Understanding the Position of Information Professionals with regards to Linked Data: A Survey of Libraries, Archives and Museums

Lucy McKenna
ADAPT Centre,
Trinity College Dublin,
Ireland
lucy.mckenna@adaptcentre.ie

Christophe Debruyne
University College Odisse,
Brussels,
Belgium
christophe.debruyne@odisee.be

Declan O'Sullivan
ADAPT Centre,
Trinity College Dublin,
Ireland
declan.osullivan@cs.tcd.ie

ABSTRACT

The aim of this study was to explore the benefits and challenges to using Linked Data (LD) in the Libraries, Archives and Museums (LAMs) as perceived by Information Professionals (IPs). The study also aimed to gain an insight into potential solutions for overcoming these challenges, with a particular focus on the idea of LD tooling for IPs as a means of doing so. Data was collected via a questionnaire which was completed by 185 Information Professionals (IPs) from a range of LAM institutions. Results indicated that there are many challenges relating to the usability and utility of LD tooling that create barriers to IPs engaging with LD. The study shows that LD tools designed with the workflows and expertise of IPs in mind could help break down these barriers.

CCS CONCEPTS

•**General and reference** → **Surveys and overviews**; •**Information systems** → **Digital libraries and archives; Resource Description Framework (RDF)**; •**Human-centered computing** → *User studies;* User interface design; Usability testing;

KEYWORDS

Linked Data, Semantic Web, library, archive, museum, cultural heritage institution, user experience, interface, survey, questionnaire

1 INTRODUCTION

Through the use of Linked Data (LD), Libraries, Archives and Museums (LAMs) have the potential to expose their collections to a larger audience, increase the use of their materials, and allow for more efficient user searches [14].

It could be argued that LD for LAMs could be generated by technical experts or via crowd-sourcing, rather than by Information Professionals (IPs). However, as keepers of large amounts of cultural heritage data, and as experts in the field of metadata creation and knowledge discovery, IPs are well positioned to play a leading role in the development of the SW. As such, facilitating and supporting their engagement with LD is of particular value.

Thus, the aim of our research is twofold. Firstly, it investigates what IPs, both with and without experience in using LD, consider to be the barriers towards LAMs publishing and consuming LD. Secondly, it explores the idea of LD tooling for IPs as a means of overcoming these barriers and facilitating the LAM community to engage with LD more frequently. The user requirements for such tooling were also investigated, which included gathering data on useful tool functions and interface usability.

In light of the above, the objectives of our research were to investigate IPs:

(1) Knowledge of the Semantic Web and Linked Data
(2) Views on the benefits and challenges of using Linked Data.
(3) Experiences of using Linked Data.
(4) Usability of cataloguing and Linked Data interfaces.
(5) Views on Linked Data tools for Information Professionals.

Data was collected via a questionnaire which was distributed to IPs working in the LAM domain.

Results indicated that, while IPs consider LD to have many benefits LAMs, such as improved resource discoverability and interoperability, IPs also felt that using LD was fraught with challenges. Challenges reported included difficulties using LD tooling and difficulties interlinking LD resources. Participants indicated that LD tools designed for IPs could make LD more accessible to the LAM domain.

Our study contributes to the LAMs domain by exploring LD tooling for IPs as a means of increasing this group's engagement with LD, and by providing an up-to-date and broader perspective on LD for LAMs by including IPs both with and without LD experience from a range of institutions.

The details of our research are discussed below in the following structure; Background, Related Work, Methodology, Findings, Discussion, Conclusions, and Future Directions.

2 BACKGROUND

The Semantic Web (SW) is an extension of the current Web where data is given well defined meaning and where the relationships between data, and not just documents, are defined in a common

machine-readable format, thus creating a Web of Linked Data [5]. Linked Data (LD) itself describes a set of best practices for publishing and interlinking structured data on the Web, as per the principles defined by the World Wide Web Consortium (W3C) [4, 6]. These principles include the use of HTTP URIs (Uniform Resource Identifiers) not only as names for entities, but also as a means for retrieving the data using the existing HTTP stack.

The Resource Description Framework (RDF) is a model, developed by the W3C, for representing and exchanging LD on the web as structured data [38]. An RDF statement takes the form of a triple, which consists of a subject, predicate (relationship) and object. RDF requires that URIs are used to identify subjects and predicates, allowing for the resulting data to be understood by computers. SPARQL is an RDF query language which allows for the retrieval and manipulation of data stored in RDF format vie a SPARQL endpoint [9].

The SW and LD have the potential to transform the Web into a globally interlinked and searchable database, rather than a disparate collection of documents [39]. This would allow for easier searching, querying, and processing of data by humans and machines alike.

2.1 Linked Data for LAMs

By freeing metadata from LAMs databases and sharing it on the SW, these institutions could increase the discoverability and re-usability of their resources [14]. Publishing records as LD also offers LAMS new opportunities to create meaningful links between objects across collections. The benefits of LAMs becoming part of the SW are discussed below.

2.1.1 Interoperability and Re-Usability. Currently a number of different metadata standards are being used across LAMs making metadata interoperability extremely challenging [1]. Publishing metadata as RDF would allow for the seamless sharing and re-use of metadata across LAMs thus increasing collaboration and reducing record duplication [1, 34]. As well as lowering costs, this would provide Information Professionals (IPs) with more time to focus on creating richer descriptions for local resources [14, 17].

2.1.2 Discoverability and Visibility. Unlike RDF, many of the metadata standards employed by LAMs cannot be processed by search engines. As such, a significant amount of relevant content is not visible in search engine results [15, 31]. As the Web is typically the first place users search for information, publishing metadata to the SW will provide LAMs with a way of enhancing the discoverability and visibility of their records [14, 26, 33]. Not only will this make it easier for current LAM patrons to find useful information; it will also provide LAMs with the opportunity to reach those who would not typically use their resources [13].

Additionally, as LD requires the use of URIs for identifying entities, such as works, people, places, and events, within and across resources; searching for specific terms will become more efficient and effective [14].

