visible on the web. Also, it is necessary to understand how current applications including web search engines perceive and treat visibility, to what extent schema.org matters, and what is the nature of the Linked Data cloud.<sup>130</sup>

An influential work may be translated into several languages, which results in multiple metadata records. Some of these are complete, and others are with missing details. Godby and Smith-Yoshimura suggest aggregating these multiple metadata records into a single record, which can be complete, link the work to its different translations and translators, and is publishable (and consumable) as Linked Data.<sup>131</sup> However, such an aggregation demands a great deal of human effort to make these records visible and consumable as Linked Data. This also includes describing all types of objects that libraries currently collect and manage, translating research findings to best practices; and establishing policies to use URIs in MARC and other types of records.<sup>132</sup> To achieve the long-term goal of making metadata consumable as Linked Data; the libraries, as well as individual researchers, should align their research with work that of the major players such as OCLC, LOC, and IFLA and follow their best practices.<sup>133</sup>

The issues in LOV needs immediate attention to make LOD more useful. These issues, according to include the following:<sup>134</sup>

- LOV publishes only a subset of RDF vocabularies with no inclusion for value vocabularies such as SKOS thesaurus;
- it provides no or almost negligible support for vocabulary authors;



- 
- it relies on third parties to get the information about vocabulary usage in published datasets;
  - it has insufficient support for multilingualism or many languages;
  - it should support multi-term vocabulary search, which is required from the ontology designers to understand and employ the complex relationships among concepts;
  - it should support vocabulary matching, vocabulary checking, and multilingualism to allow users to search and browse vocabularies using their native language. It also improves the quality of the vocabulary by translation, which allows the community to evaluate and collaborate; and
  - efforts are required to improve and make possible the long-term preservation of vocabularies.

LOD emerged to change the design and development of metadata, which has implications for controlled vocabularies, especially, the Person/Agent vocabularies that are fundamental to data linkage but suffer from the issues of metadata maintenance and verification.<sup>135</sup> Therefore, practical data management and the metadata-to-triples transition should be studied in detail to make the wider adaptation of LOD possible.<sup>136</sup> To come out of the lab environment and make LOD practically useful, the controlled vocabularies must be cleaned, and its cost should be reduced.<sup>137</sup> However, achieving this is challenging and needs to answer how knowledge artifacts could be uniquely identified and labeled across digital collections and what should be the standard practices to use them.<sup>138</sup>

Linked Data is still new to libraries.<sup>139</sup> The technological complexities, the feeling of risks in adopting new technology and limitations due to the system, politics, and economy are some of the barriers in its usage in libraries.<sup>140</sup> However, libraries can potentially overcome these barriers by learning from the use of Linked Data in other domains including, e.g., Google's Knowledge Graph and Facebook's Open Graph.<sup>141</sup> The graph interfaces could be developed to link author, publisher, and book-related information, which in turn can be linked to the other open and freely available datasets.<sup>142</sup> It is time that the Library and Information Science professionals come out of the old, document-centric approach to bibliographic metadata and adapt their thinking as more data-centric for a more meaningful consumption of bibliographic metadata by both users and machines.<sup>143</sup>

### ***Quality of Linked Bibliographic Metadata***

The use of a cataloging data defines its quality.<sup>144</sup> The quality is essential for the discovery, usage, provenance, currency, authentication, and administration of metadata.<sup>145</sup> Cataloging data or bibliographic metadata is considered fit for use based on its accuracy, completeness, logical consistency, provenance, coherence, timeliness, conformance and accessibility.<sup>146</sup> Data is commonly assessed by its quality to be used in specific application scenarios and use cases, however, sometimes, low-quality data can still be useful for a specific application as far as its quality meets the requirements of that application.<sup>147</sup> The reasons include several factors including availability, accuracy, believability, completeness, conciseness, consistency, objectivity, relevance, understandability, timeliness, and verifiability that determine the quality of data.<sup>148</sup>

The quality of Linked Data can be of two types, one is the inherent quality of Linked Data, and the other relates to its infrastructure aspects. The former can be further divided into aspects including domain, metadata, RDF model, links among data items, and vocabulary. The infrastructural

---

aspects include the server that hosts the Linked Data, Linked Data fragments, and file servers.<sup>149</sup> This typology introduces issues of their own, the issues related to the inherent quality including “linking, vocabulary usage and the provision of administrative metadata.”<sup>150</sup> The infrastructural aspect introduces issues related to naming conventions, which include avoiding blank nodes and using HTTP URIs, linking through *owl:sameAs* links, describing by reusing the existing terms and dereferencing.<sup>151</sup>

The quality cataloging definitions are mainly based on the experience and practices of the cataloging community.<sup>152</sup> Its quality falls into at least four basic categories: (1) the technical details of the bibliographic records, (2) the cataloging standards, (3) the cataloging process, and (4) the impact of cataloging on the user.<sup>153</sup> The cataloging community focuses mainly on the quality of bibliographic metadata. However, it is not sufficient enough to consider the accuracy, completeness, and standardization of bibliographic metadata, and therefore, it is necessary that they should also consider the information needs of the users.<sup>154</sup>

Van Kleeck et al. investigated issues in the quality management of metadata of electronic resources to assess in supporting user tasks of finding, selecting, and accessing library holdings as well as identifying the potential for increasing efficiencies in acquisition and cataloging workflow.<sup>155</sup> They evaluated the quality of existing bibliographic records mostly provided by their vendors and compared them with those of OCLC and found that the latter has better support users in resource discovery and access.<sup>156</sup> From the management perspective, the complexity and volume of bibliographic metadata and the method of ingesting it to the catalog emphasize the selection of highest quality records.<sup>157</sup>

From the perspective of digital repositories, the absence of well-defined theoretical and operational definitions of metadata quality, interoperability, and consistency are some of the issues for the quality of metadata.<sup>158</sup> The National Information Standards Organization (NISO) identifies several issues in creating metadata.<sup>159</sup> These include the inadequate knowledge about cataloging in both manual and automatic environments leading to inaccurate data entry, inconsistency of subject vocabularies, and limitations of resource discovery, and the development of standardized approaches to structure metadata.<sup>160</sup>

