

CONNECTING BIBLIOGRAPHIC DATA TO THE WEB: *IFLA LIBRARY REFERENCE MODEL AND SCHEMA.ORG*

Dayane Onaga Ferreira Machado

Master's of Information Science. Brazilian Institute of Information in Science and Technology (Ibict), Brasília, Brazil.

onagadayane@gmail.com.

<https://orcid.org/0000-0001-5772-418X>

Ana Carolina Simionato Arakaki

Postdoctoral research in Computer Science – University of São Paulo (USP), São Paulo, Brazil. University of Brasília (UnB), Brasília, Brazil.

anaarakaki@ibict.br.

<https://orcid.org/0000-0002-0140-9110>.

ABSTRACT

Objective: Libraries provide relevant and reliable information through digital catalogs. However, in traditional catalogs, data are neither connected nor shared on the Web. Thus, it is necessary to modernize them by using conceptual models, such as the IFLA Library Reference Model (LRM), and semantic formats for interoperability on the Web. Therefore, this research aims to discuss catalog modeling guided by the LRM and the use of Schema.org. **Method:** This study employs a qualitative approach and, in terms of objectives, is exploratory and descriptive. **Conclusions:** The LRM guides modeling by considering alignment with standards within the scope of Linked Open Data. In this context, Schema.org is a viable semantic vocabulary that can be applied to the bibliographic domain, enabling the integration of catalog data with the Web.

Keywords: Digital Catalogs. IFLA Library Reference Model. Schema.org. Semantic enrichment.

1 INTRODUCTION

In the context of libraries, digital catalogs are important tools for the search and retrieval of information resources, standing out as reliable and relevant sources. However, the methods used to organize this information are outdated in light of new technologies and the way information is currently made available, especially on the Web.

In general, relationships are not established between bibliographic records in catalogs, nor with external sources. In this regard, Serra and Santarém Segundo (2021) highlight that by replicating or creating records for their catalogs, institutions end up forming a data silo. The authors observe that the absence of connections between the collection items themselves exacerbates this problem, generating “silos within the silo” (Serra; Santarém Segundo, 2021).

Certainly, this scenario is unfavorable for libraries, given the significant effort invested in the processing of information resources to create high-quality metadata. In light of this, to meet the needs and expectations of contemporary users and facilitate the effective use of catalogs, it is essential that these tools be updated and adapted to new demands.

In this context, conceptual models play a fundamental role in the efficient structuring of systems, providing a solid foundation for the organization of complex structures. The IFLA Library Reference Model (LRM) is a high-level conceptual model that aims to guide the formulation of cataloging rules and the implementation of bibliographic systems (Riva; Boëuf; Žumer, 2017). The model provides a comprehensive conceptual framework and proposes to facilitate data reuse, considering applications in Linked Open Data (LOD). Arakaki (2020) points out that the instantiation of the model and the possibility of creating namespaces can facilitate mapping with metadata standards.

From this perspective, the use of semantic vocabularies is essential to enable the integration and reuse of data on the Web. To achieve this goal, one alternative is the use of Schema.org, a semantic vocabulary developed by major search engines. Data structured with this vocabulary are enriched, facilitating interpretation, record retrieval, and connection with other information sources.

Although several studies have addressed different aspects of catalog modernization, there is still a significant gap in the literature regarding the association between the LRM and semantic vocabularies, such as Schema.org. In view of this, adopting a qualitative approach and descriptive and exploratory research, this article aims to discuss bibliographic data modeling guided by the LRM, as well as the use of Schema.org for the modernization of digital catalogs. According to Gil (2008), exploratory research allows for the study of themes from various angles and aspects; this type of research is conducted especially when the chosen themes are under-explored.

This article is organized into five main sections. In addition to this introduction, which presents the context, objectives, and nature of the research, the second section addresses the IFLA Library Reference Model, describing the model's main features. The third section discusses the main aspects of Schema.org. The fourth

section deals with rich results, discussing how structured data can generate enriched outputs. Finally, the fifth section presents the conclusions.

2 IFLA LIBRARY REFERENCE MODEL

The IFLA Library Reference Model (LRM) is a high-level entity-relationship model developed from the Functional Requirements (FR). These include the following conceptual models: Functional Requirements for Bibliographic Records (FRBR), focused on bibliographic records; Functional Requirements for Authority Data (FRAD), specialized in authority records; and Functional Requirements for Subject Authority Data (FRSAD), for subject records (Riva; Bœuf; Žumer, 2017). Since these models originate from distinct viewpoints, relevant aspects were harmonized in the LRM, resulting in a more cohesive and comprehensive conceptual framework.