2.1.3 Interlinking and Integration. A vital component of LD is to provide links to other URIs in your data so that users can discover related information [4]. LAM metadata that has been enriched with related links would allow users to navigate seamlessly between disparate internal and external datasets [1, 10, 34]. By publishing

resources as LD, LAMs have the potential to open up and share their metadata on the Web in ways that were previously restricted by metadata models [14].

2.1.4 Reliability and Authority Control. Since anyone can publish and interlink data on the SW, as the Web of Data grows there will be an increased need to identify who completed these tasks in order to establish data authenticity and accuracy. LAMs are typically well-established trusted sources of information, as such LD generated by IPs is likely to be treated with increased credibility over data generated by non-authoritative sources [24, 26, 35].

Additionally, IPs are experts in using controlled vocabularies for the consistent description and linking of similar concepts and entities across records. A number of these vocabularies are already available as LD, and could be used to consistently describe entities across the SW [26].

2.2 Linked Data Challenges for LAMs

At present, the main successful examples of LD projects emerging from LAMs have been single-institution initiatives with limited interlinking, rather than Web-wide multi-institutional data integration projects [12, 19, 41]. As one of the fundamental prerequisites of the SW is the existence of large amounts of meaningfully interlinked resources [6], it is key that institutions not only publish RDF datasets but also interlink their data with others. These interlinking issues may be due to the fact that interlinking is one of the most challenging areas of LD. Reasons for this include that the tools required to complete data integration are limited [3], and that little usability testing of LD tools has been done with users who are non-technical experts [35]. The outcome of this is that many tools have not yet developed to the standard required for widespread adoption [1, 3].

Other challenges faced by LAMs when attempting to participate in the SW include:

- Relatively few projects that demonstrate how implementing LD can benefit LAMs [17, 25].
- Most current cataloguing software does not support LD requirements [8, 16, 25].
- Transforming existing records to RDF is time-consuming and challenging [33].
- Copyright and intellectual property issues [7, 33].

Despite the many benefits of publishing and using LD in LAMs discussed above, relatively few institutions have adopted the use of LD [17, 25]. We believe that this slow uptake is partially due to the challenges mentioned previously, and also due to a 'chicken and egg' issue, whereby LAMs are likely to be reluctant to invest time and resources on LD projects without clear signs of success from other institutions, and a variety of RDF datasets to interlink with.

3 RELATED WORK

OCLC Research conducted an International LD Survey for Implementers in both 2014 and 2015, receiving responses from a total of 90 institutions across 20 countries [27, 40]. Participating institutions, mostly libraries, were either publishing LD, consuming LD or both. The questionnaire gathered data on the types of LD projects being

implemented, the data being published and consumed, the rationale for implementing the project, and the barriers encountered.

The majority of projects were both consuming and publishing LD, second to that were projects that consumed LD, and finally projects that published LD. The chief motivations and benefits of implementing the LD projects included increased discoverability of resources, improved data re-usability, resources exposed to a larger audience, enhanced data quality, more accurate search results, and improved Search Engine Optimisation (SEO).

In projects where LD was consumed, the most commonly used resources included the Virtual International Authority File (VIAF), DBPedia, the Library of Congress LD Service (id.loc.gov), and local resources. The main barriers to consuming LD included issues with data alignment issues, vocabulary mapping, unreliable or absent URIs, lack of authority control, and lack of dataset updates.

For LD publishing projects, the most commonly used RDF vocabularies and ontologies included the System Knowledge Organisation System (SKOS), Friend of a Friend (FOAF), DCMI Metadata Terms (DCTerms), the Dubin Core Metadata Element Set (DCE) and Schema.org. The most commonly used technologies were SPARQL, Java, XSLT, Zorba, Solr and Virtuoso Universal Server. The main barriers to LD consumption included a steep learning curve for implementers, difficulties selecting ontologies, interlinking issues, lack of documentation outlining how to build LD systems, and a lack of tooling.

This study provided a detailed overview of how LD is currently being used in libraries, as well as the benefits and challenges being experienced. One area that was not covered in the study was how IPs who are not currently working with LD view the role of the SW in the library domain. However, this perspective was investigated in a study by LaPolla [20], where an online questionnaire was distributed to academic cataloguers and technical-services professionals within the library domain. This 22 question survey explored librarians' level of understanding and attitudes towards the SW, as well as their views on its role in the library catalogue and how to best move forward with SW technology within this domain.

The questionnaire received 156 responses from librarians, 35% of whom rated themselves as "Very Familiar" with the SW and 55% as "Somewhat Familiar". Also, 42% of participants reported that their institution was already exploring the use of SW catalogues.

Responses indicated that participants were interested in exploring the use of SW technologies within the library domain as they felt it could play a useful role in the future of library catalogues. Issues such as financial constraints, a lack of awareness of the SW, and a lack of SW best practices, were reported as the main barriers to libraries engaging with the SW.

Overall, the results of this study indicated that, while the majority of participants were knowledgeable and interested in the potential of the SW in the library domain, less than half of the represented institutions were actively exploring SW technologies.

Across both studies participants seemed to have a positive view of the use of LD in library domain, with many benefits being listed. However, multiple barriers to its use were also identified. With IPs clearly having an interest in LD, the importance of investigating potential solutions to the identified barriers is of significant importance.

In comparison with the two studies discussed above, the contributions made by our research include:

(1) Explores the potential of LD tooling for IPs as a means of overcoming the barriers that IPs experience when using LD.
(2) Explores the usability and utility of cataloguing and LD interfaces.
(3) Provides an up-to-date view of IPs knowledge of LD, how LD is currently being used in LAMs, as well as the benefits and challenges being experienced.
(4) Takes into account the views of IPs, both with and without LD experience, from a variety of LAM domains.

4 METHODOLOGY

The primary means of data collection for our research was a questionnaire. The structure of this questionnaire is outlined below, and information on participation criteria and sampling is also provided.

4.1 Preliminary Interviews

Two semi-structured interviews were conducted as a means of exploring whether the LD challenges identified in Section 2.2 and the proposed aims of our study were in line with the experiences and needs of working IPs [22]. A semi-structured interview is a qualitative method of inquiry in which the interviewer asks a set of pre-determined open-ended questions, but arising topics can also be explored further if pertinent to the research [2].