The poor quality of Linked Data can make its usefulness much difficult.<sup>161</sup> Datasets are created at the data level resulting in a significant variance in perspectives and underlying data models.<sup>162</sup> This also leads to errors in triplication, syntax, and data; misleading *owl:sameAs* links, and the low availability of SPARQL endpoints.<sup>163</sup> Library catalogs, because of their low quality, most often fail to communicate clear and correct information correctly to the users.<sup>164</sup> The reasons for such low quality include user’s inability to produce catalogs that are free from faults and duplicates as well as low standards and policies that drive these cataloging practices.<sup>165</sup> Although the rich collections of bibliographic metadata are available, these are rich in terms of the heaps of cataloging data and not in terms of quality with almost no bibliographic control.<sup>166</sup> These errors in and the low quality of bibliographic metadata are the result of misunderstanding the aims and functions of bibliographic metadata and adopting the “unwise” cataloging standards and policies.<sup>167</sup> Still there exist some high-quality cataloging efforts with well-maintained cataloging records, where the only quality warrant is to correctly understand the subject matter of the artifact and effectively communicate between librarians and experts in the corresponding domain knowledge.<sup>168</sup> The demand for such high quality and well-managed catalogs has increased on the web. Although



---

people are more accustomed to web search engines, the quality catalogs will attract not only libraries but the general web users as well (when published and consumed as Linked Data).<sup>169</sup>

The community must work together on metadata with publishers and vendors to approach cataloging from the user perspective and refine the skillset as well as produce quality metadata.<sup>170</sup> As library and information science professionals, we should not only be the users of the standards, instead, we must actively participate and contribute to its development and improvement so that we may effectively and efficiently connect our data with the rest of the world.<sup>171</sup> Such collaboration is required from not only the librarians and vendors but also from the users in developing an efficient cataloging environment and for a more usable bibliographic metadata, this is discussed in the next section.

## **LINKING THE SOCIALLY CURATED METADATA**

This section addresses RQ03 by reviewing the state-of-the-art literature from multiple but related domains including Library Sciences, Information Sciences, Information Retrieval, and Semantic Web. The section below discusses the importance and possible impact of making socially curated metadata as part of the bibliographic or professionally curated metadata. The next section highlights why social collaborative cataloging approaches should be adopted by librarians to work with other stakeholders in making their bibliographic data available and visible as Linked and Open Data and what is the possible impact of fusing the user-generated content with professional metadata and making it available as Linked and Open Data.

### ***The Socially Curated Metadata Matters in Cataloging***

Conventional libraries have clear and well-established classification and cataloging schemes but these are as challenging to learn, understand, and apply as they are slow and painful to consume.<sup>172</sup> Using computers to retrieve bibliographic records resulted in the massive usage of copy cataloging.<sup>173</sup> However, adopting this practice is challenging, because these records are inconsistent; incomplete; less visible, granular, and discoverable; unable to integrate metadata and content to the corresponding records; difficult to preserve with new and usable format for the consumption by users and machines; and not supportive towards integrating the user-generated content into the cataloging records.<sup>174</sup> The University of Illinois Library, through its VuFind service, offers extra features to enhance the search and exploration experience of end users by providing a book's cover image, table of contents, abstracts, reviews, comments, and user tags.<sup>175</sup> Users can contribute content such as tags, reviews, comments, and recommend books to friends. However, it is necessary to research whether this user-generated content should be integrated to or preserved along the bibliographic records.<sup>176</sup>

In their book, Alemu and Stevens mentioned several advantages of making user-generated content as part of the library catalogs.<sup>177</sup> These include (i) enhancing the functionality of professionally-curated metadata by making information objects findable and discoverable; (ii) removing the limitations posed by sufficiency and necessity principles of the professionally-curated metadata; (iii) bringing users closer to the library by “pro-actively engaging” them in ratings, tagging, and reviewing, etc., provided that users are also involved in managing and controlling metadata entries; and (iv) the resulting “wisdom of the crowd” would benefit all the stakeholders from this massively growing socially-curated metadata. However, this combination can only be utilized optimally if we can semantically and contextually link it to the internal and external resources; the resulting metadata is openly accessed, shared, and reused; users are supported in easily adding

---

the metadata and made part of the quality control by enabling them to report spamming activities to the metadata experts.<sup>178</sup>

LibraryThing for Libraries (LTFL) makes a library catalog more informative and interactive by enhancing OPAC, providing access to professional and social metadata, and enabling them to search, browse, and discover library holdings in a more engaging way (<https://www.librarything.com/forlibraries>). It is one of the practical examples of enriching library catalogs with user-generated content. This trend of merging social and professional metadata innovates library cataloging by dissolving the borders between “social sphere” and library resources.<sup>179</sup> The social media has expanded library into social spaces by exploiting tags and tag clouds as navigational tools and enriching the bibliographic descriptions by integrating the user-generated content.<sup>180</sup> It bridges the communication gaps between the library and its users, where users take active participation in resource description, discovery, and access.<sup>181</sup>

The potential role of the socially curated metadata in resource description, discovery, and access is also evident from the long long-tail Social Book Search research under the Initiative for XML Retrieval (INEX) where both professionally curated bibliographic and user-generated social metadata are exploited for retrieval and recommendation to support both known-item as well as exploratory search.<sup>182</sup> By experimenting with Amazon/LibraryThing datasets of 2.8 million book records, containing both professional and social metadata, the results conclude that enriching the professional metadata with social metadata especially tags significantly improves search and recommendation.<sup>183</sup> Koolen also noticed that the social metadata especially tags and reviews significantly improve the search performance as professionally curated metadata is “often too limited” to describe books resourcefully.<sup>184</sup>

Users add socially curated metadata with the intention of making resource re-findable during a future visit, i.e., they add metadata such as tags to facilitate themselves and allow other similar users in resource discovery and access, and therefore, form a community around the resource.<sup>185</sup> Clements found user tags (social tagging) beneficial for librarians while browsing and exploring the library catalogs.<sup>186</sup> To some librarians, tags are complementary to controlled vocabulary; however, training issues and lack of awareness of social tagging functionality in cataloging interfaces prevent their perceived benefit.<sup>187</sup>

### ***The Socially Curated Metadata as Linked Data***

Metadata is socially constructed.<sup>188</sup> It is shaping and shaped by the context in which it is developed and applied, and demands community-driven approaches, where data should be looked at from a holistic point of view rather than considering them as discrete (individual) semantic units.<sup>189</sup> The library is adopting the collaborative social aspect of cataloging that will take place between authors, repository managers, libraries, e-collection consortiums, publishers, and vendors.<sup>190</sup> Librarians should improve their cataloging skills in line with the advances in technology to expose and make visible their bibliographic metadata as Linked and Open Data.<sup>191</sup>

Currently, *linked* library data is generated and used by library professionals. Socially constructed metadata will act as a value-added in retrieving knowledge artifacts with precision.<sup>192</sup> The addition of socially constructed and community-driven metadata in current metadata structures, controlled vocabularies, and classification systems provide the holistic view of these structures as they add the *community-generated sense* to the professionally-curated metadata structures.<sup>193</sup> An