According to Riva, Bœuf, and Žumer (2017), the functional scope of the IFLA LRM is based on generic tasks performed by end-users and intermediaries, namely: Find, Identify, Select, Obtain, and Explore. The 'Find' task refers to meeting search criteria; 'Identify' concerns the ability to distinguish between similar resources; 'Select' refers to the suitability of the found resources regarding user preferences; 'Obtain' deals with accessing the resource; and, finally, 'Explore' concerns the discovery of resources based on the context and the relationships between them.

The main elements comprising the IFLA LRM are entities, attributes, and relationships. Entities are characterized as classes, attributes provide meaning to entity instances, and relationships represent the connection between them, providing context (Riva; Bœuf; Žumer, 2017). Table 1 presents the IFLA LRM entities, their hierarchies, and definitions.

Table 1-IFLA LRM entities, hierarchy, and definitions

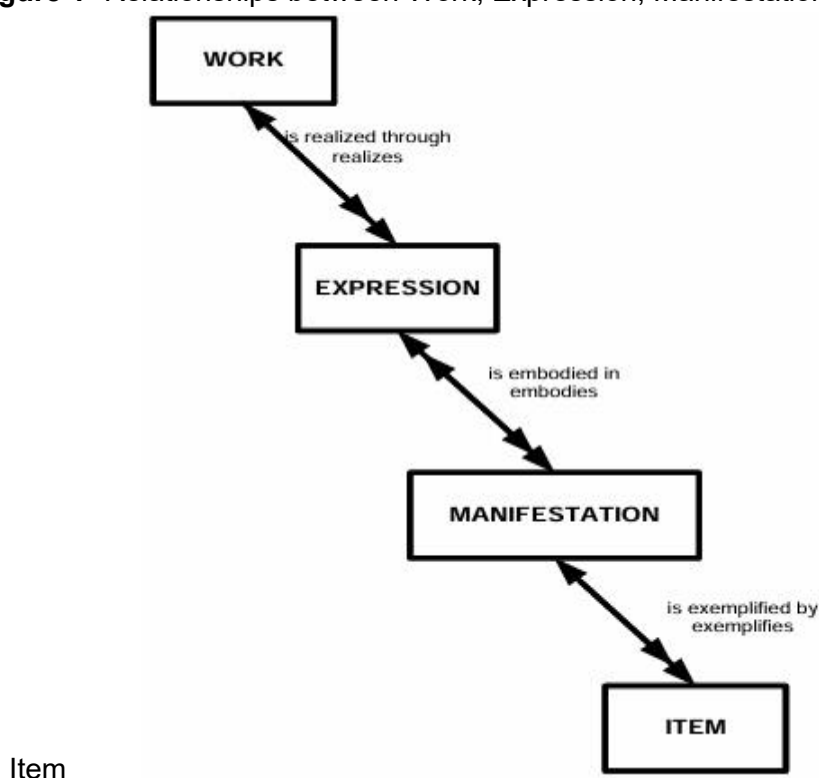
Entity	Hierarchy	Definition
LRM-E1: Res	Top Level	Any entity in the universe of discourse.
LRM-E2: Work	Second Level	The intellectual or artistic content of a distinct creation.

LRM-E3: Expression	Second Level	A distinct combination of signs conveying intellectual or artistic content.
LRM-E4: Manifestation	Second Level	A set of all carriers that are assumed to share the same characteristics as to intellectual or artistic content and aspects of physical form.
LRM-E5: Item	Second Level	An object or objects carrying signs intended to convey intellectual or artistic content.
LRM-E6: Agent	Second Level	An entity capable of deliberate actions, of being granted rights, and of being held accountable for its actions.
LRM-E7: Person	Third Level	An individual human being.
LRM-E8: Collective Agent	Third Level	A gathering or organization of persons bearing a particular name and capable of acting as a unit.
LRM-E9: Nomen	Second Level	An association between an entity and a designation that refers to it.
LRM-E10: Place	Second Level	A given extent of space.
LRM-E11: Time-span	Second Level	A temporal extent having a beginning, an end and a duration.