Participants were two IPs working in a large university library. Both had over a decade of experience working in bibliographic data management, and were familiar with the concepts of the SW and LD. Common themes emerging from the interviews included:

- Libraries can provide authority control to the SW.
- Current LD tools do not target the work processes, needs and expertise of librarians.
- LD tools for librarians should not require an in-depth knowledge of LD technicalities.
- More use cases of LD being used effectively in the LAMs required for funding and time to be dedicated to LD.

It is acknowledged that the results of these interviews are not generalisable across LAMs due to the small sample size, and that they may also contain some bias as both interviewees were highly experienced IPs with an interest in LD. However, the value of the interviews lay in the fact that both participants affirmed many of the LD challenges mentioned in Section 2.2, thus reinforcing the need for a larger, more in-depth exploration of the objectives of our study which was conducted through the use of a questionnaire.

4.2 Questionnaire

An online questionnaire consisting of 50 questions was developed using the Qualtrics service [32] [1]. In line with the aims and objectives of our study, the questionnaire was divided into six main Sections:

(1) Cataloguing Experience
(2) Usability of Cataloguing Tools
(3) Knowledge and Views on Linked Data

---

[1]Questionnaire to be made publicaly available @ http://www.tara.tcd.ie

(4) Linked Data Project Experience
(5) Usability of Linked Data Tools
(6) Linked Data for Information Professionals

Questions in Section 1, 2, and 5 of the questionnaire aimed at gathering information on the usability of LAM and LD tools. Sections 3 and 4 of the questionnaire were designed to gather data on IPs knowledge of LD, their experiences using LD, and the perceived benefits and challenges of LD. Finally, Section 6 of the questionnaire collected information on participants' perceived usefulness of LD tooling designed specifically for IPs.

Questions were mostly multiple choice, however, a small number of open-ended questions were also included. The questions and lists of answers in Sections 3, 4, and 6 of the questionnaire were partially based on the results of the OCLC Research [27, 40] and LaPolla [20] LD studies discussed in Section 3, as well as the interviews conducted above in Section 4.1.

For all questions where participants were asked to choose from a proposed list of options, they also had the opportunity to add their own additional observations. This was done so as not to limit the range of responses, thus allowing for a more comprehensive view of participants' thoughts and opinions.

Included in Sections 2 and 5 of the questionnaire was the Computer System Usability Questionnaire (CSUQ) [21] which is used for measuring software usability and utility. The CSUQ was chosen over other usability questionnaires as it does not have to be completed directly after an interaction with a system and it was designed to be administered online or via mail [37].

The CSUQ consists of 19 statements about which the user rates agreement on a seven-point scale from Strongly Agree (1) to Strongly Disagree (7) - thus lower scores indicate less usability issues. For the purpose of this questionnaire, a subset of nine statements, focusing of the efficiency, efficacy, learnability, and ease of use of tools, were used.

4.2.1 Participants. Participants in our questionnaire were IPs with experience working in the LAM domain. Also recruited were Researchers and Academics with research experience in the LAM and/or LD domain. IPs were encouraged to participate regardless of whether they had any prior experience working with the SW or LD. This was done in an attempt to recruit a broad range of participants, rather than just IPs who are highly experienced in LD. That being said, it is possible that many participants who completed the survey already had a prior interest in the SW and LD in order to be motivated to do so.

Non-probabilistic sampling methods were used to recruit participants [11]. This was done by directly contacting LAM institutions and associations with a request to distribute the questionnaire, sharing the questionnaire link on Twitter[2], and by asking participants to forward the questionnaire to potential participants. Individual IPs and researchers known to the researchers were also contacted directly.

The questionnaire received over 230 responses. Of these, 185 were deemed suitable for analysis. The questionnaires removed from the study were eliminated on the basis that participants did not indicate that they had experience working as an IP or a researcher in the LAM or LD domain. Additionally, some questionnaires were

removed due to repetitive answer patterns suggesting that the respondent may not have read the questions in their entirety.

Participants (Number of Participants (N) = 185) represented Europe (68%), North America (29%), Australia (2%), Asia (1%), Africa (1%) and South America (1%). Participants came from 20 different countries, with the majority coming from Ireland (28%), the USA (23%) and the UK (20%). Within Ireland and the UK, the types of institutions where the questionnaire was shared included large reputable universities, national libraries, archives and digital repositories, and LAM related associations, conferences and research groups. In the US the questionnaire was shared within a large reputable library centre, and by an international IP association.

Participants represented a variety of LAMs and research institutions including Academic Library (56%), Research Institution/University (7%), Public Library (7%), Special Library (6%), Archive (6%), National Library (5%), Museum (4%), and Special Archive (1%). As the majority of participants came from an Academic Library setting, results are not generalisable across all LAM domains.

Finally, 88% of participants had more than 4 years of experience working in the LAM or research domain, with the majority of participants (61%) having 10 or more years of experience.

5 FINDINGS

The results of the questionnaire are presented in the following paragraphs, and will be discussed and interpreted in corresponding Subsections of Section 6.

5.1 Cataloguing Experience

74% of participants (N = 132) reported that they are currently involved in the metadata cataloguing process in their workplace. When asked what metadata formats are applied, this subgroup of participants mentioned a total of 41 with the most commonly cited being MARC 21 (73%), Dublin Core (43%), Encoded Archival Description (EAD) (20%), MARC XML (17%), Metadata Object Description Schema (MODS) (16%). Other reported formats used included the Metadata Encoding and Transmission Standard (METS) (8%), the Text Encoding Initiative (TEI) (5%), BIBFRAME (5%), MODS-RDF (4%) and UNIMARC (4%), VRA-Core (2%), the Metadata Authority Description Schema (MADS) (2%), and MADS-RDF (2%). A Discussion of all these results is available in Section 6.1.

