Wikidata and knowledge graphs in practice: Using semantic SEO to create discoverable, accessible, machine-readable definitions of the people, places, and services in Libraries and Archives

Jason A. Clark[a], Helen K.R. Williams[b] and Doralyn Rossmann[c,*]

[a]*Professor, Head of Digital Library Initiatives, Montana State University Library, Bozeman, MT, USA*
[b]*Metadata Manager, Digital Scholarship and Innovation Group, London School of Economics Library, London, England*
[c]*Professor, Dean, Montana State University Library, Bozeman MT, USA*

Abstract. Libraries expand the access and visibility of data and research in support of an informed public. Search engines have limited knowledge of the dynamic nature of libraries - their people, their services, and their resources. The very definition of libraries in online environments is outdated and misleading. This article offers a solution to this metadata problem by redefining libraries for Machine Learning environments and search engines. Two ways to approach this problem include implementing local structured data in a knowledge graph model and "inside-out" definitions in Semantic Web endpoints. MSU Library has found that implementing a "Knowledge Graph" linked data model leads to improved discovery and interpretation by the bots and search engines that index and describe what libraries are, what they do, and their scholarly content. In contrast, LSE Library has found that contributing to Wikidata, a collaborative and global metadata source, can increase understanding of libraries and extend their reach and engagement. This article demonstrates that Wikidata can be used to push out data, the technical details of knowledge graph markup, and the practice of semantic Search Engine Optimization (SEO). It explores how metadata can represent an organization equitably and how this improves the reach of global information communities.

Keywords: Search engine optimization, semantic web, knowledge graphs, discovery, linked data, wikipedia, wikidata, metadata

1. Introduction

Traditionally, libraries and archives have focused on the internal discovery of their collections - books, journals, databases, and special collections. However, libraries and archives provide much more than collections to their users - high impact event programming, subject and functional experts, and the management of institutional outputs and quality services to name a few. These types of offerings are equally as important to a library's identity and should be just as discoverable. In 2010 Lorcan Dempsey phrased this as 'a distinction between outside-in resources, where the library is buying or licensing materials from external providers and making them accessible to a local audience (e.g., books and journals) and 'inside out' resources which may be unique to an institution (e.g., digitized images, research materials) where the audience is both local and external. Thinking about an external non-institutional

---

*E-mail: doralyn@montana.edu.

audience and how to reach it poses some new questions for the library [1]'. This 'inside-out' model, and the resulting question of how to reach an external audience, has provided a conceptual framework at both Montana State University (MSU) and London School of Economics (LSE) as each library has sought to 'pivot their expertise to organize information outward' by taking on 'broader information management challenges at their college or university [2]'. What is apparent within the library field is that definitions of libraries and library services continue to evolve [3], but libraries and archives are understood within machine environments in limited terms as either an "organized collection of resources [4]" or "an accumulation of historical records [5]". Many humans understand that today's libraries and archives are more than these definitions - machines have outdated definitions. Humans can enhance information for machines to help them interpret, learn about relationships, and make meaning.

There is an opportunity here to take the expertise of libraries and archives into new environments and contribute to the broad global learning environment. Consider some emerging examples of library and archives services and resources: Pima County Public Library in Tucson, Arizona became the first library system in the nation to hire public health nurses in its branches; San Francisco Public Library (SFPL) hired a licensed marriage and family therapist; Sanibel Public Library in Florida has a "Cooking in the Bag" program which provides users with a tote bag containing culinary items and a coordinating cookbook for making dishes such as empanadas, sushi, and dumplings. In the United Kingdom, the British Library has a Wikimedian in Residence [6] and the National Library of Scotland offers a Data Foundry for data scientists [7]. To ensure the discoverability of an ever-expanding and evolving array of services, activities, and expertise, the library and archives community needs a discovery and web-scale cataloging strategy. Trusted, structured data sources such as Wikipedia and Wikidata along with structured data-controlled vocabulary sources such as Schema.org are becoming necessary for allowing bots and Artificial Intelligence (AI) software agents to interpret meaning from the Web [8]. Building and seeding these structured data sources and applying these vocabularies with library and archive knowledge graph data about the who, what, when, where, and why of institutions is an essential act of web-scale cataloging with international reach and impact. These projects and case studies unite these disparate ideas into a coherent method of practices.

2. Problem statement and literature review

As a part of the global information community, libraries provide content and education that expands the access and visibility of data and research in support of an informed public. And yet, search engines and indexing software agents have limited knowledge of the dynamic nature of libraries - the people who make the library happen, the services provided, and the resources procured, thus the very definition of libraries is static, outdated, and misleading.

At its core, this is a metadata problem.

