

UNIVERSIDADE FEDERAL DE MINAS GERAIS
ESCOLA DE CIÊNCIA DA INFORMAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM GESTÃO & ORGANIZAÇÃO DO
CONHECIMENTO

VANESSA MARTA DE JESUS

**DADOS ABERTOS CONECTADOS A PARTIR DE CATÁLOGOS ONLINE DE
BIBLIOTECAS**

Belo Horizonte

2021

VANESSA MARTA DE JESUS

**DADOS ABERTOS CONECTADOS A PARTIR DE CATÁLOGOS ONLINE DE
BIBLIOTECAS**

Dissertação apresentada ao Programa de Pós-Graduação em Gestão & Organização do Conhecimento da Escola de Ciência da Informação da Universidade Federal de Minas Gerais para obtenção do grau de Mestre em Gestão e Organização do Conhecimento.

Área de concentração: Ciência da Informação

Orientador: Profa. Dra. Célia da Consolação Dias.

Belo Horizonte

2021

FICHA CATALOGRÁFICA

J58d	<p>Jesus, Vanessa Marta de. Dados abertos conectados a partir de catálogos online de bibliotecas / Vanessa Marta de Jesus. - 2021. 148f. : il.</p> <p>Orientadora: Célia da Consolação Dias.</p> <p>Dissertação (mestrado) – Universidade Federal de Minas Gerais, Escola de Ciência da Informação.</p> <p>1. Catalogação - Teses. 2. Bibliotecas - Teses. 3. Dados abertos conectados - Teses. 4. Inovação - Teses. 5. Web semântica - Teses. I. Dias, Célia da Consolação. II. Universidade Federal de Minas Gerais. Escola de Ciência da Informação. III. Título.</p> <p style="text-align: right;">CDU 02:007</p>
------	---

Bibliotecária responsável: Vanessa Marta de Jesus. CRB/6 - 2419



UNIVERSIDADE FEDERAL DE MINAS GERAIS
ESCOLA DE CIÊNCIA DA INFORMAÇÃO - ECI
PROGRAMA DE PÓS-GRADUAÇÃO EM GESTÃO E ORGANIZAÇÃO DO CONHECIMENTO - PPG-GOC

FOLHA DE APROVAÇÃO

DADOS ABERTOS CONECTADOS A PARTIR DE CATÁLOGOS ONLINE DE BIBLIOTECAS

VANESSA MARTA DE JESUS

Dissertação submetida à Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em GESTÃO E ORGANIZAÇÃO DO CONHECIMENTO, como requisito para obtenção do grau de Mestre em GESTÃO E ORGANIZAÇÃO DO CONHECIMENTO, área de concentração CIÊNCIA DA INFORMAÇÃO, linha de pesquisa Arquitetura e Organização do Conhecimento.

Aprovada em 01 de dezembro de 2021, pela banca constituída pelos membros:

Prof(a). Célia da Consolação Dias (Orientadora)
ECI/UFMG

Prof(a). Cíntia de Azevedo Lourenço
ECI/UFMG

Prof(a). José Eduardo Santarem Segundo
USP

Prof(a). Carlos Henrique Marcondes de Almeida
Professor Visitante ECI/UFMG

Belo Horizonte, 1 de dezembro de 2021.



Documento assinado eletronicamente por **Celia da Consolação Dias, Coordenador(a) de curso de pós-graduação**, em 02/12/2021, às 15:01, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Jose Eduardo Santarem Segundo, Usuário Externo**, em 03/12/2021, às 11:41, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Carlos Henrique Marcondes de Almeida, Professor do Magistério Superior - Visitante**, em 18/12/2021, às 10:20, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).

Documento assinado eletronicamente por **Cintia de Azevedo Lourenco, Professora do Magistério Superior**, em 21/12/2021, às 16:26, conforme horário oficial de Brasília, com fundamento no art. 5º



do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no site https://sei.ufmg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **1114800** e o código CRC **A2E1EA1C**.



UNIVERSIDADE FEDERAL DE MINAS GERAIS
ESCOLA DE CIÊNCIA DA INFORMAÇÃO - ECI
PROGRAMA DE PÓS-GRADUAÇÃO EM GESTÃO E ORGANIZAÇÃO DO CONHECIMENTO - PPG-GOC

ATA DA DEFESA DA DISSERTAÇÃO DA ALUNA

VANESSA MARTA DE JESUS

Realizou-se, no dia 01 de dezembro de 2021, às 09:00 horas, por videoconferência, da Universidade Federal de Minas Gerais, a defesa de dissertação, intitulada *DADOS ABERTOS CONECTADOS A PARTIR DE CATÁLOGOS ONLINE DE BIBLIOTECAS*, apresentada por VANESSA MARTA DE JESUS, número de registro 2020660762, graduada no curso de BIBLIOTECONOMIA, como requisito parcial para a obtenção do grau de Mestre em GESTÃO E ORGANIZAÇÃO DO CONHECIMENTO, à seguinte Comissão Examinadora: Prof(a). Célia da Consolação Dias - ECI/UFMG (Orientadora), Prof(a). Cíntia de Azevedo Lourenço - ECI/UFMG, Prof(a). José Eduardo Santarem Segundo - USP, Prof(a). Carlos Henrique Marcondes de Almeida - Professor Visitante ECI/UFMG.

A Comissão considerou a dissertação:

Aprovada

Reprovada

Finalizados os trabalhos, lavrei a presente ata que, lida e aprovada, vai assinada por mim e pelos membros da Comissão.

Belo Horizonte, 01 de dezembro de 2021.

Assinatura dos membros da banca examinadora:



Documento assinado eletronicamente por **Celia da Consolação Dias, Coordenador(a) de curso de pós-graduação**, em 02/12/2021, às 14:56, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Jose Eduardo Santarem Segundo, Usuário Externo**, em 03/12/2021, às 11:42, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Carlos Henrique Marcondes de Almeida, Professor do Magistério Superior - Visitante**, em 18/12/2021, às 10:19, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Cintia de Azevedo Lourenco, Professora do Magistério Superior**, em 21/12/2021, às 16:26, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no site https://sei.ufmg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **1114770** e o código CRC **E7AC20A4**.

AGRADECIMENTOS

Agradeço ao nosso Pai celestial pela força necessária a mim concedida para superar as dificuldades; a perseverança para superar os desafios e a possibilidade de concluir este estudo que me tornou Mestre.

À minha orientadora, Professora Célia da Consolação Dias, pelo seu amor e dedicação a esta profissão, por sua paciência e atenção aos detalhes e pelo seu profissionalismo. Obrigada!

À banca de qualificação e defesa, composta pela Professora Doutora Cíntia de Azevedo Lourenço; pelos Professores Doutor José Eduardo Santarem Segundo e Professor Doutor Carlos Henrique Marcondes. Agradeço pelo tempo despendido às considerações e por se disporem a participar e prestar valiosas contribuições à minha investigação.

A todos aqueles que estão comigo em meu cotidiano, tornando os meus dias mais leves e alegres: a minha pequena Lara, meu companheiro Gilmar Ribeiro, minha mãe Patrícia, minha irmã Kênia e minha madrinha Marília. Agradeço também aos meus avós Emilce e Benedito; e à minha amiga Caroline *[In memória]*.

Aos meus amigos, em especial Kelly, Júlio e Rafael, pelo incentivo e pela ajuda durante os estudos. Sônia e Joana Dar'c pelo carinho e amizade que perdura por muitos anos.

Aos colegas de trabalho, em especial a Jane e Jussara pelo incentivo e pelas orientações durante o processo de seleção; à Andrea, por possibilitar uma maior flexibilidade na minha escala de trabalho de acordo com a demanda das aulas do mestrado e ao Gustavo, pelo companheirismo e pelos ensinamentos na prática da catalogação, que foi a temática que me motivou a desenvolver esta pesquisa.

Por fim, e com relevada importância, aos mestrandos e doutorandos do PPGGOG pelo auxílio mútuo durante todo o tempo de convívio; a UFMG e ao PPGGOG, que me proporcionaram a oportunidade de me qualificar na esfera pública como Mestre em Gestão e Organização do Conhecimento. Muito obrigada!

Dedico esta dissertação a todos aqueles que precisaram, de alguma maneira, enfrentar as dificuldades impostas pela crise pandêmica que vivenciamos nos últimos quase dois anos.

RESUMO

Ao longo dos anos avanços significativos foram desenvolvidos para o intercâmbio de informações entre os registros bibliográficos que compõem os catálogos online de bibliotecas, como a criação do protocolo Z39.50, do *Open Archives Initiative - Protocol for Metadata Harvesting (OAI-PMH)* e do formato MARC, que permitem o intercâmbio de registro. Nesse cenário, passa a ser possível realizar a importação de dados e fazer o reuso de registros de outras bibliotecas, de forma colaborativa, diminuindo custos e retrabalhos no processo de catalogação. Uma nova possibilidade de inovação incremental no cenário das bibliotecas são os dados abertos conectados, foco principal desta pesquisa de mestrado, cujo ambiente de investigação são os catálogos online de bibliotecas. Os dados conectados podem ser entendidos como um conjunto de nós, que compreende uma união de vários dados e links os quais podem ser considerados as conexões. Seus padrões permitem a publicação de dados de tal forma que possam ser lidos por pessoas e processados por máquinas. Possibilitam que o acervo da instituição possa ficar acessível a qualquer pessoa por intermédio da interoperabilidade, contribuindo para que diferentes sistemas estejam conectados na *Web* para a busca de informações. Os atuais catálogos de bibliotecas possuem limitações considerando o contexto da *Web*. Neste cenário, **o objetivo geral** desta pesquisa é realizar o mapeamento dos desafios encontrados na implementação de práticas de dados abertos conectados em bibliotecas para a disponibilização de seus catálogos online na *Web*. Como **metodologia**, a pesquisa caracteriza-se como aplicada e exploratória, com utilização de pesquisa bibliográfica e documental. A técnica de pesquisa adotada é a análise de conteúdo de Bardin, compreendendo as fases levantamento bibliográfico, elaboração de fichamentos para apreciação inicial dos textos, criação de categorias para categorização dos dados, agrupamento dos dados nas categorias criadas e análise e interpretação dos dados. Como **resultado** verificou-se que os desafios alcançados referem-se ao formato de metadados, licenças, proveniência e qualidade, identificadores, formatos, vocabulários, acesso, preservação e enriquecimento de dados. As barreiras verificadas foram atinentes aos recursos financeiros escassos, falta de recursos humanos qualificados, baixo nível de conscientização entre a comunidade/bibliotecários e dependência do mercado externo para aquisição de software. **Conclui-se** que a tecnologia não é o principal obstáculo para publicação dos dados conectados a partir de catálogos de bibliotecas, embora alterar o modelo atual de metadados adotado não seja uma tarefa fácil e são muitos os desafios para superá-los. Observou-se, no entanto, que os percalços existem não por falta de requisitos, mas sim pela complexidade que envolve o processo de publicação de dados abertos conectados a partir de catálogos online, aliada à carência de pessoal qualificado para desempenhar a função, da escassez de recursos financeiros e da necessidade de conscientização de pessoal sobre os benefícios advindos da adoção dos dados abertos conectados pelas bibliotecas.

Palavras-chave: Catalogação. Bibliotecas. Dados abertos conectados. Inovação. *Web* semântica.

ABSTRACT

Over the years, advances have been developed for the exchange of information between bibliographic records that make up online library catalogs, such as the creation of the Z39.50 protocol and the Open Archives Initiative - Protocol for Metadata Harvesting (OAI-PMH) and the MARC format, which binds the registration exchange. In this scenario, it becomes possible to import data and reuse records from other libraries, in a collaborative way, reducing costs and rework in the cataloging process. A new possibility for incremental innovation in the library scenario is open data, the main focus of this master's research, whose research environment is the online library catalogs. Connected data can be understood as a set of nodes, which comprises a union of various data and links which can be considered as known. Its standards allow the publication of data in such a way that it can be read by people and processed by machines. They make it possible for the institution's collection to be accessible to anyone through interoperability, contributing to different systems selected on the Web to search for information. The current library catalogs have limitations considering the Web context. In this scenario, the general objective of this research is to map the challenges found in the implementation of available data practices connected in libraries to make their catalogs available online on the Web. As a methodology, the research is characterized as applied and exploratory, using bibliographical and documentary research. The research technique adopted is Bardin's content analysis, comprising the phases of elaboration of records for initial analysis of the texts, creation of categories for data categorization, grouping of data in the categories created and analysis and interpretation of data. As a result, it was found that the challenges achieved refer to the metadata format, licenses, provenance and quality, identifiers, formats, vocabularies, access, preservation, data enrichment. The barriers observed were related to scarce financial resources, lack of qualified human resources, low level of awareness among the community/librarians and dependence on the foreign market for software acquisition. It is concluded that technology is not the main obstacle for publishing connected data from library catalogs, although changing the current metadata model adopted is not an easy task and there are many challenges to overcome. It was observed, however, that the challenges exist not for lack of requirements, but for the complexity involved in the process of publishing open data connected from online catalogs, combined with the lack of qualified personnel to perform the function, the scarcity of financial resources and the need to raise awareness among staff about the benefits arising from the adoption of open data connected by libraries.

Keywords: Cataloguing. Innovation. Library. Linked open data. Semantic web.

LISTA DE FIGURAS

FIGURA 1: Relacionamentos entre as entidades do Grupo 1 do FRBR.....	32
FIGURA 2: Relacionamentos entre as entidades do Grupo 1 e 2 do FRBR.....	33
FIGURA 3: Relacionamentos de assunto do modelo FRBR.....	33
FIGURA 4: Relacionamentos do FRBR: exemplo com a obra <i>The Hobbit</i>	34
FIGURA 5: Visão geral das relações do modelo IFLA LRM.....	42
FIGURA 6: <i>Bibframe 2.0</i>	44
FIGURA 7: Grafo RDF da tripla de um livro	53
FIGURA 8: Grafo RDF genérico de uma tripla	53
FIGURA 9: Seis etapas para bibliotecas publicarem dados abertos conectados	66
FIGURA 10: Diagrama da metodologia adotada.....	99
FIGURA 11: Diagrama das categorias criadas para análise de conteúdo	109

LISTA DE TABELAS

TABELA 1: Contém as informações dos recursos do tipo livro.....	52
TABELA 2: Tripla de um recurso.....	53
TABELA 3: Comparação de pessoal envolvido nas atividades de dados conectados nas pesquisas de 2018 e 2015.....	69
TABELA 4: Respostas da pesquisa sobre como os dados conectados são usados	69
TABELA 5: Principais motivações para publicar dados conectados.....	70
TABELA 6: Os principais vocabulários/ontologias em RDF usados em 2018 em comparação com 2015.....	70
TABELA 7: Tecnologias usadas para publicar dados conectados.....	71
TABELA 8: Barreiras encontradas na publicação de dados conectados.....	72
TABELA 9: As 10 principais fontes de dados vinculadas consumidas.....	73
TABELA 10: Principais motivações para consumir dados conectados	73
TABELA 11: Barreiras encontradas no consumo de dados conectados	74
TABELA 12: Dados extraídos dos estudos selecionados.....	85
TABELA 13: Proposta de metodologia para publicação de registros bibliográficos como <i>Linked Data</i>	92
TABELA 14: Ferramentas mais utilizadas em cada um dos processos.....	93

LISTA DE QUADROS

QUADRO 1 – Regras e padrões bibliográficos	28
QUADRO 2 – Identificação das entidades no IFLA LRM	36
QUADRO 3 – Principais diferenças entre os quatro modelos	40
QUADRO 4 – Descrição com o <i>Bibframe</i> 2.0 pelo <i>Bibframe</i> editor	46
QUADRO 5 – Matriz SWOT dos fatores motivacionais e desafios de <i>Linked Data</i> em bibliotecas e centros de informação.....	90
QUADRO 6 – Categoria desafios encontrados na implementação de práticas de dados abertos conectados em bibliotecas para a disponibilização de seus catálogos online da <i>Web</i>	102
QUADRO 7 – Categoria barreiras encontrados na implementação de práticas de dados abertos conectados em bibliotecas para a disponibilização de seus catálogos online na <i>Web</i>	104
QUADRO 8 – Categoria potencialidades dos registros bibliográficos para uso em formato de dados abertos para conectar os catálogos das bibliotecas	105
QUADRO 9 – Categoria requisitos para publicação de dados abertos conectados.....	105
QUADRO 10 – Resultados atribuídos à categoria desafios encontrados na implementação de práticas de dados abertos conectados em bibliotecas para a disponibilização de seus catálogos online na <i>Web</i>	111
QUADRO 11 – Resultados atribuídos à categoria barreiras encontradas na implementação de práticas de dados abertos conectados em bibliotecas para a disponibilização de seus catálogos online na <i>Web</i>	117

QUADRO 12 – Resultados atribuídos à categoria potencialidades dos registos bibliográficos para uso em formato de dados abertos para conectar os catálogos das bibliotecas121

QUADRO 13 – Resultados atribuídos à categoria requisitos para publicação de dados abertos conectados.....124

LISTA DE GRÁFICOS

GRÁFICO 1 – Quantidade de trabalhos por programas de Pós-Graduação relacionados ao tema <i>Linked Data</i> (2006-2016)	21
GRÁFICO 2 – Descrição das 5 estrelas dos dados abertos.....	49

LISTA DE ABREVIATURAS

AAT	–	<i>Art and Architecture Thesaurus</i>
ALA	–	<i>American Library Association</i>
API	–	<i>Appliance Program Interface</i>
APP	–	<i>Aplicações para dispositivos móveis</i>
BIBFRAME	–	<i>Bibliographic Framework</i>
CC	–	<i>Creative Commons</i>
CC-BY	–	<i>CC Atribuição de autoria</i>
CC-NC	–	<i>CC Não comercial</i>
CC-ND	–	<i>CC Não permite derivação</i>
CC-AS	–	<i>CC Compartilha da mesma forma</i>
DC	–	<i>Dublin Core</i>
DCAT	–	<i>Data Catalog Vocabulary</i>
DCMI	–	<i>Dublin Core Metadata Initiative</i>
FAST	–	<i>Faceted Application of Subject Headings</i>
FOAF	–	<i>Friend of a Friend</i>
FRBR	–	<i>Functional Requirements for Bibliographic Records</i>
HTTP	–	<i>Hyper Text Transfer Protocol</i>
IFLA	–	<i>International Federation of Library Associations</i>
IRI	–	<i>International Resource Identifier</i>
ISBN	–	<i>International Standard Book Number</i>
ISNI	–	<i>International Standard Name Identifier</i>
JSON	–	<i>JavaScript Object Notation</i>
LCSH	–	<i>Library of Congress Subject Headings</i>
LD	–	<i>Linked Data</i>
LD4L	–	<i>Linked Data for Libraries</i>
LOC	–	<i>Library of Congress</i>
LOD	–	<i>Linked Open Data</i>
LOV	–	<i>Linked Open Vocabularies</i>
MARC	–	<i>Machine Readable Cataloging</i>
MeSH	–	<i>Medical Subject Headings</i>
MODS	–	<i>Metadata Object Description Schema</i>
N3	–	<i>Notation3</i>
NISO	–	<i>National Information Standards Organization</i>
NT	–	<i>N-Triples</i>
OCDE	–	<i>Organização para a cooperação e desenvolvimento econômico</i>
OCLC	–	<i>Online Computer Library Center</i>
OPAC	–	<i>Online Public Access Catalog</i>
ORCID	–	<i>Open Research and Contributor ID</i>
OWL	–	<i>Ontology Web Language</i>
PPGGOC	–	<i>Programa de Pós-Graduação em Gestão e Organização do Conhecimento</i>
RDA	–	<i>Resource Description and Access</i>
RDF	–	<i>Resource Description Framework</i>
RDFS	–	<i>Resource Description Framework Schema</i>
SKOS	–	<i>Simple Knowledge Organization System</i>
SPARQ	–	<i>SPARQL (SPARQL Protocol and RDF Query Language)</i>
TIC	–	<i>Tecnologia da Informação e da Comunicação</i>
URI	–	<i>Uniform Resource Identifier</i>
URL	–	<i>Uniform Resource Locator</i>
VIAF	–	<i>Virtual International Authority File</i>
XML	–	<i>Extensible Markup Language</i>
W3C	–	<i>World Wide Web Consortium</i>

SUMÁRIO

1 INTRODUÇÃO	11
1.1 JUSTIFICATIVA E PROBLEMA DE PESQUISA	19
1.2 OBJETIVOS	22
1.2.1 Objetivo geral	22
1.2.2 Objetivos específicos	22
1.2.3 Estrutura da pesquisa	23
2 FUNDAMENTAÇÃO TEÓRICA	24
2.1 CATALOGAÇÃO, CATÁLOGOS E REGISTROS BIBLIOGRÁFICOS	24
2.2 DADOS ABERTOS CONECTADOS E <i>WEB</i> SEMÂNTICA: CONCEITOS, APLICABILIDADE E TECNOLOGIAS NECESSÁRIAS PARA IMPLEMENTAÇÃO	47
2.2.1 Dados conectados	49
2.2.2 Dados abertos conectados	56
2.2.3 <i>Web</i> semântica	57
2.3 FONTES DE DADOS CONECTADOS	59
2.3.1 <i>OCLC Fast – Faceted Application of Subject Terminology</i>	61
2.3.2 <i>ORCID - Open Researcher and Contributor ID</i>	61
2.3.3 <i>VIAF - Virtual International Authority File</i>	61
2.3.4 <i>LC Subject Headings - LCSH</i>	62
2.3.5 <i>LC Name Authority File - LCNAF</i>	62
2.3.6 <i>AGROVOC</i>	63
3 REVISÃO DE LITERATURA: DADOS ABERTOS CONECTADOS A PARTIR DE CATÁLOGOS ONLINE DE BIBLIOTECAS	63
4 METODOLOGIA	97
4.1 LEVANTAMENTO BIBLIOGRÁFICO	99
4.2 CRIAÇÃO DAS CATEGORIAS	101
4.3 AGRUPAMENTO DOS DADOS EM CATEGORIAS	109
4.4 ANÁLISE E INTERPRETAÇÃO DOS DADOS	110

5	RESULTADOS	110
6	CONSIDERAÇÕES FINAIS.....	133
	REFERÊNCIAS	137
	APÊNDICE A – TEXTOS SELECIONADOS PARA ANÁLISE NESTE ESTUDO.....	147

1. INTRODUÇÃO

O século XXI tem como principal característica o expressivo aumento do uso da *Web* e das Tecnologias de Informação e Comunicação (TIC), que impactam diariamente várias das atividades desenvolvidas pelas pessoas no seu cotidiano, entre elas, a busca por informação. As organizações que trabalham com o processamento, manipulação e organização de dados, como as bibliotecas, também devem refletir sobre os benefícios da adoção dessas tecnologias para garantir adaptabilidade e acessibilidade à sociedade a que servem. Caso contrário, correm o risco de não terem nenhuma relação ou similaridade com uma grande parte dos usuários que apresentam um comportamento diferenciado frente à diversidade de serviços e conteúdos disponibilizados na *Web* (LEVACOV, 2005).