5.2 Usability of Cataloguing Tools

94% of participants (N = 173) reported having previous experience using a cataloguing tool. Of these participants, the most commonly reported tools can be seen in Table 1. Also included in Table 1 are the number of participants who indicated whether or not they initially required the support of a technical person to use the cataloguing tool. Participants also completed a CSUQ for each of the tools they reported having experience using - see Table 2. A Discussion of all these results is available in Section 6.2.

5.3 Knowledge and Views on LD

Participants were asked to rate their prior knowledge of the SW and LD as either Extremely Knowledgeable (EK), Very Knowledgeable (VK), Moderately Knowledgeable (MK), Slightly Knowledgeable (SK) or Not at all Knowledgeable (NK) - see Table 3 for the overall

---

[2]https://twitter.com

Tool	Technical Assistance Required				% of Participants (N = 173)
	Yes	No	Unsure	Total	%
Aleph	20	20	5	45	26%
Omeka	2	33	6	41	24%
DSpace	13	20	7	40	23%
Sierra	13	23	3	39	20%
Koha	5	18	2	25	18%
Fedora	12	8	7	27	16%
Voyager	8	12	4	24	14%
Filemaker	3	15	0	18	11%
Alma	9	20	9	20	11%
Millenium	3	9	4	16	10%

Table 1: Cataloguing Tools

results, as well as a comparison between the results of participants with LD project experience (LDExp) (54) and those with no LD project experience (NoLDExp) (131).

Table 4 outlines the results for the question, "Do you think that publishing LAM metadata as LD could add value to the SW?". Of those who responded "Yes" to this question (N = 150), the most commonly reported benefits of publishing LAM metadata as LD were; Expose data to a larger audience (89%), Improve data accessibility (82%), Easier metadata sharing (81%), More efficient data searches (73%), Increased metadata openness (71%), Improved authority control on the SW (73%), Create a research environment (55%), and by Improved Search Engine Optimisation (SEO) (51%).

Participants were also asked whether they thought LAMs faced barriers to publishing their metadata as LD - see Table 4. Of those who responded "Yes" (N = 124), the most commonly reported barriers were; Difficulty incorporating LD publication into cataloguing workflow (77%), Difficulty cleaning data (67%), Time consuming (66%), Steep learning curve (63%), Copyright issues (52%), Difficulty using LD tools (52%), Inadequate LD tools available (50%), Difficulty establishing links (43%), Difficulty using SPARQL endpoints (42%) and an Insufficient number of controlled vocabularies available as LD (41%).

Following this, participants were asked whether consuming LD could benefit LAM domains - see Table 4. Of those who responded "Yes" (N = 154), the most frequently reported benefits that consuming LD would offer LAMs were; Improved data discovery (85%), Interlinking across institutions (81%), Enriched bibliographic metadata (79%), Interlinking across datasets (75%), Harmonise data from multiple sources (73%), More efficient data searches (70%), Improved metadata quality (68%), Automated authority control (53%), and Reduced time spent cataloguing (52%).

Again, participants were asked whether they thought LAMs faced barriers to consuming LD - see Table 4. Of those who responded "Yes" (N = 119), the most significant barriers mentioned were; Difficulty ingesting into the catalogue (75%), Time consuming processes (71%), Issues with dataset reliability (58%), Difficulty using LD tools (55%), Authority control issues (53%), Lack of LD tools available (51%), Issues with data re-usability (49%), Difficulty

Tool	It was simple to use this tool	I could effectively complete my work using this tool	I was able to complete my work quickly using this tool	It was easy to learn how to use this tool	Whenever I made a mistake using the tool, I recovered easily and quickly	The organisation of information on the tool's screen was clear	The interface of this tool was pleasant	This tool had all the functions and capabilities I expected it to have	Overall, I was satisfied with this tool	Total Score	%
Aleph	4	2	3	4	4	4	4	3	3	31	49.21%
Omeka	3	2	3	2	3	3	3	4	3	26	41.27%
Dspace	3	2	3	3	3	3	4	4	3	28	44.44%
Sierra	3	3	3	3	3	3	4	4	3	29	46.03%
Koha	3	2	3	3	3	3	3	3	3	26	41.27%
Fedora	5	3	3	4	4	4	4	3	3	33	52.38%
Voyager	3	2	2	3	3	3	4	3	3	26	41.27%
Filemaker	3	2	3	3	3	3	3	3	3	26	41.27%
Alma	4	2	3	3	3	3	4	3	3	28	44.44%
Millenium	3	2	2	2	2	3	4	3	3	24	38.10%

Table 2: CSUQ for Cataloguing Tools

Topic	Rating	% of Participants (N = 185)		
		LDExp	NoLDExp	Overall
Semantic Web	EK	5%	1%	6%
	VK	9%	5%	14%
	MK	14%	24%	38%
	SK	2%	25%	26%
	NK	0%	16%	16%
Linked Data	EK	5%	1%	6%
	VK	11%	9%	20%
	MK	11%	28%	39%
	SK	2%	23%	25%
	NK	0%	10%	10%

Table 3: SW and LD Knowledge Ratings

Question	Answer	% of Participants (N = 185)		
		LDExp	NoLDExp	Overall
Publishing LAM metadata as LD could add value to the SW?	Yes	27%	54%	81%
	No	1%	1%	2%
	Unsure	2%	15%	17%
LAMs face barriers to publishing metadata as LD?	Yes	26%	41%	67%
	No	0%	1%	1%
	Unsure	3%	29%	32%
Consuming LD could benefit LAM domains?	Yes	28%	55%	83%
	No	0%	1%	1%
	Unsure	1%	15%	16%
LAMs face barriers to consuming LD?	Yes	24%	40%	64%
	No	0.5%	0.5%	1%
	Unsure	5%	30%	35%

Table 4: LD Publication and Consumption

establishing interlinks (45%), and Use of unstable URIs (45%). A Discussion of all these results is available in Section 6.3.

5.4 LD Project Experience

Of the 185 participants, 29% (N = 54) stated that they had previously been directly involved in the implementation of a LD project or service. Of this subgroup of participants, 51% indicated that the project involved both the consumption and publication of LD, 41% indicated the LD was published only, and 8% indicated that LD was consumed only.

