

DESCRIÇÃO DE ASSUNTOS EM VOCABULÁRIOS DE BIBLIOTECAS SEMÂNTICAS

DESCRIPTION OF SUBJECTS IN SEMANTIC LIBRARY VOCABULARIES

IZADORA OLIVEIRA LINHARES

Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict)

ORCID: <https://orcid.org/0000-0001-5453-9966>

ROR: <https://ror.org/006c42y96>

ANA CAROLINA SIMIONATO ARAKAKI

Universidade de Brasília (UnB)

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

ROR: <https://ror.org/02xfp8v59>

MODALIDADE: RESUMO EXPANDIDO

Resumo: O controle terminológico é essencial para a interoperabilidade de dados abertos conectados. Portanto, objetiva-se analisar a utilização de vocabulários na descrição de assuntos em bibliotecas semânticas. Caracteriza-se como um estudo exploratório, qualitativo, bibliográfico e documental. Observou-se que a Deutsche Nationalbibliothek e a Library of Congress seguem boas práticas de publicação de dados na *Web*, utilizam vocabulários controlados em comum, mas apresentam diferenças no tratamento dos assuntos. Estudos futuros poderão analisar os aspectos levantados em ambientes semânticos especializados.

Palavras-chave: Vocabulários controlados; Bibliotecas semânticas; Descrição de assuntos.

Abstract: Terminology control is essential for the interoperability of linked open data. Therefore, the objective of this study is to analyze

the use of vocabularies in the description of subjects in semantic libraries. This is an exploratory, qualitative, bibliographic, and documentary research. It was observed that the Deutsche Nationalbibliothek and the Library of Congress follow best practices for publishing data on the Web, use common controlled vocabularies, but differ in their treatment of subjects. Future studies may analyze the aspects raised in specialized semantic environments.

Keywords: Controlled vocabularies; Semantic libraries; Subject description.



1 INTRODUÇÃO

Os vocabulários atuam no controle terminológico, favorecendo interoperabilidade, descoberta, recuperação e criação de relacionamentos entre os registros bibliográficos e de autoridade de bibliotecas semânticas (Barbosa *et al.*, 2021; Serra; Santarém Segundo, 2017; Solodovnik, 2011). Todavia, considerando bibliotecas brasileiras inseridas no contexto da Web Semântica, observou-se que a adoção de vocabulários controlados carece de estudos que documentem detalhadamente o processo de escolha e aplicação desses instrumentos. Sendo assim, pretende-se analisar como bibliotecas semânticas utilizam vocabulários na descrição dos assuntos de seus recursos informacionais. Para isso, estabeleceram-se os seguintes objetivos: 1) levantar bibliotecas semânticas; 2) identificar vocabulários adotados nos campos de assunto; e 3) analisar o contexto de utilização dos vocabulários identificados.



2 FUNDAMENTAÇÃO TEÓRICA

Os vocabulários controlados consistem em listas de termos autorizados, usados para indexar e recuperar documentos (Lancaster, 2004; Lopes, 2002). Em se tratando de ambientes semânticos, vocabulários controlados e dados abertos conectados são conceitos indissociáveis. Esses ambientes, conforme explicam Serra e Santarém Segundo (2017, p. 178), são “[...] locais onde os dados estão estruturados de forma que sejam processados por máquinas e que permitam a ligação dos elementos descritivos [...]”. Nesse sentido, dados abertos conectados são dados estruturados, licenciados de forma aberta e disponibilizados na Web Semântica, apoiados por princípios arquiteturais da *web*, padrões de metadados e vocabulários controlados que, em conjunto, proporcionam uma base consistente para a interpretação e a interoperabilidade de dados bibliográficos e de autoridade, conforme recomendações do World Wide Web Consortium (W3C) (Hyland *et al.*, 2013; Isotani; Bittencourt, 2015; Lóscio; Burle; Calegari, 2017).