Esse crescimento significativo de recursos de informação e de conteúdo são fatores que acabam por dificultar a busca, localização, acesso e recuperação da informação em distintos ambientes informacionais. Diferentes áreas do conhecimento enfrentam tais dificuldades, principalmente no ambiente *Web* (SANTOS; ALVES, 2009). “Das estantes de bibliotecas aos incontáveis repositórios digitais espalhados pelo mundo, nunca foi tão difícil encontrar a informação adequada” (SCHIESSL, 2015, p.27).

A dificuldade de recuperação na *Web* em função da enorme quantidade de informações disponíveis soma-se a onda de desinformação que ficou ainda mais evidente a partir da pandemia que teve início no ano de 2020 e se estendeu no decorrer de 2021. De acordo com a OMS (2020, p.2), “o surto de COVID-19 e a resposta a ele têm sido acompanhados por [...] um excesso de informações, algumas precisas e outras não, que tornam difícil encontrar fontes idôneas e orientações confiáveis quando se precisa”.

Araújo (2021) afirma que esse cenário de desinformação tem se repetido nos campos da educação, da economia, da política, e em vários outros; somando a isso, a importância do profissional bibliotecário nunca foi tão reconhecida como nos tempos atuais, por ser um profissional com capacitação para disseminar informações oriundas de fontes seguras.

As bibliotecas podem fornecer para a *Web* informações de qualidade, disponibilizando conjuntos padronizados de dados que podem ser reutilizados e

compartilhando instrumentos consagrados na área como os catálogos de autoridades e vocabulários controlados. Como requisito para alcançar a interoperabilidade com outros sistemas e conjuntos de dados na *Web*, elas precisam adaptar seus metadados para uso no novo ambiente (BAKER, 2011; BYRNE; GODDARD, 2010; SERRA, 2019).

Ao longo da existência das bibliotecas a evolução das TIC's proporcionou mudanças significativas na maneira como os registros bibliográficos de acervos são disponibilizados, passando dos catálogos manuais para os catálogos online, e, na atualidade, caminhando para a tendência de catálogos conectados no ambiente *Web*. A origem dos catálogos é bastante antiga, remonta há 2.000 a.C., mas ele só recebeu algum tipo de normalização no séc. XIX (MEY, 1995), evoluindo ao longo dos anos e sofrendo influências pelos avanços da TI.

Os catálogos surgem, inicialmente, somente com o intuito de registrar as informações contidas nos documentos pertencentes a algum acervo. Com o surgimento da imprensa e respectivo aumento de materiais impressos, houve a necessidade que essa organização fosse otimizada e, assim, os catálogos passam a ter um objetivo mais amplo, não sendo somente o de registrar e guardar informações, mas também o de recuperar essas mesmas informações (MARTINHO; FUJITA, 2011).

Os catálogos oferecem, também, suporte às funções de gerenciamento de biblioteca: controle e desenvolvimento de coleção; aquisição e divulgação de novos materiais; gerenciamento de orçamento, entre outros. Além disso, melhora as funções de usuário, como circulação de material, gerenciamento de contas e liberação de solicitações de suspensão e desligamento (COYLE, 2010a).

De acordo com Mey (1995) os catálogos são definidos como um canal de comunicação estruturado, possibilitando assim a veiculação das mensagens contidas nos itens, e sobre os itens, de um ou vários acervos, apresentando-as de forma codificada e organizada, agrupadas por semelhanças aos usuários desses respectivos acervos. Segundo Taylor (2006), os catálogos podem ser considerados como um conjunto organizado de registros bibliográficos que representam os itens de uma coleção particular e/ou recursos acessíveis em uma localização.

Os registros bibliográficos que compõe os catálogos são definidos pela Declaração dos Princípios Internacionais de Catalogação como “um conjunto de

elementos de dados que descreve e provê acesso a um recurso bibliográfico e identifica obras e expressões relacionadas” (*INTERNATIONAL FEDERATION OF LIBRARY ASSOCIATIONS AND INSTITUTIONS*, 2009, p. 10, tradução nossa)

Na década de 1960, nos Estados Unidos, a *Library of Congress* desenvolveu um sistema de comunicação informatizada, e, a partir de então, houve uma evolução na elaboração dos registros bibliográficos. Eles passaram a ser feitos por intermédio de softwares gerenciadores de bibliotecas e deram origem aos Catálogos Públicos de Acesso Online - OPACS (MACHADO, 2003).

Comparado às antigas fichas catalográficas elaboradas na época dos catálogos impressos, o catálogo online - OPACS, que contém todos os elementos das fichas, oferece outros recursos mais avançados. É composto por um sistema informatizado que disponibiliza consulta, empréstimo e recuperação de informação, pesquisa online por autor, título e assunto, agilizando todo o processo e possibilitando ao bibliotecário a possibilidade de elaborar o registro bibliográfico apenas uma única vez, gerando automaticamente várias entradas de consultas (POZZI; FUJITA, 2012).

Avanços significativos foram desenvolvidos para o intercâmbio de informações entre os registros bibliográficos que compõe os catálogos, como, por exemplo, a criação do protocolo Z39.50, que propicia a consulta a diferentes catálogos online simultaneamente, do *Open Archives Initiative - Protocol for Metadata Harvesting* (OAI-PMH), o qual permite a coleta automática e o intercâmbio de registros e do formato *MARC* que possibilita o intercâmbio de registros (SAYÃO; MARCONDES, 2008).

Nesse cenário, passa a ser possível realizar a importação e fazer o reuso de registros de outras bibliotecas, de forma colaborativa, diminuindo custos e retrabalhos no processo de catalogação e disponibilizando o "item" com maior agilidade ao usuário. De acordo com Machado e Zafalon (2020) a representação e a recuperação da informação, agora automatizadas, possibilitam que outros metadados sejam registrados tornando o sistema robusto, com informações mais completas e dinâmicas, permitindo o intercâmbio de dados bibliográficos sem barreiras geográficas.

O desenvolvimento de padrões, formatos e protocolos que possibilitam essa troca de registros bibliográficos entre bibliotecas contribuíram para inserir a utilização de regras que favoreceram a interoperabilidade no contexto dessas organizações.

O processo de recuperação e disseminação da informação também é beneficiado com essas transformações, assim como o processo de produção de conhecimento e informação. Com os avanços tecnológicos foi possível também disponibilizar documentos digitalizados, antes contidos apenas em suporte físico, e documentos natos digitais, possibilitando a apresentação do acervo da instituição num formato acessível online, representando mudanças significativas nos formatos documentais inseridos nos catálogos e disponibilizados aos usuários.

Por quase dois séculos, desde as origens da prática moderna de catalogação de bibliotecas na década de 1830 e das 91 regras de Anthony Panizzi, as bibliotecas estão em constante transformação para atender às necessidades de seus usuários contemporâneos (COYLE, 2010a). Para Coyle (2010a) ao longo dos anos mudanças conceituais ocorreram no catálogo, ocasionadas pelas transformações relacionadas ao contexto onde ele está inserido. Algumas delas estão relacionadas às regras para a descrição de novos materiais que as bibliotecas deveriam catalogar, como, por exemplo, as modificações que ocorrem com o advento de sons e filmes gravados. Neste sentido, o fluxo de trabalho de catalogação teve que ser alterado para lidar com a produção cada vez maior e mais diversificada de recursos de informação.

Na era da *Web* as mudanças demandadas no processo de catalogação referem-se ao perfil do usuário da biblioteca de hoje. Sob essa perspectiva, Coyle (2010a) apontou que o usuário não visita mais as bibliotecas física como sua fonte primária de informações, mas sim, busca e cria informações enquanto está conectado à rede global de computadores. Em virtude disso, as mudanças que as bibliotecas precisarão fazer deve incluir a transformação do catálogo da biblioteca de um banco de dados independente de registros bibliográficos em um conjunto de dados altamente hiperlinkado que pode interagir com recursos de informação na *World Wide Web*.

No contexto da *Web*, Serra e Santarem (2017) postularam que os tradicionais catálogos das bibliotecas são vistos como ilhas de informação, não sendo interligados, exigindo a elaboração de pesquisas nos sítios das instituições ou em projetos de bibliotecas digitais, sem compartilhamento de elementos descritivos.

De acordo com Coyle (2010b), é preciso fornecer serviços de biblioteca por meio da *Web*, visto que os usuários vão à rede para fazer pesquisas, interagir com outras pessoas e criar seus trabalhos. Portanto, para que seja possível atender a esse novo formato é necessário fornecer serviços de biblioteca aos usuários através da *Web*.

Desde meados do século XX, o compartilhamento de dados entre instituições vem ganhando força por meio da interoperabilidade e da disponibilização dos catálogos na *Web*, de acordo com os princípios dos dados abertos conectados, bem como pelo uso de tecnologias da *Web* semântica (MARCONDES, 2016). O uso e a troca de informações são as características da interoperabilidade. Com o objetivo de compreender o termo, apresentam-se a seguir a sua definição extraída da ISO, da *National Information Standard Organization* (NISO) e do *Committee on Cataloging, Description & Access*. O conceito de interoperabilidade é definido pela ISO 25964 como a “capacidade de dois ou mais sistemas ou componentes trocarem informações e usarem as informações que foram trocadas” (ISO 25964-2: 2013, tradução nossa).

Para a NISO interoperabilidade é a “capacidade de vários sistemas com diferentes plataformas de *hardware*, *software*, estruturas de dados e interfaces trocarem dados com perda mínima de conteúdo e funcionalidade” (NISO, 2004, p. 2, tradução nossa). Outra definição acrescenta o uso e a troca de informações: “Interoperabilidade é a capacidade de dois ou mais sistemas ou componentes de trocarem informações e usar as informações trocadas sem esforço especial em qualquer um dos sistemas” (*COMMITTEE ON CATALOGING, DESCRIPTION & ACCESS*, 2000, não paginado, tradução nossa). Por fim, Marcondes (2016) afirma que a interoperabilidade e a integração de acervos de diferentes instituições como bibliotecas, arquivos, museus e centros de documentação está baseada nas tecnologias da *Web Semântica*¹ e dos dados abertos conectados.

¹ A *Web* semântica surgiu em 2001 e foi proposta por Tim Berners-Lee, James Hendler e Ora Lassila em um artigo publicado na revista *Scientific American*, intitulado: *The Semantic Web - A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities* (*Web Semântica: um novo formato de conteúdo para a Web que tem significado para computadores vai iniciar uma revolução de novas possibilidades*) (BERNERS-LEE, HENDER; LASSILA, 2001).

Entende-se que a *Web* semântica é “uma extensão da *Web* atual, onde a informação possui um significado claro e bem definido, possibilitando uma melhor interação entre computadores e pessoas” (BERNERS-LEE; HENDLER; LASSILA, 2001, p. 2, tradução nossa). Segundo o W3C (2013) a *Web* semântica “fornece uma estrutura comum que permite que os dados sejam compartilhados e reutilizados entre aplicativos, empresas e comunidades, tendo como base o *Resource Description Framework – RDF*” (W3C, 2013, tradução nossa).

A importância da *Web* semântica está na possibilidade de se ter dados definidos e conectados. Estes requisitos podem ser usados para descoberta, automação, integração e reutilização mais eficazes dos dados, permitindo a vinculação de bancos de dados, compartilhamento de conteúdo entre aplicativos e a combinação de serviços da *Web* (HENDLER; BERNERS-LEE; MILLER, 2002).

Para isso, é requerido a adoção de novas tecnologias e modelos de dados que estruturam as informações de forma que a máquina possa compreender seu significado, em vez de simplesmente exibi-lo. A habilitação dessas ferramentas tornará mais fácil a criação de conteúdos amplamente disponíveis e legíveis por máquinas (BERNERS-LEE; HENDLER, 2001). De acordo com Marcondes (2016), pensando nos sistemas gerenciadores de bibliotecas, pode-se inferir que é fundamental o uso de tais tecnologias para que os registros bibliográficos cumpram o papel de serem mais interoperáveis e passíveis de integração.

A proposta da *Web* semântica evolui com o passar dos anos, desenvolvendo conceitos e tecnologias mais sólidas, embasadas em aplicações capazes de explorar bases de dados com maior precisão e compreender o significado das informações. Para Luz, Coneglian e Santarem Segundo (2019, p.5) “essa fase em que aplicações de uso são construídas e utilizadas em massa, está sendo chamado de materialização da *Web* semântica, processo que tem o *Linked Data* como principal aplicação”.

O *Linked Data* foi proposto por Berners-Lee em 2006 e fornece diretrizes para a criação de bases de dados de acordo com regras que tornam mais fácil localizar e inserir significado nesses dados. Baseadas nessa proposta, várias bases de dados conectadas e abertas foram criadas seguindo a proposta de *Linked Data* e usando as tecnologias recomendadas pela *Web* semântica (ISOTANI; BITTENCOURT, 2015).

O termo *Linked Data*, traduzido para português como “dados conectados” é adotado com essa tradução pelo Guia Prático para publicação de dados conectados na *Web* (2018), por Isotani e Bittencourt (2015); Laufer (2015), pelo Manual dos dados abertos (2011), entre outros.

Outros pesquisadores usam os termos "dados interligados", "dados ligados" e "dados conectados" para se referirem ao termo “*Linked Data*”. Isotani e Bittencourt (2015, p.31) chamam a atenção para a semântica do termo “conectar” em virtude do termo ligado não ser o mais adequado para transmitir o significado atribuído ao termo “*linked*”, visto que ao observar a palavra “ligar”, a mesma pode apresentar muitos cenários que não estão adequadamente aplicados ou relacionados a *Linked Data*. No entanto, isso não acontece ao considerarmos a palavra “conectar”.

Além disso, outro ponto fundamental para a escolha é com relação aos princípios de *Linked Data*, principalmente ao se observar os dois últimos, em que fica claro que o objetivo proposto foi literalmente conectar dados na *Web*. Tratando-se do termo interligado, existe um problema, visto que a tradução inversa deste termo para o inglês é *interlinked*.

Em virtude dessas informações, foi adotado então a tradução do termo *Linked Data* para dados conectados. Por não haver tradução oficial do termo em português, e de forma a compartilhar as mesmas ideias propostas por Isotani e Bittencourt (2015), este estudo adota o termo “dados conectados” para tradução do termo “*Linked Data*”.

O consórcio da *World Wide Web* - W3C (W3C, 2015) indica algumas tecnologias para que a *Web* semântica se torne implementável. Entre elas estão o XML (*Xtensible Markup Language*), RDF (*Resource Description Framework*), OWL (*Web Ontology Language*), SPARQL (*SPARQL Protocol and RDF Query Language*) entre outras tecnologias descritas pelo consórcio.

Na biblioteconomia, Byrne e Goddard (2010) destacaram que os dados abertos só podem se transformar numa ferramenta potencial se sua utilização for crescente, exigindo um preparo das bibliotecas, e, acima de tudo, de seus profissionais, para que essas instituições sejam agentes no processo dos novos meios de informação do mundo de dados ligados e abertos. Em comparação com a prática atual de criação e transferência de dados entre bibliotecas, os benefícios do uso de dados conectados nessas organizações têm vantagens significativas

enquanto ampliam naturalmente o modelo de compartilhamento colaborativo que as bibliotecas tradicionalmente adotam (BAKER et al., 2011).

Os autores ainda observaram que os dados conectados (especialmente os dados abertos conectados) são compartilháveis, extensíveis e reutilizáveis. Além desses aspectos, os autores apresentaram outros elementos que apontam para o benefício do uso dos dados conectados que oferecem suporte a funções multilíngues para dados e serviços do usuário usando identificadores *Uniform Resource Identifier* (URI). Esses recursos são inerentes ao modelo de dados conectados e podem ser obtidas usando os dados estruturados pelas tecnologias da *Web* semântica. Os recursos podem ser descritos em cooperação com outras bibliotecas e conectados a dados fornecidos por outras comunidades ou outros indivíduos.

Os dados do catálogo da biblioteca podem vir a ser a conexão entre a instituição bibliotecária e o espaço de conhecimento na *Web*. De acordo com Coyle (2010b), a ideia de que os metadados da biblioteca serão amplamente utilizados na *Web* modifica o significado do processo de catalogação, não se limitando mais à criação de registros para o catálogo da biblioteca. Além disso, os dados também servirão a outras funções e usuários que nunca poderão usar diretamente o referido catálogo. Esse processo é visto como uma verdadeira expansão do papel dos dados da biblioteca, a ponto de poderem ser usados para qualquer função bibliográfica.

A *Stanford University*, em junho 2011, compartilhou uma visão de *Linked Data* como tecnologia disruptiva, com o potencial de mover bibliotecas e outros provedores de informação além das restrições de metadados baseados em MARC, bem como as restrições de muitas formas variantes desses metadados gerados para uma ampla variedade de gêneros em uso na comunicação científica. É visto que especialistas e estudiosos da área defenderam a necessidade do desenvolvimento crescente de ferramentas e serviços que apresentem “uma variedade de meta-serviços emergentes, abertos e direcionados por link, a fim de habilitar totalmente o *Linked Data* como uma tecnologia suspensiva para processos de descoberta e navegação (BE PART OF THE WEB, 2011, não paginado, tradução nossa).

O desenvolvimento de estudos e projetos de dados conectados tem crescido nos últimos anos, assim como as ferramentas e recursos para bibliotecas implementarem a tecnologia, como os conversores MARC em RDF para plataformas

semânticas. A tecnologia não é mais o principal obstáculo para conectar dados na biblioteca, embora alterar o modelo atual de metadados não seja uma tarefa fácil. Os projetos da *Web* semântica implementados com sucesso por bibliotecas estão sendo cada vez mais difundidos na literatura. Nesse caso, pode-se observar que um dos principais desafios no desenvolvimento de dados conectados é a necessidade de aumentar a conscientização da comunidade bibliotecária (BYRNE; GODDARD, 2010; GALVÃO; LOPES, 2012).

Para fomentar esse cenário dos dados abertos conectados entre a comunidade bibliotecária, é importante o aumento da realização de estudos que contribuam com reflexões sobre quais os desafios são encontrados pelas bibliotecas para a disponibilização dos catálogos online como dados abertos conectados. Este estudo caminha, nesse sentido, tendo como temática a aplicação dos dados abertos conectados a partir dos catálogos online de bibliotecas e apresentará agora a justificativa e os problemas de pesquisa abordados nesta pesquisa.

1.1 JUSTIFICATIVA E PROBLEMA DE PESQUISA

As razões para o desenvolvimento dessa pesquisa se deram em virtude da potencialidade dos registros bibliográficos frente às limitações dos atuais catálogos online de bibliotecas no ambiente *Web*.

O século XXI centra-se na comunicação em ambiente *Web*, na difusão de novos modos de acesso à informação, e, no âmbito da biblioteconomia, na resolução de um paradoxo: o de dados bibliográficos ricos em informação, estrutura e invisíveis no ambiente *Web* (GALVÃO E LOPES, 2012). Além disso, as bibliotecas, ao agregarem dados bibliográficos à *Web* semântica, fornecem informações valiosas para a comunidade em geral, devido à qualidade dos dados oriundos dos seus registros bibliográficos (SHIEH; REESE, 2015).

Nesse contexto, essas instituições precisam reconhecer-se como potenciais provedoras de dados confiáveis. Para tanto, precisam adequar seus metadados ao ambiente da *Web*, permitindo interoperabilidade com outros sistemas e conjuntos de dados, conforme já foi reportado em outros estudos (SERRA, 2019). É válido destacar que a *Web* semântica possibilita diversas aplicações no ambiente de bibliotecas, podendo proporcionar o enriquecimento do catálogo de forma dinâmica,

vinculando dados com informações complementares dispostas na Web (SERRA, 2019).

Coyle (2010b) evidenciou em seus estudos diversos benefícios da disponibilização dos registros bibliográficos no ambiente *Web*, tais como a utilização dos vastos recursos de informação existentes para aprimorar os dados, criando relacionamentos entre os dados da biblioteca e os recursos de informação. A autora apresentou como vantagem a ampliação das oportunidades para os usuários descobrirem a biblioteca e seus recursos, além de aumentar também o valor e a riqueza dos dados, permitindo seu uso em uma ampla variedade de contextos.

Os dados abertos conectados são uma possibilidade de inovação incremental no cenário das bibliotecas, onde o acervo da instituição fica acessível a qualquer pessoa por intermédio da interoperabilidade, contribuindo para que diferentes sistemas possam estar interligados na *Web* para a busca de informações (SOUZA; ALVARENGA, 2014). Assim, as tecnologias contribuem para melhorar esse cenário ao estabelecer uma linguagem comum para interligar conteúdos informatizados, fazendo a conexão de vários dados de instituições distintas numa mesma pesquisa (BIZER, 2009).

Conforme Coyle (2010a), para atender às necessidades dos usuários contemporâneos, as bibliotecas modernas estão em constante transformação. Esse processo compreende, também, a disponibilização de registros bibliográficos em ambiente *Web*, cada vez mais utilizado para pesquisas, significando assim a utilização de uma nova abordagem. Iniciativas nesse sentido são desenvolvidas com êxito no exterior.

Senso e Arroyo Machado (2018) apresentaram projetos pioneiros de dados conectados que são basilares para o entendimento do processo de conversão dos atuais catálogos de bibliotecas para o formato de dados conectados, visto que quando são verificadas as diferentes contribuições voltadas para a análise da situação atual do *Linked Data* em bibliotecas, é possível observar que todos coincidem quando se fala de um conjunto de projetos que são considerados paradigmáticos dentro deste ambiente (SENSO; ARROYO MACHADO, 2018).