With regards to all projects where LD was consumed (N = 31), results indicate that the most frequently used datasets included the Art and Architecture Thesaurus (AAT) (55%), DBpedia (52%), VIAF (52%), Wikidata (26%), GeoNames (38%), Getty Thesaurus of Geographic Names (TGN) (35%), Library of Congress - NACO Authority File (NAF) (29%), WorldCat.org (29%), and Europeana (26%).

Tool	Technical Assistance Required				% of Participants (N = 54)
	Yes	No	Unsure	Total	%
OpenRefine	2	12	1	15	28%
Protégé	3	6	4	9	20%
Fedora	6	2	2	10	18%
Apache Fuseki	1	2	3	6	15%
RDF Refine	1	3	2	6	11%
Virtuoso	1	4	0	5	11%
Pubby	1	3	0	4	11%
Blazegraph	1	3	0	4	11%

Table 5: LD Tools

In all projects where LD was published (N = 49), participants indicated that the most common types of data published were Bibliographic data (55%), Digital collections (55%), Authority Files (35%), Controlled vocabularies (32%), Ontologies (32%) and Holdings data (22%). Data was also gathered on the RDF Vocabularies and Ontologies that were used as part of these projects. The most frequently mentioned were Schema.org (50%), SKOS (45%), DCTerms (43%), DCE (37%), FOAF (33%) and Local vocabularies (31%). A Discussion of all these results is available in Section 6.4.

5.5 Usability of LD Tools

The most commonly reported LD tools that the above subgroup (N = 54) had experience using can be seen in Table 5. Also included in Table 5 are the number of participants who indicated whether or not they initially required the support of a technical person to use the tool. Participants also completed a CSUQ for each of the tools they reported having experience using - see Table 6. A Discussion of all these results is available in Section 6.5.

5.6 LD for IPs

When participants (N= 185) were asked to provide their thoughts on the concept of LD tooling specifically for IPs, a number of common themes emerged including that such tools would:

- Enable more LAMs to become part of LD community.
- Allow LAMs to fully engage in the LD ecosystem.
- Make LD creation and usage more accessible for IPs.
- Enable LD to be incorporated into cataloguing workflows.
- Make it easier for IPs to understand the benefits of LD.
- Help reduce the technological barrier.
- Be more appealing/likely to be used in LAMs.

Requirements included that such a tool should be:

- Easily integrated with Library Management Systems.
- Incorporable into cataloguing workflows.
- Available in the public domain.
- Considerate of IPs point of view.
- Tuned to IPs working environment.
- Usable without having to understand LD technicalities or requiring the help of an IT professional.
- Standards compliant with a user friendly, low-tech interface.

Tool	It was simple to use this tool	I could effectively complete my work using this tool	I was able to complete my work quickly using this tool	It was easy to learn how to use this tool	Whenever I made a mistake using the tool, I recovered easily and quickly	The organisation of information on the tool's screen was clear	The interface of this tool was pleasant	This tool had all the functions and capabilities I expected it to have	Overall, I was satisfied with this tool	Total Score	%
OpenRefine	4	3	3	4	4	4	4	3	3	31	49.21%
Protégé	4	3	3	3	2	3	3	3	3	27	42.86%
Fedora	5	3	4	4	4	4	3	3	3	33	52.38%
Apache Fuseki	5	3	3	4	4	4	3	3	3	32	50.79%
RDF Refine	3	2	3	3	3	3	3	3	3	26	41.27%
Virtuoso	4	3	4	4	4	5	5	4	3	36	57.14%
Pubby	3	2	2	3	3	3	4	3	3	26	41.27%
BlazeGraph	3	3	3	3	3	3	4	4	5	31	59.21%

Table 6: CSUQ for LD Tools

Rating	% of Participants (N = 185)		
	LDExp	NoLDExp	Overall
Extremely Useful	16%	24%	40%
Moderately Useful	5%	25%	30%
Slightly Useful	2%	5%	7%
Neither Useful nor Useless	14%	5%	19%
Slightly Useless	0%	0%	0%
Moderately Useless	.5%	.5%	1%
Extremely Useless	1.5%	1.5%	3%

Table 7: Usefulness of a LD Interlinking Tool for IPs

Concerns raised were:

- The technologies and principles of LD are the same across different domains; what LAMs require is knowledge of how to apply existing tools to their domain.
- Having bespoke tools may limit how LAMs could interact with allied communities not using these tools, potentially limiting the use of their work.
- Would tools be able to interact with closed vendor systems.
- If too bespoke it may be difficult to adapt the tool to the individual needs of specialised teams.
- Workflows and data processes differ across institutions.

As highlighted in Section 2.2, one of the main challenges with LAM LD projects is the process of interlinking LD resources. Table 7 outlines participants' responses when asked to rate the usefulness of a LD Interlinking Tool developed specifically for IPs. Participants also provided feedback on why they provided their selected rating - see Table 8. Participants who rated the tool as Slightly Useless or under did not provide feedback.

Participants were then asked to indicate which functions they thought such a tool should have. The most commonly stated functions included; Data enriching (61%), Awareness of common data sources (58%), Automatic linking to controlled vocabularies (54%), Configurable to the institution's workflow, (54%), Ability to integrate LD datasets into the catalogue (52%), Data cleaning (49%), Link validation (48%), Link discovery (46%), Review data source quality (43%), Vocabulary alignment/reconciliation (43%), Automatically link to ontologies (42%), Remove the need for understanding LD technicalities (40%), and to Create controlled vocabularies in SKOS (64%).

Also documented were the LD datasets that participants would find most useful to interlink with. The most frequently selected datasets were; Workdcat.org (72%), id.loc.gov (59%), NAF (45%), ORCID (44%), British National Bibliography (BNB) (43%), VIAF (41%), AAT (40%), Europeana (40%), The British Museum's Semantic Web Collection (34%), and GeoNames (33%).

Finally, participants were asked to indicate what quality criteria [42] they apply when using, or searching for, external data sources - see Table 9. Participants indicated that Trustworthiness (66%) was by far the most important criteria, followed by Interoperability (51%) and Licensing Issues (49%). A Discussion of all these results is available in Section 6.6.