Source: Riva, Bœuf and Žumer (2017).

Table 1 presents all the entities of the model. The top-level entity is Res, which has the following subclasses: Work, Expression, Manifestation, Item, Agent, Person, Collective Agent, Nomen, Place, and Time-span. The relationships between the main entities of the model are illustrated in Figure 1.

Figure 1 -Relationships between Work, Expression, Manifestation, and



Source: Riva; Bœuf and Žumer (2017, p. 83).

As illustrated in the figure above, cardinality is indicated by the arrowheads; a single arrowhead signifies a cardinality of one (1), while a double-headed arrow indicates a cardinality of many (M). Thus, it is understood that a Work is realized through one or more Expressions; an Expression is the realization of only one Work; an Expression can be embodied in more than one Manifestation; and a Manifestation can be exemplified by one or more Items, which represent only one Manifestation. In other words, a Work is the intellectual or artistic creation idealized by an individual or organization; an Expression is the realization of this Work, which can be expressed in different formats and languages; and a Manifestation is the materialization of the Expression into one or more Items, in physical or digital format, for example.

The IFLA LRM was designed for the context of the Semantic Web, aiming to provide meaning to data and facilitate interpretation by computational systems. Thus, the model enables the implementation of more efficient bibliographic systems, reducing ambiguity and allowing for the assignment of namespaces for Linked Open Data (LOD) applications (Riva; Bœuf; Žumer, 2017).

In addition to data linking, it is essential to consider the discovery of such data on the Web. The Statement of International Cataloguing Principles (ICP) highlights, in principle 2.10, the use of vocabularies that facilitate automatic translation and disambiguation to promote data exchange and discovery tools (Galeffi et al., 2016).

Thus, connecting data and using semantic vocabularies provide context to the description process and create connections that can lead to serendipity, while simultaneously streamlining retrieval. To this end, standardizing data according to vocabularies recognized by search engines is essential. From this perspective, the next section discusses Schema.org, a semantic vocabulary that promotes data discovery on the Web.

3 SCHEMA.ORG

Schema.org was developed by the search engines Google, Microsoft, Yahoo, and Yandex. It is a collaborative, community-based initiative aimed at creating, maintaining, and providing a set of schemas for structured data, with the goal of optimizing information interpretation and retrieval by Web search engines (Schema.org, 2024a). Currently, the vocabulary is in version 28.0 and has been widely adopted across the Web. As indicated by the official Schema.org website, it is estimated that in 2024, more than 45 million pages have adopted the set of schemas. According to Schema.org (2024a), the adoption of a common vocabulary facilitates schema selection and the achievement of optimal results for webmasters and developers.

The vocabulary is composed of a set of 'types' or 'classes', associated with a set of properties. Figure 2 illustrates a part of the main hierarchy, presenting the subclasses of the Thing and CreativeWork classes.

Figure 2 -Main hierarchy of Schema.org



Source: Schema.org (2024c).

These elements are organized hierarchically, with Thing being the top-level class, which includes the subclasses: Action, BioChemEntity, CreativeWork, Event, Intangible, MedicalEntity, Organization, Person, Place, Product, and Taxon. All classes inherit properties from the superclass and possess their own specific subclasses and properties. There is also a parallel hierarchy of data types, which defines the types of values that can be specified for properties. Figure 3 illustrates the hierarchy for data types.

Figure 3- Hierarchy of data types

DataTypes:



Source: Schema.org (2024c).

The hierarchy includes the following data types: Boolean, with the values false and true, expressing two types of conditions or states; Date, which includes dates in the ISO 8601 format (year-month-day); DateTime, representing the combination of date and time; Number, for numeric values, including specific types such as Float and Integer; Text, representing values in text format; and Time, used to express specific times.

Hitzler (2021) highlights that Schema.org, as well as ontology modeling applications and Wikidata, emerged within the context of the Semantic Web, aiming to facilitate data sharing, discovery, integration, and reuse. Thus, Schema.org can be considered one of the examples of results stemming from various studies on the realization of the Semantic Web. The vocabulary is compatible with semantic formats recommended by the World Wide Web Consortium (W3C), such as Resource Description Framework in Attributes (RDFa), Microdata, and JavaScript Object Notation for Linking Data (JSON-LD). The use of these formats enables search engines to better understand the content and context of the information, providing more satisfactory search experiences.