2.1. Knowledge Graph as a concept

In essence, Knowledge Graphs describe various instances, concepts, and relationships between them in the real world as kinds of semantic knowledge bases, and they are a significant method to realize substantive knowledge management [9]. A history of the concept of knowledge graphs reflects an initial use by Google in 2012 and subsequent knowledge graphs have taken both open and proprietary forms. In this article, a knowledge graph is defined as a data model for representing and understanding information
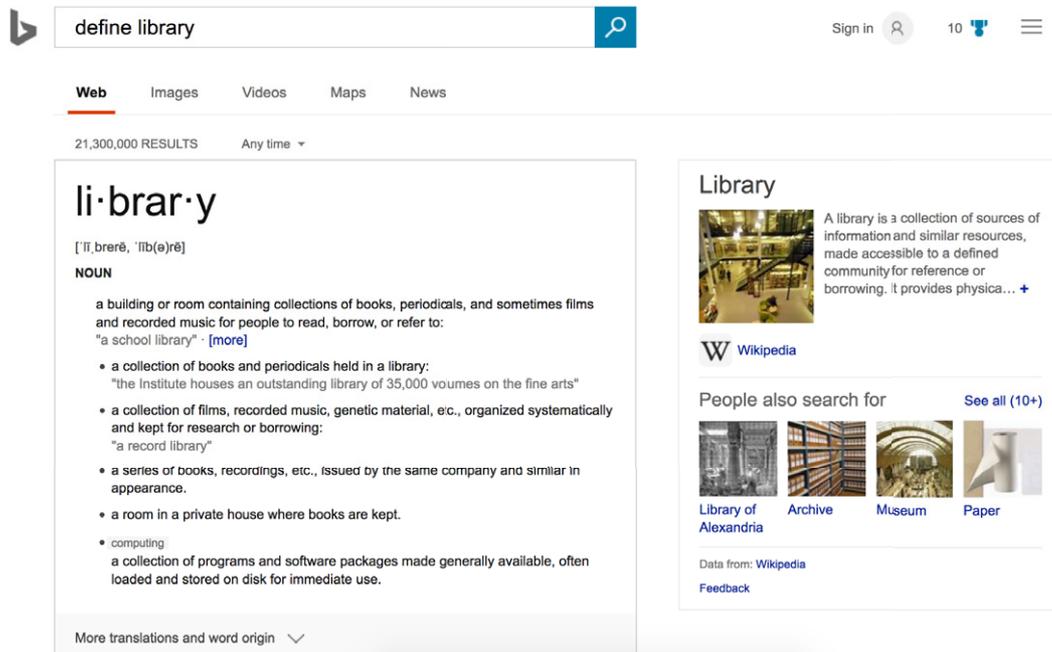
Fig. 1. Bing Knowledge Panel showing limited understanding of library as a concept.

as real-world entities and the relationships between those entities. In practical terms, they encode facts related to people, processes, applications, data and things, and the relationships among them [10].

2.2. Wikidata

Libraries traditionally collaborate with each other through shared networks and development tools. The literature reflects the potential for Wikidata to expand the library's reach to the global stage. Through the application of linked data, Wikidata can be used to expand into the broader metadata community [11]. This reach extends with data made freely available for reuse using Wikidata's automatically applied CC0, public domain license [12].

Many libraries are drawn to linked data, and Wikidata particularly, for its potential to build on existing metadata efforts in library systems [13,14]. It offers more powerful tools for search and discovery, better access for users, and more complete and consistent data, particularly in an era of disinformation [15]. It also has the potential to create a worldwide dataset which includes connected and aligned data between institutions [16]. Wikidata's collaborative and open elements empower libraries and archives to work together with partners and communities to create, enhance, and research data in new ways, including means such as crowdsourcing [17]. This collaborative approach mitigates some of the challenges the adoption of linked data presents and allows libraries and archives to pool resources in order to improve data, and ultimately to continue to facilitate and serve as essential providers of open knowledge [18].

2.3. Library definitions and visibility

Considering the opening sentence for the term "library" in Wikipedia, "A library is a collection of materials, books or media that are easily accessible for use and not just for display purposes", there is an

insufficient description of the library as a dynamic entity beyond the book warehouse [19]. Research by the MSU Library finds that implementing a knowledge graph linked data model within HTML markup leads to improved discovery and interpretation by the bots and search engines that index and describe what libraries are, what they do, and their scholarly content [20,21].

Similarly, Library-provided content may not be surfaced to users through search engine results. SEO and social media optimization (SMO) practices can lead to greater understanding of library content by search engines and social media networks [22]. Likewise, resources such as Wikipedia and Wikidata can allow libraries to give new information for search engine understanding [23]. "Wikidata not only has the power to surface information that doesn't exist outside of library catalogs, but it also has the power to amplify the reach of existing collections through Wikipedia and digital assistants (such as Siri and Alexa) which employ Wikidata's structured data to answer questions [24]".

3. Methods and case studies

By working with knowledge graph vocabularies that already exist and seeding global Semantic Web resources, such as Wikidata, libraries and archives have an opportunity to reach beyond their own organizational silos to a broad global audience by sharing their knowledge in ways that can be understood and used in the Semantic Web environment. This could relate to collections, content, and the entities within them, or to faculty and their research and data, or any other institutional information that a library or archive wants to expose more broadly.

The methods and key concepts followed here are important to unpack. Semantic search engine optimization, the act of encoding graph relationships in HTML markup to improve findability and understanding of concepts by search engines and other browsing agents (e.g., social media), is the part of this method which unites the theory of knowledge graphs to the practice of discovery. Semantic search relies on a network of related entities such as contextually related concepts, ideas, people, places, and things to determine what a web page is about and the "intent" which it might fulfill. Optimization adds coding and indexing cues to web pages to give search engines and other software agents more information, proactively.