---

example of the possibilities of making user-generated content as part of cataloging and Linked Open Data is the semantic book mashup (see “Consuming Linked Bibliographic Metadata” above) which demonstrates how the commercially [and socially] curated metadata could be retrieved and linked with bibliographic descriptions.<sup>194</sup> While enumerating the possible applications of this mashup, they argue that book reviews from different websites could be aggregated using Linked Data principles by extending the *Review* class of BIBFRAME 2.0.<sup>195</sup>

From the analysis of twenty-one in-depth interviews with LIS professionals, Alemu discovered four metadata principles, namely metadata enrichment, linkage, openness, and filtering.<sup>196</sup> This analysis revealed that the absence of socially curated metadata is sub-optimal for the potential of LOD in libraries.<sup>197</sup> Their analysis advocates for a *mixed-metadata approach*, in which social metadata (tags, ratings, and reviews) augments the bibliographic metadata by involving users proactively and by offering a social collaborative cataloging platform. The metadata principles should be reconceptualized, and Linked Data should be exploited to address the existing library metadata challenges. Therefore, the current efforts in Linked Data should fully consider social metadata.<sup>198</sup> Library catalogs should be enriched by mixing the professional and social metadata as well as semantically and contextually interlinked to internal and external information resources to be optimally used in different application scenarios.<sup>199</sup> To fully exploit this linkage, the duplication of metadata should be reduced. It must be made openly accessible so that its sharing, reuse, mixing, and matching could be made possible. The enriched metadata must be filtered per user requirements using an interface that is flexible, personalized, contextual, and re-configurable.<sup>200</sup> Their analysis suggests a “*paradigm shift*” in metadata’s future, i.e., from simple to enriched; from disconnected, invisible and locked to well-structured, machine-understandable, interconnected, visible, and more visualized metadata; and from single OPAC interface to reconfigurable and adaptive metadata interfaces.<sup>201</sup> By involving users in the metadata curation process, the mixed approach will bring diversity in metadata and make resources discoverable, usable, and user-centric with the wider and well-supported platform of Linked and Open Data.<sup>202</sup>

In conclusion, the fusion of socially curated metadata with the standards-based professional metadata is essential from the perspective of the user-centric paradigm of cataloging, which has the potential to aid resource discovery and access and open new opportunities for information scientists working in Linked and Open Data as well as catalogers who are transiting to the Web of Data to make their metadata visible, reusable, and linkable to other resources on the web. From the analysis and scholarly discussions of Alemu, Stevens, Farnel, and others as well as from the initial experiments of Kalou et al.<sup>203</sup> it becomes apparent that the application of Linked Data principles for library catalogs is future-proof and promising towards more user-friendly search and exploration experience with efficient resource description, discovery, access, and recommendations.

## CONCLUSIONS

In this paper, we presented a brief yet holistic review of the current state of Linked and Open Data in cataloging. The paper identified the potentials of LOD and LOV in making the bibliographic descriptions publishable, linkable, and consumable on the web. Several prominent challenges, issues, and future research avenues were identified and discussed. The potential role of socially-curated metadata for enriching library catalogs and the collaborative social aspect of cataloging were highlighted. Some of the notable points include the following:

- Publishing, linking, and consuming bibliographic metadata on the web using Linked Data principles brings several benefits for libraries.<sup>204</sup> The library community should improve their skills regarding this paradigm shift and adopt the best practices from other domains.<sup>205</sup>
- Standards have a key role in cataloging, however, we are living in a “jungle of metadata standards” with varying complexity and scale, which makes it difficult to select, apply and work with.<sup>206</sup> To be part of global scale activity of making bibliographic data available on the web as Linked and Open Data, these standards should be considered and re-envisioned.<sup>207</sup>
- The quality of bibliographic metadata depends on several factors including accuracy, completeness, logical consistency, provenance, coherence, timeliness, conformance and accessibility.<sup>208</sup> However, achieving these characteristics is challenging because of several reasons including cataloging errors; limited bibliographic control; misunderstanding the role of metadata; and “unwise” cataloging standards and policies.<sup>209</sup> To ensure high-quality and make data visible and reusable as Linked Data, the library community should contribute to developing and refining these standards and policies.<sup>210</sup>
- Metadata is socially constructed and demands community-driven approaches and the social collaborative aspect of cataloging by involving authors, repository managers, librarians, digital collection consortiums, publishers, vendors, and users.<sup>211</sup> This is an emerging trend, which is gradually dissolving the borders between the “social sphere” and library resources and bridging the communication gap between libraries and their users, where end users contribute to the bibliographic descriptions resulting in a diversity of metadata and making it user-centric and usable.<sup>212</sup>
- Adopting a “mixed-metadata approach” by considering bibliographic metadata and the user-generated content complementary and essential for each other suggests a “*paradigm shift*” in the metadata’s future from simple to enriched; from human-readable data silos to machine understandable, well-structured, and reusable; from invisible and restricted to visible and open; and from single OPAC to reconfigurable interfaces on the web.<sup>213</sup>

Several researchers including the ones cited in this article agree that the professionally curated bibliographic metadata supports mostly the known-item search and has little value to open and exploratory search and browsing. They believe that not only the collaborative social efforts of the cataloging community are essential but also the socially curated metadata, which can be used to enrich bibliographic metadata and support exploration and serendipity. This is not only evident from the wider usage of LibraryThing and its LTFL but also from the long-tail INEX Social Book Search research where both professionally curated bibliographic and user-generated social metadata are exploited for retrieval and recommendation to support both known-item as well as exploratory search.<sup>214</sup> Therefore, this aspect should be considered for further research to make cataloging more useful for all the stakeholders including libraries, users, authors, publishers, and for the general consumption as Linked Data on the web.

The current trend of social collaborative cataloging efforts is essential to fully exploit the potential of Linked Open Data. However, if we look closely, we find four groups including librarians, Linked Data experts, Information Retrieval (IR) and Interactive IR researchers; and users, all going on their separate ways with minimal collaboration and communication. More specifically, they are not benefiting from each other to a greater extent, which could result in better possibilities of



---

resource description, discovery, and access. For example, the library community should consider the findings of INEX SBS track, which have demonstrated that professional and social metadata, are essential for each other to facilitate end users in resource discovery and access and support not only known-item search but also exploration and serendipity. The current practices of LibraryThing, LTFL, and social web in general advocate for user-centric cataloging, where users are not only the consumers of bibliographic descriptions but also the contributors to metadata enrichment. Linked Open Data experts have achieved significant milestones in other domains including, e.g., e-Government, they should understand the cataloging and resource discovery & access practices in libraries to make the bibliographic metadata not only visible as Linked Data on the web but also shareable, re-usable, and beneficial to the end-users. The social collaborative cataloging approach by involving the four mentioned groups actively is significant to make bibliographic descriptions more useful not only for the library community and users but also for their consumption on the web as Linked and Open Data. Together we can, and we must.