4 RESULTS

Structured data are fundamental for search engines to crawl, interpret, and classify content available on the Web. Google, for example, recommends the use of Schema.org. Many of the vocabulary elements are not essential for Google Search, but may be useful for other search engines, services, and platforms (Google Search Central, 2024b).

Schema.org facilitates the integration and opening of data on the Web, allowing, for example, the inclusion of structured data in the Knowledge Graph. This database organizes and relates facts about entities, using information from reliable sources (Google, 2024). Thus, data can be retrieved as enriched fragments, called 'rich snippets'; that is, they can be retrieved as enriched results with additional information and visual elements. This approach not only improves the quality of information but also provides a richer and more informative experience, offering the user relevant contextual information, such as historical data, biographies, and connections between themes.

In this context, Wallis (2022) highlights that Schema.org is an open Linked Data vocabulary for the Web, based on the Resource Description Framework (RDF), which facilitates interoperability and the integration of structured data across different platforms. To ensure data are retrieved by search engines, they can be embedded directly into a page's HTML code using JSON-LD, RDFa, or Microdata formats. The JSON-LD format is the most recommended, as it separates structured data from the HTML, facilitating maintenance and development. In contrast, RDFa and Microdata integrate information directly into the HTML code via specific attributes (Google Search Central, 2024b).

As a generic vocabulary, Schema.org can be used in diverse contexts, as evidenced by its application in distinct platforms: Pinterest applies tags to implement 'Rich Pins' (Pinterest, 2024); Facebook enriches shared links (Meta, 2024); and Amazon employs structured data to interpret and respond to user queries via Alexa (Amazon, 2023). In the academic context, recent research has explored the vocabulary for representing research data (Wu et al., 2022; Payne; Verhey, 2022;

Alrashed et al., 2021; Silva, 2020), in the area of e-commerce (Belz, 2022; Kejriwal et al., 2021; Peeters et al., 2020), and for social networks (Jones et al., 2021; Tomoyse; Santos; Simionato, 2019; García; Gil, 2019), among other fields of study.

In the bibliographic domain, Schema.org has been applied for the semantic enrichment of catalog data, as evidenced by the studies of Wallis (2023) and Godby (2016). Adapting the data model used in catalogs can establish communication with the Web. Opening data can facilitate both integration with other datasets in Knowledge Graphs and the retrieval of bibliographic records directly on the search results page. From this perspective, the next section discusses Schema.org in the context of libraries.

4.1 Results for Bibliographic Resources

To ensure interoperability and the efficient integration of catalog data on the Web, it is essential to adopt an appropriate reference model. The LRM is an internationally recognized model, developed considering data reuse in the context of Linked Open Data and integration with semantic applications. Riva, Bœuf, and Žumer (2017) explain that the current landscape has changed since the original creation of the FRBR model, bringing new demands, especially regarding data reuse in Semantic Web applications (Riva; Bœuf; Žumer, 2017).

The model provides a detailed approach to the representation of various types of resources, establishing connections between entities and attributes within a domain. As Arakaki (2020) highlights, the semantic viability of attributes depends on the definition of the metadata that instantiate the model; that is, metadata will shape the structure proposed by the IFLA LRM, which should serve as a basis for the creation of representation tools, such as standards and cataloging codes.

From this perspective, the Schema.org data model is broad and flexible, assigning meaning to data. The vocabulary's versatility enables its application in distinct contexts. Thus, if applied to catalog data, records can be retrieved in response to a Google search, for example. Furthermore, the results may display purchase or borrowing options. Structured data aligned with the Schema.org vocabulary enable the retrieval of results similar to those presented in Figure 4.

As shown in Figure 4, a search was performed for the book *Wuthering Heights* by the author Emily Brontë, and the structure presented in the image was displayed directly on the results page. The available options include: overview, characters, reviews, more information about the author, videos, and the option to buy the book. Additionally, the content description, reviews, store suggestions for purchase, and the option to borrow from a library are displayed. The 'borrow' option is activated by selecting 'buy book', providing a field (highlighted in red in the figure) for the user to enter their ZIP code and check if the item is available at a nearby library.

These features can be enabled by following Google guidelines. In the case of the borrowing option, it is necessary to create a data structure called a 'feed', in JSON-LD, Microdata, or RDFa formats, and use the properties defined by Schema.org (Google Search Central, 2024a). Figure 5 illustrates an example of a JSON-LD feed for 'BorrowAction', which enables the borrowing option.