In practice creating expressions of relationships and encoding them into web pages involves using a vocabulary and additional HTML markup. The vocabulary works by defining a language set of properties and types that can be understood by machine agents. In this scenario an HTML page with standard markup that looks like this:

```
<div >
<ul class="aboutList">
<li class="photo"><img
src="http://www.lib.montana.edu/people/meta/img/photos/anika-200px.jpg"
alt="photo of Anika" /></li>
<li class="infoContact">
<h3 >Anika </h3>
```

Becomes a richly defined entity with types and properties:

```
<div id="person" typeof="schema:Person"
about="http://www.lib.montana.edu/people/about/64">
<ul class="aboutList">
<li class="photo"><img property="schema:image"
```

Fig. 2. Knowledge Graph information influencing a Search Engine Result Page.

src="http://www.lib.montana.edu/people/meta/img/photos/anika-200px.jpg"
alt="photo of Anika" />
<li class="infoContact">
<h3 property="schema:name" >Anika </h3>

This refinement of definitions and relationships embedded as structured data in the HTML attributes or other markup like JSON-LD creates Search Engine Optimization (SEO) and impacts the interpretability and visibility of the web page leading to new forms of access, downloads, organic search results, etc. [25].

Wikidata is a structured database operating as the central data store for Wikimedia projects. It is a 'free and open knowledge base, multilingual and it can be read and edited by humans and machines [26]'. Google Knowledge Graphs, digital assistants, and Wikipedia infoboxes are all populated, in part, with information harvested from Wikidata, so its content influences Web-based discovery. By minting unique Wikidata identifiers pertaining to libraries and archives institutions, the resources and research of those institutions, and the entities within their data, become part of the linked open data ecosystem. Links and relationships are established between entities and connections are made with unique identifiers from a myriad of external knowledge systems. This assists machines in interpreting library and archives resources and creates bridges between previously siloed domains, in turn impacting search engine results by providing a fuller picture of globally available data.

There is incredible potential to apply these linked data models and optimizations to impact discovery and understanding of library and archival work and content. As just one example, consider figure 1 below which demonstrates the impact of knowledge graph structured data implemented to describe Web of Science, one of the MSU Library's subscribed library databases.

Not only is the MSU Library's Web of Science database appearing in the top set of results, but it is also styled differently with ratings, breadcrumbs showing page hierarchy, and a clear definition of the database. This is just one example of successful machine interpretation enabled by knowledge graph implementations. In this case, there is the addition of structured data markup using enhanced HTML to describe the MSU Library Web of Science database and define the facts and the relationship of the database to the library. This same knowledge graph structured data can be applied to define a library's people, services, physical spaces, events, etc. This method picks up on this idea of machine user experience to introduce the potential for a library knowledge graph that can be applied and reused by multiple institutions and organizations. Each of the case studies below approaches the machine user experience using different methods, and yet, yielding similar results.

3.1. Case study 1: Wikidata and Global Metadata

A 2019 ARL White Paper [27] highlights Wikidata as a 'means of documenting and surfacing researchers, publications and research data' and 'sharing faculty scholarship on an open and accessible platform [28]'. LSE Library wanted to further unpack the potential that Wikidata offered to take the collaborative creation and management of metadata beyond the Library to a global landscape. The Metadata team sits in the Library's Digital Scholarship and Innovation Group and as part of the digital shift sought to expand its existing focus on the management of scholarly content to include the exploration and development of new ways in which metadata can support research, teaching, and learning. The approach taken drew on an 'inside-out' model with a particular focus on scholarly content and the way in which metadata can extend its access and visibility by pushing data beyond organizational silos to increase its discoverability. By making content more widely accessible and enabling potential new connections and discoveries, the team hoped to bring benefits to LSE and beyond that to global research.

The Library's Wikidata journey began from scratch. Some institutions can draw on the expertise of a Wikimedian in Residence, which was not the situation at LSE, so local learning began by reading articles and watching presentations. Various barriers had to be overcome, including the technical skills required for bulk uploading content and getting to grips with Wikimedia's policies and procedures. There was also the necessity of articulating the value of Wikidata to justify the staff time and resources that would need to be spent overcoming the aforementioned barriers. Following Dan Scott's blog post *Creating and editing libraries in Wikidata* [29] the existing item for the British Library of Political and Economic Science [30] was edited to include new properties and identifiers. New items were then created for component parts of the Library, such as LSE Digital Library [31], which could be linked to the Library using 'has part or parts [32]' and 'part of [33]' properties.

Different content types in the Library were mapped to Wikidata, and data models were created for LSE Digital Library collections, blogs, open access journals and online exhibitions as well as for corporate bodies related to the Archives. Finally corresponding Wikidata items were created for each data model. All this activity provided ideas for potential avenues of work, where the focus could be organizational, community, research, theses (inspired by Martin Poulter's work at Oxford University), open access, digital or archival [34]. The possibilities for extending the institutional web presence far outweighed the staff time and resources available, and in discussion with colleagues it was decided to have an initial focus on content in LSE Theses Online, (commonly referred to in the institution as LSETO) which is an online archive of just over four thousand born-digital and digitized PhD theses [35]. It is already indexed by Google, but value could be added by bringing the data into the linked open data ecosystem and contextualizing it by linking it to related data and content that is not created or managed by LSE.