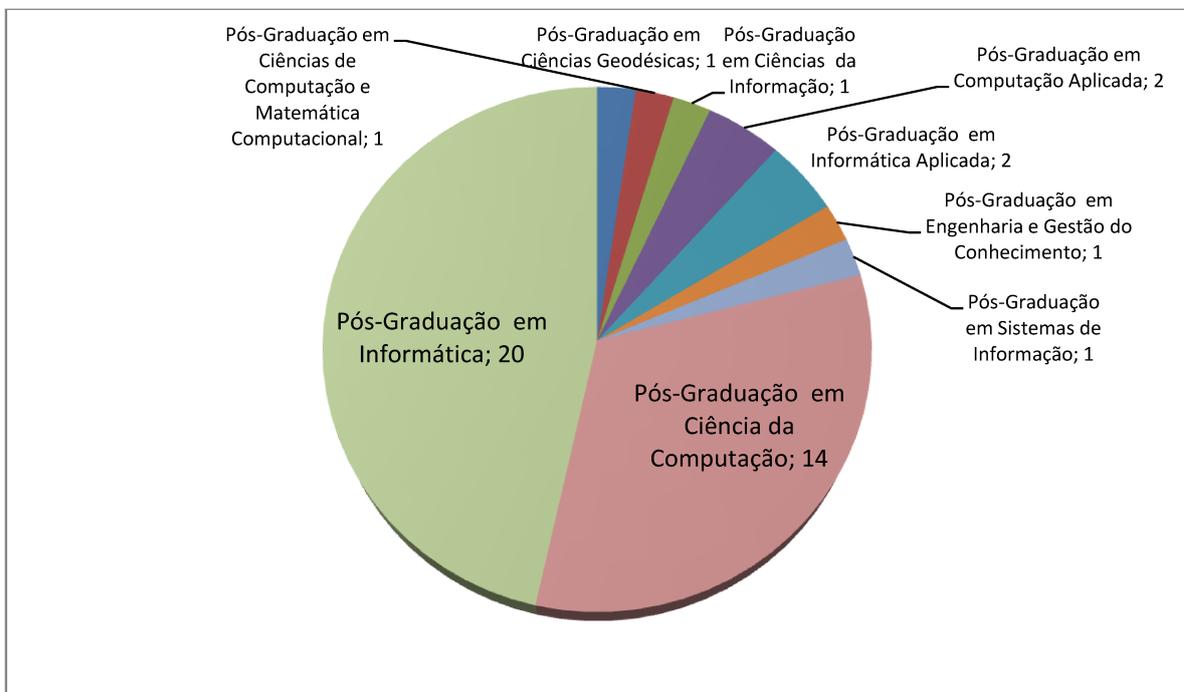
Senso e Arroyo Machado (2018) observaram também a predominância de iniciativas desenvolvidas por bibliotecas nacionais no exterior, como os projetos desenvolvidos na Biblioteca Nacional da Espanha, na Biblioteca Britânica, na

Biblioteca Nacional da França, na Europeia e na Biblioteca do Congresso Norte Americano e identificaram que elas fornecem o conhecimento necessário para saber quais são as metodologias mais utilizadas atualmente para realizar a conversão de registros bibliográficos para *Linked Data*.

Por outro lado, as bibliotecas universitárias estão se movimentando mais lentamente em direção a tais iniciativas. Rio Branco (2020) destacou que as bibliotecas universitárias ainda apresentam dificuldades quanto à disseminação de vocábulos semânticos, gerando isolamento entre elas, prejudicando a interoperabilidade com outros vocábulos.

Na área da pesquisa, Byrne e Goddard (2010) apontaram que é pequena a publicação de estudos sobre dados abertos conectados na literatura da biblioteconomia, o que pode ser observado nos anais de conferências de bibliotecas e nos principais bancos de dados de artigos da área. O gráfico a seguir demonstra que essa afirmativa também se confirma no Brasil.

GRÁFICO 1 - Quantidade de trabalhos por Programas de Pós-graduação relacionados ao tema *Linked Data* (2006-2016)



Fonte: Adaptado de Arakaki, 2016.

Dentro desse cenário, a pergunta que norteou a realização desta pesquisa foi: *Quais os desafios são encontrados pelas bibliotecas para a disponibilização dos catálogos online como dados abertos conectados?*

O presente estudo buscou responder a essa questão podendo assim abrir novos horizontes de reflexão para a evolução no processo de compartilhamento de dados no âmbito da biblioteconomia. Para responder ao questionamento proposto, foi seguido à idealização dos objetivos gerais e específicos, que serão evidenciados nos próximos tópicos, assim como também a estrutura da pesquisa.

1.2 OBJETIVOS

1.2.1 Objetivo geral

O objetivo geral desta pesquisa é mapear os desafios encontrados na implementação de práticas de dados abertos conectados em bibliotecas para a disponibilização de seus catálogos online na *Web*.

O termo “desafio”, na definição do dicionário Michaelis (2021) denota o “ato de desafiar; desafiação”. O dicionário apresenta também a definição em sentido figurado de que “desafio” é uma “situação ou problema cujo enfrentamento demanda esforço e disposição firme”, definição esta que representa o uso do termo desafio nessa pesquisa.

Também deve ser esclarecido que, enquanto o termo "dados conectados" se refere à interoperabilidade técnica dos dados, "dados abertos" se concentra em sua interoperabilidade legal. Em princípio, os dados abertos são livremente utilizáveis, reutilizáveis e redistribuíveis - sujeitos, no máximo, aos requisitos de atribuição e compartilhamento semelhantes. Dados conectados em si não exige que os dados sejam abertos, embora o potencial da tecnologia seja melhor explorado quando os dados são publicados como dados abertos conectados (BAKER et al., 2011).

1.2.2 Objetivos específicos

Para alcançar o objetivo geral, foram definidos como objetivos específicos:

Identificar as barreiras encontradas na implementação de práticas de dados abertos conectados em bibliotecas para a disponibilização de catálogos online na *Web*;

- 1) Apresentar as potencialidades dos registros bibliográficos para uso em formato de dados abertos para conectar os catálogos das bibliotecas;
- 2) Identificar na literatura os requisitos para publicação de dados abertos conectados.

1.2.3 Estrutura da pesquisa

Para a apresentação, desenvolvimento e alcance dos objetivos propostos para esta pesquisa, este trabalho está subdividido em seis capítulos, a saber:

Capítulo 1: Introdução – descrevem-se os temas gerais e os fundamentos que foram abordados no desenvolvimento desta pesquisa, bem como o problema, as justificativas de investigação, os objetivos e a estrutura formal do conteúdo desta dissertação.

Capítulo 2: Fundamentos teórico-conceituais – apresentam-se as bases teóricas e conceituais desta pesquisa, principalmente as ideias gerais e autores que dialogaram e enriqueceram o estudo.

Capítulo 3: Antecedentes teórico-metodológicos: discorre-se sobre os principais estudos mapeados na literatura sobre os desafios enfrentados por bibliotecas para disponibilização de seus catálogos como dados abertos conectados, identificando autores e trabalhos que orientaram os procedimentos utilizados para alcançar os objetivos propostos nesta pesquisa.

Capítulo 4: Metodologia e procedimentos – detalha-se a caracterização e delimitação da pesquisa; o percurso metodológico, incluindo detalhes de como foi realizada a pesquisa bibliográfica e documental, assim como a seleção dos trabalhos e o conjunto de procedimentos gerais realizados ao longo desta pesquisa.

Capítulo 5: Resultados – apresentação e discussão dos resultados alcançados pela pesquisa.

Capítulo 6: Considerações finais – apresentam-se algumas reflexões sobre os objetivos propostos por esta pesquisa e sugerem-se possibilidades para futuras investigações. Apresentada essa conjectura da estrutura do estudo, será seguido agora para a fundamentação teórica.

2. FUNDAMENTAÇÃO TEÓRICA

Este capítulo destina-se a apresentar os fundamentos teóricos e conceituais que sustentam as discussões para a análise do problema de pesquisa proposto neste estudo, que estão relacionados à aplicabilidade de dados abertos conectados a partir dos catálogos online de bibliotecas.

Portanto, o presente capítulo tem a seguinte propositura organizacional: Catalogação, catálogos e registros bibliográficos; Dados abertos conectados e *Web* semântica: conceitos, aplicabilidade e tecnologias necessárias para implementação; e Fontes de dados conectados. Nessa perspectiva, discorre sobre o histórico da evolução dos catálogos e da catalogação, os elementos e conceitos da *Web* semântica e dos dados abertos conectados e as possibilidades de aplicação em bibliotecas, refletindo os resultados nos catálogos.

2.1 CATALOGAÇÃO, CATÁLOGOS E REGISTROS BIBLIOGRÁFICOS.

A essência do trabalho do bibliotecário, em termos mais simplificados, pode ser resumida em organizar, tratar e disseminar conhecimentos registrados para diferentes categorias de usuários a partir de uma série de características dessas mesmas categorias (MEY; SILVEIRA, 2009).

Nesse contexto, as bibliotecas são formadas por coleções de fontes de informações armazenadas em variados suportes e formatos, cuja representação e disponibilização ao público são realizadas por meio do catálogo, objeto da catalogação (SERRA; SANTAREM SEGUNDO, 2017).

Sob esse prisma, a catalogação, na visão de Mey e Silveira (2009, p.7), consiste em:

Um conjunto de informações que simbolizam um registro do conhecimento. [...] Pode definir catalogação como o estudo, preparação e organização de mensagens, com base em registros do conhecimento, reais ou ciberespaciais, existentes ou passíveis de inclusão em um ou vários acervos, de forma a permitir a interseção entre as mensagens contidas nestes registros do conhecimento e as mensagens internas dos usuários.

O objetivo e a finalidade da catalogação, assim como dos catálogos oriundos dessa técnica é atender ao público ao que se destina. Essa visão reflete as Cinco Leis da Biblioteconomia, formuladas por Ranganathan em 1931, que estabelece o

seguinte: livros são para usar; a cada leitor seu livro; a cada livro seu leitor; poupe o tempo do leitor e, por último, a biblioteca é um organismo em crescimento (RANGANATHAN, 2009).

Em 1953, Seymour Lubetzky teve publicada sua obra intitulada “*Cataloging Rules and Principles*”. A teoria supracitada serviu de base para as discussões da Conferência Internacional sobre os princípios da Catalogação, realizada em Paris em 1961 (FIUZA, 1987; MACHADO; ZAFALON, 2020).

A Declaração dos Princípios Internacionais de Catalogação, promulgada em 1961, ampliada em 2009 e atualizada em 2016, tem o objetivo de servir como base para a normalização internacional da catalogação. Esse objetivo foi atingido, tendo em vista que grande parte dos códigos de catalogação desenvolvidos mundialmente, desde então, vem seguindo os princípios propostos no documento, ou mesmo, a sua maioria (IFLA, 2016).

Os Princípios definidos pela IFLA (2016) são os seguintes:

- a) Interesse do usuário;
- b) Uso Comum;
- c) Representação;
- d) Precisão;
- e) Suficiência e Necessidade;
- f) Significação;
- g) Economia;
- h) Coerência e Normalização;
- i) Integração;
- j) Interoperabilidade;
- k) Abertura;
- l) Acessibilidade; e
- m) Racionalidade.

Segundo a IFLA (2016, p. 5), os princípios acima “regem a criação e o desenvolvimento de códigos de catalogação, as decisões que tomam os catalogadores e as políticas sobre o acesso e intercâmbio de dados”.

No entanto, apesar da primeira publicação dos Princípios ter ocorrido em 1961, os primeiros teóricos da Catalogação surgem a partir do século XIX trazendo

contribuições para as práticas até então realizadas com a formulação de princípios, conforme apresentado por Ortega (2011). É também no Século XIX que, segundo Galvão e Lopes (2000, p. 2) foi dado início a “catalogação moderna, a catalogação baseada em axiomas e normas”.

Em 1831 Anthony Panizzi, refugiado político italiano e advogado de profissão, foi contratado como bibliotecário assistente pelo Museu Britânico para coordenar as atividades de revisão do catálogo da instituição. Após quatro anos de trabalho, a proposta apresentada por Panizzi foi aceita por uma comissão encarregada de aprovar o novo código, que ficou conhecido como “as noventa e uma regras” de Panizzi (MEY, 1995).

Para Ortega (2011), Panizzi implementou a noção de obra como modo de reunir documentos contidos na sua ideologia sobre a catalogação. As 91 regras de Panizzi estabeleceram as funções que um catálogo deve realizar para atingir seus objetivos, sendo as principais destacadas por Fiuza (1987, p. 47):

1. Um livro deve ser considerado e representado no catálogo, não como uma entidade separada, mas como uma edição de determinada obra, de um determinado autor;
2. Todas as obras de um autor, e suas edições, devem ter entradas sob um nome definido, usualmente o nome original do autor, independentemente dos diferentes nomes que aparecem nas diferentes obras e edições;
3. Todas as edições e traduções de uma obra, independentemente de seus títulos individuais, devem ter entradas sob seu título original, numa ordem prescrita (edições cronologicamente, traduções por língua, etc.) de maneira que a pessoa em busca de um livro em particular, encontre-o junto com as outras edições, dando ensejo a uma escolha da edição que melhor sirva a seus objetivos;
4. Referências apropriadas devem ser feitas para auxiliar o usuário a encontrar a obra desejada.

Para Fiuza (1987), o catálogo deve ser visto como um todo. O livro procurado por uma pessoa não é realmente, na maioria das vezes, o objeto de seu interesse, mas a obra nele contida; esta obra pode ser encontrada em outras edições, traduções e versões, publicada sob diferentes nomes do autor e diferentes títulos. Nessa perspectiva, para servir bem ao usuário, o catálogo deve ser planejado para revelar todas as edições, versões, etc. das obras, bem como outras obras geneticamente relacionadas que existem em uma biblioteca (FIUZA, 1987).

As regras de Panizzi influenciaram Charles Coffin Jewett que, em 1850, nos Estados Unidos, teve seu código aceito para o catálogo do *Smithsonian Institution*. Alguns dos princípios estabelecidos por Jewett estão relacionados aos cabeçalhos

de responsabilidade e obras anônimas e são seguidos até os dias atuais (MEY, 1995).

Em 1856, Charles Ami Cutter publicou sua obra intitulada “*Rules for a Dictionary Catalogue*” (Regras para um Catálogo Dicionário), definindo os objetivos e funções do catálogo. Segundo Barbosa (1978), Cutter preconizava que o catálogo deve ser um instrumento que permita encontrar um livro do qual se conheça o autor, o título ou o assunto e deve mostrar o que existe em uma coleção de um determinado autor ou assunto.

Mey e Silveira (2009, p. 71) complementam que a obra de Cutter é um exemplo de catalogação prática adequada ao usuário, ressaltando que:

Cutter, considerado por Ranganathan como um gênio da Biblioteconomia, não elaborou apenas um código de catalogação e sua obra é na verdade uma declaração de princípios. As regras são entremeadas com os porquês das soluções e com observações diversas, [...]. De fácil leitura, constitui um código muito completo, incluindo a catalogação de assunto e de materiais especiais, normas de transliteração e elaboração de catálogos auxiliares.

Os catálogos de bibliotecas são ferramentas de mediação entre o usuário e o acervo. É por intermédio dos catálogos que se realizam as pesquisas nos registros bibliográficos em busca da informação que se pretende obter, seja ela em texto completo, seja em referencial, que consiste em direcionar o usuário a localização física do material na estante ou para uma base de dados.

Para Mey e Silveira (2009, p. 12) o catálogo pode ser definido como:

Um meio de comunicação, que veicula mensagens sobre os registros do conhecimento, de um ou vários acervos, reais ou ciberespaciais, apresentando-os com sintaxe e semântica próprias e reunindo os registros do conhecimento por semelhanças para os usuários desses acervos. O catálogo explicita, por meio das mensagens, os atributos das entidades e os relacionamentos entre elas.

A criação dos registros bibliográficos é feita pelo bibliotecário durante a catalogação. Este processo consiste na técnica de extrair de determinada obra, descritores com base em regras e normas pré-estabelecidas. Tem como objetivo realizar a representação descritiva e temática dos itens e alimentar o catálogo, com vistas a garantir o acesso e a recuperação da informação por parte do usuário.

Nos períodos mais remotos a informação não era organizada de forma sistematizada e unificada. Os suportes informacionais da época eram os tabletes de

argila, madeira, papiro, pergaminho e papel. Os primeiros registros bibliográficos datam de 1.300 a.C. e eram representados em tabletes, com as primeiras informações bibliográficas de descrição física encontrados em escavações hititas (atual Turquia). Continham o número do tablete em uma série, o título e o nome do escriba, conforme Mey (1995).

Com a invenção da imprensa, os catálogos passaram a ser criados por bibliógrafos e livreiros, interessados, apenas, na compilação de seus catálogos e bibliografias (BARBOSA, 1978). Machado (2003, p.42) ao citar Melo (1981) e Pinto (1987) explicita que:

Os primeiros catálogos e bibliografias são puramente listas inventariais e não instrumentos bibliográficos. Nas bibliografias, a ênfase é dada aos autores e não aos livros, são biobibliográficas; nos catálogos, únicos tipos de listas bibliográficas, a caracterização dá-se pela técnica pouco elaborada, falta de arranjo e transcrição sucinta e pouco precisa dos títulos.

Lubetzky destaca dois objetivos principais do catálogo:

1) mostrar se a biblioteca possui ou não um determinado item, publicado sob um nome de autor ou sob um título determinado; e 2) identificar o autor e a obra representados pelo item e relacionar as várias obras do autor e as várias edições da obra (FIUZA, 1987, p. 50).

Com o início da catalogação moderna, que contou com a contribuição das obras de Panizzi, Jewett, Cutter, Lubetzky e Ranganathan; é percebido o desenvolvimento da prática e da teoria da catalogação que passa a ser regida por regras e padrões (MACHADO; ZAFALON, 2020). Essas estão vinculadas a uma categorização que, embora possam parecer divergentes entre elas, foram definidas sob diferentes propósitos e variáveis, cabendo ao pesquisador escolher a categoria que melhor se ajusta aos seus objetivos (ASSUMPÇÃO, 2012).

O Quadro 1 destaca, em linhas gerais, regras e padrões bibliográficos, bem como à categoria que pertencem e o ano em que foram implantados.

QUADRO 1 – Regras e padrões bibliográficos

Ano de Implantação	Padrão Bibliográfico	Categorização
1968	MARC	Formato de Metadados
1971	ISBD	Regras de Catalogação
1977	UNIMARC	Formato de Metadados

1978	AACR-2	Regras de Catalogação
1981	ISO 2709	Sintaxe de Codificação
1997	MARC-21	Formato de Metadados
1998	XML 1.0	Sintaxe de Codificação
1998	FRBR	Modelo Conceitual
2002	MODS	Formato de Metadados
2009	FRAD	Modelo Conceitual
2010	FRSAD	Modelo Conceitual
2010	RDA	Regras de Catalogação
2017	LRM	Modelo Conceitual

Fonte: Adaptado de Assumpção, 2012.

Com o passar dos anos, alguns novos desafios relacionados à catalogação e à necessidade de evolução por parte das bibliotecas foram surgindo, produtos do advento da internet. A evolução resultou, entre outras mudanças, na substituição das fichas catalográficas manuscritas e impressas para os registros bibliográficos legíveis por máquina; dos catálogos manuais para os online até a migração dos catálogos para o ambiente da *Web* onde a introdução da interoperabilidade se faz cada dia mais necessária.

Para Isotani e Bittencourt (2015), o uso de tecnologia de dados abertos permite que diferentes organizações e sistemas trabalhem juntos. Isso ocorre porque essas organizações e esses sistemas podem interoperar com dados abertos, expandindo assim a comunicação e melhorando o desenvolvimento eficiente de sistemas complexos. Para isso, os dados devem ser acessíveis, legíveis por máquina, em formato aberto e informações geradas por todos.

Na década de 1960 quando a Biblioteca do Congresso Norte Americano começou a usar computadores ela desenvolveu o formato LC MARC, um sistema que usava números, letras e símbolos breves dentro do próprio registro de catalogação para marcar diferentes tipos de informação. Esse formato original evoluiu para o MARC 21 e é um padrão de metadados usado por muitos programas de computadores de bibliotecas. O MARC 21, bem como toda a sua documentação oficial é mantido pela Biblioteca do Congresso e publicado como Formato MARC 21 para Dados Bibliográficos (*LIBRARY OF CONGRESS*, 2009).

Senso e Arroyo Machado (2018) observaram que trocar registros bibliográficos não é o mesmo que trocar dados. Esse é um dos fatores que inviabiliza a utilização do formato MARC neste novo ambiente de dados conectados na *Web*. Sobre a composição dos registros bibliográficos os autores afirmaram:

Um registro bibliográfico é composto por inúmeros dados (autor, título, local, editora, data e outros dados) que, conectados a dados diferentes de outros conjuntos de dados, poderiam fornecer informações individuais sobre cada um deles, oferecendo ao usuário infinitas novas informações (onde aquele autor nasceu, história local daquele lugar, dados contemporâneo) e, para a biblioteca, a possibilidade de expandir seus horizontes para além do catálogo. E isso é algo que não é possível fazer utilizando o formato MARC (SENSO; ARROYO MACHADO, p.2, tradução nossa).

Os dados no formato MARC 21 precisam ser convertidos para o modelo de dados RDF antes de serem usados como dados abertos conectados. No entanto, o processo de conversão ocasiona muitos problemas, como, por exemplo, a perda da consistência dos dados (ALI; WARRAICH, 2018).

Observa-se a evolução da catalogação e a introdução de novos conceitos e padrões que procuram adequar a organização da informação às novas realidades que o ambiente *Web* demanda. “A publicação de guias, tutoriais e de estudos que apontam tendências e perspectivas da catalogação no século XXI constituem reflexo inequívoco dessa evolução” (BAPTISTA, 2006, p. 4).

A IFLA, motivada pela necessidade de estudos a respeito de novas propostas para padrões bibliográficos, desenvolveu o modelo conceitual denominado: Requisitos Funcionais para Registros Bibliográficos (FRBR), do Inglês “*Functional Requirements for Bibliographic Records*”.

Os FRBR surgiram com dois objetivos principais: fornecer uma estrutura para relacionar os dados registrados às necessidades dos usuários; e recomendar um nível básico de funcionalidade para os registros bibliográficos (IFLA, 2009). O modelo FRBR objetiva, portanto, “reestruturar os registros bibliográficos a fim de reorganizar os elementos por meio da análise lógica dos dados refletidos nesses registros, gerando a descrição dos elementos básicos do Modelo E-R” (SANTOS, 2011, p.11).

O Modelo Entidade- Relacionamento (E-R) que embasa a estrutura do FRBR foi proposto por Chen (1976) e define a representação da informação

baseada em entidades, atributos e em relacionamentos entre as entidades (SANTOS, 2011).