Figure 5 -BorrowActionfeed

```
{
  "@context": "https://schema.org",
  "@type": "DataFeed",
  "dataFeedElement": [
    {
      "@context": "https://schema.org",
      "@type": "Book",
      "@id": "https://example.com/work/the_catcher_in_the_rye",
      "url": "https://example.com/work/the_catcher_in_the_rye",
      "name": "The Catcher in the Rye",
      "author": {
        "@type": "Person",
        "name": "J.D. Salinger"
      },
      "sameAs": "https://en.wikipedia.org/wiki/The_Catcher_in_the_Rye",
      "workExample": [
        {
          "@type": "Book",
          "@id": "https://example.com/edition/the_catcher_in_the_rye_paperback",
          "isbn": "9787543321724",
          "bookEdition": "Mass Market Paperback",
          "bookFormat": "https://schema.org/Paperback",
          "inLanguage": "en",

```

Source:Google Search Central (2024a)

Book-related actions enable users to discover authors and books and find where they can access them, whether through purchase or borrowing. The

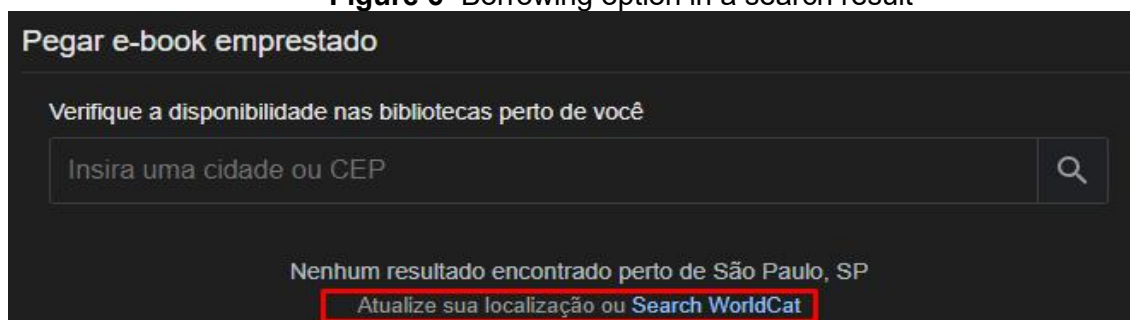
'ReadAction' specification, from Schema.org, enables the purchase option, while 'BorrowAction' enables borrowing (Google Search Central, 2024a). In addition to these specifications, it is necessary to create a feed with the indicated properties for the Library entity and another for the Book entity.

In the context of search engines and Schema.org, a “work” is considered a tangible object that can be accessed and acquired by the user. As described in Schema.org (2024b), “[...] something may be simultaneously a Book and a Product and be usefully described with properties from both types.”

In the library domain, Schema.org was implemented in the WorldCat catalog, managed by the Online Computer Library Center (OCLC). In this context, the vocabulary was aligned with the conceptualizations of FRBR. As described by Godby (2016), OCLC experts considered structuring the CreativeWork class into RDF triples to build a semantic model for WorldCat data. The class includes metadata for the description of books, articles, scholarly works, paintings, sculptures, manuals, photographs, musical recordings, films, series, games, among others.

It is possible to visualize the results of OCLC's work in practice. Catalog records can appear directly on the search results page, enriched with additional information. Since results are dynamic, the manner in which they are displayed may vary depending on the time of the search. Figure 6 presents a snapshot of the same display format presented previously in Figure 4, and includes the recommendation (highlighted in red) for the user to access WorldCat.

Figure 6 -Borrowing option in a search result



Source: Google (2024).

As shown in Figure 6, upon entering "São Paulo, SP" in the search field, a message is displayed stating that no results were found for the searched resource.

However, the message suggests updating the ZIP code or searching for the book directly on WorldCat. Thus, even if the search does not return results originating from the catalog, the user is directed to it, allowing them to explore the collections and discover other resources of interest.

Still regarding the application of Schema.org in the context of libraries, a more recent initiative presented at the 2023 LD4 Conference on Linked Data is also noteworthy. At this conference, Richard Wallis, one of the main contributors to Schema.org, exposed the data structuring project developed for the National Library Board Singapore. During his presentation, Wallis addressed how linked data were implemented using Bibframe at the National Library Board (NLB) – a government institution whose mission is to promote reading, develop information literacy, and stimulate knowledge about Singapore's history and cultural heritage (National Library Board, 2024).