This would test whether inclusion in Wikidata extended the reach and engagement of this set of content, thereby increasing potential for new connections and discoveries. It was a boundaried dataset of a suitable size for an experimental project and would have real world benefits to the institution by offering value to early career researchers and alumni whose work would be promoted by the Library's efforts.

The theses data was modeled and mapped in detail and OpenRefine was used for name reconciliation with Wikidata to start creating relationships between entities by making connections with data outside LSE [36]. QuickStatements and OpenRefine were used to bulk upload content to Wikidata, before undertaking further reconciliation work, this time to match the theses with identifiers in the external datasets EThOS, CORE, DART-Europe, and ProQuest Dissertations and Theses Global, making further connections between the LSE theses content and the linked open data cloud [37]. Finally, 'roundtripping' work was carried out, adding statements to create links between authors, supervisors, and theses, and show their relationships to each other and to LSE.

Learning new skills was rewarding, but it was important to assess whether the work was valuable to the institution. The Metadata team mission supports the institutional strategy by facilitating discovery of LSE Library collections and LSE research for the LSE community and the wider world, and the vision for achieving this is to 'create and manage comprehensive and authoritative metadata which adds value to LSE's outstanding collections, contributes to the global web of data, and facilitates wide use of the collections. An interim analysis, when approximately a quarter of the data had been added to Wikidata, would seek to investigate whether bringing the metadata into the linked open data ecosystem was extending the reach and engagement of this specific set of content.

3.2. *Case study 2: A Library Knowledge Graph*

At the MSU Library, interest in employing linked data and SEO markup stemmed from the pieces that make up this network. The presence of Google Knowledge Panels, the emergence of Wikipedia and Wikidata as information resources, the broader adoption of Schema.org and linked data as methods of describing entities, and the products and services they offer provided a robust testing ground for libraries and archives, including the MSU Library to insert their metadata expertise for greater machine understanding.

Initial efforts at MSU using the inside-out concept, as described earlier, intended to put information about the library out for greater understanding by search engines. This work included creating a Business Profile for the library on Google Search as well as populating a Google Knowledge Panel. Additional efforts went towards adding Schema.org and SMO tags to digital collections and later, to pages linking to library-subscribed-resources.

The focus of this case study extends towards MSU Library's most recent efforts to give search engines a better understanding of the MSU Library as a suite of offerings - of knowledgeable people, useful places, and beneficial services. Using the knowledge graph concept led to rethinking the library's website as a graph of pages which tell a story about what the library is, as a whole. The dozen top-level pages were identified for the library's website as potential for both Semantic SEO and SMO markup.

Schema.org has a searchable, structured vocabulary. A review of options led to selecting the closest types of categories to describe the library - offering services, and what types of services, for example. Then, these types of services and resources were matched with Wikipedia and Wikidata entries to connect these offers and types to their definitions. In this phase of the project, 6 primary webpages from the MSU Library website were given enhanced structured data including: the library homepage, the library about page, the library services page, the library people page, the library resources page, and the library spaces

Fig. 3. Semantic SEO and SMO markup in place on MSU Library Services webpage.

page. (The complete structured data model markup is available in the library's GitHub account.) The final step was to encode the knowledge graph model that would define the relationships between these pages. The knowledge graph encoding was done using the Schema.org website type and hasPart properties to identify the MSU Library website as a whole and then linking each of the 6 primary pages using the Schema.org webpage type and isPartOf properties. An example of this markup encoding in place on the library about page appears below.

```
<script type="application/ld+json" >
{
 "@context": "https://schema.org",
 "@type": "AboutPage",
 "@id": "https://www.lib.montana.edu/about/",
 "isPartOf": [
  {"@type": "WebSite", "url": "https://www.lib.montana.edu/"}
 ],
...
}
<script >
```

4. Findings and discussion

Within the varied approaches in the case studies, the authors expected to see some levels of improvement in indexing rates, overall visibility of scholarly materials or web pages, reuse and downloads of material,

and traffic pattern referrals from search engines to the LSE and MSU websites as part of the results. This expectation held and both case studies found levels of growth in the areas above which is encouraging. It was also encouraging to see how each distinct optimization method - for LSE, seeding existing global knowledge bases with LSE links and resources, for MSU, creating new descriptive markup definitions in local web pages and defining the relationships between those pages as a graph - had similar impacts. This finding suggests there are multiple ways to approach optimization work and provide two types of solutions for library and archives staff looking to create visibility and redefine their institutions for machine environments. One note of import: 2021 was a year where many library and archives patrons were exclusively online. This is a factor in the numbers we see in the findings when comparing 2021 to 2022. Case study findings are discussed individually below.

4.1. Case study 1 - Findings

LSE's initial analysis of working with Wikidata focused on the visibility of scholarship. There has been an expansion of access and visibility through increased downloads of LSE theses, increased referrals from Wikipedia, increased reach and engagement and new opportunities to visualize the data.