3 PROCEDIMENTOS METODOLÓGICOS

A pesquisa consiste em um estudo exploratório, de abordagem qualitativa, que utiliza pesquisas bibliográfica e documental e aplica o método dedutivo na análise de dados. O *corpus* bibliográfico foi constituído a partir de revisão integrativa¹, da qual se analisou um recorte específico de publicações sobre o uso de vocabulários nas descrições de recursos informacionais de bibliotecas semânticas. A seleção dos materiais considerou menções a vocabulários ou controle terminológico nos campos de título, palavras-chave e resumo dos materiais, e em descritores extras, atribuídos após leitura preliminar dos documentos da revisão. Para o *corpus* documental, utilizaram-se os *websites* do W3C e das bibliotecas selecionadas. O percurso da pesquisa compreendeu a listagem das bibliotecas semânticas mais citadas nas publicações avaliadas (objetivo específico 1), seguida da verificação, em seus respectivos ambientes, dos vocabulários adotados nos campos de assunto (objetivo específico 2). Por fim, a análise documental (objetivo específico 3) foi realizada por meio das seguintes perguntas (P) norteadoras, elaboradas em alinhamento às recomendações do W3C sobre uso de vocabulários: P1) "O(s) campo(s) de assunto apresentam *links* ou conteúdo textual?"; P2) "Todos os campos de assunto estão preenchidos com termos provenientes de vocabulário(s)?"; P3) "Qual(is) vocabulário(s) foi(ram) adotados na descrição dos assuntos?"; e P4) "Qual a área de cobertura do vocabulário?".

1 A revisão investigou a publicação de dados abertos conectados por bibliotecas, com coleta de artigos, teses, dissertações, livros e capítulos, anais, relatórios técnicos e resumos expandidos, publicados entre 2006 e 2026, em inglês, português, espanhol e italiano, pesquisados nas bases de dados Scopus, Web of Science, Association for Computing Machinery (ACM), Digital Library e Institute of Electrical and Electronics Engineers (IEEE) Xplore e nos repositórios Biblioteca Digital Brasileira de Teses e Dissertações (BDTD), Networked Digital Library of Theses and Dissertations (NDLTD) e Open Access Theses and Dissertations (OATD).



4 RESULTADOS E DISCUSSÕES

Para o levantamento das bibliotecas semânticas, a partir da análise integral dos documentos (Barbosa *et al.*, 2021; Bergamin; Lucarelli, 2013; Carrasco; Candela; Marco-Such, 2025; Miranda; Kutt; Nalepa, 2025; Serra; Santarém Segundo, 2017; Silveira; Martínez-Ávila, 2025; Solodovnik, 2011), foram listados todos os ambientes de dados conectados. Em seguida, foram agrupados os nomes dos ambientes e conferidas as ocorrências. Cada uma das seguintes bibliotecas foram mencionadas apenas uma vez no grupo de documentos: Austrian National Library, Biblioteca Nacional do Brasil, Universiteitsbibliotheek Gent, Hungarian National Library, Jagiellonian Digital Library, Biblioteca Virtual Miguel de Cervantes, National Central Library of Florence, National Library of Medicine, National Library of Finland, Bibliothèque nationale du Luxembourg, National Library of Scotland e University of Michigan Digital Library. As bibliotecas Europeana, Bibliothèque Nationale de France e Koninklijke Bibliotheek foram mencionadas duas vezes, cada uma. As bibliotecas British Library e Biblioteca Nacional de España foram mencionadas três vezes, cada uma. Já as bibliotecas Deutsche Nationalbibliothek (DNB) e Library of Congress (LC) foram citadas quatro vezes, cada uma delas, sendo as bibliotecas com maior número de menções e, conseqüentemente, as selecionadas para análise. Os vocabulários foram identificados a partir do exame dos registros e de informações disponíveis nos *websites* das bibliotecas. Os resultados da análise do contexto de uso dos vocabulários foram sumarizados nos Quadros 1 e 2.