O desenvolvimento e a evolução da catalogação vêm sendo impactados expressivamente pelo modelo FRBR. Segundo Coyle (2016, p. 65, tradução nossa):

O modelo de dados bibliográficos FRBR é a mudança de pensamento mais radical sobre os catálogos bibliográficos desde que Panizzi desenvolveu suas noventa e uma regras para o catálogo do British Museum. [...] O FRBR é alternadamente visto como uma análise das necessidades do usuário, como uma descrição do fluxo de catalogação e como um modelo de dados para um futuro formato de registro bibliográfico.

A mudança impactante trazida pelo modelo também é enfatizada por Santos (2011) ao explicitar que a inovação trazida pelo FRBR se reflete no fornecimento de um quadro estruturado para associação de dados em registros bibliográficos e na proposta de um nível funcional básico. Os catálogos baseados nesse modelo mostram as relações bibliográficas de uma maneira cada vez mais clara, tornando-se uma ferramenta que auxilia os usuários no uso de sistema de informação automatizado.

O modelo FRBR visa aprimorar a construção de registros bibliográficos, dos catálogos e da catalogação. Eles foram criados para estabelecer o conceito de múltiplos tipos de materiais descritos em bases de dados pertencentes a bibliotecas ou unidades de informação. Destinam-se também a tornar os catálogos automáticos mais simplificados aos olhos dos usuários e a promover a compatibilidade internacional de seu uso, de forma que a recuperação de registros bibliográficos se torne cada vez mais efetiva e universal, ajudando a reexaminar os princípios básicos da catalogação descritiva (SANTOS, 2011).

Mey e Silveira (2009) destacaram que os FRBR não são um código de catalogação e, dessa forma, não descrevem a forma de apresentação dos elementos descritivos: trata-se de um modelo conceitual.

Um modelo conceitual, chamado de modelo abstrato no mundo dos sistemas, é convenientemente usado para entendermos uma determinada coisa, para automatizar ou organizar processos que o circundam, compreendendo suas características essenciais e em detalhes (PACHECO; ORTEGA, 2015, p. 65).

Os requisitos estabelecidos pelo FRBR a partir do modelo de análise Entidade - Relacionamento (E-R) para bancos de dados são assim definidos:

- 1) As entidades dos registros bibliográficos: abrem um novo tipo de abordagem dos registros bibliográficos;

2) Para os FRBR, as entidades representam os principais objetos de interesse para os usuários de dados bibliográficos. A entidade é o fator primordial de um banco de dados E-R; Os atributos de cada uma das entidades: são características das entidades; e As relações entre as entidades: fundamentos dos catálogos e dos bancos de dados se fazem entre as entidades (MEY; SILVEIRA, 2009, p.18).

O modelo FRBR é composto por 10 entidades que são apresentadas em três grupos distintos:

Grupo 1 (o que): representa o recurso que está sendo descrito, produto de esforço intelectual ou artístico e possui quatro entidades: Obra, Expressão, Manifestação e Item;

Grupo 2 (quem): representa os agentes que têm relações com as entidades do Grupo 1, ou seja, entidades responsáveis pelo conteúdo intelectual ou artístico, pela produção física e disseminação ou pela guarda das entidades do Grupo 1: Pessoas e Entidades Coletivas, e;

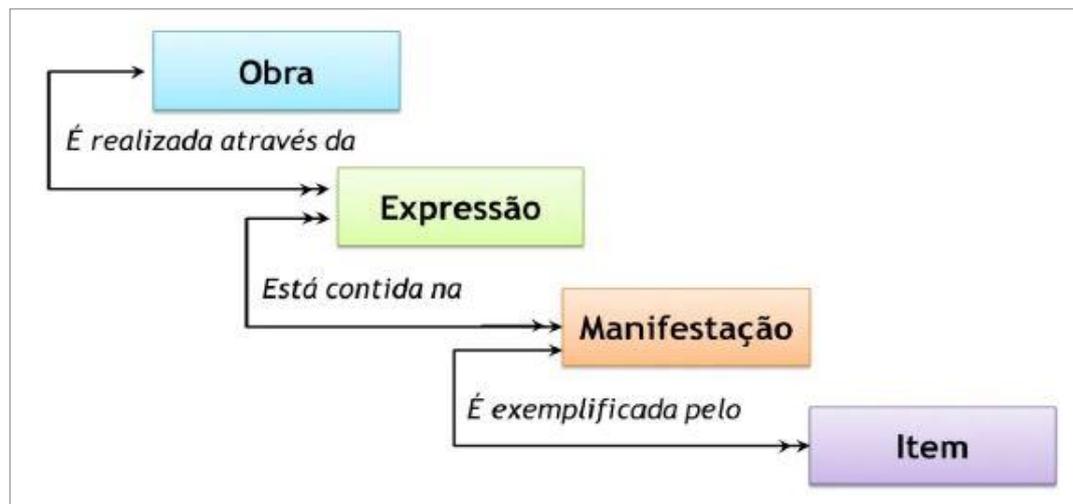
Grupo 3 (sobre): abarca os elementos adicionais de representação de tema das obras: Conceito, Objeto, Lugar e Evento (PACHECO e ORTEGA, 2015, p. 65).

O modelo FRBR permite que os atributos das entidades sejam organizados nas bases de dados por meio da relação bibliográfica das 10 entidades propostas no modelo, agrupando-os e conectando-os entre si (PACHECO; ORTEGA, 2015). O modelo define os relacionamentos entre as entidades que, segundo a IFLA (1998):

Servem como um meio para descrever ligações entre uma entidade e outra, e assim como meio para ajudar o usuário a "navegar" no universo que é representado por uma bibliografia, catálogo ou base de dados bibliográficos (IFLA, 1998, p. 55).

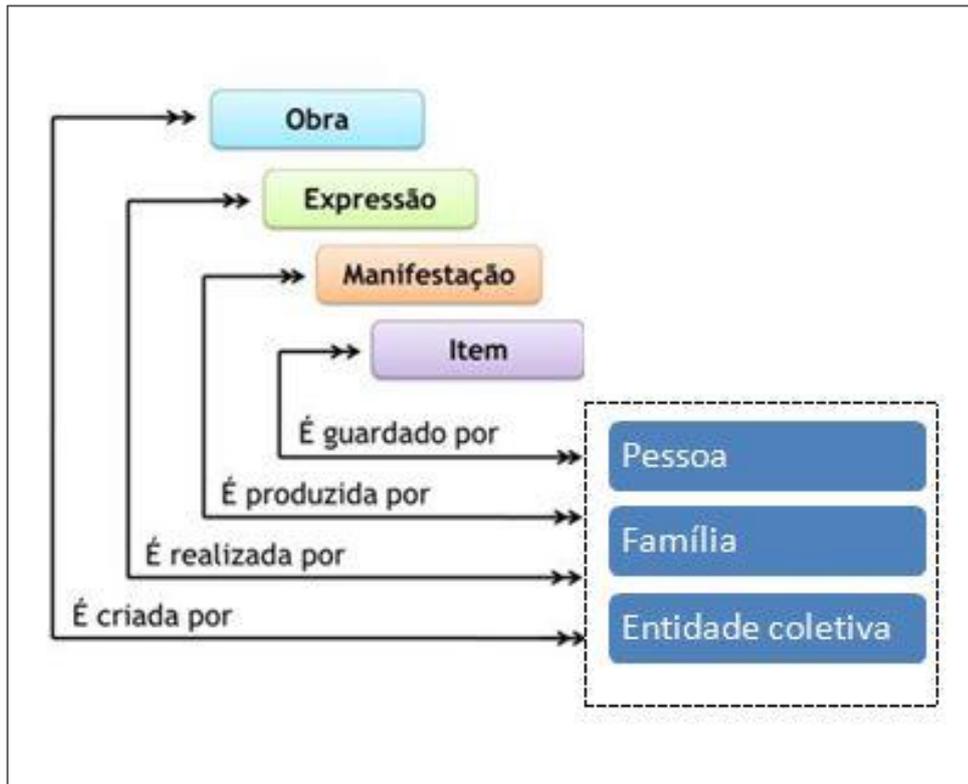
As figuras 1, 2 e 3 ilustram estes relacionamentos:

FIGURA 1 – Relacionamentos entre as entidades do Grupo 1 do FRBR



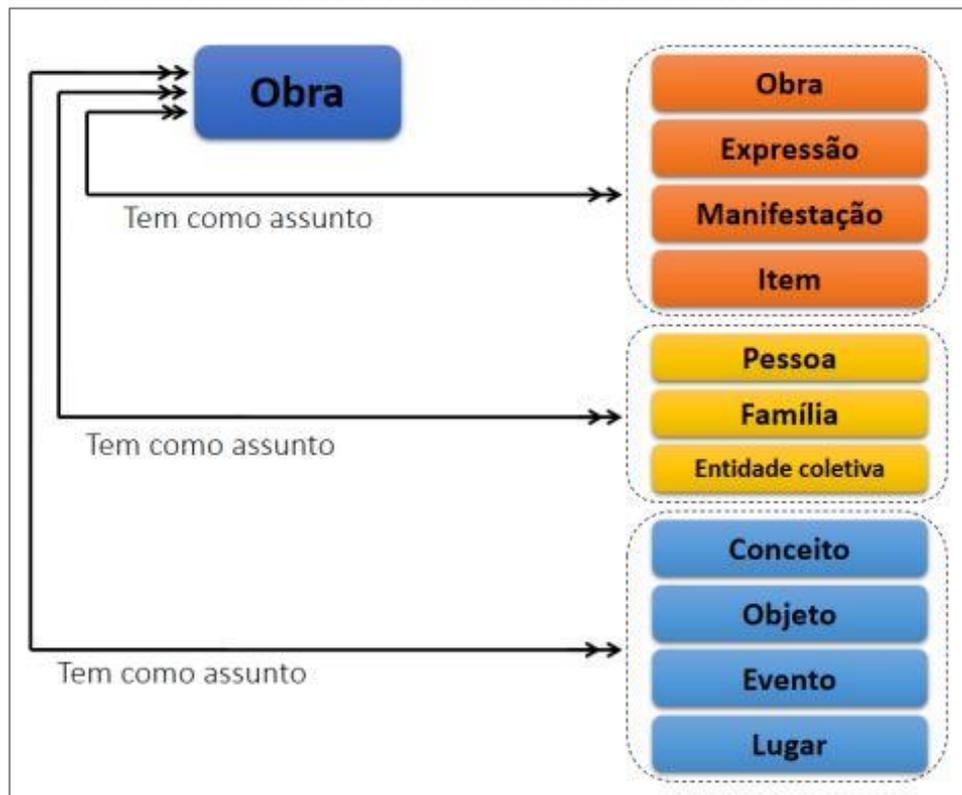
Fonte: Assumpção, 2012.

FIGURA 2 – Relacionamentos entre as entidades do Grupo 1 e 2 do FRBR



Fonte: Adaptado de Assumpção, 2012.

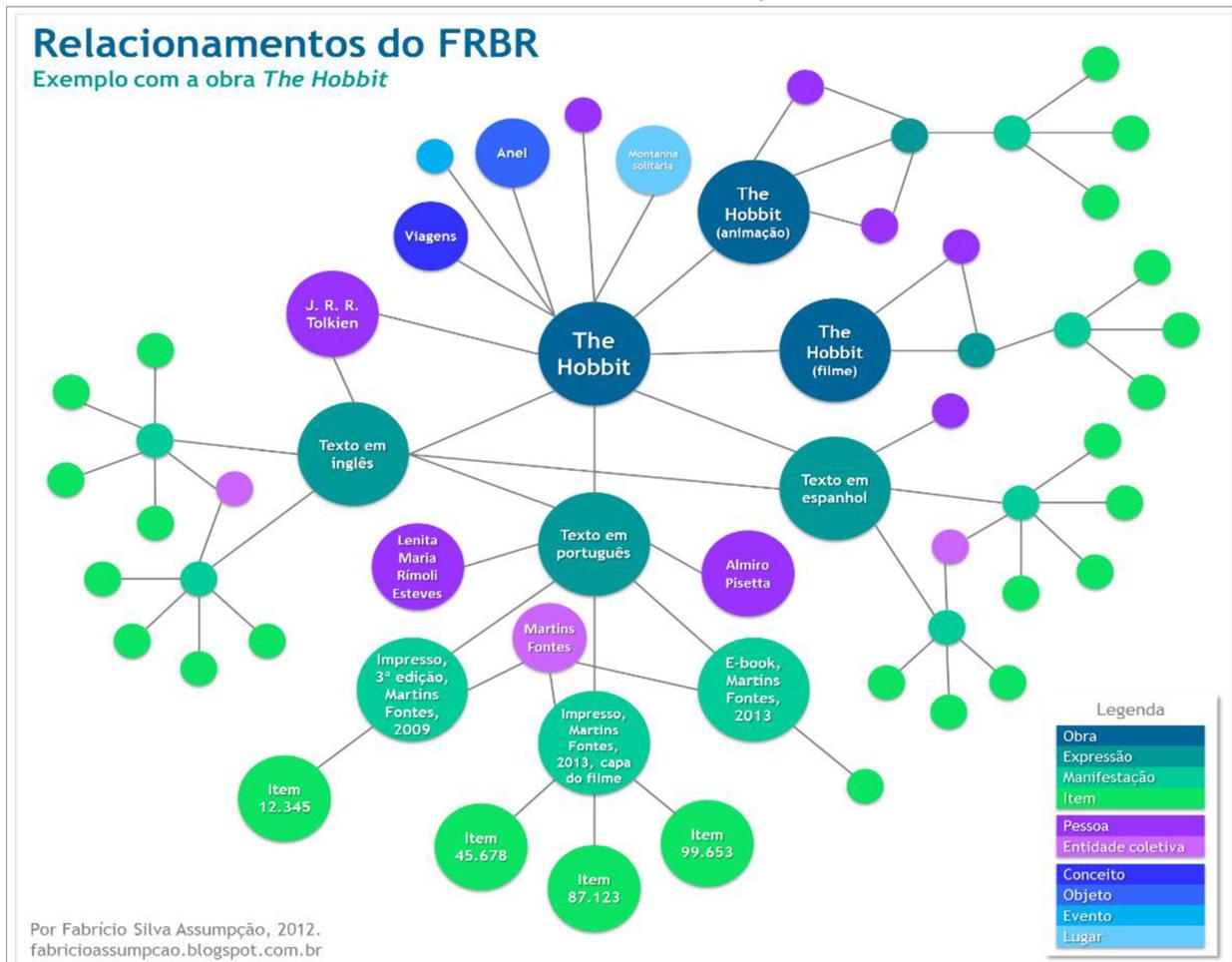
FIGURA 3 – Relacionamentos de assunto do modelo FRBR



Fonte: Adaptado de Assumpção, 2012.

Assumpção (2012) utiliza a obra *The Hobbit* para exemplificar as entidades, os atributos e os relacionamentos existentes na obra, baseado no modelo conceitual FRBR, conforme mostrado na figura 4:

FIGURA 4 – Relacionamentos do FRBR: exemplo com a obra *The Hobbit*



Fonte: Assumpção, 2012.

O FRBR define também, além das entidades, atributos e relacionamentos, quatro tarefas dos usuários, que são ações realizadas pelo usuário durante a utilização de um catálogo ou de dados bibliográficos, e associam cada atributo e relacionamento com uma (ou mais) dessas tarefas (ASSUMPÇÃO, 2012). As tarefas dos usuários definidas são encontrar, identificar, selecionar e obter, especificadas a seguir:

- 1) Encontrar entidades que correspondem ao critério de busca do usuário (localizar uma única entidade ou um grupo de entidades em um arquivo ou banco de dados como o resultado de uma busca usando um atributo ou relacionamento da entidade);

- 2) Identificar uma entidade (confirmar que a entidade descrita corresponde à entidade procurada ou distinguir entre duas ou mais entidades com características similares);
- 3) Selecionar uma entidade apropriada às necessidades do usuário (escolher uma entidade que satisfaça os requisitos do usuário com respeito ao conteúdo, forma física, etc., ou para rejeitar uma entidade por ser inapropriada às necessidades do usuário);
- 4) Adquirir ou obter acesso à entidade descrita (adquirir uma entidade por comprar, empréstimo, etc., ou acessar uma entidade eletronicamente por uma conexão online com um computador remoto) (ASSUMPÇÃO, 2012, não paginado).

O modelo conceitual FRBR apresentado, sua terminologia, os relacionamentos e as tarefas de usuários auxiliam na revisão das tradições catalográficas atuais à luz dos ambientes digitais, contribuindo para a padronização internacional e redução de custos com a catalogação em escala global (TILLET, 2003).

A família inicial de modelos conceituais FR incluía, além do FRBR para dados bibliográficos, os *Functional Requirements for Authority Data* (FRAD) e os *Functional Requirements for Subject Authority Data* (FRSAD). O FRBR, O FRAD e o FRSAD foram desenvolvidos de maneira independente por diferentes grupos de trabalho por muitos anos, conforme síntese apresentada a seguir (IFLA, 2017):

- **O FRBR** foi apresentado no relatório final do IFLA *Study Group on the Functional Requirements for Bibliographic Records*. O *Study Group* foi constituído em 1992 e o relatório foi aprovado pelo *Standing Committee of the Section on Cataloguing* em 5 de setembro de 1997.
- **O FRAD** foi o resultado do IFLA *Working Group on Functional Requirements and Numbering of Authority Records* (FRANAR). O FRANAR foi estabelecido em abril de 1999 pela *Division of Bibliographic Control and the Universal Bibliographic Control and International MARC Programme* (UBCIM). O relatório foi aprovado pelos *Standing Committees of the Cataloguing Section and the Classification and Indexing Section* em março de 2009.
- **O FRSAD** foi apresentado no relatório do IFLA *Working Group on the Functional Requirements for Subject Authority Records* (FRSAR), formado em 2005. O relatório foi aprovado pelo *Standing Committee of the IFLA Section on Classification and Indexing* junho de 2010 (IFLA, 2017, p.5).

Apesar de esses três modelos FR serem todos criados na estrutura de modelagem de dados entre entidades, eles adotam diferentes pontos de vista e diferentes soluções para problemas comuns. Embora as modelagens sejam necessárias em um sistema bibliográfico completo, para adotá-las em um único sistema é necessário resolver problemas complexos de forma temporária com o mínimo de orientação do modelo. Mesmo quando o FRAD e o FRSAD estavam

sendo finalizados em 2009 e 2010, ficou claro que seria necessário mesclar a família FR em um modelo unificado para estabelecer a compreensão do modelo comum e remover os obstáculos para a adoção do modelo (IFLA, 2017).

Criou-se, dessa forma, o IFLA LRM, um modelo conceitual de referência para dados de bibliotecas, resultado da junção dos três modelos conceituais da família FR (FRBR, FRAD e FRSAD) e é, ao mesmo tempo, um modelo distinto dos outros três modelos (MACHADO; ZAFALON, 2020).

Considerando que o IFLA LRM surgiu com o propósito de resolver as inconsistências entre os modelos da família FR, houve a junção de algumas entidades, atributos e relacionamentos dos três modelos. Do modelo FRBR foram trazidas para o IFLA LRM as entidades obra, expressão, manifestação e item; do FRAD, a entidade pessoa passa a ser subordinada à entidade agente que, por sua vez, também tem a ela subordinada a entidade agente coletivo; do FRSAD, advém a entidade nome. Além das entidades agente e agente coletivo, a intervalo de tempo também é nova no IFLA LRM (MACHADO; ZAFALON, 2020). No total são onze as entidades definidas pelo modelo, identificadas no próximo quadro:

QUADRO 2 – Identificação das entidades no IFLA LRM

Identificação	Entidades	Definição	Nota de escopo (síntese)	Exemplo(s)
LRM-E1	<i>Res</i>	Qualquer entidade no universo do discurso	Res (“coisa” em latim) é a entidade superior do modelo. Res inclui coisas e conceitos materiais ou físicos. Tudo o que é considerado relevante para o universo bibliográfico, que é o universo do discurso neste caso, está incluído. Res é uma superclasse de todas as outras entidades explicitamente definidas, bem como de quaisquer outras entidades não especificamente identificadas.	{Homer’s Odyssey} [obra grega antiga]; {Agatha Christie} [pessoa, autora de romances policiais]

LRM-E2	Obra	O conteúdo intelectual ou artístico de uma criação específica	Uma <i>obra</i> é uma entidade abstrata que permite o agrupamento de expressões que são consideradas equivalentes funcionais ou quase equivalentes. Uma <i>obra</i> é um objeto conceitual, nenhum objeto material único pode ser identificado como a <i>obra</i> .	{Homer's <i>Odyssey</i> }
LRM-E3	Expressão	Uma combinação específica de sinais que transmitem conteúdo intelectual ou artístico	Uma expressão é uma combinação específica de sinais de qualquer forma ou natureza (incluindo sinais visuais, auditivos ou gestuais) destinada a transmitir conteúdo intelectual ou artístico e identificável como tal. O termo "sinal" é entendido aqui tal qual o significado utilizado na semiótica. Uma expressão é uma entidade abstrata distinta dos suportes utilizados para gravá-la.	A tradução para o inglês de Homer's <i>Odyssey</i> , por Robert Fagles, copyright 1996
LRM-E4	Manifestação	Um conjunto de todos os suportes que supostamente compartilham as mesmas características do conteúdo intelectual ou artístico e aspectos da forma física. Esse conjunto é definido pelo conteúdo geral e pelo plano de produção para seu suporte ou suportes.	Uma manifestação resulta da captura de uma ou mais Expressões em um suporte ou conjunto de suportes. Como uma entidade, a manifestação representa as características comuns compartilhadas por esses suportes, tanto no conteúdo intelectual ou artístico, quanto na forma física.	Homer. <i>The Odyssey / translated by Robert Fagles</i> , Penguin Classics, Deluxe edition published in New York by Penguin Books in 1997, ISBN 0-670-82162-4 [manifestação contendo o texto completo da tradução De Robert Fagles do poema grego]

LRM-E5	Item	Um ou vários objetos que carregam sinais destinados a transmitir conteúdo intelectual ou artístico	Em termos de conteúdo intelectual ou artístico e forma física, um item exemplificando uma manifestação normalmente reflete todas as características que definem a própria manifestação.	O manuscrito conhecido como <i>Codex Sinaiticus</i>
LRM-E6	Agente	Uma entidade capaz de ações deliberadas, de cessão de direitos e de ser responsabilizada por suas ações	A entidade agente é uma superclasse estritamente equivalente à união das entidades pessoa e do agente coletivo. Ela é definida para reduzir a redundância no modelo, fornecendo uma única entidade para servir como domínio ou abrangência de certos relacionamentos que se aplicam a todos os tipos específicos de agentes.	{Rainha Victoria}; {Orquestra Sinfônica da BBC}
LRM-E7	Pessoa	Um ser humano individual	A entidade <i>pessoa</i> é restrita a pessoas reais que vivem ou presume-se que tenham vivido.	{Pythagoras}; {Marco Polo}; {Agatha Christie}
LRM-E8	Agente coletivo	Uma reunião ou organização de pessoas com um nome específico e capaz de atuar como uma unidade	A entidade agente coletivo designa uma ampla variedade de grupos nomeados de pessoas que possuem um nome específico e têm o potencial de agir em conjunto como uma unidade. Além de famílias, instituições comerciais ou corporativas e outros órgãos legalmente registrados, a entidade agente coletivo inclui organizações e associações, grupos musicais, artísticos ou performáticos, governos e qualquer uma de suas subunidades. A associação de muitos tipos de agentes coletivos continuará a evoluir ao longo do tempo.	{a <i>International Federation of Library Associations and Institutions</i> } [uma associação]

LRM-E9	<i>Nomen</i>	Uma associação entre uma entidade e uma designação que se refere a ela	Um nomen associa qualquer denominação (isto é, combinação de sinais) utilizada para se referir a uma instância de qualquer entidade encontrada no universo bibliográfico com essa entidade. Qualquer entidade mencionada no universo do discurso é nomeada por meio de, no mínimo, um nomen.	<i>Nomens</i> para uma <i>pessoa</i> : 'Agatha Christie' como uma maneira de se referir {a <i>pessoa</i> Dame Agatha Christie, Lady Mallowan}
LRM-E10	Lugar	Uma extensão de espaço determinada	A entidade lugar, relevante em um contexto bibliográfico, é uma construção cultural; é a identificação humana de uma área geográfica ou extensão do espaço. Lugares são geralmente identificados por meio de um objeto físico (uma característica geográfica ou um objeto feito pelo homem) ou devido à sua relevância em relação a um determinado agente (entidades geopolíticas como países, cidades) ou como a localização de um evento. O lugar como uma extensão do espaço é diferente dos órgãos governamentais que exercem a jurisdição nesse território. O governo responsável por um território é um agente coletivo. Lugares podem ser contemporâneos ou históricos, terrestres ou extraterrestres. Lugares imaginários, lendários ou fictícios não são instâncias da entidade lugar.	{Itália}; {África}; {Marte}

LRM-E11	Intervalo de Tempo	Uma extensão temporal que possui um início, um fim e uma duração	Um intervalo de tempo é um período que pode ser identificado especificando seu início e fim. A duração resultante pode ser associada a ações ou ocorrências que aconteceram durante esse período de tempo. Mesmo um intervalo de tempo muito preciso tem uma duração mensurável, por mais breve que seja.	{Século XX}; {2015-03-01} [intervalo de tempo de um dia expresso no calendário gregoriano no formato AAAA-MM-DD]
---------	--------------------	--	---	--

Fonte: Adaptação de IFLA, 2017.