A SPARQL query ranking institutions according to the number of theses they have in Wikidata showed that LSE began somewhere between 287th and 467th place, in company with all the other institutions that had one single thesis in Wikidata, but by the time of the interim analysis LSE was 9th on the list, and at the time of writing was 5th [38]. At the time of publication LSE has moved to 12th following a similar large scale project in New Zealand to add thesis data to Wikidata

An analysis of LSETO data between February and May 2021 revealed that downloads for that period were 14% higher than the same period in 2020, when for comparison the same time period over the previous three years had seen an increase in downloads of 6.9 % in 2019 and decreases of 5% and 12% in the previous two years, so there was a notable difference in the downloads. Those figures are for downloads across the whole of LSETO, so the analysis then focused on eighty titles which had already been added to Wikidata, investigating the downloads for those individual titles in the six months before and after addition to Wikidata. The results showed that on average downloads in the six months after the content was added to Wikidata were 47% higher than the preceding six months, which was an encouraging uplift.

Google Analytics for LSETO was another source of data for analyzing the impact of the Wikidata work. There was not an expectation of seeing referrals to LSETO directly from Wikidata, because putting the data in Wikidata is about other sources using that data to drive traffic, but as part of the project where a thesis author had a Wikipedia page their thesis title was added with a citation to LSETO, so it was reasonable to expect an increase in referrals from Wikipedia. Readers will need a bit of context for the figures; the primary referral source into LSETO is Google Scholar, and during the period of analysis that accounted for approximately 40% of referrals. The second referral source was Twitter at approximately 10%. Another ten sources were referring between 1% and 6% of traffic each and after that a long tail of approximately three hundred sites were referring 0.x or 0.0x % of traffic. In the six months before the Wikidata work began LSETO received an average of 3.82% of its traffic from Wikipedia. In the following six months it increased to 9.31% (with the most recent week before analysis being 13.61% of traffic). This moved Wikipedia from the 5th referral source in the six months before Wikidata work began, to the 3rd referral source in the six months since (still following Google Scholar and Twitter). Finally, looking at Twitter there were thirty-eight mentions of e-theses.lse.ac.uk between February and May 2020, increasing to seventy-four during the same period for 2021. There were some delays in adding the full set of theses data, and further analysis will take place when enough time has elapsed to obtain meaningful data.

Wikidata also offers opportunities to visualize the data in new ways through the SPARQL query service, such as relationships between authors and supervisors, or a graph of awards won by LSE thesis authors and supervisors, or a table of related employers. A full set of queries can be seen at LSE's WikiProject page for the thesis project. [39]. Working with Histropedia LSE was also able to create a timeline of people who are part of the LSE thesis project and of theses authored at LSE, with filters which make use of additional data held in Wikidata (such as birth and death dates, birth country, citizenship and occupation). [40,41].

The success of this initial Wikidata project in extending the reach and engagement of theses content has led to expansion of the Wikidata work, both with theses content, and with other content and data unique to LSE Library. Work has begun to establish an automated process for the addition of theses authors and supervisors who do not already have a Qid on Wikidata. Alongside this work has taken place to model metadata for content on the institution's University Press for addition to Wikidata, and an initial process to automate that is in the early stages of use. The visualization of University Press data in Scholia is being examined, for example data relating to LSE Public Policy Review and graphs can be created via Wikidata's SPARQL query service to show author collaborations, again using LSE Public Policy Review as an example [42,43]. Further options under discussion include:

- Special collections focus whereby under-exposed or under-represented content could benefit from enhanced search engine discoverability via inclusion in Wikidata.
- Digital Library focus including using Wikidata to create a collections map of LSE digital content.
- Researcher focus whereby the potential of Wikidata as an identifier hub is utilized to support the management of names related to LSE, enhancing them for use by search engines.

4.2. Case study 2 - Findings

Within website usage metrics, MSU was expecting to see a general improvement of the referrals from search engines to the library's pages, an increase in user visits, and an increase in user sessions. (Note: A session is the period of time a user is actively engaged with your website, app, etc. All usage data (Screen Views, Events, Ecommerce, etc.) is associated with a session.) MSU was particularly interested in understanding web traffic from search engines and monitored the Acquisitions and Source/Medium metric in Google Analytics to gather this data. Within website indexing metrics, MSU was expecting to see growth in total clicks, a rise in impressions, and an increase in clickthrough rates (CTR). (Note: An impression means that a user has seen (or potentially seen) a link to your site in search engine result pages.)

MSU's initial analysis of the knowledge graph website implementation shows growth in these common website usage metrics. The analytics MSU compares in this analysis applies a matching range of four months (January to May) from 2021 to 2022. The date range was January 24 to May 24. Knowledge graph optimization was put into place on January 3. The 21 days before monitoring allowed the optimizations to take effect.

A Google Analytics property segment identifying only the six knowledge graph optimized pages was used to isolate and monitor impact. Specific analytics of note include:

- A 20% increase in Users including a 29% increase in New Users.
- An 8.6% increase in Sessions.

  Overall traffic patterns from search engines showed growth as well.