According to Wallis (2023), Schema.org played a fundamental role in the development of the data model in the NLB project. The adopted strategy involved aligning data in MARC-XML, DC-XML, and CSV with Bibframe, allowing for the preservation of both the integrity and detail of bibliographic records. In parallel, Schema.org was implemented for entity description, which optimized their discovery and interpretation by search engines. The author reports that the obtained results demonstrated the success in transforming and integrating NLB data into a structured format, making them suitable for analysis and exploration in Knowledge Graphs on the Web (Wallis, 2023).

When a domain is modeled based on appropriate reference models, such as the LRM or FRBR, and structured using semantic formats and vocabularies like Schema.org, it enables the integration of library data into rich search results. By accessing these structures, search engines can interpret the context and meaning of the data, making information retrieval more precise and relevant. Thus, libraries can connect their collections to a broader network.

The LRM presents advancements over FRBR, making it more aligned with the current requirements of open data and semantic applications. This characteristic of the model can facilitate mapping with the Schema.org vocabulary.

5 CONCLUSIONS

The LRM provides the basis for a consistent conceptual framework, offering context to bibliographic data and enabling the establishment of relationships between them. Beyond promoting these connections, it is essential that data be structured according to recommended Web formats and vocabularies, aiming to open up catalogs and integrate them with other data. In this way, resources can be found by users with greater ease.

It is posited that catalog modeling based on a conceptual model like the LRM, in alignment with semantic vocabularies such as Schema.org, can establish a comprehensive and flexible structure for the representation and sharing of catalog data on the Web. Thus, while the LRM promotes a consistent conceptual framework, Schema.org enables the integration and discovery of such data by search engines.

The results observed regarding OCLC's work demonstrate the effectiveness of applying semantic technologies in the library context. Therefore, these examples can inspire and guide the modernization of services and information processing in Brazilian libraries, highlighting the importance of adopting innovative approaches to enhance the connection, openness, and discovery of records on the Web.

Thus, this article presented an initial overview of the LRM and Schema.org topics. The research conducted provides a foundation for future studies, enabling in-depth investigation into how Schema.org metadata can be effectively integrated and mapped according to LRM guidelines.

REFERENCES

AMAZON. **CreativeWork Intents**. 2023. Available at: <https://developer.amazon.com/ja-JP/docs/alexa/custom-skills/creativework-intents.html>. Accessed on: July 11, 2024.

ALRASHED, Tarfah. *et al.* Dataset or not? A study on the veracity of semantic markup for dataset pages. *In: INTERNATIONAL SEMANTIC WEB CONFERENCE*, 20., 2021, [s.l.]. **Proceedings [...]**. [S.l.]: IBM, 2021. p. 2-17. Available at: <http://people.csail.mit.edu/tarfah/papers/dataset.pdf>. Accessed on: December 12, 2024.

ARAKAKI, Ana Carolina Simionato. O Modelo IFLA Library Reference Model e o Linked Data. **Informação & Informação**, Londrina, v. 25, n. 3, p. 163–186, 2020. Available at: <http://dx.doi.org/10.5433/1981-8920.2020v25n3p163>. Accessed on: August 7, 2024.

BELZ, Olexandra. Use of Schema.org micro-markup in e-commerce projects. **Three Seas Economic Journal**, Riga, v. 3, n. 4, 2022. Available at: <http://baltijapublishing.lv/index.php/threeseas/article/view/1964>. Accessed on: December 12, 2024.

GALEFFI, Agnese et al. **Statement of International Cataloguing Principles (ICP)**. 2016. Edition with minor revisions, 2017. Den Haag: IFLA, 2017. Available at: <https://www.ifla.org/g/cataloguing/international-cataloguing-principles-icp>. Accessed on: July 8, 2024.

GARCÍA, Roberto; GIL, Rosa. Social media copyright management using Semantic Web and Blockchain. *In: INTERNATIONAL CONFERENCE ON INFORMATION INTEGRATION AND WEB-BASED APPLICATIONS & SERVICES*, 21., 2019, Munich. **Proceedings [...]**. Munich: Association for Computing Machinery, 2019. p. 339-343. Available at: <https://dl.acm.org/doi/10.1145/3366030.3366128>. DOI 10.1145/3366030.3366128. Accessed on: December 12, 2024.

