Colombian Libraries Directory

A growing dataset of the Colombian GLAM sector, using Linked Open Data (LOD) technology

Martha Elena Muñoz Diaz David Ramírez-Ordóñez

Abstract

This project aims to create a directory of Colombian libraries with linked open data in Wikidata, to help the Colombian library community recognize the potential of centralized structured data storage that supports information systems, whether internal or external. We present an approach strategy to WIKIMEDIA communities and tools, through uploading data to WIKIDATA from sites with open data about libraries in Colombia. In this way, any user can satisfy their interest in finding, identifying, selecting, obtaining, navigating, and exploring data about Colombian libraries. We also seek, through the participation of those involved in a community of practice, to recognize the FAIR principles—Findable, Accessible, Interoperable, and Reusable—of data. In Colombia, we have found that the partial adoption of current standards in bibliographic information systems generates incomplete records, deteriorating the process of querying and visualizing data. We hope to immerse Colombian librarians in WIKIDATA so that they incorporate current linked open data standards into their work, which will optimize information services by meeting user interests.

Keywords: data quality, information service functions, Linked Open Data - LOD, data standards, WIKIDATA

Introduction

The Colombian Government is implementing data publication strategies without meeting basic standards, such as ensuring data has licenses that allow reuse instead of using total copyright protection. On the other hand, the library community increasingly uses tools or systems that benefit from their data and confine them to walled gardens. Through the state of services, which will be identified through the Colombian libraries directory in WIKIDATA, the free knowledge base of Wikimedia (Proffitt, 2018), we will corroborate the need to train the library sector in new information technologies offered by linked open data and highlight the dangers of working with proprietary software that limits access to knowledge and restricts the application of the public domain concept. The publishing sector is requiring libraries to invest large amounts of money in annual updates and improvements paid through subscription, taking knowledge to platforms that are designed to limit access to the knowledge hosted on them without prior payment, and also extract data about their uses, which allows publishers to know search interests and generate enormous profits to offer access (Chan, 2019). They have monetized most of the access keys to databases (Brown, 2021). Libraries, often due to lack of knowledge about the potential and development of platforms to make knowledge available, pay for each type of persistent identifiers (PID), which forces them to allocate a significant percentage of their budget to maintain subscriptions to software marketing companies

and PID providers where they host their catalogs and/or collections, or to identify their items (Publishers, 2017). Additionally, copyright licenses limit the bibliographic loan service of new acquisitions of printed works made by libraries.

Librarians share with WIKIMEDIA FOUNDATION that knowledge should be recognized as a common good. Therefore, we see it necessary to make a common front to teach the concept of public domain, the advantages of using and promoting open source, and support all dissemination campaigns on the need to open data, recognize that standards are non-negotiable, that if working with linked open data, input costs are lower as well as facilitating accurate information retrieval. The application of linked open data (LOD) is essential to guarantee the rediscovery of information and eliminate disambiguation problems. Thus, authority data should be managed in tools like WIKIDATA. Colombia does not have a collective state project to keep a standardized and public record for improving the identification of resources and people, a situation that proprietary software owners take advantage of to extract and use the data hosted on their platforms, increasing their power over their clients, forcing them to renew their product subscriptions, also generating biases in the way of categorizing knowledge because only what these companies have commercialized will be registered. This point is critical in regions of Colombia that are underrepresented and have valuable collections that cannot be adequately described. Either because they do not have the money to acquire the software, or they do not have trained people, or on these platforms have no possibilities to make improvements to properly register their collections without prior payment for improvements in the versions of the software where the data is registered. WIKIMEDIA communities and tools are a solution, and WIKIDATA in particular, as a tool where all perspectives can be registered, is constantly growing, multilingual, and continuously improving, facilitates data registration and links to properties to ensure precision without raising costs but improving access and navigation in consultation by collections and/or authors.

Dates: We could start on July 1, 2025, and finish on June 30, 2026.

Related work

Similar research or projects to bring the Latin American library community closer to the concept of linked open data (LOD) are the workshops held by AGRIS (Subirats, 2014) and the one conducted by Desiree Rodríguez Calvo at the National Library of Costa Rica (Rodríguez Calvo, 2023). They have not worked directly with WIKIDATA but with the conceptualization and implementation of the LOD concept for bibliographic data.

On the other hand, we worked on infographics to map the access to libraries and bookstores in Colombia, noticing challenges to share linked open data from the Colombian government, taking data available on PDF or images and using it to explain to decision makers what does the copyright update means in Colombia (Fundación Karisma et al., 2014). The website of the National Cultural Information System of the Ministry of Culture of Colombia SINIC reports 1563 libraries (Ministerio de las Culturas, las Artes y los Saberes, República de Colombia, 2025). They share quantitative data but do not authorize its reuse. In contrast, the Colombian government's open data portal has about 298 libraries, not counting the satellites of the National Library of Colombia, the Bank of the Republic, and the National Library System of the National University of Colombia, but without updating or standardizing the data for each region (Ministerio de Tecnologías de la

Información y las Comunicaciones, República de Colombia, n.d.). With this panorama, we can move forward to demonstrate how necessary it is for any researcher or person interested in consulting sites that jointly offer either quantitative and/or qualitative data, which must be published under current international standards, with usage licenses in accordance with the provisions of the national government so that the general public can consult on the dimensions of services and their characterizations.