- A 10% increase in organic search result referrals from Google.
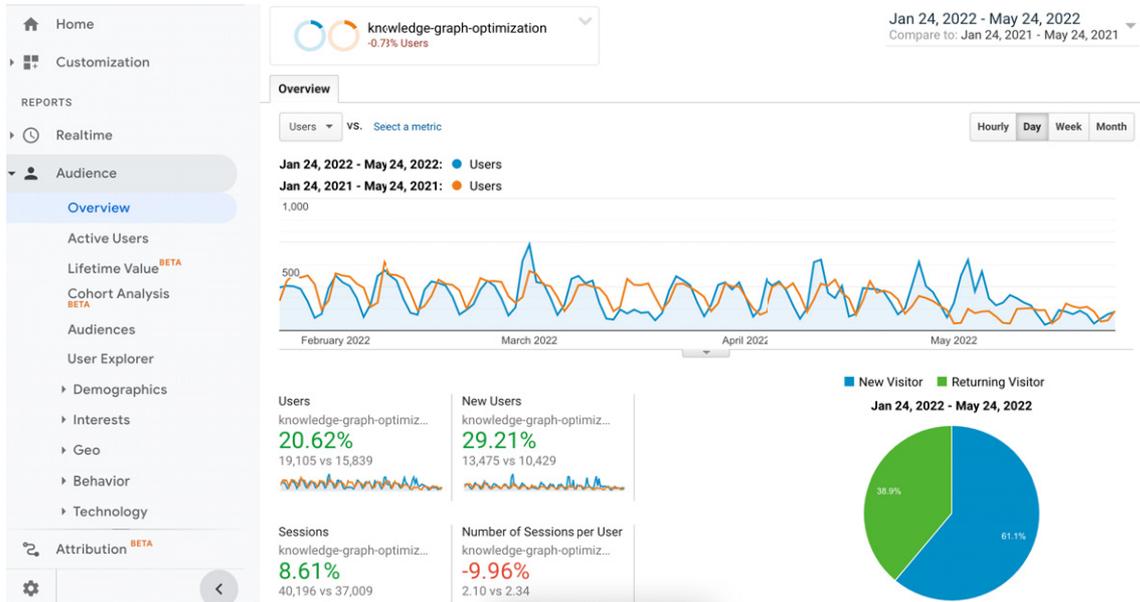- A 34% increase in organic search result referrals from Bing.

Fig. 4. Comparison of users and sessions in google analytics - pre and post optimization.

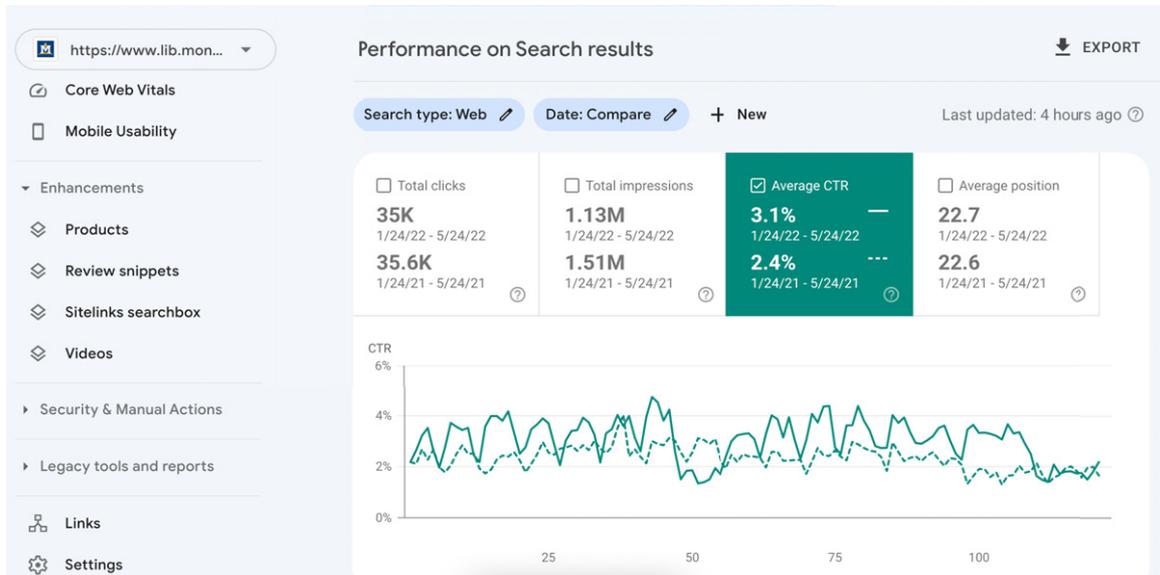Fig. 5. Search performance benchmark in Google search console - pre and post optimization.

Website indexing metrics and search engine analytics were gathered from a Google Search Console property to understand performance of the optimized web pages in Google's search engine results. The same date ranges were assessed in this tool and an initial comparison provided a benchmark for how the website was performing in search settings.

Upon initial review, impressions and clicks held to the same numbers from 2021 to 2022. There was an uptick in the Click Through Rate (CTR) from 2.3% to 3.1% which suggests some more engagement from users in clicking from a search result page into the library website. In this particular assessment, MSU also noted a 3.6% increase in CTR to the library homepage. (CTR for the homepage in 2022 was 8.9% while it was 5.4% in 2021.) This analysis gets more interesting if one looks into the performance of the individual optimized pages. As noted above the overall website indexing metrics and search engine analytics are maintained. However, once we consider individual pages, some growth metrics can be observed. Among the highlights:

- The library about page impressions moved from 22,380 to 31,047 impressions.
- The library spaces page impressions moved from 1,504 to 5,565 impressions.
- The library service page impressions moved from 4,822 to 10,150 impressions.
- The library resources page impressions moved from 46,869 to 50,817 impressions.