## REFERENCES

- <sup>1</sup> María Hallo et al., “Current State of Linked Data in Digital Libraries,” *Journal of Information Science* 42, no. 2 (2016):117–27, <https://doi.org/10.1177/0165551515594729>.
- <sup>2</sup> Tim Berners-Lee, “Design Issues: Linked Data,” W3C, 2006, updated June 18, 2009, accessed November 09, 2018, <https://www.w3.org/DesignIssues/LinkedData.html>; Hallo, “Current State,” 117.
- <sup>3</sup> Yuji Tosaka and Jung-ran Park, “RDA: Resource Description & Access—A Survey of the Current State of the Art,” *Journal of the American Society for Information Science and Technology* 64, no. 4 (2013): 651–62, <https://doi.org/10.1002/asi.22825>.
- <sup>4</sup> Hallo, “Current State,” 118; Angela Kroeger, “The Road to BIBFRAME: The Evolution of the Idea of Bibliographic Transition into a Post-MARC Future,” *Cataloging & Classification Quarterly* 51, no. (2013): 873–90. <https://doi.org/10.1080/01639374.2013.823584>; Martin Doerr et al., “The Europeana Data Model (EDM).” Paper presented at the World Library and Information Congress: 76th IFLA General Conference and Assembly, Gothenburg, Sweden, August 10–15, 2010.
- <sup>5</sup> Getaneh Alemu and Brett Stevens, *An Emergent Theory of Digital Library Metadata—Enrich then Filter*, 1st Edition (Waltham, MA: Chandos Publishing, Elsevier Ltd. 2015).
- <sup>6</sup> Hallo, “Current State,” 118 .
- <sup>7</sup> Berners-Lee, “Design Issues.”
- <sup>8</sup> Kim Tallerås, “Quality of Linked Bibliographic Data: The Models, Vocabularies, and Links of Data Sets Published by Four National Libraries,” *Journal of Library Metadata* 17, no. 2 (2017):126–55, <https://doi.org/10.1080/19386389.2017.1355166>.
- <sup>9</sup> Becky Yoose and Jody Perkins, “The Linked Open Data Landscape in Libraries and Beyond,” *Journal of Library Metadata* 13, no. 2–3 (2013): 197–211, <https://doi.org/10.1080/19386389.2013.826075>.

- 
- <sup>10</sup> Robert Fox, "From Strings to Things," *Digital Library Perspectives* 32, no. 1 (2016): 2–6, <https://doi.org/10.1108/DLP-10-2015-0020>.
- <sup>11</sup> Stanislava Gardašević, "Semantic Web and Linked (Open) Data Possibilities and Prospects for Libraries," *INFOtheca—Journal of Informatics & Librarianship* 14, no. 1 (2013): 26–36, [http://infoteka.bg.ac.rs/pdf/Eng/2013-1/INFOTHECA\\_XIV\\_1\\_2014\\_26-36.pdf](http://infoteka.bg.ac.rs/pdf/Eng/2013-1/INFOTHECA_XIV_1_2014_26-36.pdf).
- <sup>12</sup> Thomas Baker, Pierre-Yves Vandenbussche, and Bernard Vatant, "Requirements for Vocabulary Preservation and Governance," *Library Hi Tech* 31, no. 4 (2013): 657–68, <https://doi.org/10.1108/LHT-03-2013-0027>.
- <sup>13</sup> Pierre-Yves Vandenbussche et al., "Linked Open Vocabularies (LOV): A Gateway to Reusable Semantic Vocabularies on the Web," *Semantic Web* 8, no. 3 (2017): 437–45, <https://doi.org/10.3233/SW-160213>.
- <sup>14</sup> Tosaka, "RDA," 651, 652.
- <sup>15</sup> Amanda Sprochi, "Where Are We Headed? Resource Description and Access, Bibliographic Framework, and the Functional Requirements for Bibliographic Records Library Reference Model," *International Information & Library Review* 48, no. 2 (2016): 129–36, <https://doi.org/10.1080/10572317.2016.1176455>.
- <sup>16</sup> Brigid M. Gonzales, "Linking Libraries to the Web: Linked Data and the Future of the Bibliographic Record," *Information Technology and Libraries* 33, no. 4 (2014): 10, <https://doi.org/10.6017/ital.v33i4.5631>.
- <sup>17</sup> Shoichi Taniguchi, "Is BIBFRAME 2.0 a Suitable Schema for Exchanging and Sharing Diverse Descriptive Metadata about Bibliographic Resources?," *Cataloging & Classification Quarterly* 56, no. 1 (2018): 40–61, <https://doi.org/10.1080/01639374.2017.1382643>.
- <sup>18</sup> Shoichi Taniguchi, "BIBFRAME and Its Issues: From the Viewpoint of RDA Metadata," *Journal of Information Processing and Management* 58, no. 1 (2015): 20–27, <https://doi.org/10.1241/johokanri.58.20>.
- <sup>19</sup> Shoichi Taniguchi, "Examining BIBFRAME 2.0 from the Viewpoint of RDA Metadata Schema," *Cataloging & Classification Quarterly* 55, no. 6 (2017): 387–412, <https://doi.org/10.1080/01639374.2017.1322161>.
- <sup>20</sup> Nosheen Fayyaz, Irfan Ullah, and Shah Khusro, "On the Current State of Linked Open Data: Issues, Challenges, and Future Directions," *International Journal on Semantic Web and Information Systems (IJSWIS)* 14, no. 4 (2018): 110–28, <https://doi.org/10.4018/IJSWIS.2018100106>.
- <sup>21</sup> Asim Ullah, Shah Khusro, and Irfan Ullah, "Bibliographic Classification in the Digital Age: Current Trends & Future Directions," *Information Technology and Libraries* 36, no. 3 (2017): 48–77, <https://doi.org/10.6017/ital.v36i3.8930>.
- <sup>22</sup> Tosaka, "RDA," 659.