Quanto aos usuários o “modelo se preocupa [...] com os dados e a funcionalidade exigidos pelos usuários finais (e intermediários que trabalham em nome dos usuários finais) para atender às suas necessidades informacionais” (IFLA, 2017, p. 15).

As tarefas do usuário definidas pelo modelo IFLA LRM são encontrar, identificar, selecionar, obter e explorar. Elas constituem um conjunto de tarefas específicas do usuário e “refletem os modelos conceituais anteriores, que consideraram tais tarefas como objetivo a ser alcançado pelos dados elaborados conforme os requisitos funcionais estabelecidos para o universo bibliográfico” (MACHADO; ZAFALON, 2020, p. 100). Veja as principais diferenças entre os quatro modelos conceituais:

QUADRO 3 – Principais diferenças entre os quatro modelos

Modelo Conceitual	FRBR (1998)	FRAD (2009)	FRSAD (2010)	IFLALRM (2017)
Entidades	Obra Expressão Manifestação Item Pessoa Instituição Conceito Objeto Evento Lugar	Obra Expressão Manifestação Item Pessoa Instituição Conceito Objeto Evento Lugar Família Nome Identificador Pontos de acesso controlado	<i>Thema</i> <i>Nomen</i>	<i>Res</i> Obra Expressão Manifestação Item Agente (pessoa e agente coletivo) Nomen Lugar Intervalo de tempo

		Regras Agência		
Tarefas de Usuário	Encontrar Identificar Selecionar Obter	Encontrar Identificar Contextualizar Justificar	Encontrar Identificar Selecionar Explorar	Encontrar Identificar Selecionar Obter Explorar

Fonte: Machado e Zafalon, 2020.

Além dos elementos entidades e relacionamentos, os atributos também compõem o modelo entidade-relacionamento e, conseqüentemente, o modelo IFLA LRM. Os atributos são os dados que caracterizam instâncias de entidades (IFLA, 2017).

Nos modelos entidade-relacionamento, as entidades definem a estrutura do modelo e funcionam como nós, ao passo que os relacionamentos conectam as entidades entre si. Os atributos dependem das entidades e fornecem informações sobre elas (IFLA, 2017, p.19).

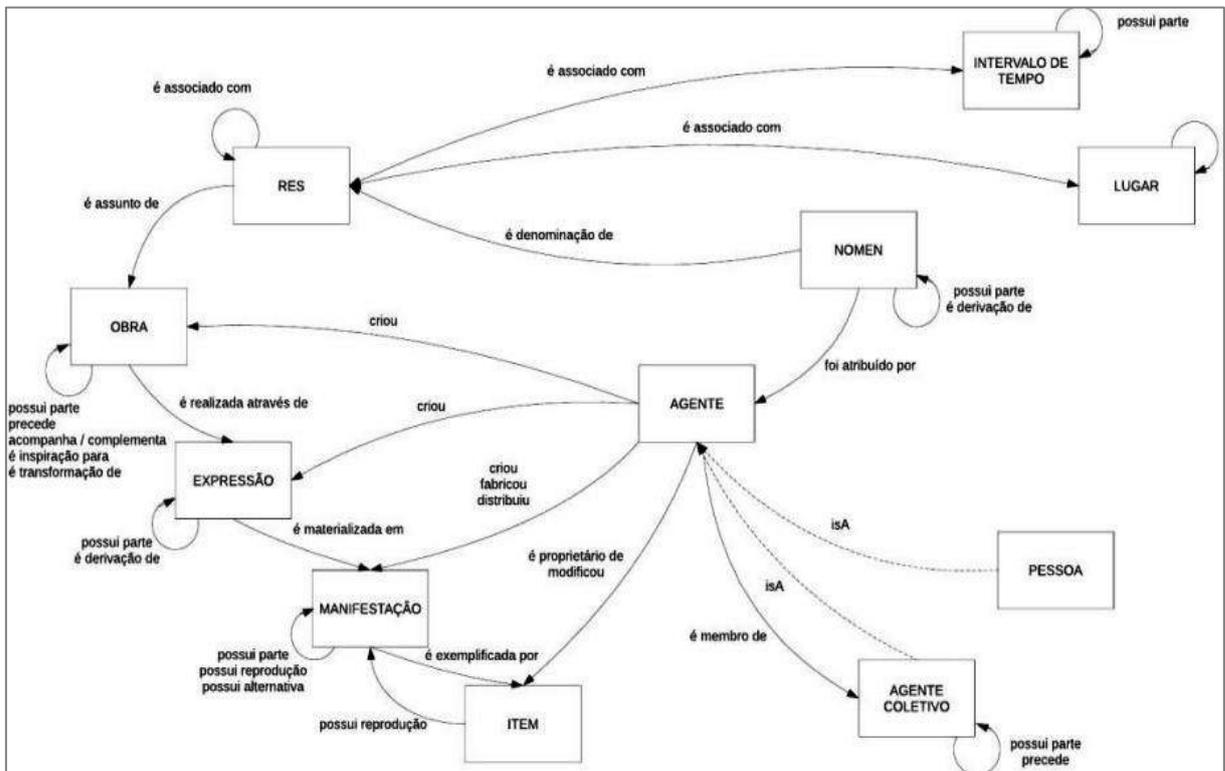
Os atributos que compõe o modelo são representados a seguir, obedecendo ao relacionamento com a entidade a qual faz parte (MACHADO; ZAFALON, 2020, p.105):

- LRM-E1 - Res
- LRM-E1-A1 Categoria
- LRM-E1-A2 Nota
- LRM-E2 – Obra
- LRM-E2-A1 Categoria
- LRM-E2-A2 Atributo de expressão representativa
- LRM-E3 – Expressão
- LRM-E3-A1 Categoria
- LRM-E3-A2 Extensão
- LRM-E3-A3 Público-alvo
- LRM-E3-A4 Direitos de uso
- LRM-E3-A5 Escala cartográfica
- LRM-E3-A6 Idioma
- LRM-E3-A7 Tonalidade
- LRM-E3-A8 Meio de execução
- LRM-E4 – Manifestação
- LRM-E4-A1 Categoria do suporte
- LRM-E4-A2 Extensão
- LRM-E4-A3 Público-alvo
- LRM-E4-A4 Declaração de Manifestação
- LRM-E4-A5 Condições de acesso
- LRM-E4-A6 Direitos de uso
- LRM-E5 – Item
- LRM-E5-A1 Localização
- LRM-E5-A2 Direitos de uso
- LRM-E6 – Agente
- LRM-E6-A1 Informações de contato
- LRM-E6-A2 Campo de atividade
- LRM-E6-A3 Idioma

- LRM-E7 Pessoa
- LRM-E7-A1 Profissão/Ocupação
- LRM-E9 – Nomen
- LRM-E9-A1 Categoria
- LRM-E9-A2 Nomen string
- LRM-E9-A3 Esquema
- LRM-E9-A4 Público-alvo
- LRM-E9-A5 Contexto de uso
- LRM-E9-A6 Fonte de referência
- LRM-E9-A7 Idioma
- LRM-E9-A8 Escrita
- LRM-E9-A9 Conversão de escrita
- LRM-E10 – Lugar
- LRM-E10-A1 Categoria
- LRM-E10-A2 Localização
- LRM-E11 – Intervalo de tempo
- LRM-E11-A1 Início
- LRM-E11-A2 Término

Uma visão geral do modelo IFLA LRM, de suas entidades e relacionamentos podem ser resumidas por meio do diagrama de relacionamento da entidade. Os atributos não aparecem nessas figuras porque cada um corresponde a uma característica associada à entidade que está relacionada. A figura 5 apresenta as relações do modelo IFLA LRM:

FIGURA 5 - Visão geral das relações do modelo IFLA LRM



Fonte: IFLA, 2017.

Esta é uma apresentação concisa do modelo IFLA LRM e de seus principais componentes. O modelo na íntegra é encontrado no site da IFLA² (2017). O IFLA LRM representa um importante avanço para a catalogação por ser “um modelo completo do universo bibliográfico, que pode e deve servir de base para o desenvolvimento de regras de catalogação e formatos bibliográficos” (ZUMER, 2018, p. 318, tradução nossa). As próximas etapas do modelo irão contemplar:

A declaração de *namespaces*, o que permitirá implementações compatíveis com a *Web* semântica e mapeamento para *namespaces* existentes. Tarefas futuras importantes incluem extensões para tipos de materiais específicos, diferentes públicos-alvo e outras circunstâncias importantes para o projeto de sistemas de informação bibliográfica (ZUMER, 2018, p. 318, tradução nossa).

Com IFLA LRM temos, enfim, um modelo conceitual no âmbito bibliográfico compatível com a *Web* semântica. O desenvolvimento de novos catálogos de biblioteca baseados nesse modelo possibilitará que a riqueza dos dados da biblioteca seja explorada em totalidade e que a tendência de uso reduzido dos atuais catálogos da biblioteca seja diminuída (ZUMER, 2018, p. 318, tradução nossa).

A importância dos dados conectados no escopo de bibliotecas aumentou em 2004, quando o WWW Consortium recomendou que as bibliotecas publicassem seus dados usando tecnologias da *Web* semântica para aumentar seu impacto e utilidade social (HALLO et al., 2016, tradução nossa).

Em 2010 surge o *W3C Library Linked Data Incubator Group*, que desenvolveu um estudo no período de maio de 2010 a agosto de 2011, com o intuito de “ajudar a aumentar a interoperabilidade global de dados de biblioteca na *Web*” (BAKER et al., 2011, não paginado, tradução nossa). A iniciativa do W3C coincidiu com o anúncio feito pela Biblioteca do Congresso em 2011 sobre o lançamento do *Bibframe*, proposto como a evolução do formato MARC 21 para a *Web* semântica e dados conectados (KROEGER, 2013).

O *Bibframe* fornece uma base para o futuro da descrição bibliográfica, tanto na *Web* quanto no mundo em rede mais amplo, baseado em técnicas de Linked Data. Um dos principais focos da iniciativa é determinar um caminho de transição para os formatos MARC 21 enquanto preserva uma troca de dados robusta que tem apoiado o compartilhamento de recursos e economia de custos de catalogação nas últimas décadas (LIBRARY OF CONGRESS, 2016, não paginado, tradução nossa).

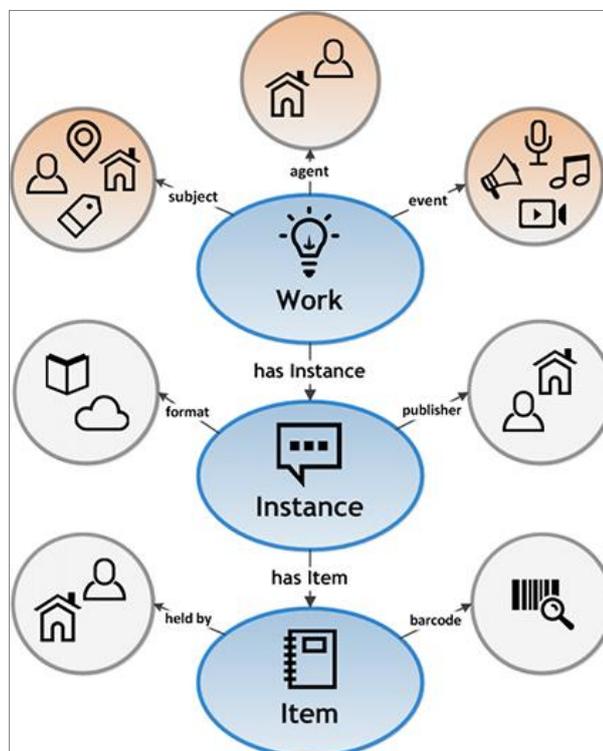
² https://www.ifla.org/files/assets/cataloguing/frbr-lrm/ifla-lrm-august-2017_rev201712-por.pdf

O modelo *Bibframe* “simplifica as relações entre Obra, Expressão, Manifestação e Item do modelo FRBR, prevendo somente três entidades, ou três níveis de abstração: *Work* (Obra), *Instance* (equivalente a Manifestação) e *Item*” (MARCONDES, 2021, p.97). O caráter abstrato entre os relacionamentos dessas três entidades acarreta um modelo mais simplificado:

Esta simplificação se deve ao caráter bastante abstrato que possuem os relacionamentos entre Obra, como a criação artística abstrata de um autor e sua Expressão, a codificação de uma Obra em um sistema simbólico (um idioma), que por sua vez se relaciona então com uma Manifestação concreta, uma determinada edição em um idioma. Além dessas entidades estão também previstas no BIBFRAME entidades como *Agent*, *Subject* e *Event* (MARCONDES, 2021, p.97).

O *Bibframe* 1.0 foi lançado em 2011 e era composto inicialmente por quatro classes principais: *Work*, *Instance*, *Authority* e *Annotation*. A configuração atual, que compreende os três níveis de abstração (*Work*, *Instance* e *Item*) foi lançada em abril de 2016, dando origem a versão 2.0 do modelo, conforme apresentado a seguir (ARAKAKI et.al., 2017):

FIGURA 6 - *Bibframe* 2.0



Fonte: Library of Congress, 2016

A *Library of Congress* (2016) estabelece as seguintes definições para os três níveis de abstração: obra, instância e item, conforme apresentado a seguir:

- **Obra:** é o mais alto nível de abstração. Uma Obra, no contexto BIBFRAME, reflete a essência conceitual do recurso catalogado: autores, línguas e assunto tratado.
- **Instância:** uma Obra pode ter uma ou mais modalidades materiais individuais, por exemplo, uma forma publicada particular. Estas são instâncias da obra. Uma instância reflete informações como seu editor, local e data de publicação e formato.
- **Item:** Um item é uma cópia real (física ou eletrônica) de uma Instância. Ele reflete informações como sua localização (física ou virtual), marca de prateleira e código de barras (*LIBRARY OF CONGRESS*, 2016, não paginado, tradução nossa).

As entidades Agentes, Assuntos e Eventos possuem relacionamentos com as classes principais descritas (Obra, Instância e Item) e são assim definidas:

- **Agentes:** agentes são pessoas, organizações, jurisdições, etc., associados a uma Obra ou Instância por meio de funções como autor, editor, artista, fotógrafo, compositor, ilustrador, etc.
- **Assuntos:** um Trabalho pode ser “sobre” um ou mais conceitos. Tal conceito é considerado um “sujeito” da Obra. Os conceitos que podem ser assuntos incluem tópicos, lugares, expressões temporais, eventos, obras, instâncias, itens, agentes, etc.
- **Eventos:** ocorrências, cujo registro pode ser o conteúdo de uma Obra (*LIBRARY OF CONGRESS*, 2016, não paginado, tradução nossa).

Além das três classes principais apresentadas, o *Bibframe* também possui outras classes e subclasses, que estão descritas no vocabulário *Bibframe* e são usadas para “[...] definir formalmente os atributos e relacionamentos e representar os metadados dos recursos” (RAMALHO, 2016, p. 299).

A *Library of Congress* (2016) assim define o Vocabulário *Bibframe*:

O vocabulário *BIBFRAME* consiste em classes e propriedades RDF. As classes incluem as três classes principais listadas acima, bem como várias classes adicionais, muitas das quais são subclasses das classes principais. As propriedades descrevem as características do recurso que está sendo descrito, bem como os relacionamentos entre os recursos. Por exemplo: uma Obra pode ser uma “tradução de” outra Obra; uma instância pode ser uma “instância de” uma obra *BIBFRAME* específica. Outras propriedades descrevem atributos de trabalhos e instâncias. Por exemplo: a propriedade “assunto” de *BIBFRAME* expressa um atributo importante de uma Obra (do que trata a Obra), e a propriedade “extensão” (por exemplo, número) expressa um atributo de uma Instância (*LIBRARY OF CONGRESS*, 2016, não paginado, tradução nossa).

A *Library of Congress* fornece ferramentas para promover o uso do *Bibframe* em seu site, como o editor *Bibframe* e o *software* para converter registros

MARCXML em *Bibframe*. O próximo quadro apresenta um exemplo da descrição do livro "O *Silmarillion*" escrito por J.R.R. Tolkien, que foi produzido no *Bibframe* editor usando dados de autoridade do VIAF e da *Library of Congress* (ARAKAKI et al., 2017):

QUADRO 4 - Descrição com o Bibframe 2.0 pelo Bibframe editor

	BIBFRAME Vocabulary	Valores	Identificador único do valor	
Work - Monograph (Otra)	Title	The Silmarillion	FakeID 000	
	Author (Person)	Tolkien, J. R. R.	VIAF ID: 95218067 (Personal)	
	Date of work	1977		
	Place of creation	England	VIAF ID: 142995804 (Geographic)	
	Editor (Person)	Tolkien, Christopher	VIAF ID: 12315790 (Personal)	
	Translator (Person)	Barcellos, Waldeá	VIAF ID: 98796962 (Personal)	
	DDC Classification	823.91		
	Language of work	English		
	Related work	The lord of the rings	VIAF ID: 184093601 (Work)	
	Related work	The hobbit	VIAF ID: 309569545 (Work)	
	Subject (Topic)	Fantasy fiction	lc/authorities/sh85047114	
	Instance of Work - Monograph	The Silmarillion	VIAF ID: 183065413 (Work)	
	Instance/ Manifestation - Monograph (Manifestação/ Instância)	Title	O Silmarillion	FakeID 001
Copyright date		1999		
Publication (Provider Entity)		Provider name (Organization)	Martins Fontes	http://id.loc.gov/authorities/names/no2002002677
		Provider place	São Paulo	VIAF ID: 126223335 (Geographic)
Media type		unmediated	http://id.loc.gov/vocabulary/mediaTypes/n	
Carrier type		volume	http://id.loc.gov/vocabulary/carriers/nc	
Page nos.		460		
ISBN10	85-336-1165-X			
Holding - Monograph (Item)	Holding for Instance/ Manifestation - Monograph	O Silmarillion	FakeID 001	
	Held by (Organization)	Universidade Estadual Paulista	VIAF ID: 122646643 (Corporate)	
	Barcode	0300130480		
	Shelf location	823.91 T649si SP/23210		

Fonte: Arakaki et al., 2017

Para Arakaki et al., (2017) o uso do *Bibframe* torna mais clara a proposta de catalogação recomendada pelo FRBR, já que pode ser dividida em múltiplas partes, e não em um único registro, como sugere a estrutura do MARC 21. A interconexão de dados fornecida pelo uso de RDF permite que a máquina crie uma compreensão do registro para distinguir cada entidade separadamente. Portanto, uma vez descrita a obra de um recurso de informação, basta estabelecer um vínculo com o registro da manifestação, agilizando o processo de catalogação dos recursos de informação posteriores (ARAKAKI et al., 2017).

Apesar do *Bibframe* não ser o único modelo de dados conectados existente, pelo fato de estar sendo desenvolvido pela Biblioteca do Congresso Americano,

provavelmente terá um maior nível de aceitação para a adaptação dos dados das bibliotecas (STEELE, 2018).

O uso de modelos conceituais da área da computação, como o modelo de Entidade-Relacionamento (E-R) ilustra claramente a importância da interdisciplinaridade e ampliação do universo bibliográfico. O surgimento dessa nova visão que incorpora modelos de dados é uma evolução na forma conceber, descrever e disponibilizar objetos bibliográficos, o que difere do modelo tradicional de descrição praticado pela comunidade bibliotecária. É preciso basear-se nas necessidades atuais dos usuários, adequar o catálogo, considerar o ponto de vista da recuperação da informação com lógica atualizada e focar na urgência de aumentar sua relevância num campo cada vez mais competitivo dos serviços de informação em rede, no contexto da *Web Semântica* com aplicações em *linked data* (GALVÃO; CORDEIRO, 2010).