These individual page findings show potential correlation to the optimizations MSU Library put into place. MSU plans to continue with this pilot and place more knowledge graph markup on additional primary landing pages to see if the impact and performance continues to improve.

5. Conclusion

Extending MSU Library's descriptions of its people, resources, and services using Wikipedia and Wikipedia definitions shows initial gains in search engine traffic to its website. A next step could include applying similar practices to other parts of MSU Library's website as well as looking to other libraries and archives as potential comparison points. Schema.org offers many types and properties and testing different ones of these options could yield more or less traffic. At this point, having more libraries participate in and share their data would yield greater understanding for how humans can send more accurate signals to machines about who and what today's libraries do and offer.

By contributing to Wikidata as a global and collaborative metadata source LSE Library has extended the reach and engagement of a specific set of content. The work has been shared with the institution's PhD Academy, furthering the role of the Library in research dissemination and demonstrating the value of metadata in expanding the access and visibility of libraries and their resources to ensure that they are understood in the Semantic Web environment. The Wikidata theses project is already being used as the basis to establish further work developing the role of metadata in supporting research, teaching, and learning and in supporting inclusive and equitable access to that research and data, improving the accessibility and reach of global information communities.

Graph data models and sources have an impact for libraries. These methods and forms of cataloging create new understandings for humans and machines. Moreover, they allow information professionals to apply expertise. And finally, they enhance discovery and usage of libraries and archives content.

About the Corresponding Author

Doralyn Rossmann is a Professor and Dean of the Library at Montana State University in Bozeman, MT, USA, where she has worked in a variety of capacities since 2001. She graduated from the University of North Carolina Chapel Hill in 1992 with a B.A. in Political Science and English where she also obtained her M.S in Library and Information Science two years later. She obtained a Master's in Public Administration from Montana State University, Bozeman, MT in 2008. For more information go to: https://doralyn.org/curriculum-vitae/. Phone: (406)-993-6978; E-mail: doralyn@montana.edu.

References

[1] L. Dempsey, Outside-in and inside-out [cited 2022, March 12]. Available from: https://www.lorcandempsey.net/orweblog/outside-in-and-inside-out/, accessed September 3, 2022.

[2] M. Dahl, "Inside-out library services", Challenging the "Jacks of all trades but masters of none" librarian syndrome, *Advances in Library Administration and Organization* **39**: (2018), 15–34.

[3] Socialworker.com. Public libraries add social workers and social programs. [cited 2022, May 29]. Available from: https://www.socialworker.com/feature-articles/practice/public-libraries-add-social-workers-and-social-programs/, accessed September 3, 2022.

[4] Wikidata. library. [cited 2022, May 29]. Available from: https://www.wikidata.org/wiki/Q7075, accessed September 3, 2022.

[5] Dbpedia. Archive [cited 2022, May 29]. Available from: http://dbpedia.org/page/Archive, accessed September 3, 2022.

[6] The British Library. Welcome to the British Library's new Wikimedian in residence, The British Library Blog (2021, March). Available from: https://blogs.bl.uk/digital-scholarship/2021/03/welcome-to-the-british-librarys-new-wikimedian-in-residence.html,%20accessed, accessed September 3, 2022.

[7] National Library of Scotland. Digital resources: Data foundry [cited 2022, May 29]. Available from: https://www.nls.uk/digital-resources/data-foundry/, accessed September 3, 2022.

[8] Wired. Inside the Alexa friendly world of Wikidata [cited 2022, May 29]. Available from: https://www.wired.com/story/inside-the-alexa-friendly-world-of-wikidata/, accessed September 3, 2022.

[9] H. Yu, H. Li, D. Mao and Q. Cai, A domain knowledge graph construction method based on Wikipedia, *Journal of Information Science* **47**(6) (2021), 783–793.

[10] D. Fensel, U. Şimşek, K. Angele, E. Huaman, E. Kärle, O. Panasiuk et al. Introduction: What is a knowledge graph? in: *Knowledge Graphs* Springer, 2020.

[11] Association of Research Libraries. ARL White Paper on Wikidata. 2019. Available from: https://www.arl.org/resources/arl-whitepaper-on-wikidata/, accessed September 3, 2022.

[12] J. Evans, Library data as linked open data, *Catalogue & Index* (2020), 199.

[13] OCLC, Hanging together: Wikimedia. 2020. Available from: https://hangingtogether.org/category/wikimedia/, accessed September 3, 2022.

[14] The Science Museum (UK), Wikidata and Cultural Heritage Seminar. 2020, June 22. Available from: https://thesciencemuseum.github.io/heritageconnector/events/2020/06/22/wikidata-and-cultural-heritage-collections-webinar/, accessed September 3, 2022.

[15] M. Poulter and N. Sheppard, Wikimedia and universities: Contributing to the global commons in the Age of Disinformation, *Insights* **33**(1) (2020), 14.

[16] N. Sheppard, Wikimedia in universities, Leeds University Library Blog. 2019, February 1. Available from: https://leedsunilibrary.wordpress.com/2019/02/01/wikimedia-in-universities/, accessed September 3, 2022.

[17] E. Kapsalis, Wikidata: Recruiting the crowd to power access to digital archives, *Journal of Radio & Audio Media* **26**(1) (2019), 134–142.