- 
- <sup>23</sup> Tosaka, “RDA,” 651, 652, 659.
- <sup>24</sup> Tosaka, “RDA,” 653, 660.
- <sup>25</sup> The first author used the trial version of RDA Toolkit to report these facts about RDA (<https://access.rdatoolkit.org>). RDA Toolkit is co-published by American Library Association (<http://www.ala.org>), Canadian Federation of Library Associations (<http://cfla-fcab.ca/en/home-page>), and Facet Publishing (<http://www.facetpublishing.co.uk>).
- <sup>26</sup> IFLA, “IFLA Conceptual Models,” The International Federation of Library Associations and Institutions (IFLA), 2017, updated April 06, 2009, accessed November 12, 2018, <https://www.ifla.org/node/2016>.
- <sup>27</sup> Tosaka, “RDA,” 651, 652, 655.
- <sup>28</sup> Michael John Khoo et al., “Augmenting Dublin Core Digital Library Metadata with Dewey Decimal Classification,” *Journal of Documentation* 71, no. 5 (2015): 976–98. <https://doi.org/10.1108/JD-07-2014-0103>; Ulli Waltinger et al., “Hierarchical Classification of Oai Metadata Using the DDC Taxonomy,” in *Advanced Language Technologies for Digital Libraries*, edited by Raffaella Bernardi, Frederique Segond and Ilya Zaihrayeu. Lecture Notes in Computer Science (Lncs), 29–40: Springer, Berlin, Heidelberg, 2011; Aaron Krowne and Martin Halbert, “An Initial Evaluation of Automated Organization for Digital Library Browsing,” Paper presented at the Proceedings of the 5th ACM/IEEE-CS joint conference on Digital libraries, Denver, CO, USA, June 7–11, 2005 2005; Waltinger, “DDC Taxonomy,” 30.
- <sup>29</sup> Khoo, “Dublin Core,” 977, 984 .
- <sup>30</sup> LOC, “MARC Standards: MARC21 Formats,” Library of Congress (LOC), 2013, updated March 14, 2013, accessed January 2, 2014, <http://www.loc.gov/marc/marcdocz.html>.
- <sup>31</sup> Philip E Schreur, “Linked Data for Production and the Program for Cooperative Cataloging,” *PCC Policy Committee Meeting*, 2017, accessed May 18, 2018, <https://www.loc.gov/aba/pcc/documents/Facil-Session-2017/PCC and LD4P.pdf>.
- <sup>32</sup> Sarah Bull and Amanda Quimby, “A Renaissance in Library Metadata? The Importance of Community Collaboration in a Digital World,” *Insights* 29, no. 2 (2016): 146–53, <http://doi.org/10.1629/uksg.302>.
- <sup>33</sup> Philip E. Schreur, “Linked Data for Production,” *PCC Policy Committee Meeting*, 2015, accessed November 09, 2018, <https://www.loc.gov/aba/pcc/documents/PCC-LD4P.docx>.
- <sup>34</sup> Vandenbussche, “Linked Open Vocabularies,” 437, 438, 450.
- <sup>35</sup> Hallo, “Current State,” 120.
- <sup>36</sup> Hallo, “Current State,” 118.
- <sup>37</sup> Hallo, “Current State,” 120, 124.

- 
- <sup>38</sup> Hallo, “Current State,” 120, 124.
- <sup>39</sup> Hallo, “Current State,” 124.
- <sup>40</sup> Bull, “Community Collaboration,” 147.
- <sup>41</sup> Sam Gyun Oh, Myongho Yi, and Wonghong Jang, “Deploying Linked Open Vocabulary (LOV) to Enhance Library Linked Data,” *Journal of Information Science Theory and Practice* 2, no. 2 (2015): 6–15, <http://dx.doi.org/10.1633/IISTaP.2015.3.2.1>.
- <sup>42</sup> Carlo Bianchini and Mauro Guerrini, “A Turning Point for Catalogs: Ranganathan’s Possible Point of View,” *Cataloging & Classification Quarterly* 53, no. 3-4 (2015): 341–51, <http://doi.org/10.1080/01639374.2014.968273>.
- <sup>43</sup> Bianchini, “Turning Point,” 350.
- <sup>44</sup> LOC, “Library of Congress Linked Data Service,” The Library of Congress, accessed March 24, 2018, <http://id.loc.gov/about/>.
- <sup>45</sup> LOC, “Linked Data Service.”
- <sup>46</sup> LOC, “Linked Data Service.”
- <sup>47</sup> LOC, “Linked Data Service.”
- <sup>48</sup> LOC, “Linked Data Service.”
- <sup>49</sup> Margaret E Dull, “Moving Metadata Forward with BIBFRAME: An Interview with Rebecca Guenther,” *Serials Review* 42, no. 1 (2016): 65–69, <https://doi.org/10.1080/00987913.2016.1141032>.
- <sup>50</sup> LOC, “Overview of the BIBFRAME 2.0 Model,” Library of Congress, April 21, 2016, accessed November 09, 2018, <https://www.loc.gov/bibframe/docs/bibframe2-model.html>.
- <sup>51</sup> Taniguchi, “BIBFRAME 2.0,” 388; Taniguchi, “Suitable Schema,” 40.
- <sup>52</sup> OCLC. 2016, “OCLC Linked Data Research,” Online Computer Library Center (OCLC), <https://www.oclc.org/research/themes/data-science/linkedata.html>.
- <sup>53</sup> OCLC, “Linked Data Research.”
- <sup>54</sup> Jeff Mister, “Turning Bibliographic Metadata into Actionable Knowledge,” *Next Blog—OCLC*, February 29, 2016, <http://www.oclc.org/blog/main/turning-bibliographic-metadata-into-actionable-knowledge/>.
- <sup>55</sup> Mister, “Turning Bibliographic Metadata.”
- <sup>56</sup> George Campbell, Karen Coombs, and Hank Sway, “OCLC Linked Data,” *OCLC Developer Network*, March 26, 2018, <https://www.oclc.org/developer/develop/linkedin-data.en.html>.
- <sup>57</sup> Campbell, “OCLC Linked Data.”



- 
- <sup>58</sup> Roy Tennant, “Getting Started with Linked Data,” *NEXT Blog—OCLC*, February 8, 2016, <http://www.oclc.org/blog/main/getting-started-with-linked-data-3/>.
- <sup>59</sup> Tennant, “Linked Data.”
- <sup>60</sup> DBLP, “DBLP Computer Science Bibliography: Frequently Asked Questions,” Digital Bibliography & Library Project (DBLP), updated November 07, 2018, accessed 08 November 2018. <http://dblp.uni-trier.de/faq/>.
- <sup>61</sup> DBLP, “Frequently Asked Questions.”
- <sup>62</sup> Jörg Diederich, Wolf-Tilo Balke, and Uwe Thaden, “Demonstrating the Semantic Growbag: Automatically Creating Topic Facets for Faceteddblp,” Paper presented at the Proceedings of the 7th ACM/IEEE-CS joint conference on Digital libraries, Vancouver, Canada, June 17–22, 2007.
- <sup>63</sup> Jörg Diederich, Wolf-Tilo Balke, and Uwe Thaden, “About FacetedDBLP,” 2018, accessed November 09, 2018, <http://dblp.l3s.de/dblp++.php>.
- <sup>64</sup> Tennant, “Linked Data.”
- <sup>65</sup> In this Section, LOV catalog or portal refers to the LOV platform available at <http://lov.okfn.org/dataset/lov/>, whereas the abbreviation LOV, when used alone (without the term catalog/portal), refers to Linked Open Vocabularies in general; Vandenbussche, “Linked Open Vocabularies,” 437.
- <sup>66</sup> Vandenbussche, “Linked Open Vocabularies,” 443, 450.
- <sup>67</sup> Vandenbussche, “Linked Open Vocabularies,” 437.
- <sup>68</sup> Vandenbussche, “Linked Open Vocabularies,” 437, 438, 450.
- <sup>69</sup> Vandenbussche, “Linked Open Vocabularies,” 438.
- <sup>70</sup> Vandenbussche, “Linked Open Vocabularies,” 437, 438, 443–46.
- <sup>71</sup> Baker Thomas, Pierre-Yves Vandenbussche, and Bernard Vatant, “Requirements for Vocabulary Preservation and Governance,” *Library Hi Tech* 31, no. 4 (2013): 657–68, <https://doi.org/10.1108/LHT-03-2013-0027>.
- <sup>72</sup> Thomas, “Vocabulary Preservation,” 658.
- <sup>73</sup> Oh, “Deploying,” 9.
- <sup>74</sup> Oh, “Deploying,” 9.
- <sup>75</sup> Oh, “Deploying,” 9, 10.