Perante o contexto apresentado, na próxima seção apresenta-se o conceito dos dados abertos conectados no contexto da *Web Semântica* e as tecnologias necessárias para sua implementação, assim como as fontes de dados disponíveis para uso como dados abertos conectados.

2.2 DADOS ABERTOS CONECTADOS E *WEB SEMÂNTICA*: CONCEITOS, APLICABILIDADE E TECNOLOGIAS NECESSÁRIAS PARA IMPLEMENTAÇÃO

Na *Web*, os dados são publicados por diferentes pessoas e armazenados em repositórios distintos ao redor do mundo. Para facilitar a construção de uma base de dados global é necessário estabelecer uma forma de conexão padrão entre esses dados. Apresenta-se a seguir a conceituação e aplicação dos dados abertos conectados no contexto da *Web semântica*.

No contexto de dados e conteúdos abertos, “aberto significa que qualquer pessoa pode acessar, usar, modificar e compartilhar livremente para qualquer finalidade, sujeito, no máximo, aos requisitos que preservam a procedência e a abertura” (OPEN KNOWLEDGE FOUNDATION, 2010, não paginado).

Dessa forma, “Dados e conteúdo abertos podem ser usados, modificados e compartilhados livremente por qualquer pessoa para qualquer propósito” (OPEN KNOWLEDGE FOUNDATION, 2010, não paginado).

A abertura de dados está interessada em evitar um mecanismo de controle e restrições sobre os dados [...] permitindo que tanto pessoas físicas quanto jurídicas possam explorar estes dados de forma livre (ISOTANI; BITTENCOURT, 2015, p. 20).

Para que o dado seja considerado aberto, ele deve apresentar três aspectos que estão descritos a seguir:

- 1) O dado precisa estar disponível em sua totalidade, em um formato modificável e conveniente;
- 2) O dado precisa ser fornecido com termos que possibilitam sua reutilização e redistribuição, além de poder ser combinado com outros dados;
- 3) O dado pode ser utilizado, reutilizado e redistribuído por qualquer pessoa, grupo ou área, sem restrições (OPEN KNOWLEDGE FOUNDATION, 2010, não paginado).

Isotani e Bittencourt (2015, p. 21) destacaram a importância do trabalho colaborativo entre as organizações com vistas a aumentar a comunicação, reforçando a importância dos dados em formato aberto:

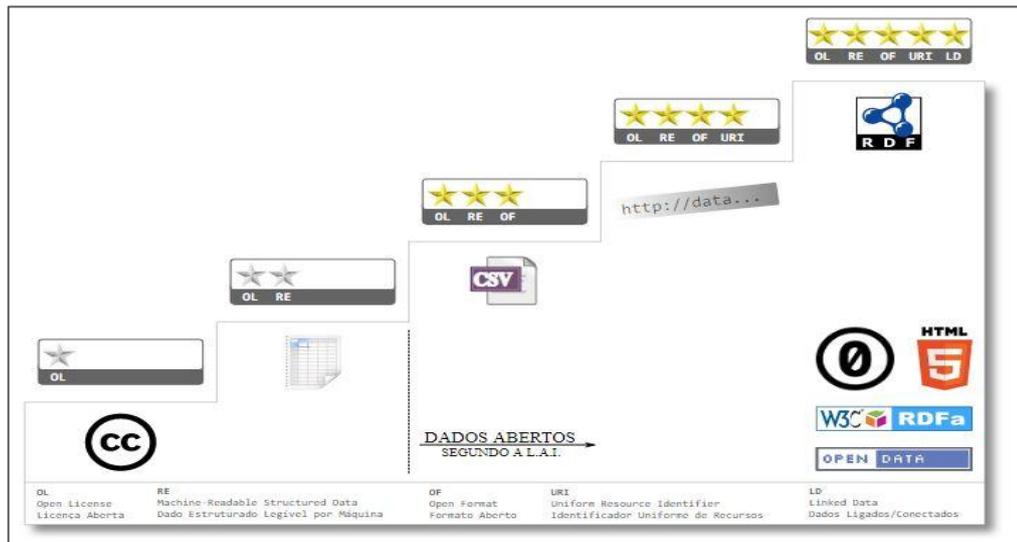
A partir do momento que há um movimento de abrir dados, em que as três normas fundamentais supracitadas são respeitadas, é possível que diferentes organizações e sistemas possam trabalhar de forma colaborativa. Isso ocorre devido à capacidade dessas organizações e desses sistemas de interoperar os dados que foram abertos, ampliando assim a comunicação e potencializando o desenvolvimento eficiente de sistemas complexos. Para tal, os dados devem ser acessíveis, legíveis por máquinas, ter formato aberto e informação produzida por todos e para todos.

Nesse contexto, Berners-Lee (2006) classificou os dados abertos, por uma escala, baseada em cinco estrelas. Tal classificação exprime a inserção das informações no universo dos dados conectados:

- 1) 1 estrela, licença aberta que permita seu uso e reutilização sem restrições.
- 2) 2 estrelas, formato que permite que os dados possam ser estruturados automaticamente por *software*.
- 3) 3 estrelas, formato estruturado e aberto que ocorre quando os dados são publicados em formatos não proprietários.
- 4) 4 estrelas, dados possuem identificadores URI, quando os dados recebem uma identificação que é única e podem ser conectados com outros.
- 5) 5 estrelas, dados conectados com outros dados - quando se conectam com outros dados já disponíveis na *Web*.

A classificação dos dados baseado na escala de cinco estrelas é ilustrada por Diraol (2013) e apresentada no gráfico 2:

GRÁFICO 2 - Descrição das 5 estrelas dos dados abertos



Fonte: Diraol, 2013.

Nessa contextualização da Web semântica, assim como também um panorama introdutório do tópico apresentado, serão apresentadas a seguir teorias que versam estudos e conceituações de dados conectados.

2.2.1 Dados conectados

A evolução da *Web* de documentos para a *Web* de dados, através do avanço das tecnologias e da introdução da *Web* semântica no modelo inicial da *Web*, tornou possível criar inúmeras oportunidades para a integração dos dados, motivando o desenvolvimento de novos tipos de aplicações e ferramentas, como navegadores e motores de busca (ISOTANI; BITTENCOURT, 2015).

A *Web* de documentos se baseia na navegação entre documentos dispostos em uma estrutura de hipertexto, usando o padrão HTML para publicação dos dados. Objetiva utilizar a internet para consultas e atualizações de documentos que é composta por 03 (três) elementos básicos: URI (identificador dos documentos, dos nós das estruturas), HTML (linguagem de marcação para a descrição dos documentos) e o HTTP (que consiste no protocolo de comunicação para acesso aos documentos, mecanismo de acesso universal). A *Web* de dados possui um modelo padrão para a representação de dados que consiste no RDF, uma linguagem de consulta para acesso aos dados, o SPARQL, mecanismo de identificação e acesso universal que consiste em URIs e HTTP (ISOTANI; BITTENCOURT, 2015).

Para criar um contexto, os dados abertos precisam estar conectados a outros dados e para isso alguns requisitos são requeridos:

Porém isso somente é possível com a utilização de um modelo padrão para trocas de dados. Nesse sentido, é recomendado o uso da linguagem RDF, a qual permite a semantificação dos dados, facilitando sua ligação, mesmo entre dados com esquemas diferentes. Além disso, ao adotar a linguagem RDF é permitida a evolução desses esquemas ao longo do tempo, sem a necessidade de alteração pelos consumidores dos dados (GUIA PRÁTICO PARA PUBLICAÇÃO DE DADOS ABERTOS CONECTADOS NA WEB, 2018, p. 468).

Assim, os dados conectados podem ser entendidos como um conjunto de nós, que compreende uma união de vários dados e links que podem ser considerados as conexões. Seus padrões permitem a publicação de dados de uma forma que possam ser lidos por pessoas e processados por máquinas.

Isso é possível porque os dados que antes estavam “escondidos” na *Web* de Documentos estão agora acessíveis graças à utilização dos padrões supracitados para a conexão de dados. Esta conexão de dados permite que todos (homens e máquinas) possam trabalhar conjuntamente de forma mais eficiente [...] (ISOTANI; BITTENCOURT, 2015, p. 33).

Dados conectados é definido pelo W3C (2015) como:

No contexto da *Web* semântica os dados disponibilizados na *Web* precisam ser disponibilizados utilizando um formato padrão, acessível e gerenciável por ferramentas da *Web* semântica. Além disso, a *Web* semântica não apenas precisa de acesso aos dados, mas os relacionamentos entre os dados também devem ser disponibilizados para criar uma *Web* de dados (em oposição a uma coleção simples de conjuntos de dados). Essa coleção de conjuntos de dados inter-relacionados na *Web* também pode ser chamada de dados conectados (WE3, 2015, não paginado, tradução nossa).

Baker et al., (2011) apresenta a seguinte definição:

São considerados dados em *Linked Data*, dados conectados, aqueles publicados de acordo com os princípios definidos para facilitar o vínculo entre *datasets*, ou seja, utilizando URIs como identificadores únicos de recursos, expressos sob o modelo RDF, especificando as relações existentes entre as coisas (BAKER et al., 2011, não paginado, tradução nossa).

De acordo com Isotani e Bittencourt (2015, p.34) o termo dados conectados “se refere a um conjunto de boas práticas para publicação e conexão de dados estruturados na *Web*, usando padrões internacionais recomendados pelo W3C”.

Percebe-se que não há distinção entre as definições. Alguns autores a fazem de forma mais concisa, porém sem perder o sentido do termo.

Destacam-se também a importância de se esclarecer que os dados conectados não necessariamente precisam ser dados abertos, pois há dados que são conectados, mas são fechados, ou seja, com acesso privado, por vários motivos, dentre eles por segurança e para manter a privacidade. Para a criação de um banco de dados global, além de se aplicar semântica aos dados é preciso que esses dados sejam conectados. De acordo com Berners-Lee (2006) há 4 princípios para a conexão desses dados:

- 1) Utilização de URIs (Uniform Resource Identifier) para identificar recursos (documentos, arquivos, imagens, serviços, etc.) de maneira única na web;
- 2) Utilização de HTTP URIs (http://...) para que os “recursos” possam ser localizadas por pessoas ou aplicativos;
- 3) Fornecimento de informações relevantes utilizando os padrões RDF, SPARQL para facilitação da busca por uma URI;
- 4) Inclusão de links para outras URIs, para possibilitar que mais recursos sejam descobertos.

A aplicação dos princípios para conexão de dados na internet cria, portanto, um rico contexto para os dados. O modelo de dados padrão para troca de informações na internet é o RDF, definido pelo Manual dos Dados Abertos como (p.40, 2011):

O RDF é um modelo padrão para troca de informações na rede que facilita tarefas como a mescla de dados, testando *schemas* diferentes e propiciando a evolução dos dados e seus *schemas*. Ele estende o modelo de estrutura de ligação (links) da web e chega ao relacionamento de documentos via nomes de URIs, interligando-os. Dessa forma, é possível criar estruturas de dados que podem ser misturados, compilados, interpretados e compartilhados utilizando-se aplicativos diversos.

Em RDF, o método de identificação de recursos e atributos de uma forma única e universal dá-se pelo uso de URIs, mais especificamente, HTTP URIs, definido por Laufer (p. 39, 2015) como:

URIs são uma forma mais abrangente de URL, pois não estão necessariamente ligadas à localização do recurso. Têm o mesmo formato de uma URL, mas são utilizadas com o intuito de identificar as coisas, enquanto uma URL identifica um endereço para a recuperação de uma informação, um documento.

A base do RDF é a "tripla":

A tripla “define o sujeito, o predicado e o objeto utilizados para descrever uma informação na web e o relacionamento entre elas. É assim que será descrito o RDF: um conjunto de dados e seus relacionamentos em forma de triplas” (MANUAL DOS DADOS ABERTOS, p.40, 2011).

Como forma de ilustrar a ideia da definição dos dados por meio de um conjunto de triplas, recomenda-se o uso das tabelas exemplificativas criadas por Laufer (2015). A tabela inicial contém informações sobre recursos do tipo livros. Cada linha da tabela 1 tem informações sobre um determinado livro. Cada um desses livros é um recurso. Cada coluna da tabela define um tipo de propriedade relacionada aos livros e cada célula da tabela define uma tripla.

TABELA 1: Contém as informações dos recursos do tipo livro

ISBN	Título	Author	Editora_id	pg.
9788535912388	Gabriela, Cravo e Canela	Jorge Amado	1243	424
...
9788501067340	Vidas Secas	Graciliano Ramos	3244	176
...
9788535921199	Antologia Poética	Carlos Drummond de Andrade	1243	344

Fonte: Laufer, 2015.

As linhas da tabela 2, apresentada a seguir, representam os recursos, e as colunas representam as propriedades dos recursos. Cada célula da tabela define uma propriedade (coluna) de um recurso (linha), formando assim uma tripla de um recurso.

TABELA 2: Tripla de um recurso

ISBN	Título	Author	Editora_id	pg.
9788535912388	Gabriela, Cravo e Canela	Jorge Amado	1243	424
...
9788501067340	Vidas Secas	Graciliano Ramos	3244	176
...
9788535921199	Antologia Poética	Carlos Drummond de Andrade	1243	344

Fonte: Laufer, 2015.

Uma tripla é representada por um grafo RDF, do sujeito para o objeto, conforme indicadas nas figuras 7 e 8.

FIGURA 7 - Grafo RDF da tripla de um livro



Fonte: Laufer, 2015.

FIGURA 8 - Grafo RDF genérico de uma tripla



Fonte: Laufer, 2015.

O “RDF é um modelo de dados abstrato, não importando como ele seja representado, desde que a representação permaneça fiel às suas propriedades

abstratas” (LAUFER, 2015, p.40), o que torna necessário o uso de notações para representar o RDF de forma legível ao computador. Isotani e Bittencourt (2015) elencam os formatos de notações mais utilizados para descrever recursos em RDF, descritos a seguir:

RDF/XML: foi a primeira serialização feita para RDF e gera um documento XML válido;

RDFa (*Resource Description Framework in Attributes*): o formato visa embutir o código RDF na estrutura HTML. Isso é feito adicionando significado aos atributos do elemento. Por esse motivo, RDFa é considerado o RDF em "atributos". A maior vantagem de usar RDFa é que os mecanismos de busca podem melhorar os resultados da pesquisa, melhorando a precisão do significado real de um determinado documento. Portanto, o mecanismo de pesquisa pode agregar os dados de um documento com os dados de outro documento, enriquecendo os resultados da pesquisa;

JSON-LD: surgiu como uma extensão da proposta JSON, onde o objetivo é converter o código JSON para RDF com o mínimo de esforço. Para programadores que já estão familiarizados com a sintaxe JSON, este formato é muito intuitivo;

N-Triples: Faz parte da série de formatos *Turtle*. É o formato de serialização mais simples e intuitivo entre os formatos existentes. Uma vez que a principal característica do RDF é a estrutura de triplas <sujeito> <predicado> <objeto>, onde N-Triples estrutura o código na forma de triplas, sendo o resultado um documento (texto), onde cada linha representa uma tripla;

Turtle: O formato foi criado para expandir as possibilidades de descrição de documentos N-Triples, sendo capaz de descrever prefixos e IRIs relativos à estrutura do documento. O Turtle é ainda mais fácil e simples de ler comparado ao formato N-Triples;

TriG: Este formato foi criado para representar vários grafos, sendo uma extensão do formato Turtle, mantendo a mesma simplicidade e facilidade de leitura, características do *Turtle*. A diferença entre o código *Turtle* e o código TriG é que os

sujeitos descritos são encapsulados dentro de uma palavra-chave chamada GRAPH e significa que um conjunto de triplas dentro do elemento GRAPH representa um catálogo de dados (ou *dataset*);

N-Quads: O formato *N-Quads* é usado para permitir a troca de catálogo de dados. É uma extensão do *N-Triples* e cada linha tem a identificação de uma tripla;

Notation3 (N3): o resultado é um documento (texto) mais compacto e legível do que RDF / XML e *N-triples*.

Estando os dados disponíveis em RDF, um padrão é indispensável para pesquisá-los de forma inteligente e uma das alternativas é o uso da linguagem SPARQL (*Simple Protocol and RDF Query Language*). O W3C recomendou o SPARQL em 2008 e o definiu como uma linguagem de consulta RDF e protocolo usado para expressar consultas a bancos de dados em que são armazenadas várias fontes de dados no formato RDF (W3C, 2008).

Marcondes (2021, p.50) apresenta a seguinte definição para a linguagem:

SPARQL - SPARQL Protocol and RDF Query Language – é uma linguagem de consulta a dados em RDF, uma linguagem de consulta semântica. Um gerenciador de triplas RDF pode ser acessado na *Web* através de um “link” chamado de “*end point*”. Nesse “*end point*” podem ser feitas consultas ao conjunto de triplas através da linguagem SPARQL. SPARQL é tanto uma linguagem de consulta a conjuntos de triplas quanto um protocolo que permite a um usuário em um navegador cliente submeter consultas a vários “*end points*” hospedados em servidores distintos simultaneamente.

Para Laufer (2015, p. 55) SPARQL é:

A linguagem de consulta da *Web Semântica*. Podemos fazer uma analogia entre SPARQL e a linguagem SQL de consulta a bancos de dados relacionais, considerando que SPARQL tem uma sintaxe adequada a consultas a dados representados como um conjunto de triplas RDF.

Para o W3C (2015) às consultas SPARQL baseiam-se em padrões triplos, semelhantes aos triplos RDF. O uso da linguagem SPARQL e dos protocolos que a acompanham possibilitam a realização de consultas e o recebimento de resultados via HTTP, por exemplo. Por intermédio dessa linguagem, os usuários da *Web semântica* podem extrair informações complexas, incluindo referências de recursos existentes e seus relacionamentos, que podem ser retornadas em formato de tabela

e incorporadas em outra página da Web. Desse modo, a linguagem permite a construção de mecanismos de pesquisa que contêm dados oriundos da *Web* semântica.

Complementando os elementos da Web semântica, em 2004 o W3C formalizou a OWL - *Ontology Web Language*, definida como uma linguagem para instanciação de ontologias. Por ontologia entende-se como sendo “uma especificação de um conceito dentro de um determinado domínio de interesse” (LAUFER, 2015, p. 49). O autor afirma ainda que a OWL é uma linguagem que estende o modelo de dados RDF e fornece uma gama mais ampla de tipos de restrição para um conjunto de triplos definidos. Desta forma, o número de diferentes interpretações possíveis para um determinado conteúdo pode ser limitado (LAUFER, 2015).

Dessa maneira, vê-se que:

O uso de padrões e de modelos de dados comuns torna possível a criação de aplicações genéricas capazes de trabalhar por todo o espaço de dados global, sendo esse o objetivo principal da criação de dados conectados (GUIA PRÁTICO PARA PUBLICAÇÃO DE DADOS ABERTOS CONECTADOS NA WEB, 2018, p. 492).

Nessa perspectiva, assim como as afirmações de Luz, Coneglian e Santarem Segundo (2019, p.7), a linguagem OWL pode deixar um domínio formal que busque definir classes, propriedades, indivíduos que sejam passíveis de explicar as sequências lógicas. Assim, os autores ponderaram que estes fatos não se apresentam no âmbito ontológico, mas, todavia, são expressos pela semântica.

2.2.2 Dados abertos conectados

Dados abertos conectados é a junção das características de dados abertos e dados conectados, assim como também a publicação de dados usando as ferramentas da *Web* para compartilhar dados e informações estruturadas em escala global, sem restrições a pessoas e aplicações. No Guia prático para publicação de dados abertos conectados na *Web* (2018, p. 492), o documento infere que “os dados abertos conectados representam um dos mecanismos de gerenciamento e interoperabilidade de informações mais importantes na atualidade favorecendo a inovação e a geração de novos conhecimentos”.

Segundo Isotani e Bittencourt (2015, p.39), o grupo de trabalho do W3C de dados abertos governamentais definiu e publicou um conjunto de boas práticas divididas em dez etapas para a publicação de dados abertos conectados, conforme indicado a seguir:

- 1) Preparação de *stakeholders* que irão criar e manter os dados abertos conectados;
- 2) Seleção das fontes de dados, para definição do que será aberto e conectado a outros dados e disponibilizado para reutilização;
- 3) Definição de como serão representados os dados e como os mesmos se relacionam com os outros dados e de maneira independente;
- 4) Definição e especificação das licenças que serão usadas;
- 5) Definição e uso das boas práticas para URIs;
- 6) Uso de vocabulários padrões para conexão dos dados;
- 7) Conversão dos dados da fonte original para representação correta aos dados conectados;
- 8) Definição de quais serão as formas de acesso aos dados, por parte das pessoas e máquinas;
- 9) Divulgação para a sociedade de novo conjunto de dados conectados;
- 10) Permanência dos dados publicados ao longo do tempo.

Dessa forma, percebe-se que as etapas possuem processos criteriosos para a publicação dos dados abertos conectados. E, para se compreender melhor, é interessante entender a relação com a Web semântica, que será apresentada na subseção a seguir.

2.2.3 Web Semântica

Como descrever informações de tal forma, que tanto as pessoas como as máquinas possam compreender? Na *Web* inicial, a *Web* de documentos, o conteúdo das páginas era inserido nas máquinas de maneira sintática e a interpretação dos dados era feita pelos usuários que visualizavam as páginas. Não havia uma semântica das informações por parte das máquinas. Com a evolução para a *Web* de dados está sendo possível que as informações sejam compreendidas também pelas máquinas onde se tem acesso a recursos abrangentes da *Web* e de diferentes desenvolvedores, que podem criar aplicações agrupando os dados de diferentes maneiras (LAUFER, 2015).

A *Web* semântica, segundo Laufer (2015) pode ser definida como:

Uma teia de informações construída de forma a ser facilmente processável por máquinas em uma escala global. A ideia geral é a de criar uma maneira eficiente para representar dados na *World Wide Web* de forma a construir um banco global de dados conectados (LAUFER, 2015, p.7).

Nesse contexto, para que se possa identificar essa linguagem por parte das máquinas, é necessário inserir informações adicionais nesses dados, mais especificamente dentro do código HTML. Tais informações são chamadas de metadados, descrito por Isotani e Bittencourt (2015, p.43) como “informações adicionais que descrevem os dados contidos nos documentos [...], ou seja, dados sobre dados”. Outra denominação para metadados descritas pelos autores é a seguinte: “campos de cabeçalhos utilizados para identificar o formato de representação” (ISOTANI; BITTENCOURT, 2015, p.59).