Na DNB (Quadro 1) verificou-se o uso dos vocabulários *LC Subject Headings* (LCSH) e *Répertoire d'autorité-matière encyclopédique et alphabétique unifié* (RAMEAU). No âmbito geral, a biblioteca disponibiliza publicamente dados bibliográficos e de autoridade, em *Resource Description Framework* (RDF), por meio do Serviço de dados conectados (ativo desde 2010) e mantém mapeamentos entre o Arquivo Integrado de Autoridades (GND) e outros vocabulários controlados² (DNB, [2026?]).

2 AGROVOC, MeSH, Standard-Thesaurus Wirtschaft (STW), Thesaurus Sozialwissenschaften (TheSoz), Encabezamientos de materia de la Biblioteca Nacional de España (EMBNE), Nuovo Soggettario (NSogg), Thesaurus der Provenienzbegriffe (TPRO), Europäischer Thesaurus für Internationale Beziehungen und Länderkunde (ET) e Art & Architecture Thesaurus (AAT®) (DNB, [2026?]).



Quadro 1 – DNB: Livro *Energy and Sustainable Development*

P1) Assuntos com links ou conteúdo textual?	P2) Termos provenientes de vocabulário?	P3) Vocabulário(s) adotado(s)	P4) Área de cobertura do(s) vocabulário(s)
<i>Energiepolitik</i> , linkável	Sim	LCSH e RAMEAU	Geral (ambos)
<i>Energieerzeugung</i> , linkável	Idem	Idem	Idem
<i>Nachhaltigkeit</i> , linkável	Idem	Idem	Idem

Fonte: Elaboração própria (2026).

Na LC (Quadro 2) observou-se a utilização dos vocabulários LCSH, *Medical Subject Headings* (MeSH), *Faceted Application of Subject Terminology* (FAST), *LC Children's and Young Adults' Subject Headings* e *LC Genre/Form Terms*. A biblioteca disponibiliza informações sobre controle terminológico por meio do Serviço de dados conectados "ID.LOC.GOV", onde também é possível acessar metadados bibliográficos e de autoridade. O ambiente permite realizar buscas por valores de vocabulários específicos e fazer *download*, em RDF, dos *datasets* (LC, [2026?]).

Quadro 2 – LC: livro *Globalización y derechos humanos*

P1) Assuntos com links ou conteúdo textual?	P2) Termos provenientes de vocabulário?	P3) Vocabulário(s) adotado(s)	P4) Área de cobertura do(s) vocabulário(s)
<i>Human rights</i> , linkável	Sim	<i>LC Children's and Young Adults' Subject Headings</i>	Crianças e jovens adultos, vocabulário independente do LCSH
		MeSH	Área médica
		LCSH	Geral
		FAST	Geral (derivado do LCSH)



P1) Assuntos com <i>links</i> ou conteúdo textual?	P2) Termos provenientes de vocabulário?	P3) Vocabulário(s) adotado(s)	P4) Área de cobertura do(s) vocabulário(s)
Globalization, linkável	Idem	LC Children's and Young Adults' Subject Headings	Idem
		LCSH	Idem
		FAST	Idem
		LC Genre/Form Terms	Dicionário de gênero/forma

Fonte: Elaboração própria (2026).

Observa-se que ambas as bibliotecas estão alinhadas às boas práticas do W3C, fornecendo acesso público aos *datasets*. Quanto à descrição dos assuntos, enquanto na DNB são incluídos estritamente os vocabulários utilizados como fonte para os descritores, na LC são também reunidos termos relacionados. Ademais, a LC apresenta uma maior variedade temática de vocabulários, em comparação à DNB.



5 CONSIDERAÇÕES FINAIS

As bibliotecas semânticas analisadas seguem boas práticas internacionais na disponibilização de dados bibliográficos e de autoridade e, embora utilizem vocabulários em comum (como o LCSH), apresentam diferenças no tratamento dos assuntos. Em pesquisas futuras, poderão ser analisadas amostras maiores de registros e investigada a utilização tanto de vocabulários, como de ontologias, adotados em ambientes semânticos especializados.