Models to enter library metadata in countries such as Spain use a set of properties and data from public portals (Obregón-Sierra, 2022), where data from 7861 libraries have been uploaded to Wikidata. In the same way, projects mapping the Colombian public domain presents where you can access the Colombian heritage and which institutions hold the works in the public domain (Fundación Conector, 2024).

The Colombian library field had participated in communities of practice, to learn about copyright and libraries and discussing the proper implementation for a law update (Ramírez-Ordóñez, 2014). We want to use the communities of practice theory to close the gap between librarians and Wikimedians.

Methods

The methodology we will follow for data collection will be qualitative exploratory through content analysis of the different web portals that publish data about libraries in Colombia. For dissemination, we will base on the theory of communities of practice.

For data ingestion, we will propose webquests that motivate workshop participants to search the web for the information we need to register to fill in the information gaps in WIKIDATA, subsequently leading them to register the data and sources that increase the relevant statements to improve the records that allow in said tool to characterize a legal or natural person, a concept, an object, an event, or a place. The result of the group effort will be valued by the community itself through a grid that allows them to recognize the fulfillment of learning objectives and results when conducting queries that make visible the possibilities of discovery and generation of information products with the data registered in WIKIDATA.

To identify which, how many, and where libraries are located in Colombia, we have reviewed Colombia's open data portal, the directory of public libraries of Colombia, and the Colombian National Library Association (ASCOLBI) database. We found there that there are no minimum agreements to upload the data, but it allows us to collect information
to build a directory that can at least have concrete and reliable data to uniquely identify each of the libraries that have published their data on the portal or have been collected by the National Library of Colombia in a project to build a collective catalog for consultation by its users.

The data collected there will be migrated through OPEN REFINE to WIKIDATA, keeping in mind RECOIN's suggestions to describe the data of each institution.

Once we have the data collected in WIKIDATA, we will conduct a persuasion campaign to librarians, inviting them to participate in a workshop where we will present the advantages of WIKIDATA and propose to them to edit the data of the institution they represent to complete what is needed. Work groups will be formed to improve the registered information, and the result of the exercise will be presented in a common presentation of the directory in WIKIDATA.

Through the participation of those involved in a community of practice in Wikimedia Colombia. If there is a continuation in data updating, the project can be sustainable and will have been successful.

This activity will be carried out in stages:

1. Normalization and structuring of the dataset
   • Identification of portals with open data on characterization of Colombian libraries
   • Usage licenses
   • Dataset selection
   • Data modeling
   • Use and editing of controlled vocabularies included in WIKIDATA
   • Recognition of URIs (QID and PID) application of the concept of linked data in each record, property, and statement
2. Data conversion
   • Data ingestion
      • Manual -> RECOIN
         - Batch -> OPENREFINE and QUICKSTAMENTS
   - WIKIDATA queries
   - Conclusions and recommendations

Expected output

- **Data in Wikidata**: We will offer through the various applications developed by the WIKIMEDIA community, specifically in WIKIPEDIA, WIKIDATA, and WIKICOMMONS, the consultation of the data of each participating institution. This will allow the library sector to demonstrate that the more refined and precise the record is, the greater the possibility of consulting, making visible, and using the collections it is offering in its information services, and that participating in cooperative projects will reduce maintenance costs and will be able to invest with others in improving their data until achieving their positioning as entities in charge of the custody of heritage assets and quality and reliable information services.
- **Wikiproject with data visualizations:** A detailed record of the activity carried out so that the library community has a source of consultation on how to continue contributing data. Recommendations will also be offered to take into account with the intention of continuing to advance in the construction and strengthening of the spanish-speaking library community that is aligned with the principle of offering knowledge for consumption, in an open environment. Also seeking to offer examples that encourage other groups of organizations that work with knowledge such as archivists and museographers.
- **Virtual workshop**: We will hold four virtual workshops of two hours per session where we will explain the process of data collection, ingestion on the platform, either manually or using free software tools, and the options for consultation and download of their data.
- **Scientific article:** We will produce an article on the experience and findings we have detected, such as the structure of the data, the description standards, the value vocabularies used, the typology of files, the licenses used vs. the recommended ones, and the updating of the data that are shared in the data portals to characterize the libraries. We hope to publish it in open access journals that wellcome library science research developments.
- **Events :** Every two years, the National Library of Colombia prepares a seminar

aimed at Colombian heritage libraries, where we hope that in this 2025 they will welcome our initiative to be presented in said community, which will allow us to receive feedback and we hope to receive requests to work in the community of practice that we will announce would be held in 2026.

Risks

The risk we see is the scarce participation of areas far from large cities due to the intermittence of connectivity, the precarious infrastructure conditions to house the collections, and the scarce competent personnel to respond in a timely manner to this call, for which we will have to propose subsequent approaches, with prior support from the National Library of Colombia and through their communication channels. With this, we hope to combine institutional efforts to offer in a second stage the workshops that, being conceived asynchronously but with the support of online teaching assistants, we will manage to raise the level of training and competence of those who work with goodwill but without the adequate infrastructure to advance in the construction of structured data on open source platforms, using new technologies and current standards.