Desse modo, os metadados auxiliam as máquinas a entenderem o conteúdo dos recursos. Outrossim, são utilizados para a qualidade de dados, licenças de uso, origem de quem gerou os dados, entre outras características.

Nesse sentido, a *Web* semântica ajuda a criar uma forma eficiente para representar dados na *Web*, de maneira a gerar um banco global de dados conectados, onde um conjunto crescente de informações tenha acesso livre e aplicações com os mais diferentes propósitos (LAUFER, 2015).

A *Web* semântica está em crescimento e algumas áreas já se utilizam de suas potencialidades. Marcondes (2016), no seu artigo sobre interoperabilidade, cita a importância da *Web* semântica e dos dados abertos conectados para o desenvolvimento de instituições como bibliotecas, museus e arquivos. Essas instituições são responsáveis desde o surgimento em preservar, resguardar e disseminar memórias e culturas da sociedade que até então se utilizava apenas de catálogos para divulgação desses acervos. Porém, com o início da *Web*, houve um grande avanço na publicação desses catálogos, sendo os mesmos digitalizados e disponibilizados para consultas. Mas, apresentando ainda grandes restrições de acesso e interligação dos dados, principalmente, tendo em vista que cada uma das instituições apresenta várias formas de gestão e curadoria dos seus acervos.

Com o surgimento da *Web* semântica e das tecnologias que propiciam os dados abertos conectados, surgiu um novo método de expressão e representação de conteúdo, permitindo que as máquinas processem essas informações de forma mais eficaz. Desse modo, é possível publicar acervos digitais diretamente na *Web*, sem usar um sistema de gerenciamento de catálogos (MARCONDES, 2016).

Segundo Marcondes (2016, p.68), interoperabilidade pode ser entendida como:

Pode-se compreender interoperabilidade como a propriedade de sistemas diferentes (por ex. sistemas de gestão de bibliotecas digitais, instrumentos de pesquisa arquivísticos automatizados, sistemas de gestão de acervos museológicos), através de padrões tecnológicos, acordos ou propostas, de serem capazes de operar em conjunto, visando a execução de uma tarefa. As diferentes soluções de interoperabilidade utilizadas até hoje enfatizam o aspecto da troca ordenada de conteúdos de forma significativa; formatos de metadados como MARC e Dublin Core têm um papel fundamental nestas soluções de interoperabilidade.

Nesse sentido, pode-se verificar que a interoperabilidade em relação aos acervos digitais evoluiu desde as soluções baseadas na troca e agregação de metadados, até a publicação de conteúdo diretamente na *Web* com interligações através dos links.

Marcondes (2016, p.78), aponta, ainda, que:

A *Web* pode contribuir assim para democratizar o acesso a estes conteúdos, antes só disponíveis ao público presencial. A possibilidade de disseminar seus acervos através da *Web* amplia os papéis e justificativas sociais, educacionais e culturais das instituições de memória. Hoje se torna cada vez mais necessário que estas levem em conta a *Web* e suas possibilidades em suas políticas.

Desse modo, percebe-se a importância da implementação da *Web* semântica e dos dados abertos conectados na interoperabilidade entre os acervos digitais de bibliotecas, arquivos e museus. O que possibilita novos tipos de relações culturalmente significativas que podem ser estabelecidas entre objetos digitais pertencentes a estes acervos (MARCONDES, 2016). Isso irá proporcionar às “instituições tirarem partido das sinergias e complementaridades entre seus acervos e oferecerem melhores serviços - prover informação cultural curada, tratada, contextualizada e de alta qualidade aos seus usuários” (MARCONDES, 2016, p. 65).

Nesse sentido, com a intenção de discorrer sobre as fontes de dados conectados e as teorias correspondentes a temática serão, a seguir, introduzidas a partir de então, haja vista a importância de difundir um construto teórico para um Sistema de Organização do Conhecimento (SOC).

2.3 FONTES DE DADOS CONECTADOS

As fontes de dados conectados abordadas nessa seção são representadas em grande maioria por *Knowledge Organization System* (KOS), no português traduzido como Sistema de Organização do Conhecimento (SOC). O SOC é um

termo genérico utilizado para se referir a uma ampla gama de itens, como, por exemplo, cabeçalhos de assuntos, tesouros, esquemas de classificação, ontologias, e que estão disponíveis para uso em conjunto com as tecnologias da *Web* semântica (MARCONDES, 2021; MAZZOCCHI, 2018).

Os SOC foram concebidos para finalidades distintas em diferentes momentos históricos. Eles são caracterizados por estruturas e funções específicas, constituídos por tecnologias variáveis e usados por comunidades em diversos ambientes. No entanto, o que eles têm em comum é que são projetados para apoiar a organização do conhecimento e da informação, a fim de tornar mais fácil a sua gestão e recuperação (MAZZOCCHI, 2018).

Existem várias tipologias diferentes de SOC, entre elas estão os Arquivos de autoridade, Glossários, Dicionários, *Gazetteers* (dicionário geográfico), Cabeçalhos de assuntos, Esquemas de classificação, Taxonomias, Tesouros, Redes semânticas e Ontologias (MAZZOCCHI, 2018).

Hodge (2000) afirma que o termo KOS, conforme conhecido hoje, foi estabelecido no primeiro Grupo de Trabalho de Sistemas de Organização de Conhecimento em Rede (*First Networked Knowledge Organization Systems Working Group - NKOS WG*), que aconteceu em 1998 na Conferência de Bibliotecas Digitais ACM '98 em *Pittsburgh* (*ACM Digital Libraries '98 Conference in Pittsburgh*).

Segundo Miles e Bechhofer (2008) o termo passou a ser usado também pela W3C desde o início da especificação do SKOS (*Simple Knowledge Organization System*) em 2008.

Os requisitos para que um SOC possa ser usado com as tecnologias da *Web* semântica, de acordo com Marcondes (2021, apud CARACCILO et al., 2012; ZENG, 2019) são os seguintes:

- Os conceitos [...] devem ter escopo global, isto é, serem independentes de qualquer sistema local, terem semântica global e serem isentos, o mais possível de particularidades regionais;
- [...] deve ser multilíngue, seus conceitos devem ter termos traduzidos em diversos idiomas;
- Cada conceito deve ter um identificador único, além de um identificador persistente, um IRI/URI que permita que ele possa ser acessado e “dereferenciado” através da *Web* e usado em triplas RDF;
- Cada conceito deve poder ser representado em linguagem de máquina segundo os padrões usados com as tecnologias de dados abertos interligados: RDF/XML, TURTLE, JASON; O vocabulário deve poder ser exportado, ou ser feito “download”, no todo ou em parte, em formato SKOS (MARCONDES, 2021, p. 62).

O *Linked Data for Libraries (LD4L) project* elenca em seu portal uma lista com algumas fontes de dados que estão disponíveis para serem usadas com as tecnologias de dados abertos conectados. As definições a seguir trazem as especificações dessas fontes de dados conectados (CONFLUENCE, 2020), tais como OCLC *Fast*; ORCID; VIAF; LC *Subject Headings*; LC *Name Authority File* (LCNAF) e AGROVOC que serão apresentados a seguir.

2.3.1 OCLC Fast – *Faceted Application of Subject Terminology*

A *Faceted Application of Subject Terminology* (FAST) é um vocabulário de assuntos de nove facetas com um universo de aproximadamente 1,8 milhões de cabeçalhos, derivada da *Library of Congress Subject Headings* (LCSH), que é um dos cabeçalhos de assuntos mais usados no âmbito das bibliotecas (OCLC, 2019). O FAST é desenvolvido desde o final de 1998 pela OCLC *Research* em parceria com a Biblioteca do Congresso com o propósito de tornar a LCSH mais fácil de entender, controlar, aplicar e usar, criando, assim, uma sintaxe mais simplificada e, ao mesmo tempo, mantendo a riqueza do vocabulário LCSH. O FAST mantém compatibilidade com a LCSH e com qualquer conjunto de cabeçalhos de assunto da LC, que podem ser convertidos em cabeçalhos FAST (OCLC, 2019).

2.3.2 ORCID - *Open Researcher and Contributor ID*

O ORCID fornece a cada pesquisador um identificador digital persistente (ID ORCID) para distingui-lo de todos os outros pesquisadores, substituindo as variantes de nome por um único código numérico. Com ele cada indivíduo pode vincular o seu ID a outras informações profissionais, como afiliação, bolsas, publicações, revisão por pares e assim por diante. Este ID pode ser usado para compartilhar essas informações com outros sistemas. O ORCID, portanto, remove a ambiguidade entre os pesquisadores e conecta as pessoas às suas atividades de pesquisa (ORCID, 2021).

2.3.3 VIAF - Virtual International Authority File

O VIAF é um projeto implementado pela OCLC em conjunto com várias organizações que vincula atualmente dados de autoridades de 37 agências em 29

países. Trata-se, portanto, de um arquivo de autoridades que possibilita acesso a nomes conectados para a mesma entidade nos principais arquivos de autoridade de nomes do mundo, incluindo variações nacionais e regionais de idioma, conjunto de caracteres e ortografia (OCLC, 2019).

2.3.4 LC *Subject Headings* – LCSH

A LCSH também é usada internacionalmente, em seu idioma oficial ou de forma traduzida (*LIBRARY OF CONGRESS*, 2021). A LCSH é uma lista de cabeçalhos de assuntos mantida pela LC desde 1898, criada originalmente para catalogar os materiais armazenados na Biblioteca do Congresso. Como resultado da catalogação colaborativa, outras bibliotecas nos Estados Unidos também usam a LCSH para fornecer acesso aos assuntos de suas coleções.

2.3.5 LC *Name Authority File* – LCNAF

O LCNAF é um arquivo de autoridades de nomes, mantido pela LC, que fornece acesso uniforme aos recursos bibliográficos para nomes de pessoas, incluindo autores e outros produtores, como editores, tradutores, fotógrafos e artistas. O LCNAF formaliza também nomes de pessoas jurídicas, nomes de lugares, nomes de conferências, entidades coletivas, órgãos governamentais, conferências e jurisdições (*LIBRARY OF CONGRESS*, 2021; HARPRING, 2016).

As entradas da LCNAF são estabelecidas em parceria com várias instituições, estando entre elas bibliotecas dos Estados Unidos, a *British Library*, a *National Library of New Zealand*, a *National Library of South Africa* e a *National Library of Australia*. A Biblioteca do Congresso participa do *Program for Cooperative Cataloging* (PCC), uma colaboração internacional projetada para fornecer catalogação que atenda aos padrões de biblioteconomia mundialmente reconhecidos (*LIBRARY OF CONGRESS*, 2021; HARPRING, 2016).

2.3.6 AGROVOC

O AGROVOC é um vocabulário multilíngue controlado relacionado à alimentação e agricultura, publicado na forma de dados abertos conectados. É o maior conjunto de dados agrícolas abertos conectados à disposição do público, e

seu maior impacto é promover o acesso e a visibilidade dos dados em diferentes regiões e idiomas, promovendo assim a interoperabilidade e a reutilização de dados. Desde 1980 é coordenado pela *Food and Agriculture Organization of the United Nations* (FAO) e passou de um catálogo impresso para uma tecnologia de *Web* semântica. AGROVOC está disponível online e conectado a outros sistemas de organização do conhecimento multilíngue (FAO, 2021).

Após fundamentação teórica apresentada nesse capítulo, que deu embasamento científico a essa pesquisa, é necessário discorrer o que tem sido investigado sobre os dados abertos conectados no âmbito dos catálogos de bibliotecas. Desse modo, o capítulo seguinte, reservado à revisão de literatura, apresenta os resultados de algumas pesquisas correlatas ao tema deste estudo.

3. REVISÃO DE LITERATURA: DADOS ABERTOS CONECTADOS A PARTIR DE CATÁLOGOS ONLINE DE BIBLIOTECAS

Os estudos de Frosterus et al. (2021), Jesus Castro (2019), Smith-Yoshimu (2018), Ali; Warraich (2018), Senso e Arroyo Machado (2018), Serra e Santarem Segundo (2017), Vila-Suero e Gómez-Pérez (2013), Baker et al. (2011) foram selecionados para compor este capítulo, que discorre sobre trabalhos correlatos ao tema estudado. As produções deram subsídios para os resultados mapeados nesta pesquisa, conforme os critérios de seleção e estratégia de busca descrita no capítulo que versa sobre a metodologia.

Uma apresentação desses estudos é feita a seguir, apresentando as recomendações por eles apontadas, assim como as ações, vantagens, recursos e desafios relacionados à utilização dos dados abertos conectados em bibliotecas.

Um dos estudos mais recentes identificados foi desenvolvido pelo Grupo de trabalho de dados abertos conectados da LIBER (*Ligue des Bibliothèques Européennes de Recherche* - Associação de Bibliotecas de Pesquisa Européias).

O grupo publicou em fevereiro de 2021 um guia prático para bibliotecas que desejam publicar dados abertos conectados, intitulado *Best Practices for Library Linked Open Data Publication* (Melhores práticas para publicação de dados abertos conectados para bibliotecas).

O guia é baseado em uma pesquisa de projetos de dados conectados implementados ou em implementação, desenvolvido pelas seguintes bibliotecas:

- *ZBW Leibniz Information Center for Economics*
- *Hochschulbibliothekszenrum des Landes Nordrhein-Westfalen (hbz)*
- *Fundacion Ignacio Larramendi (Ignacio Larramendi Foundation)*
- *BCU Lausanne*
- *Herzog August Bibliothek Wolfenbüttel*
- *Bibliothèque Nationale de France*
- *University of Alberta Libraries*
- *Frankfurt University Library*
- *National Library of Finland*
- *Learning and Research Resources Centre (CRAI) of the University of Barcelona*
- *National Library of Scotland*
- *The British Library*
- *University of Venice*
- *Karolinska Institutet*
- *National Library of Estonia*

O grupo conduziu a pesquisa entre 2018 e 2020, que, posteriormente, foi aprimorada em um *Workshop* durante a Conferência Anual LIBER 2020. Os integrantes do grupo de trabalho da LIBER e autores do guia são Matias Frosterus, David Hansson, Maral Dadvar, Ilias Kyriazis, Sofia Zapounidou e Friedel Grant (FROSTERUS, 2021).

O estudo realizado por este grupo examinou as atividades LOD atualmente realizadas pelas Bibliotecas de pesquisa europeias. Os pontos principais trazidos pelos resultados desse estudo são (FROSTERUS, 2021):

- Projetos de dados conectados são diversos em caráter e escopo;
- A despesa mais notável relacionada à publicação de dados conectados é o trabalho humano;
- Não existe uma ferramenta que sirva para todos. Uma grande variedade de ferramentas como as comerciais, abertas e fontes especializadas são usadas junto com ferramentas desenvolvidas localmente;
- Os vocabulários mais comumente usados são GeoNames, VIAF, ISNI e Wikidata;
- Os esquemas de dados usados são frequentemente relacionados ao LOD, principalmente SKOS e Schema.org, com FOAF e Dublin Core também mencionados;
- As bibliotecas desejam cooperar e trocar ideias (FROSTERUS, 2021, p. 5, tradução nossa).

Frosterus et al. (2021) destacaram os motivos pelos quais as bibliotecas deveriam publicar seus dados em um formato aberto e conectado:

O principal benefício de publicar dados de biblioteca como LOD é que torna os dados prontamente disponíveis e mais fácil de usar para pesquisadores, desenvolvedores de sistemas, bibliotecários, etc. O formato LOD torna os dados mais atraentes e mais fáceis de analisar, combinar, e integrar. Além disso, tornar os dados 'vinculáveis' permite ao usuário enriquecê-los com a ajuda de recursos externos. Este enriquecimento pode ser realizado adicionando, por exemplo, dados faltantes (por exemplo, o ano faltante da morte de uma pessoa) ou novas informações (por exemplo, coordenadas geográficas). Links também podem ser usados para encontrar discrepâncias nos dados e, assim, corrigir erros (FROSTERUS et al., 2021, p.5, tradução nossa).

Um dos desafios sobre o sucesso de um projeto LOD, apontado por Frosterus et al. (2021), diz respeito ao controle dos dados publicados: os dados quando abertos dificultam o rastreamento de quem faz uso deles. Mesmo sendo possível rastrear os downloads, uso de API e outras estatísticas da *Web* não é possível se ter ideia de como os dados são usados.

Incentivar os usuários a dizerem como eles utilizaram os dados é uma alternativa, para que esses feedbacks possibilitem a elaboração de um relatório para um melhor entendimento quanto ao uso dos dados. Procurar links para os identificadores criados também é uma opção e pode mostrar quem está usando esses dados, mas medir o sucesso ainda é um desafio.

O Grupo de trabalho LIBER LOD desenvolveu 6 (seis) etapas para publicar dados abertos conectados às bibliotecas. As etapas foram assim definidas (FROSTERUS et al., 2021):

Etapa 1 - Planejando o processo de publicação;

Etapa 2 - Escolha e curadoria do conjunto de dados;

Etapa 3 - Identificar os recursos para vincular;

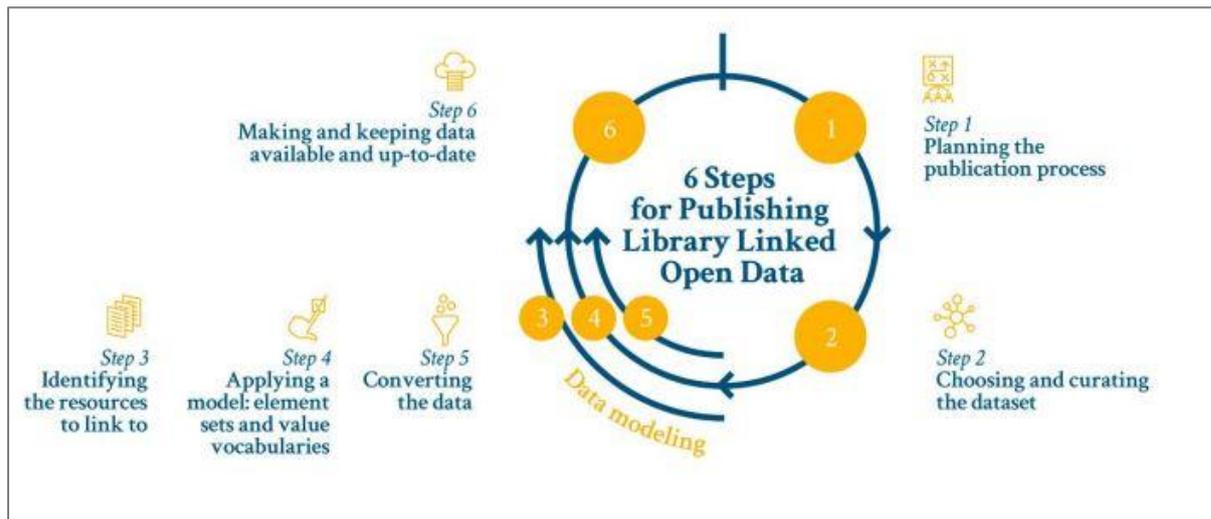
Etapa 4 - Aplicação de um modelo: conjuntos de elementos e vocabulários de valores;

Etapa 5 - Convertendo os dados;

Etapa 6 - Disponibilizando e mantendo os dados disponíveis e atualizados.

A figura 9 seguinte sintetiza essas etapas:

FIGURA 9 - Seis etapas para bibliotecas publicarem dados abertos conectados



Fonte: Frosterus; et al, 2021

Sobre as 6 (seis) etapas identificadas na figura acima, Frosterus et al. (2021) enfatizam e esclarecem que:

É importante notar que a progressão descrita das etapas acima não é estritamente linear. Iterações para revisitar etapas podem ocorrer e novos ajustes podem ser feitos. Este aspecto se aplica especificamente às três etapas do processo de 'modelagem de dados', que são altamente interligadas e podem ser executados simultaneamente em vez de sequencialmente. Além disso, deve-se notar que nem todas as seis etapas são sempre necessárias. Se, por exemplo, você está construindo um modelo LOD do zero, obviamente não há dados anteriores que exijam conversão. Portanto, as etapas servem como uma diretriz que pode ser adaptada de acordo com as necessidades de LOD (FROSTERUS et al., 2021, p. 6, tradução nossa).

Os autores concluem fazendo indicações de fontes importantes para quem deseja aprofundar-se no assunto. Uma das fontes citadas é o projeto LD4L (*Linked Data for Libraries*), que há vários anos trabalha na coleta e produção de recursos e ferramentas para auxiliar na publicação de dados abertos conectados para bibliotecas (FROSTERUS et al., 2021).

As conferências de bibliotecas também são citadas como importantes nesse contexto de aprendizagem por possibilitarem a troca de experiências e ideias. A SWIB (*Semantic Web in Libraries*) e ELAG (*European Library Automation Group*) são as mais importantes para questões de dados conectados às bibliotecas. Merece destaque também a *Linked Open Data Cloud* ou Nuvem LOD, que apesar de não ter como alvo específico às bibliotecas coleta um grande número de conjuntos de dados

interconectados e é um ambiente conhecido por enviar conjuntos de dados e ganhar visibilidade (FROSTERUS et al., 2021).

Os autores finalizaram enfatizando que o sucesso da implantação dos dados abertos conectados em bibliotecas está relacionado ao projeto que será desenvolvido:

É preciso dizer que tudo depende dos dados e do escopo do projeto. Se o projeto é pequeno, ou se está seguindo os passos de quem já fez algo muito semelhante, algumas das etapas podem ser quase triviais. Por outro lado, se houver muitos tipos diferentes de dados ou se a ambição do projeto é alta, algumas das etapas podem ser bastante trabalhosas. Então, planejamento adequado percorre um longo caminho e, considerando as etapas que descrevemos aqui, simplifique um pouco o seu planejamento (FROSTERUS et al., 2021, p. 17, tradução nossa).

O estudo descrito a seguir constitui uma das pesquisas mais abrangentes identificadas na literatura. Esse estudo foi realizado em 2018 e desenvolvido pela *OCLC Research Library Partnership*; liderado por Karen Smith-Yoshimura, pesquisadora e membro sênior do programa. A pesquisadora trabalha com instituições de pesquisa afiliadas da OCLC e se concentra em questões relacionadas aos metadados necessários para descrever e fornecer acesso aos recursos multilíngues gerenciados por bibliotecas, arquivos, museus e outras organizações de patrimônio cultural (SMITH-YOSHIMUR, 2018).

O estudo pesquisou acerca de dados conectados direcionados àquelas instituições que já haviam implementado um projeto ou serviço de dados conectados, ou estavam em processo de fazê-lo, em sua grande maioria composto por bibliotecas. Foram feitas perguntas sobre a publicação de dados conectados e o consumo desses mesmos dados (SMITH-YOSHIMUR, 2018).