Rating	Feedback
Extremely Useful	Reconciliation is one of LOD's greatest problems
Moderately Useful	There needs to be a means for IPs to establish connections as they create/interact with data, rather than leaving it for someone (or no one) else to do
	More institutions would get involved if there were less barriers to entry
	Overcome the technical knowledge gap of content experts that create metadata
	Create potential for new interdisciplinary research opportunities
	Reduce the need for IPs to learn to use different interfaces, search strategies, and vocabularies
	A bespoke/dedicated approach would mean that more IPs would "buy into" using this system
	Time saving
Slightly Useful	A tool with awareness of the sources that IPs trust/prefer will be more efficient
	Potentially useful within particular LAMs but may have limited use within sectors that do similar work but are not 'cultural heritage' institutions.
	More valuable if there was a way to automate interlinking across collection silos
Neither Useful nor Useless	Need useful case study to be convinced
	Unsure if IPs will take kindly to doing more or having roles changed
	Until there is more evidence that LD is the future of information systems, there may not be a big buy in from the financially strapped heritage sector

Table 8: Rationale for Usefulness Rating

6 DISCUSSION

A discussion and interpretation of the results from Section 5 has been provided in a corresponding Subsection below.

6.1 Cataloguing Experience

With the majority of participants coming from the library domain, unsurprisingly, the most frequently reported metadata format used was MARC 21. As MARC does not inherently allow for linking, the standard is incompatible with LD. Converting MARC 21 to RDF, though possible, is extremely challenging [8]. Although MARC is still the most commonly used library metadata model, there seems to be a consensus that it is no longer the most effective means of encoding library metadata, however the future of the bibliographic record still remains unclear [18, 36].

One future possibility is BIBFRAME [30], a LOC initiative aimed at evolving bibliographic description standards to a LD model. With

Criteria	% of Participants (N = 185)		
	LDExp	NoLDExp	Overall
Trustworthiness	44%	22%	66%
Interoperability	21%	30%	51%
Licensing Issues	17%	32%	49%
Completeness	9%	32%	41%
Understandability	13%	27%	40%
Provenance	13%	27%	39%
Timeliness	13%	25%	38%
Syntactic Validity	11%	25%	36%
Availability	13%	19%	32%
Conciseness	8%	15%	23%
Versatility	4%	7%	11%
Other	4%	2%	6%
None/Unsure	1%	8%	9%

Table 9: Criteria Used for Dataset Quality Evaluation

BIBFRAME currently in development, and with libraries having reservations about moving away from MARC and towards LD, one possible approach would be to use a schema that is compatible with both. One such format is MODS - an XML schema for a bibliographic element set that can be used for the purpose of cataloguing digital resources [29]. Being derived from MARC 21 and having a MODS-RDF ontology already [28], MODS could be a viable format for institutions wishing to move towards LD [23]. Interestingly, 16% of participants currently involved in the cataloguing process in their workplace (N = 132) reported using MODS, with 5% reporting using MODS-RDF.

6.2 Usability of Cataloguing Tools

Across all of the tools that participants mentioned they had experience using, in the majority of instances, participants did not require the support of a technical person in order to be able to use the tool. Looking at the individual tools presented in Table 1, only Fedora[3] had more participants requiring assistance, and only Aleph[4] had equal across both groups. Interestingly, both Fedora and Aleph had the highest scores on the CSUQ, indicating the most usability problems. That being said, the overall scores for all tools indicated moderate usability issues. Higher scores were noted on questions regarding the usability of the tool's interface, suggesting increased issues in this area. Also noted was that, while participants indicated that they could complete their work effectively using the tool, tools received higher scores on questions concerning their ease of use, learnability, and number of available functions.

6.3 Knowledge and Views on LD

Results show that the vast majority participants had some prior knowledge of the SW (84%) and LD (90%). With the majority rating themselves as at least Moderately Knowledgeable (MK) for both the SW (58%) and LD (65%) - see Table 3.

Looking at the individual subgroups, i.e. participants with (LD-Exp) and without (NoLDExp) LD project experience:

---

[3]http://fedorarepository.org
[4]http://www.exlibrisgroup.com/category/Aleph

- 95% of the LDExp group (N = 54) rated their knowledge of the SW as MK or above, as compared with 44% of the NoLDExp group (N = 131). 86% of the NoLDExp group rated themselves as Slightly Knowledgeable (SK) or above.
- 95% of the LDExp group (N = 54) rated their knowledge of LD as MK or above, as compared with 54% of the NoLD-Exp group (N = 131). 87% of the NoLDExp group rated themselves as SK or above.
- Of all participants (N = 185), only 16% considered themselves to have no prior knowledge of the SW and only 10% as having no prior knowledge of LD. All of these participants were part of the NoLDExp group.

These high knowledge ratings could allow for the results of this study to be treated with increased confidence. It is important to bear in mind that participants were asked to rate their own level of knowledge of the SW and LD, without having to demonstrate this knowledge, thus running the risk of participants being more or less knowledgeable than they rated themselves to be.

6.3.1 Benefits of LD. Results regarding the benefits of LAMs publishing and consuming LD aligned with those suggested in past research (See Sections 2.1 & 3).

Regarding LAMS publishing LD, the overarching benefits were:

(1) Improved Data Searching
(2) Improved Data Accessibility
(3) Improved Data Sharing
(4) Improved Authority Control

Regarding LAMs consuming LD, the overarching benefits were:

(1) Improved Data Searching
(2) Enriched Metadata
(3) Improved Data Interlinking
(4) Improved Cataloguing Efficiency

6.3.2 Challenges of LD. With regards to the challenges of publishing and consuming LD, the results affirm those mentioned in Section 2.2. Additional issues were also identified.

The overarching challenges to LAMs publishing LD were:

(1) Difficulty Integrating and Interlinking
(2) Issues with LD Tooling
(3) Insufficient LD Resources Available

The overarching challenges to LAMs consuming LD were:

(1) Difficulty Integrating and Interlinking
(2) Issues with LD Tooling
(3) Issues with LD Resource Quality.