- 
- <sup>76</sup> Erik Radio and Scott Hanrath, "Measuring the Impact and Effectiveness of Transitioning to a Linked Data Vocabulary," *Journal of Library Metadata* 16, no. 2 (2016): 80–94, <https://doi.org/10.1080/19386389.2016.1215734>.
- <sup>77</sup> Radio, "Transitioning," 81.
- <sup>78</sup> Robert, "Strings to Things," 2.
- <sup>79</sup> Robert, "Strings to Things," 2, 4, 6.
- <sup>80</sup> Vandenbussche, "Linked Open Vocabularies," 438.
- <sup>81</sup> As of April 23, 2018, the Schema.org vocabulary is now available at <http://lov.okfn.org/dataset/lov/>; Alberto Nogales et al., "Linking from Schema.org Microdata to the Web of Linked Data: An Empirical Assessment," *Computer Standards & Interfaces* 45 (2016): 90-99. <https://doi.org/10.1016/j.csi.2015.12.003>.
- <sup>82</sup> Bull, "Community Collaboration," 146.
- <sup>83</sup> Bull, "Community Collaboration," 146.
- <sup>84</sup> Bull, "Community Collaboration," 147.
- <sup>85</sup> Bull, "Community Collaboration," 147.
- <sup>86</sup> Bull, "Community Collaboration," 147, 148.
- <sup>87</sup> Schreur, 2015. Linked Data for Production.
- <sup>88</sup> Yhna Therese P. Santos, "Resource Description and Access in the Eyes of the Filipino Librarian: Perceived Advantages and Disadvantages," *Journal of Library Metadata* 18, no. 1 (2017): 45–56, <https://doi.org/10.1080/19386389.2017.1401869>.
- <sup>89</sup> Santos, "Filipino Librarian," 51–55.
- <sup>90</sup> Philomena W. Mwaniki, "Envisioning the Future Role of Librarians: Skills, Services and Information Resources," *Library Management* 39, no. 1, 2 (2018): 2–11, <https://doi.org/10.1108/LM-01-2017-0001>.
- <sup>91</sup> Mwaniki, "Envisioning the Future," 7, 8.
- <sup>92</sup> Taniguchi, "BIBFRAME 2.0," 410, 411 .
- <sup>93</sup> Taniguchi, "Suitable Schema," 52–58 .
- <sup>94</sup> Taniguchi, "Suitable Schema," 59, 60.
- <sup>95</sup> Taniguchi, "Suitable Schema," 60.
- <sup>96</sup> Sprochi, "Where Are We Headed?," 129, 134.



- 
- <sup>97</sup> Sprochi, "Where Are We Headed?," 129.
- <sup>98</sup> Sprochi, "Where Are We Headed?," 134.
- <sup>99</sup> Sprochi, "Where Are We Headed?," 134.
- <sup>100</sup> Sprochi, "Where Are We Headed?," 134, 135.
- <sup>101</sup> Sprochi, "Where Are We Headed?," 134.
- <sup>102</sup> Caitlin Tillman, Joseph Hafner, and Sharon Farnel, "Forming the Canadian Linked Data Initiative," Paper presented at the the 37th International Association of Scientific and Technological University Libraries 2016 (IATUL 2016) Conference, Dalhousie University Libraries in Halifax, Nova Scotia, June 5–9, 2016.
- <sup>103</sup> Carol Jean Godby, Shenghui Wang, and Jeffrey K Mixer, *Library Linked Data in the Cloud: OCLC's Experiments with New Models of Resource Description. Vol. 5, Synthesis Lectures on the Semantic Web: Theory and Technology*, San Rafael, California (USA), Morgan & Claypool Publishers, 2015, <https://doi.org/10.2200/S00620ED1V01Y201412WBE012>.
- <sup>104</sup> Sofia Zapounidou, Michalis Sfakakis, and Christos Papatheodorou, "Highlights of Library Data Models in the Era of Linked Open Data," Paper presented at the The 7th Metadata and Semantics Research Conference, MTSR 2013, Thessaloniki, Greece, November 19–22, 2013; Timothy W. Cole et al., "Library MARC Records Into Linked Open Data: Challenges and Opportunities," *Journal of Library Metadata* 13, no. 2–3 (2013): 163–96, <https://doi.org/10.1080/19386389.2013.826074>; Kim Tallerås, "From Many Records to One Graph: Heterogeneity Conflicts in the Linked Data Restructuring Cycle," *Information Research* 18, no. 3 (2013) paper C18, accessed November 10, 2018.
- <sup>105</sup> Fabiano Ferreira de Castro, "Functional Requirements for Bibliographic Description in Digital Environments," *Transinformação* 28, no. 2 (2016): 223–31. <https://doi.org/10.1590/2318-08892016000200008>.
- <sup>106</sup> Castro, "Functional Requirements," 223, 224.
- <sup>107</sup> Castro, "Functional Requirements," 224, 230.
- <sup>108</sup> Castro, "Functional Requirements," 223, 228–30.
- <sup>109</sup> Gardašević, "Possibilities and Prospects," 35.
- <sup>110</sup> Godby, *OCLC's Experiments*, 112.
- <sup>111</sup> Gonzales, "The Future," 17.
- <sup>112</sup> Karim Tharani, "Linked Data in Libraries: A Case Study of Harvesting and Sharing Bibliographic Metadata with Bibframe," *Information Technology and Libraries* 34, no. 1 (2015): 5–15. <https://doi.org/https://doi.org/10.6017/ital.v34i1.5664>.
- <sup>113</sup> Tharani, "Harvesting and Sharing," 16.