Community impact plan

The Wikimedia community, upon learning about the efforts made to reach underrepresented communities, can join to detect gaps and errors in the records that can be registered in lists that will be distributed and assigned according to their complexity to improve the data and information registered in WIKIMEDIA products in this effort to advance in supporting volunteers who want to populate reliable data about Colombia.

Evaluation

We hope that the results will serve for Wikimedia Colombia to implement its 2025 - 2027 action plan.

With this project, we seek to produce linked open data sets, facilitating training, evaluation, and comparison by the Colombian library community. In two public seminars, we will seek to connect the Colombian library community with the Wikimedia community, presenting the developed models and integrating feedback.

The workshops will promote dialogue, facilitate discussion of findings, and also seek to internalize in participants the potential of working together to advance in strengthening their data. The culmination of these efforts will be an academic publication.

Activity reports will be carried on the Outreach dashboards that WIKIMEDIA proposes to know the participation index of each person in the platforms for monitoring projects endorsed by WIKIMEDIA.

Budget

Project budget

References

Brown, J., Jones, P., Meadows, A., Murphy, F., & Clayton, P. (2021). *UK PID Consortium: Cost-Benefit Analysis* (Versión 1.0). Zenodo. https://doi.org/10.5281/ZENODO.4772626

Chan, L. (2019). *Platform Capitalism and the Governance of Knowledge Infrastructure*. https://doi.org/10.5281/zenodo.2656601

Fundación Conector. (2024). *Visibilizando el dominio público colombiano en Wikidata* [Wiki]. Wikimedia Colombia. https://co.wikimedia.org/wiki/Dominio_p%C3%BAblico/Proyecto

Fundación Karisma, Fundación Conector, & ASCOLBI. (2014). *Where the books are in Colombia (and another facts)*. https://karisma.org.co/wp-content/uploads/2014/07/Where-the-books-are-in-Colombia.pdf

Gili Sampol, E. (2019). *Las normas RDA según el modelo IFLA-LRM*.

Ministerio de las Culturas, las Artes y los Saberes, República de Colombia. (2025). *Sinic* [Portal]. Sistema Nacional de Información Cultural. https://www.sinic.gov.co/

Ministerio de Tecnologías de la Información y las Comunicaciones, República de Colombia. (s. f.). *Datos Abiertos Colombia* [Repositorio digital]. La plataforma de datos abiertos del gobierno colombiano. Recuperado 25 de junio de 2020, de https://www.datos.gov.co/

Obregón-Sierra, Á. (2022). Inserción de metadatos de las bibliotecas españolas en Wikidata: Un modelo de datos abiertos enlazados. *Revista Española de Documentación Científica*, *45*(3), Article 3. https://doi.org/10.3989/redc.2022.3.1870

Park, H., & Kipp, M. (2019). Library Linked Data Models: Library Data in the Semantic Web. *Cataloging & Classification Quarterly*, *57*(5), 261–277. https://doi.org/10.1080/01639374.2019.1641171

Proffitt, M. (Ed.). (2018). *Leveraging Wikipedia: Connecting communities of knowledge*. ALA Editions. https://www.oclc.org/content/dam/research/publications/2018/Proffitt_LeaveragingWikipedia.pdf

Publishers increasingly in control of scholarly infrastructure and this is

why we should care. (2017, septiembre 20). The Knowledge G.A.P. http://knowledgegap.org/index.php/sub-projects/rent-seeking-and-financialization-of-the-academic-publishing-industry/preliminary-findings/

Ramírez-Ordóñez, D. (2014). *Propuesta de formación e implementación en derecho de autor, bibliotecas y normatividad en Colombia: Estudio de casos para aprender haciendo*. http://wiki.nomono.co/index.php?title=Archivo:DerechoAutor-Bibliotecas-NormatividadV1-0.pdf

Riva, P., Le Bœuf, atrick, & Žumer, M. (2017). *IFLA Library Reference Model: A Conceptual Model for Bibliographic Information*. 101.

Rodríguez Calvo, D. (2023). Creación de datos abiertos enlazados en la Biblioteca Nacional de Costa Rica. *Investigación Bibliotecológica: archivonomía, bibliotecología e información*, *37*(97), 75-95. https://doi.org/10.22201/iibi.24488321xe.2023.97.58809

Subirats, I. (2014, julio 8). *AGRIS 2.0: Datos bibliográficos enlazados en el contexto de la información agrícola* [Portal]. Food and Agricultural Organization of the United Nations. https://aims.fao.org/webinars/agris-20-datos-bibliogr%C3%A1ficos-enlazados-en-el-contexto-de-la

Subirats, I., & Zeng, M. L. (2020). *Linked Open Data Enabled Bibliographical Data (LODE-BD) 3.0: A practical guide on how to select appropriate encoding strategies for producing Linked Open Data Enabled Bibliographical Data*. FAO. https://doi.org/10.4060/cb2209en