The challenges experienced across both LD publication and consumption appear to be quite similar. A more in depth exploration of the Integration and Interlinking issue indicated that participants feel the main barriers lie in the area of data reconciliation. More specifically, enabling IPs to align URIs from related controlled vocabularies, used across different datasets to identify persons, places, dates, and concepts, with greater ease, efficiency and efficacy.

With regards to the barriers experienced with LD Tooling, participants specifically mentioned that tools are often challenging to learn and use, inadequate for use in the LAMs, and difficult to incorporate into cataloguing workflows. The fact that all the LD

tools reviewed in Section 6.5 were found to have moderate usability issues could be seen as evidence of this.

Finally, in relation to challenges experienced with LD resources, participants highlighted concerns with the reliability of available LD resources, and copyright issues.

Other concerns that were frequently mentioned included the cost, both financial and time-wise, of publishing and consuming LD. This included the time and cost of training staff current staff on LD and hiring new IT staff. Participants indicated that in order to invest time and finances into LD, more useful examples of its applications would need to be seen.

6.4 LD Project Experience

Almost one third (29%) of participants had experience working on a LD project previously. The vast majority of these projects involved LD Publication (92%), with 59% involving LD Consumption. Judging by this discrepancy, it is possible that IPs are experiencing issues with LD resources that have already been published to the SW - a challenge that was highlighted by participants in Section 6.3 above.

6.5 Usability of LD Tools

Similar to Section 6.2, across all of the LD tools that participants mentioned having experience using, in the majority of instances participants did not require the support of a technical person in order to be able to use the tool. Again, Fedora was the only tool where more participants stated that they required assistance. Fedora was given one of the highest CSUQ scores in this instance also, however both Virtuoso[5] and BlazeGraph[6] received higher scores. That said, scores across all tools suggest moderate usability issues. Participants indicated that they had mild-moderate difficulties completing their work effectively using the tools, with interface design, ease of use, error recovery and learnability receiving some of the highest scores, suggesting increased usability issues in these areas.

6.6 LD for IPs

Participants had a generally positive response to the idea of LD tooling for IPs, and indicated that such tools could make LD more accessible for IPs. Participants suggested that, being more attuned to the needs and workflows of IPs, such tools could increase the number of LAMs engaging with LD.

Concerns with LD tooling for IPs were also raised, including the importance of ensuring such tools would be flexible and adaptable to the needs and processes of different institutions and teams. It was also highlighted that there are already existing LD Tools that LAMs could use, and, whilst this is true, based on the data discussed in Section 6.3 it could be argued that LAMs find these tools challenging to use and adapt to their workflows.

When asked to rate the usefulness of a LD Interlinking Tool for IPs, the vast majority of participants (89%) indicated that they thought it to be a useful idea. When asked to explain why they rated the idea as useful, see Table 8, the most commonly cited reasons included that, because interlinking resources provides tremendous value to LAMs in terms of enriching data and improving resource discoverability, a bespoke tool could help overcome the technical

---

[5]https://virtuoso.openlinksw.com
[6]https://www.blazegraph.com

knowledge gap of IPs who find interlinking to be one of the most challenging areas of creating LD. Again, concerns regarding how adaptable such a tool would be to the individual needs of institutions were raised, as well as a need for useful case studies if time and finances were to be invested in using such tooling.

7 CONCLUSIONS

Nowadays, with the Web being the first, and often only, place where users search for information, it is of great importance that LAMs make their data available online where it can be found by search engines and interact with other information resources. LD offers a means for LAMs to achieve this.

The results of our study identified a need amongst IPs to make LD tooling more accessible and attuned to their workflows. The response to the idea of a LD interlinking tool for IPs was positive, and a detailed set of requirements for such a tool was established. Developing this tool has the potential to facilitate increased engagement amongst IPs with LD, something which would benefit the LAM domain as a whole.

8 FUTURE DIRECTIONS

Following this study, the next step of our research will be to develop a LD interlinking tool specifically for IPs and the LAM community. This will be done by following the set of tool requirements identified by IPs during the course of this study as well as by taking into account the results of the CSUQ scores.

It is important to note that our research provides the LAM community with data outlining where the barriers lie in relation to IPs engaging with LD - see Section 6.3.2. While it is not within the scope of this research to address all of these issues, LAMs could use this information to address and provide potential solutions to other identified challenges.

REFERENCES

[1] G. Alemu, B. Stevens, P. Ross, and J. Chandler. 2012. Linked Data for Libraries: Benefits of a conceptual shift from library-specific record structures to RDF-based data models. *New Library World* 113, 11/12 (2012), 549–570.

[2] L. Ayres. 2008. *The SAGE Encyclopaedia of Qualitative Research Methods* (1st ed.). SAGE Publications Ltd., CA, Chapter Semi-Structured Interview, 811–812.

[3] M. Bergman. 2014. A Decade in the Trenches of the Semantic Web. (2014). http://www.mkbergman.com/1771/a-decade-in-the-trenches-of-the-semantic-web/.

[4] T. Berners-Lee. 2006. Linked Data - Design Issues. (2006). https://www.w3.org/DesignIssues/LinkedData.

[5] T. Berners-Lee, J. Hendler, and O. Lassila. 2001. The Semantic Web. *Scientific American* 284, 5 (2001), 1–5.

[6] C. Bizer, T. Heath, and T. Berners-Lee. 2009. Linked Data - The Story So Far. *International Journal on Semantic Web and Information Systems* 5, 3 (2009), 1–22.

[7] G. Byrne and L. Goddard. 2010. The Strongest Link: Libraries and Linked data. *D-Lib Magazine* 16 (2010), 11–12.

[8] T.W. Cole, M.J. Han, W.F. Weathers, and E. Joyner. 2013. Library Marc Records Into Linked Open Data: Challenges and Opportunities. *Journal of Library Metadata* 13, 2-3 (2013), 163–196.

[9] K. Coyle. 2012. Semantic Web and Linked Data. *Library Technology Reports* 48, 4 (2012), 10–14.