REFERÊNCIAS

- BARBOSA, E. R. *et al.* Thesaurus and subject heading lists as Linked Data. **Transinformação**, Campinas, v. 33, p. 1-13, 2021. Disponível em: <https://doi.org/10.1590/2318-0889202133e200077>. Acesso em: 15 abr. 2026.
- BERGAMIN, G.; LUCARELLI, A. The Nuovo soggettario as a service for the linked data world. **JLIS.it**, [Itália], v. 4, n. 1, p. 213-226, 2013. Disponível em: <https://jlis.fupress.net/index.php/jlis/article/view/265>. Acesso em: 27 abr. 2026.
- CARRASCO, R. C.; CANDELA, G.; MARCO-SUCH, M. Measuring the diversity of data and metadata in digital libraries. **Int J Digit Libr**, [s. l.], v. 26, n. 5, p. 1-14, 2025. Disponível em: <https://doi.org/10.1007/s00799-025-00411-1>. Acesso em: 13 abr. 2026.
- DEUTSCHE NATIONALBIBLIOTHEK. **Metadatendienste**. Alemanha: DNB, [2026?]. Disponível em: https://www.dnb.de/DE/Professionell/Metadatendienste/metadatendienste_node.html. Acesso em: 30 abr. 2026.
- HYLAND, B. *et al.* **Linked data glossary**. [S. l.]: W3C, 2013. Disponível em: <https://www.w3.org/TR/ld-glossary/>. Acesso em: 13 abr. 2026.
- ISOTANI, S.; BITTENCOURT, I. I. **Dados Abertos Conectados**: em busca da Web do conhecimento. São Paulo: Novatec Editora, 2015. Disponível em: <https://ceweb.br/livros/dados-abertos-conectados/>. Acesso em: 15 mar. 2026.
- LANCASTER, F. W. **Indexação e resumos**: teoria e prática. 2. ed. Brasília: Briquet de Lemos, 2004.
- LIBRARY OF CONGRESS. **ID.LOC.GOV - Linked Data Service**. Estados Unidos da América: LC, [2026?]. Disponível em: <https://id.loc.gov/>. Acesso em: 28 abr. 2026.
- LOPES, I. L. Uso das linguagens controlada e natural em bases de dados: revisão da literatura. **Ci. Inf.**, Brasília, v. 31, n. 1, p. 41-52, 2002. Disponível em: <https://www.scielo.br/j/ci/a/WhYch5gHnWYPCPY3NPwbkcR/?format=html&lang=pt>. Acesso em: 27 jun. 2026.
- LÓSCIO, B. F.; BURLE, C.; CALEGARI, N. **Data on the web best practices**. [S. l.]: W3C, 2017. Disponível em: <https://www.w3.org/TR/2016/PR-dwbp-20161215/>. Acesso em: 15 mar. 2026.



MIRANDA, L. do V.; KUTT, K.; NALEPA, G. J. CIDOC-CRM and the first prototype of a semantic portal for the CHExRISH Project. **CEUR-WS**, Eslovênia, v. 4009, 2025. Disponível em: https://ceur-ws.org/Vol-4009/paper_14.pdf. Acesso em: 12 mar. 2026.

SERRA, L. G.; SANTARÉM SEGUNDO, J. E. O catálogo da biblioteca e o linked data. **Em Questão**, Porto Alegre, v. 23, n. 2, p. 167-185, 2017. Disponível em: <https://seer.ufrgs.br/index.php/EmQuestao/article/view/67162>. Acesso em: 27 abr. 2026.

SILVEIRA, R. C.; MARTÍNEZ-ÁVILA, D. Linked data inter-vocabulários: uso do modelo SKOS para a ligação de dados conceituais. **RICI**, Brasília, v. 18, n. 1, p. 27-46, 2025. Disponível em: <https://doi.org/10.26512/rici.v18.n1.2025.56616>. Acesso em: 17 mar. 2026.

SOLODOVNIK, I. Metadata issues in Digital Libraries: key concepts and perspectives. **JLIS.it**, [s. l.], v. 2, n. 2, p. [1-27], 2011. Disponível em: <https://www.jlis.it/index.php/jlis/article/view/304>. Acesso em: 27 abr. 2026.