GIL, Antonio Carlos. **Métodos e técnicas de pesquisa social**. 6. ed. São Paulo: Atlas, 2008.

GODBY, Carol Jean. A division of labor: the role of Schema.org in a Semantic Web model of library resources. *In: GODBY, Carol Jean. **Linked data for cultural heritage***. Chicago: Amer Library Assn Editions, 2016. 152 p. Available at: <https://www.oclc.org/content/dam/research/publications/2017/godby-division-of-labor-2017.pdf>. Accessed on: January 21, 2024.

GOOGLE. **How Google's Knowledge Graph works**. 2024. Available at: <https://support.google.com/knowledgepanel/answer/9787176?hl=en>. Accessed on: July 20, 2024.

GOOGLE SEARCH CENTRAL. **Book actions (Book) structured data**. 2024a. Available at: <https://developers.google.com/search/docs/appearance/structured-data/book?hl=pt-br#create-your-feed>. Accessed on: July 12, 2024.

GOOGLE SEARCH CENTRAL. **Introduction to structured data markup in google search**. 2024b. Available at: <https://developers.google.com/search/docs/appearance/structured-data/intro-structured-data>. Accessed on: October 14, 2024.

HITZLER, Pascal. A review of the semantic web field. **Communications of the ACM**, Nova York, v. 64, n. 2, 2021. Available at: <https://dl.acm.org/doi/abs/10.1145/3397512>. Accessed on: August 9, 2024.

JONES, Shawn, M. *et al.* It's all about the cards: sharing on social mediaproably encouraged HTML metadata growth. **arXiv**, Nova York, 2021. Available at: <https://arxiv.org/abs/2104.04116>. Accessed on: December 12, 2024.

KEJRIWAL, Mayank. *et al.* Empirical best practices on using product-specific Schema.org. *In*: AAAI CONFERENCE ON ARTIFICIAL INTELLIGENCE, 35., 2021, Palo Alto. **Proceedings [...]**. Palo Alto: Association for the Advancement of Artificial Intelligence, 2021. p. 15452-15457. Available at: <https://ojs.aaai.org/index.php/AAAI/article/view/17816>. Accessed on: December 12, 2024.

META. **Tags de microdados**. 2024. Available at: <https://developers.facebook.com/docs/marketing-api/catalog/guides/microdata-tags>. Accessed on: July11, 2024.

NATIONAL LIBRARY BOARD. **About us**. 2024. Available at: <https://www.nlb.gov.sg/main/about-us>. Accessed on: July11, 2024.

PAYNE, Karen; VERHEY, Chantelle. Schema.org for research data managers: a primer. **International Journal of Big Data Management**, Genebra, v. 2, n. 2, 2022. Available at: <https://www.inderscienceonline.com/doi/10.1504/IJBDM.2022.128449>. Accessed on: December12, 2024.

PEETERS, Ralph. *et al.* Using Schema.org annotations for training and maintaining product matchers. *In*: INTERNATIONAL CONFERENCE ON WEB INTELLIGENCE, 10., 2020, Biarritz. **Proceedings [...]**. Biarritz: Association for Computing Machinery, 2020. p. 195-204. Available at: <https://dl.acm.org/doi/10.1145/3405962.3405964>. Accessed on: December 12, 2024.

PINTEREST. **Createrich Pins**. 2024. Available at: <https://help.pinterest.com/en/business/article/rich-pins>. Accessed on: July11, 2024.

RIVA, Pat; LE BOEUF, Patrick; ŽUMER, Maja. **IFLA Library Reference Model: a conceptual model for bibliographic Information**. Netherlands: International Federation of Library Associations and Institutions, 2017. Available at: https://www.ifla.org/files/assets/cataloguing/frbr-lrm/ifla-lrm-august-2017_rev201712.pdf. Accessed on: July08, 2024.

SERRA, Liliane Giust; SANTARÉM SEGUNDO. Dos silos de dados à web dos dados: bibliotecas e o linked data. **Informação & Informação**, Londrina, v. 26, n. 2, p. 625-645, abr./jun. 2021. Available at: <https://ojs.uel.br/revistas/uel/index.php/informacao/article/view/41603>. Accessed on: October 8, 2024.