- 
- 114 Gonzales, "The Future," 16.
- 115 Karen Smith-Yoshimura, "Analysis of International Linked Data Survey for Implementers," *D-Lib Magazine*, 2016, July/August 2016.
- 116 Smith-Yoshimura, "Analysis."
- 117 Smith-Yoshimura, "Analysis."
- 118 Aikaterini K. Kalou, Dimitrios A. Koutsomitropoulos, and Georgia D. Solomou, "Combining the Best of Both Worlds: A Semantic Web Book Mashup as a Linked Data Service Over CMS Infrastructure," *Journal of Library Metadata* 16, no. 3–4 (2016): 228–49, <https://doi.org/10.1080/19386389.2016.1258897>.
- 119 Cole, "MARC," 163, 165, 175.
- 120 Cole, "MARC," 163, 164, 191.
- 121 Cole, "MARC," 164, 191.
- 122 IFLA, "Linked Open Data: Challenges arising," The International Federation of Library Associations and Institutions (IFLA), 2014, accessed March 03, 2018, <https://www.ifla.org/book/export/html/8548>.
- 123 Hallo, "Current State," 124.
- 124 Hallo, "Current State," 126.
- 125 Hallo, "Current State," 124.
- 126 Karen Smith-Yoshimura, "Linked Data Survey results 4–Why and What Institutions are Publishing (Updated)," *Hanging Together the OCLC Research blog*, September 3, 2014, accessed November 12, 2018, <https://hangingtogether.org/?p=4167>.
- 127 Bull, "Community Collaboration," 148.
- 128 Tallerås, "One Graph."
- 129 Karen Smith-Yoshimura, "Linked Data Survey Results 3–Why and What Institutions are Consuming (Updated)," *Hanging Together the OCLC Research blog*, September 1, 2014, accessed November 12, 2018, <http://hangingtogether.org/?p=4155>.
- 130 Godby, *OCLC's Experiments*, 116.
- 131 Carol Jean Godby and Karen Smith-Yoshimura, "From Records to Things: Managing the Transition from Legacy Library Metadata to Linked Data," *Bulletin of the Association for Information Science and Technology* 43, no. 2 (2017): 18–23, <https://doi.org/10.1002/bul2.2017.1720430209>.
- 132 Godby, "From Records to Things," 23.



- 
- 133 Godby, "From Records to Things," 22.
- 134 Vandenbussche, "Linked Open Vocabularies," 449, 450.
- 135 Silvia B. Southwick, Cory K Lampert, and Richard Southwick, "Preparing Controlled Vocabularies for Linked Data: Benefits and Challenges," *Journal of Library Metadata* 15, no. 3–4 (2015): 177–190, <https://doi.org/10.1080/19386389.2015.1099983>.
- 136 Southwick, "Controlled Vocabularies," 177.
- 137 Southwick, "Controlled Vocabularies," 189, 190.
- 138 Southwick, "Controlled Vocabularies," 183.
- 139 Robin Hastings, "Feature: Linked Data in Libraries: Status and Future Direction," *Computers in Libraries* (Magzine Article), 2015, <http://www.infotoday.com/cilmag/nov15/Hastings--Linked-Data-in-Libraries.shtml>.
- 140 Hastings, "Status and Future."
- 141 Hastings, "Status and Future."
- 142 Hastings, "Status and Future."
- 143 Hastings, "Status and Future."
- 144 Tallerås, "National Libraries," 129 (by quoting from van Hooland 2009; Wang and Strong 1996).
- 145 Jung-Ran Park, "Metadata Quality in Digital Repositories: A Survey of the Current State of the Art," *Cataloging & Classification Quarterly* 47, no. 3–4 (2009): 213–28, <https://doi.org/10.1080/01639370902737240>.
- 146 Tallerås, "National Libraries," 129 (by quoting from Bruce & Hillmann, 2004).
- 147 Park, "Metadata Quality," 213, 224; Tallerås, "National Libraries," 129, 150.
- 148 Park, "Metadata Quality," 213, 215, 218–21, 224, 225; Tallerås, "National Libraries," 141.
- 149 Tallerås, "National Libraries," 129.
- 150 Tallerås, "National Libraries," 129.
- 151 Tallerås, "National Libraries," 129.
- 152 Karen Snow, "Defining, Assessing, and Rethinking Quality Cataloging," *Cataloging & Classification Quarterly* 55, no. 7–8 (2017): 438–55, <https://doi.org/10.1080/01639374.2017.1350774>.
- 153 Snow, "Quality Cataloging," 445.
- 154 Snow, "Quality Cataloging," 451, 452.

- 
- 155 David Van Kleeck et al., "Managing Bibliographic Data Quality for Electronic Resources," *Cataloging & Classification Quarterly* 55, no. 7-8 (2017): 560-77, <https://doi.org/10.1080/01639374.2017.1350777>.
- 156 Van Kleeck, "Data Quality," 560, 575, 576.
- 157 Van Kleeck, "Data Quality," 575.
- 158 Park, "Metadata Quality," 214, 216-18, 225.
- 159 NISO, *A Framework of Guidance for Building Good Digital Collections*, ed. NISO Framework Advisory Group, 3rd ed (Baltimore, MD: National Information Standards Organization, 2007), <https://www.niso.org/sites/default/files/2017-08/framework3.pdf>.
- 160 Park, "Metadata Quality," 214, 215; NISO. *Guidance*; Jane Barton, Sarah Currier, and Jessie MN Hey, "Building Quality Assurance into Metadata Creation: An Analysis Based on the Learning Objects and E-Prints Communities of Practice," Paper presented at the Proceedings of the International Conference on Dublin Core and Metadata Applications: Supporting Communities of Discourse and Practice—Metadata Research & Applications, Seattle, Washington, September 28-October 2, 2003.
- 161 Pascal Hitzler and Krzysztof Janowicz, "Linked Data, Big Data, and the 4th Paradigm," *Semantic Web* 4, no. 3 (2013): 233-35, <https://doi.org/10.3233/SW-130117>.
- 162 Hitzler, "4th Paradigm," 234.
- 163 Hitzler, "4th Paradigm," 234.
- 164 Alberto Petrucciani, "Quality of Library Catalogs and Value of (Good) Catalogs," *Cataloging & Classification Quarterly* 53, no. 3-4 (2015): 303-13. <https://doi.org/10.1080/01639374.2014.1003669>.
- 165 Petrucciani, "Quality," 303, 305.
- 166 Petrucciani, "Quality," 303, 309, 311.
- 167 Petrucciani, "Quality," 303, 309.
- 168 Petrucciani, "Quality," 309, 310.
- 169 Petrucciani, "Quality," 310.
- 170 Bull, "Community Collaboration," 147.
- 171 Bull, "Community Collaboration," 148.
- 172 Han, Myung-Ja, "New Discovery Services and Library Bibliographic Control," *Library Trends* 61, no. 1 (2012):162-72, <https://doi.org/10.1353/lib.2012.0025>.
- 173 Han, "Bibliographic Control," 162.