[10] K. Coyle. 2013. Linked Data: An Evolution. *Italian Journal of Library, Archives and Information Science* 4, 1 (2013), 53.

[11] J. Daniel. 2011. *Sampling Essentials: Practical Guidelines for making Sampling Choices* (1st ed.). SAGE Publications Ltd., CA, Chapter Choosing Between Non-Probability Sampling and Probability Sampling, 66–80.

[12] C. Deliot. 2014. Publishing the British National Bibliography as Linked Open Data. *Catalogue and Index* 174 (2014), 13–18.

[13] T. Fons. 2016. Making Libraries Visible on the Web. *Library Journal* 141, 13 (2016), 44–46.

[14] B.M. Gonzales. 2014. Linking Libraries to the Web: Linked Data and the Future of the Bibliographic Record. *Information Technology and Libraries* 33, 4 (2014), 10–22.

[15] M. Guerrini and T. Possemato. 2016. From Record Management to Data Management: RDA and New Application Models BIBFRAME, RIMMF, and OliSuite/WeCat. *Cataloging and Classification Quarterly* 54, 3 (2016), 179–199.

[16] M. Hallo, S. Lujan Mora, and J. C. Trujillo Mondejar. 2014. Transforming library catalogs into Linked Data.. In *Proceedings of the 7th International Conference of Education, Research and Innovation*.

[17] R. Hastings. 2015. Linked Data in Libraries: Status and Future Direction. *Computers in Libraries* 35, 9 (2015), 12–16.

[18] A. Kroeger. 2013. The Road to BIBFRAME: The Evolution of the Idea of Bibliographic Transition into a Post-MARC Future. *Cataloging and Classification Quarterly* 51, 8 (2013), 873–890.

[19] C. K. Lampert and S. B. Southwick. 2015. Leading to Linking: Introducing Linked Data to Academic Library Digital Collections. *Journal of Library Metadata* 13, 2-3 (2015), 230–253.

[20] F. LaPolla. 2013. Perceptions of Librarians Regarding Semantic Web and Linked Data Technologies. *Journal of Library Metadata* 13, 2-3 (2013), 114–140.

[21] R.J. Lewis. 1995. IBM computer usability satisfaction questionnaires: psychometric evaluation and instructions for use. International Journal of Humanfifft-Computer Interaction. *International Journal of HumanfiComputer Interaction* 7, 1 (1995), 57–78.

[22] L. McKenna. 2017. Engaging Librarians in the Process of Interlinking RDF Resources. In *Proceedings of the 14th European Semantic Web Conference*.

[23] L. McKenna, M. Bustillo, T. Keefe, C. Debruyne, and D. O'Sullivan. 2017. Development of an RDF-Enabled Cataloguing Tool. In *Proceedings of the 21st International Conference on Theory and Practice of Digital Libraries*.

[24] E. Miller and M. Westfall. 2011. Linked Data and Libraries. *The Serials Librarian* 60, 1-4 (2011), 17–22.

[25] E.T. Mitchell. 2016. Library Linked Data: Early Activity and Development. *Library Technology Reports* 52, 1 (2016), 5–33.

[26] P. Neish. 2015. Linked data: what is it and why should you care? *The Australian Library Journal* 64, 1 (2015), 3–10.

[27] OCLC. 2016. OCLC Linked Data Research. (2016). http://www.oclc.org/research/themes/data-science/linkeddata.html.

[28] Library of Congress. 2013. MODS RDF Ontology: Primer. (2013). https://www.loc.gov/standards/mods/modsrdf/primer.html.

[29] Library of Congress. 2016. Metadata Object Description Schema: MODS. (2016). http://www.loc.gov/standards/mods/.

[30] Library of Congress. 2016. Overview of the BIBFRAME 2.0 Model. (2016). http://www.loc.gov/bibframe/docs/bibframe2-model.html.

[31] O. Pesch and E. Miller. 2016. Using BIBFRAME and Library Linked Data to Solve Real Problems: An Interview with Eric Miller of Zepheira. *The Serials Librarian* 71, 1 (2016), 1–8.

[32] Qualtrics. 2018. Qualtrics. (2018). https://www.qualtrics.com.

[33] V. Schilling. 2012. Transforming Library Metadata into Linked Library Data. American Library Association. (2012). http://www.ala.org/alcts/resources/org/cat/research/linked-data.

[34] D. Seeman and L. Goddard. 2015. Preparing the Way: Creating Future Compatible Cataloging Data in a Transitional Environment. *Cataloging and Classification Quarterly* 53, 3-4 (2015), 331–340.

[35] P. Shvaiko and J. Euzenat. 2013. Ontology matching: State of the Art and Future Challenges. *IEEE Transactions on Knowledge and Data Engineering* 25, 1 (2013), 158–176.

[36] A. Sprochi. 2016. Where Are We Headed? Resource Description and Access, Bibliographic Framework, and the Functional Requirements for Bibliographic Records Library Reference Model. *International Information and Library Review* 48, 2 (2016), 129–136.

[37] T. Tullis and W. Albert. 2013. *Measuring the User Experience, Second Edition: Collecting, Analyzing, and Presenting Usability Metrics* (2nd ed.). Morgan Kaufmann Publishers Inc., MA, Chapter Self-Reported Metrics, 121–161.

[38] W3C. 2014. RDF: Resource Description Framework. (2014). https://www.w3.org/RDF/.

[39] W3C. 2015. Semantic Web. (2015). https://www.w3.org/standards/semanticweb/.

[40] K. S. Yoshimura. 2016. Analysis of an International Linked Data Survey for Implementers. *D-Lib Magazine* 22, 7 (2016), 6.

[41] F. Zablith, M. Fernandez, and M. Rowe. 2015. Production and Consumption of University Linked Data. *Interactive Learning Environments* 23, 1 (2015), 55–78.

[42] A Zaveri, A. Rula, A Maurino, R. Pietrobon, J. Lehmann, and S. Auer. 2015. Quality Assessment for Linked Data: A Survey. *Semantic Web* 7, 1 (2015), 63–93.