[18] M. Lemus-Rojas and L. Pintscher, Wikidata and libraries: Facilitating open knowledge. 2017, March 3. Available from: https://scholarworks.iupui.edu/bitstream/handle/1805/16690/Lemus-Rojas-Pintscher-Wikidata-2017-07-03.pdf, accessed September 3, 2022.

[19] Wikipedia, Library [cited 2022, May 29]. Available from: https://en.wikipedia.org/wiki/Library, accessed September 3, 2022.

[20] J.A. Clark and S.W.H. Young, Linked data is people: Building a knowledge graph to reshape the library staff directory, Code4lib Journal. 2017; 36.

[21] Wikipedia. Knowledge Graph [cited 2022, May 29]. Available from: https://en.wikipedia.org/wiki/Knowledge-Graph, accessed September 3, 2022.

[22] J.A. Clark and D. Rossmann, The Open SESMO (Search Engine & Social Media Optimization) project: Linked and structured data for library subscription databases to enable web-scale discovery in search engines, Journal of Web Librarianship. 2017; 11(3-4): 1-22.

[23] K. Arlitsch and J. Shanks, Wikipedia and Wikidata Help search engines understand your organization: Using semantic web identity to improve recognition and drive traffic. in: *Leveraging Wikipedia: Connecting Communities of Knowledge*, M. Proffitt (ed.), 2018.

[24] Wikiedu. Expanding the reach of your library collections through Wikidata, Wikiedu Blog. 2019, August, 20. Available from: https://wikiedu.org/blog/2019/08/20/expanding-the-reach-of-your-library-collections-through-wikidata/, accessed September 3, 2022.

[25] Google. Introduction to structured data [cited 2022, May 29]. Available from: https://developers.google.com/search/docs/advanced/structured-data/intro-structured-data, accessed September 3, 2022.

[26] Wikidata. Wikidata: Main page [cited 2022, May 29]. Available from: https://www.wikidata.org/wiki/Wikidata:Main-Page, accessed September 3, 2022.

[27] Association of Research Libraries. ARL white paper on Wikidata. 2019. Available from: https://www.arl.org/resources/arl-whitepaper-on-wikidata/#.XOaxXdNKh-X, accessed September 3, 2022.

[28] Association of Research Libraries. ARL white paper on Wikidata: opportunities and recommendations. 2019; 36. Available from: https://www.arl.org/wp-content/uploads/2019/04/2019.04.18-ARL-white-paper-on-Wikidata.pdf, accessed September 3, 2022.

[29] D. Scott, Creating and editing libraries in Wikidata. 2018. Available from: https://coffeecode.net/creating-and-editing-libraries-in-wikidata.html, accessed September 3, 2022.

[30] Wikidata. British Library of Political and Economic Science [cited 2022, May 5]. Available from: https://www.wikidata.org/wiki/Q2371017, accessed September 3, 2022.

[31] Wikidata. LSE Digital Library [cited 2022, May 5]. Available from: https://www.wikidata.org/wiki/Q96354844, accessed September 3, 2022.

[32] Wikidata. Has part or parts [cited 2022, May 5]. Available from: https://www.wikidata.org/wiki/Property:P527, accessed September 3, 2022.

[33] Wikidata. Part of [cited 2022, May 5]. Available from: https://www.wikidata.org/wiki/Property:P361, accessed September 3, 2022.

[34] M. Poulter, A step forward in the sharing of open data about theses, The Bodleian Library Blog. 2017, July 19. Available from: http://blogs.bodleian.ox.ac.uk/digital/2017/07/19/a-step-forward-in-the-sharing-of-open-data-about-theses/, accessed September 3, 2022.

[35] The London School of Economics Library. LSE Theses Online [cited 2022, May 5]. Available from: https://etheses.lse.ac.uk/.

[36] OpenRefine [homepage on the internet]. [cited 2022, May 2]. Available from: https://openrefine.org/, accessed September 3, 2022.

[37] QuickStatements [homepage on the internet]. [cited 2022, May 5]. Available from: https://quickstatements.toolforge.org/#/, accessed September 3, 2022.

[38] Wikidata query service. [Count of doctoral theses submitted to institutions]. [cited 2022, May 5]. Available from: https://w.wiki/jwZ, accessed September 3, 2022.

[39] Wikidata query service. [Author and supervisor relationships in LSE doctoral theses.] [cited 2022, May 5]. Available from: https://www.wikidata.org/wiki/Wikidata:WikiProject_LSEThesisProject, accessed September 3, 2022.

[40] Histropedia. Women educated at LSE [cited 2022, May 5]. Available from: https://tinyurl.com/yckfwhxa, accessed September 3, 2022.

[41] Histropedia [homepage on the internet]. [cited 2022, May 5]. Available from: https://tinyurl.com/ycyu8nwu, accessed September 3, 2022.

[42] Scholia. [B] Public Policy Review (Q97011661). [cited 2022, May 5]. Available from https://scholia.toolforge.org/venue/Q97011661, accessed September 3, 2022.

[43] Wikidata query service. [Author collaborations in [B] Public Policy Review] [cited 2022, May 5]. Available from: https://w.wiki/4wQG, accessed September 3, 2022.