SCHEMA.ORG. **Schema.org**. 2024a. Available at: <https://schema.org>. Accessed on: July 10, 2024.

SCHEMA.ORG. **Data model**. 2024b. Available at: <https://schema.org/docs/datamodel.html>. Accessed on: July14, 2024.

SCHEMA.ORG. **Full hierarchy**. 2024c. Available at:
<https://Schema.org/docs/full.html>. Accessed on: August 17, 2024.

SILVA, L. C. da. **Publicação de dados de pesquisa científica**: proposta de estruturação semântica de cadernos abertos de pesquisa frente às dimensões da e-Science. 2020. 243 f. Tese (Doutorado em Ciência da Informação) - Faculdade de Filosofia e Ciências, Universidade Estadual Paulista Júlio de Mesquita Filho, Marília, 2020. Available at: <https://repositorio.unesp.br/handle/11449/194341>. Accessed on: December 12, 2024.

TOMOYOSE, Kazumi; SANTOS, Amanda Azevedo dos; ARAKAKI, Ana Carolina Simionato. Schema.org para recuperação da informação em redes sociais. *In*: BRAGATO, T. H. B.; TOGNOLI, N. B. (org.). **Organização do conhecimento responsável**: promovendo sociedades democráticas e inclusivas. Belém: ISKO-Brasil; UFPA, 2019, v. 5, p. 176-182. Available at:
<https://brapci.inf.br/index.php/res/v/123290>. Accessed on: December 12, 2024.

WALLIS, Richard. Follow me to the library! bibliographic data in a discovery driven world. **Italian Journal of Library, Archives and Information Science**, Macerata, v. 13, n. 1, p. 38-44, jan. 2022. Available at:
<https://www.jlis.it/index.php/jlis/article/view/417>. Accessed on: December 12, 2024.

WU, Mingfanget *al*. An Analysis of Crosswalks from Research Data Schemas to Schema.org. **Data Intelligence**, Pequim, v. 5, n. 1, 2023. Available at:
<https://direct.mit.edu/dint/article/5/1/100/113281/An-Analysis-of-Crosswalks-from-Research-Data>. Accessed on: December 12, 2024.

ARTICLE NOTES AND CREDITS

- **Acknowledgements:** The authors would like to thank the Brazilian Institute of Information in Science and Technology (IBICT) and the National Council for Scientific and Technological Development (CNPq).
- **Funding:** This study was funded by the National Council for Scientific and Technological Development (CNPq) and the Brazilian Institute of Information in Science and Technology (IBICT).
- **Conflicts of interest:** Not applicable.
- **Ethical approval:** Not applicable.
- **Data and materials availability:** Not applicable.
- **Manuscript published as a preprint:** The manuscript was originally published as a full paper at the *National Seminar on Cataloging and Technology (SNCat)* in 2024. It subsequently underwent a new double-blind peer review process and received content revisions and updates.
- **Author contributions:**

Contribuição	MACHADO, D. O. F.	ARAKAKI, A. C. S.
Concepção do estudo	X	
Conceitualização	X	X
Metodologia	X	X
Coleta de dados / investigação	X	
Curadoria de dados	X	
Análises dos dados	X	
Discussão dos resultados	X	
Visualização (gráficos, tabelas e outros)	X	
Rascunho original	X	
Revisão e edição final	X	X
Supervisão e administração	X	X
Aquisição de financiamento		X

- **License to Use**

The authors grant *Ciência da Informação Express* – *CIExpress* exclusive rights of first publication, with the work simultaneously licensed under the Creative Commons Attribution License (CC BY) 4.0 International. This license allows third parties to remix, adapt, and build upon the published work, attributing proper credit for authorship and initial publication in this journal.

- **Publisher**

Federal University of Lavras (UFLA).

The ideas expressed in this article are his own responsibility and do not necessarily represent the opinion of the editors or the university

Editor of the communication and scientific dissemination channel *Ciência da*

Informação Express

Nivaldo Calixto Ribeiro, Universidade Federal de Lavras (UFLA).

Linguistics Reviewer

From the authors.

Reference Reviewer

From the authors.

English version

Dayane Onaga Ferreira Machado

• Historical

Received on: 2024-11-14

Accepted on: 2024-12-21

Posted on: 2025-01-20