- 
- 174 Han, "Bibliographic Control," 169–71.
- 175 Han, "Bibliographic Control," 163.
- 176 Han, "Bibliographic Control," 167–70.
- 177 Alemu, *Emergent Theory*, 29–33, 43–65.
- 178 Alemu, *Emergent Theory*, 29–65.
- 179 Lorri Mon, *Social Media and Library Services*, Synthesis Lectures on Information Concepts, Retrieval, and Services, ed. Gary Marchionini, 40, San Rafael, California (USA), Morgan & Claypool Publishers, 2015), <https://doi.org/10.2200/S00634ED1V01Y201503ICR040>.
- 180 Mon, *Social Media*, 50.
- 181 Mon, *Social Media*, 24.
- 182 Marijn Koolen et al., "Overview of the CLEF 2016 Social Book Search Lab," Paper presented at the 7th International Conference of the Cross-Language Evaluation Forum for European Languages, Évora, Portugal, September 5–8, 2016; Koolen et al., "Overview of the CLEF 2015 Social Book Search Lab," Paper presented at the 6th International Conference of the Cross-Language Evaluation Forum for European Languages, Toulouse, France, September 8–11, 2015; Patrice Bellot et al., "Overview of INEX 2014," Paper presented at the International Conference of the Cross-Language Evaluation Forum for European Languages, Sheffield, UK, September 15–18, 2014; Bellot et al., "Overview of INEX 2013," Paper presented at the International Conference of the Cross-Language Evaluation Forum for European Languages, Valencia, Spain, September 23–26, 2013.
- 183 Bo-Wen Zhang, Xu-Cheng Yin, and Fang Zhou, "A Generic Pseudo Relevance Feedback Framework with Heterogeneous Social Information," *Information Sciences* 367–68 (2016): 909–26, <https://doi.org/10.1016/j.ins.2016.07.004>; Xu-Cheng Yin et al., "ISART: A Generic Framework for Searching Books with Social Information," *PLOS ONE* 11, no. 2 (2016): e0148479, <https://doi.org/10.1371/journal.pone.0148479>; Faten Hamad and Bashar Al-Shboul, "Exploiting Social Media and Tagging for Social Book Search: Simple Query Methods for Retrieval Optimization," in *Social Media Shaping E-Publishing and Academia*, edited by Nashrawan Taha et al., 107–17 (Cham: Springer International Publishing, 2017).
- 184 Marijn Koolen, "User Reviews in the Search Index? That'll Never Work!" Paper presented at the 36th European Conference on IR Research (ECIR 2014), Amsterdam, The Netherlands, April 13–16, 2014.
- 185 Alemu, *Emergent Theory*, 29–33, 43–65.
- 186 Lucy Clements and Chern Li Liew, "Talking about Tags: An Exploratory Study of Librarians' Perception and Use of Social Tagging in a Public Library," *The Electronic Library* 34, no. 2 (2016): 289–301, <https://doi.org/10.1108/EL-12-2014-0216>.
- 187 Clements, "Talking about Tags," 291, 297–99.

- 
- <sup>188</sup> Sharon Farnel, "Understanding Community Appropriate Metadata through Bernstein's Theory of Language Codes," *Journal of Library Metadata* 17, no. 1 (2017): 5–18, <https://doi.org/10.1080/19386389.2017.1285141>.
- <sup>189</sup> Farnel, "Bernstein's Theory," 5, 6.
- <sup>190</sup> Mwaniki, "Envisioning the Future," 8.
- <sup>191</sup> Mwaniki, "Envisioning the Future," 8, 9.
- <sup>192</sup> Getaneh Alemu et al., "Toward an Emerging Principle of Linking Socially-Constructed Metadata," *Journal of Library Metadata* 14, no. 2 (2014): 103–29, <https://doi.org/10.1080/19386389.2014.914775>.
- <sup>193</sup> Farnel, "Bernstein's Theory," 15–16.
- <sup>194</sup> Kalou, "Book Mashup."
- <sup>195</sup> Kalou, "Book Mashup," 242, 243.
- <sup>196</sup> Alemu, "Socially-Constructed Metadata," 103, 107.
- <sup>197</sup> Alemu, "Socially-Constructed Metadata," 103.
- <sup>198</sup> Alemu, "Socially-Constructed Metadata," 103, 104, 120, 121.
- <sup>199</sup> Getaneh Alemu, "A Theory of Metadata Enriching and Filtering: Challenges and Opportunities to Implementation," *Qualitative and Quantitative Methods in Libraries* 5, no. 2 (2017): 311–34, <http://www.qqml-journal.net/index.php/qqml/article/view/343>
- <sup>200</sup> Alemu, "Metadata Enriching and Filtering," 311.
- <sup>201</sup> Alemu, "Socially-Constructed Metadata," 125.
- <sup>202</sup> Alemu, "Metadata Enriching and Filtering," 319, 320.
- <sup>203</sup> Alemu, "Metadata Enriching and Filtering"; Alemu, *Emergent Theory*; Alemu, "Socially-Constructed Metadata"; Farnel, "Bernstein's Theory"; Kalou, "Book Mashup."
- <sup>204</sup> Hallo, "Current State," 120.
- <sup>205</sup> Alemu, "Socially-Constructed Metadata," 125; Hastings, "Status and Future."
- <sup>206</sup> Bull, "Community Collaboration," 147.
- <sup>207</sup> Bull, "Community Collaboration," 152; Bull, "Community Collaboration," 152; Schreur, 2015. Linked Data for Production.
- <sup>208</sup> Tallerås, "National Libraries," 129.
- <sup>209</sup> Petrucciani, "Quality," 303, 309.



---

<sup>210</sup> Bull, “Community Collaboration,” 147, 152.

<sup>211</sup> Farnel, “Bernstein's Theory,” 5, 6, 12, 13, 15, 16; Mwaniki, “Envisioning the Future,” 8.

<sup>212</sup> Mon, *Social Media*, 3; Alemu, “Metadata Enriching and Filtering,” 320.

<sup>213</sup> Alemu, “Socially-Constructed Metadata,” 125.

<sup>214</sup> Koolen, “CLEF 2016”; Koolen, “CLEF 2015”; Bellot, “INEX 2014”; Bellot, “INEX 2013.”