O estudo se propôs a “apresentar uma visão geral dos projetos de dados conectados ou instituições de serviços que implementaram ou estão implementando; quais dados eles publicam e consomem; as razões apresentadas para implementar dados conectados e as barreiras encontradas” (SMITH-YOSHIMURA, 2018, não paginado, tradução nossa).

A pesquisa inicial foi feita entre os meses de julho e agosto do ano de 2014, utilizando um instrumento de pesquisa desenvolvido pela OCLC e instituições afiliadas, testado por vários implementadores de dados conectados. O link para a

pesquisa foi distribuído por meio de mala direta e no *Twitter*. O instrumento de pesquisa não foi disponibilizado no artigo em estudo.

Algumas lacunas foram apontadas pelos participantes da pesquisa inicial realizada em 2014, e, para preencher essas lacunas, ela foi realizada novamente entre junho e julho de 2015.

O segundo estudo foi entre 17 de abril e 25 de maio de 2018 para apurar as mudanças ocorridas desde a última pesquisa. Um total de 143 instituições em 23 países representa o número de organizações que participaram da pesquisa de 2018, somadas aos participantes das pesquisas anteriores.

Em 2018 foram analisados 104 projetos ou serviços de dados conectados, descritos pelas 81 instituições participantes, envolvendo tanto aquelas que publicam dados conectados, quanto as que consomem dados conectados ou ambas.

A pesquisa de 2018 contou com a participação de 34 instituições dos Estados Unidos, representando 42 % dos participantes; seguido pela Espanha (12); Reino Unido (8); Holanda (4). Participaram também com três respostas o Canadá, Alemanha e Noruega; duas respostas da Itália; uma resposta da Austrália, Áustria, China, República Tcheca, Finlândia, França, Hungria, Japão, Luxemburgo, Portugal, África do Sul e Suíça.

O tipo de instituições participantes foi categorizado como Bibliotecas de pesquisa que representam 28 % dos participantes de 2018 (23); seguidas por 13 Bibliotecas Nacionais (16 %); 11 Instituições de Pesquisa (14 %); 08 Redes de Bibliotecas; 08 Instituições Governamentais (10 %); 06 Provedores de serviços (7 %); 05 Bibliotecas Públicas (6 %); 04 Museus (5 %); e ainda “outros” 02 (uma sala de concertos e uma editora).

Os resultados alcançados com o desenvolvimento da pesquisa podem ser visualizados nas 9 (nove) tabelas a seguir, identificadas de 3 a 11:

TABELA 3: Comparação de pessoal envolvido nas atividades de dados conectados nas pesquisas de 2018 e 2015

Pessoal	2018		2015	
As atividades de dados conectados foram somadas às responsabilidades da equipe atual.	86	85 %	98	92 %
Temos uma equipe dedicada a projetos de dados conectados.	30	30 %	20	19 %
A instituição adicionou ou está adicionando uma nova equipe com experiência em dados conectados.	10	10 %	4	4 %
A instituição adicionou ou está adicionando pessoal temporário com experiência em dados conectados.	13	13 %	13	12 %
A instituição contratou consultores externos com experiência em dados conectados.	12	12 %	17	16 %
Quantidade de entrevistados que responderam às questões	101		107	

Fonte: Smith-Yoshimur, 2018, tradução nossa.

TABELA 4: Respostas da pesquisa sobre como os dados conectados são usados

Como os dados conectados são usados	Pesquisa de 2018	Pesquisa de 2015
A instituição consome dados conectados	34	38
A instituição publica dados conectados	5	10
A instituição consome e publica dados conectados	65	64

Fonte: Smith-Yoshimur, 2018, tradução nossa.

TABELA 5: Principais motivações para publicar dados conectados

Principais motivações para publicar dados conectados	2018	2015
Expor os dados a um público maior na web.	74 %	73 %
Demonstrar o que pode ser feito com conjuntos de dados como dados conectados.	65 %	64 %
Ouvi falar sobre dados conectados e queria experimentar expondo alguns dados locais como dados conectados.	45 %	47 %
Necessário para publicar dados conectados para consumi-los.	25 %	
Explorar-se a publicação de dados como dados conectados melhorará a otimização do mecanismo de busca (SEO) para recursos locais.	24 %	30 %
A pedido da administração que solicitou que exponhamos nossos dados como dados conectados.	11 %	5 %

Fonte: Smith-Yoshimur, 2018, tradução nossa.

TABELA 6: Os 8 principais vocabulários / ontologias em RDF usados em 2018 em comparação com 2015

Vocabulários / ontologias em RDF usados	2018		2015	
Schema.org	44	46 %	30	30 %
SKOS	42	44 %	59	60 %
Dublin Core Terms	39	41 %	51	52 %
FOAF	36	38 %	55	56 %
DCMI Metadata Terms	35	37 %	49	49 %
RDF Schema	35	37 %	45	45 %

BibFrame	26	27 %	15	15 %
Vocabulário Local	21	22 %	19	19 %

Fonte: Smith-Yoshimur, 2018, tradução nossa.

TABELA 7: Tecnologias usadas para publicar dados conectados

Nº de projetos que utilizam	Tecnologia (ordem de frequência)
Mais que 20	SPARQL, Java
10-20	Python, XSLT, RDF Store, Solr, Jena Applications, Virtuoso Universal Server (fornece endpoint SPARQL),
2-9	Google Refine, Apache Fuseki, Blazegraph, GraphDB (anteriormente OWLIM da Ontotext Software), DIGIBIB for Libraries, 4store, Fedora Commons, Map / Reduce, Metafacture, Django, Elasticsearch, AllegroGraph, Drupal7, OpenRDF, Pubby
1	Plataforma 4store Seme4, Amazon Neptune, Apache Spark, ARC2 em PHP, Arches, bib-lod-ui (aplicativo da Web para publicação de dados abertos conectados bibliográficos), <i>bib-rdf-pipeline</i> (conversão de MARC em RDF), <i>Blacklight</i> , Catamandu, Cliopatria, <i>Cubicweb</i> , bibliotecas D3, conversor FAST, modelos <i>FreeMarker</i> , <i>Government Site Builder</i> [Alemanha], Hbase / Hadoop, JAX-RS, MARC Report e MARC Global (do MARC of Quality), MongoDB, <i>Mapping Memory Mapper (3M)</i> , <i>MarkLogic Semantics</i> , <i>Orbean Xforms</i> , identificadores permanentes para a Web, RDFLib para <i>Python</i> , <i>ResearchSpace</i> , <i>Ruby on Rails</i> , biblioteca <i>Skosmos</i> <i>Skosify EasyRdf</i> para PHP, visualizador de resultados SPARQL, editor <i>Squebi</i> SPARQL, <i>Stardog</i>

Fonte: Smith-Yoshimur, 2018, tradução nossa.

Ressalvasse que os resultados descritos na tabela 7 apresentam SPARQL como linguagem para publicar dados conectados. No entanto, trata-se de uma linguagem de consulta a dados em RDF (W3C, 2008).

TABELA 8: Barreiras encontradas na publicação de dados conectados

Barreiras – Desafios	2018		2015	
Curva de aprendizado íngreme para a equipe	41	51 %	40	51 %
Inconsistência em dados antigos (<i>legacy data</i>)	38	48 %	33	42 %
Seleção de ontologias apropriadas para representar nossos dados	26	33 %	31	39 %
Falta de recursos	23	29 %		
Pouca documentação ou orientação sobre como construir os sistemas	23	29 %	21	27 %
Estabelecendo os links	22	28 %	27	34 %
Falta de ferramentas	18	23 %	15	19 %
Software imaturo	17	21 %	11	14 %
Verificação sobre quem é o proprietário dos dados	4	5 %	10	13 %
Outro	19	24 %	21	27 %
Entrevistados para a pergunta	80		79	

Fonte: Smith-Yoshimur, 2018, tradução nossa.

TABELA 9: As 10 principais fontes de dados vinculadas consumidas

Dez principais fontes de dados conectados consumidas	2018		2015	
id.loc.gov	39	57 %	35	51 %
VIAF	36	51 %	41	60 %
Dbpedia	30	43 %	36	53 %
<i>GeoNames</i>	29	42 %	35	51 %
<i>Wikidata</i>	28	41 %	6	9 %
<i>WordCat.org</i>	28	41 %	15	22 %
Vocabulários <i>Getty</i>	23	33 %	16	24 %
FAST	17	25 %	15	22 %
ISNI	17	25 %	8	12 %
Recursos convertidos em dados conectados por conta própria	13	19 %	17	25 %

Fonte: Smith-Yoshimur, 2018, tradução nossa.

TABELA 10: Principais motivações para consumir dados conectados

Principais motivações para consumir dados conectados	2018	2015
Oferecer aos usuários locais uma experiência mais rica	78 %	75 %
Aprimorar os dados locais consumindo dados conectados de outras fontes	71 %	74 %
Ouvi falar sobre dados conectados e queria experimentá-lo usando recursos de dados conectados	33 %	25 %
Gerenciamento de metadados interno mais eficaz	30 %	47 %
Experimentar combinar diferentes tipos de dados em um único armazenamento triplo	29 %	25 %
Maior precisão e abrangência nos resultados da pesquisa local	28 %	40 %
Explorar se o consumo de dados conectados de fontes externas melhora o <i>Search Engine Optimization</i> (SEO) para recursos locais	10 %	28 %

Fonte: Smith-Yoshimur, 2018, tradução nossa.

TABELA 11: Barreiras encontradas no consumo de dados conectados

Barreiras – Desafios	2018		2015	
Combinar, eliminar a ambiguidade e alinhar os dados de origem aos recursos de dados conectados	28	48 %	23	39 %
O que é publicado na internet como dados conectados nem sempre é reutilizável ou necessita de URIs	18	31 %	16	27 %
Tamanho dos RDF <i>dumps</i> (despejos RDF)	16	28 %	12	20 %
Endpoints instáveis	16	28 %	10	17 %
Confiabilidade do serviço	15	26 %	9	15 %
Mapeamento de vocabulário	15	26 %	17	29 %
Compreender como os dados são estruturados antes de usá-los	14	24 %	12	20 %
Falta de ferramentas prontas para uso	14	24 %	10	17 %
Conjuntos de dados desatualizados	13	22 %	14	24 %
Falta de controle de autoridade	11	19 %	15	25 %
Volatilidade de formatos de dados	10	17 %	11	19 %
A desambiguação de termos em diferentes idiomas é difícil	8	14 %	6	10 %
É difícil fazer com que outras instituições façam sua própria harmonização entre objetos e conceitos	7	12 %	9	15 %

Fonte: Smith-Yoshimur, 2018, tradução nossa.

O estudo de Smith-Yoshimur (2018), apontou que a atividade e o uso entre serviços de dados conectados aumentaram e que o surgimento de provedores de serviços pode resultar em menos organizações individuais lançando seus próprios projetos de dados conectados. Entre os entrevistados na pesquisa de 2018, 37% são, pelo menos parcialmente, dependentes de fornecedores de sistemas, empresas, consultores ou desenvolvedores externos.

Algumas diferenças foram identificadas em comparação com a pesquisa realizada em 2015:

- Ainda menos projetos ou serviços que publicam apenas dados conectados, com um aumento espelhado daqueles que publicam e consomem dados conectados;
- Mais pessoal dedicado a dados conectados;
- Um aumento na publicação de dados conectados em *schema.org* e *Bibframe*, refletido por uma diminuição em SKOS;
- A ascensão do *Wikidata* como uma fonte de dados conectados. Observou-se que na Biblioteca Nacional da Finlândia o *Wikidata* está se tornando cada vez mais significativo para a biblioteca.

A maioria dos projetos e serviços conectados descritos foram implementados nos últimos dois anos, ou pelo menos está em fase de implementação. Portanto, essas diferenças podem refletir novas tendências em vez de mudanças nas implementações. Uma constatação importante é que a maioria dos projetos e serviços de dados conectados permanecem de natureza experimental ou educacional. Como exceção a pesquisa cita a Biblioteca Pública de Oslo: “Pelo que posso ver, a biblioteca pública de Oslo ainda é a primeira e única biblioteca com seu catálogo de produção e fluxos de trabalho de catalogação originais feitos diretamente com dados conectados” (SMITH-YOSHIMUR, 2018, não paginado, tradução nossa).

A pesquisa apresentada descreve, portanto, os projetos de dados conectados implementados ou em implementação por organizações, quais dados eles publicam e usam; explicam as razões para a implementação de dados conectados e os obstáculos encontrados; fornece informações e sugestões destinadas aqueles que estão pensando em implementar projetos ou serviços de dados conectados; e ainda incentiva o aumento de iniciativas de implementação de dados conectados.

Outra pesquisa identificada foi desenvolvida pelo *W3C Library Linked Data Incubator Group*³, no período de maio de 2010 a agosto de 2011. O grupo *Library Linked Data Incubator Group* contou com a participação de especialistas nas áreas de Ciência da Informação e Ciência da Computação, incluindo representantes de

³ A missão do grupo incubador *Library Linked Data* é ajudar a aumentar a interoperabilidade global de dados de biblioteca na Web, reunindo pessoas envolvidas em atividades da Web Semântica - com foco em *Linked Data* - na comunidade de bibliotecas e além, com base em iniciativas existentes, e identificando caminhos de colaboração para o futuro (W3C INCUBATOR GROUP REPORT, 2011, sem paginação, tradução nossa).

instituições que têm conhecimento reconhecido no desenvolvimento de tecnologias para bibliotecas, arquivos e museus (ARAKAKI, 2016).

Os autores que conduziram o estudo são Thomas Baker, da *Dublin Core Metadata Initiative* (EUA); Emmanuelle Bermès, do *Centre Pompidou* (França); Karen Coyle, Consultora (EUA); Gordon Dunsire, Consultor (Reino Unido); Antoine Isaac, da *Europeana e Vrije Universiteit Amsterdam* (Holanda); Peter Murray, da LYRISIS (EUA); Michael Panzer e Jeff Young da OCLC *Online Computer Library Center Inc.* (EUA); Jodi Schneider, *Deri Galway da National University of Ireland Galway* (Irlanda); Ross Singer, Talis Group Ltd. (Reino Unido); Ed Summers, da Biblioteca do Congresso (EUA); William Waites, da Universidade de Edimburgo (Reino Unido); e Marcia Zeng, da *Kent State University* (EUA) (BAKER et.al., 2011, não paginado, tradução nossa).

Os dados que serão apresentados a seguir foram retirados do relatório final elaborado pelo W3C *Library Linked Data Incubator Group* contendo os resultados dos estudos e as considerações realizadas.

O foco dos estudos baseou-se na exploração de instrumentos já utilizados por bibliotecas como modelos de metadados, esquemas, padrões e protocolos para a interoperabilidade. Esta atividade almejou tornar as bibliotecas provedoras de fontes confiáveis para a *Web* de dados, principalmente em virtude da possibilidade de disponibilização de conjuntos padronizados que podem ser consumidos, compartilhando instrumentos como catálogos de autoridades, vocabulários controlados, entre outros (BAKER et al., 2011; SERRA, 2019).

O escopo do relatório "*Library Linked Data*" definiu os seguintes entendimentos para os termos: biblioteca, dados de biblioteca, *Linked Data*, *Open Data* e *Library Linked Data*, conforme detalhamento a seguir:

1) Biblioteca: o termo biblioteca refere-se ao conjunto de instituições do âmbito cultural e histórico, incluindo bibliotecas, museus e arquivos, independentemente se suas coleções são pequenas ou grandes, se são formadas por recursos analógicos ou digitais, ou sob custódia de instituições públicas ou privadas;

2) Dados de biblioteca (*library data*): se refere a qualquer tipo de informação digital produzida e sob curadoria de bibliotecas. Foram consideradas três categorias de dados de biblioteca:

a) Conjunto de dados (*datasets*): são os que descrevem recursos relacionados à biblioteca, como a Bibliografia Nacional Britânica, o catálogo da biblioteca nacional húngara, a Biblioteca Aberta, CrossRef e Europeana;

b) Conjuntos de elementos (*element sets*): como termos de metadados DCMI, os elementos de RDA, descrição e acesso de recursos, sistema de organização de conhecimento simples (SKOS) e o vocabulário FOAF;

c) Vocabulários de valores (*value vocabularies*): como o Cabeçalhos de assuntos da Biblioteca do Congresso, AGROVOC, Arquivo de Autoridade Internacional Virtual (VIAF), Classificação Decimal de *Dewey* e *GeoNames*.

3) *Linked Data*: são considerados dados em *Linked Data*, dados conectados, aqueles publicados de acordo com os princípios definidos para facilitar o vínculo entre datasets, ou seja, utilizando URIs como identificadores únicos de recursos, expressos sob o modelo RDF, especificando as relações existentes entre as coisas;

4) *Open Data*: são considerados dados abertos aqueles que não possuem restrição legal para uso, reuso ou distribuição. O *Linked Data* não precisa necessariamente ser aberto, porém ao usar dados desta natureza, potencializam-se as possibilidades de uso.

5) *Library Linked Data*: é qualquer tipo de dados da biblioteca (conforme definido acima) que é expresso como *Linked Data* (BAKER et al., 2011, não paginado, tradução nossa).

Observou-se que os autores apontaram vários benefícios relacionados à abordagem do uso dos dados conectados em bibliotecas. Destacaram-se a seguir as seguintes vantagens:

1) Ao usar identificadores únicos para especificar obras, locais, pessoas, eventos, tópicos e outros objetos ou conceitos de interesse, a biblioteca permitirá que recursos sejam referenciados a partir de uma ampla gama de fontes de dados, tornando assim as descrições de metadados mais acessíveis. As bibliotecas e instituições de memória se destacam nesse sentido, sendo as mais indicadas para fornecer metadados confiáveis para Web, devido sua longa experiência na descrição de metadados bibliográficos.

2) Os dados de autoridade da biblioteca para nomes e assuntos ajudarão a reduzir a redundância de descrições bibliográficas na Web, ao identificar claramente as principais entidades que são compartilhadas entre os dados conectados. Isso também ajudará a reduzir a redundância de metadados que representam coleções de bibliotecas.

3) À medida que os links entre os dados estruturados se tornam conectados, os usuários terão recursos aprimorados para descobrir e usar dados. A navegação conjunta entre recursos de bibliotecas e outras instituições será aprimorada e os usuários terão um conjunto mais rico de caminhos de navegação (BAKER et al., 2011, não paginado, tradução nossa).

Outros aspectos verificados foram os principais problemas apontados pelo relatório referentes ao atual formato do registro bibliográfico das bibliotecas. São eles:

1) Os dados da biblioteca não são integrados aos recursos da *web*, mantendo a estrutura de ilhas ou silos de informação observados até então.

2) Muitos padrões adotados pela biblioteca, como o MARC 21, ISO 2709, protocolo Z39.50 e etc., são usados apenas pela comunidade bibliotecária. Além disso, grande parte dos recursos bibliográficos é expressa em linguagem natural ou em códigos MARC. Alguns dos identificadores carregados em registros MARC, como ISBNs para livros, poderiam, em princípio, ser usados para vinculação, mas somente após serem extraídos dos campos de texto nos quais estão embutidos e, em seguida, normalizados. Alguns campos de dados, como nomes e assuntos controlados por autoridade, têm registros relacionados em arquivos separados e esses registros têm identificadores que podem ser usados para representar essas entidades nos metadados da biblioteca. No entanto, os formatos de dados em uso nem sempre suportam a inclusão desses identificadores nos registros. Portanto, muitos dos sistemas de bibliotecas atuais não suportam adequadamente seu uso. Esses identificadores também tendem a ser gerenciados localmente em vez de globalmente e, portanto, não são expressos como URIs que permitiria links para eles na *web*. A ausência de links ou suporte insuficiente para eles nos sistemas de biblioteca levanta questões importantes. As alterações nos identificadores dos registros de autoridades exigem que todos os registros bibliográficos relacionados sejam recuperados para alterar suas sequências de texto, um processo disruptivo e caro que muitas vezes impede as bibliotecas de implementar alterações em tempo hábil.

3) A abordagem de dados abertos em bibliotecas pode ser prejudicada por diferenças de conceitos e terminologia entre a biblioteca e a comunidade da *Web* semântica. Poucos bibliotecários falam sobre "declarações" de metadados, e a comunidade da *Web* Semântica desconhece conceitos que sejam claramente equivalentes a "cabeçalhos de assuntos" ou "controle de autoridade". Cada comunidade tem seu próprio vocabulário, o que reflete as diferenças em suas opiniões. Ambas as partes devem aumentar o entendimento mútuo, pois os dois grupos trazem conhecimentos importantes para a construção de redes de dados.

4) Mudanças na tecnologia de bibliotecas dependem do desenvolvimento de sistemas de fornecedores que desenvolvem e fornecem *softwares* gerenciadores para as bibliotecas. Isso significa que se as bibliotecas pretendem adotar iniciativas de dados conectados existe essa dependência em relação ao fornecedor, que precisa estar disposto a adaptar seus produtos às necessidades das bibliotecas. As bibliotecas não contam com uma infraestrutura própria quando desejam adotar o *Linked Data* (BAKER et al., 2011, não paginado, tradução nossa).

Em relação aos dados conectados disponíveis para bibliotecas, na ocasião da realização do estudo, o grupo observou que eram poucas as iniciativas de publicação de datasets (conjunto de dados) de bibliotecas. A Bibliografia Nacional Britânica é uma iniciativa pioneira nesse contexto, servindo de referência para que seja entendido o esforço necessário para resolver desafios como licenciamento, modelagem de dados, manipulação de dados e colaboração com várias comunidades de usuários.

Em contrapartida, muitos conjuntos de elementos de metadados (*element sets*) e vocabulários de valor (*value vocabularies*) foram publicados como *Linked Data*, abrangendo vocabulários emblemáticos, como *Library of Congress Subject Headings* e *Dewey Decimal Classification*. Conjuntos de elementos foram publicados como dados conectados ou em um formulário compatível com os esses dados, como os termos de metadados DCMI, e também, estruturas de referência, como Requisitos funcionais para registros bibliográficos (FRBR).

O grupo produziu um inventário que originou um relatório a parte de recursos úteis para criar ou consumir dados conectados no domínio da biblioteca, intitulado *Datasets, Value Vocabularies and Metadata Element Sets (VOCABDATASET)*⁴.

O relatório mostra que em muitas áreas a adoção inicial dos princípios e técnicas da *Web* semântica e dos dados conectados levaram ao desenvolvimento de conjuntos de dados e vocabulários maduros. Além disso, aponta áreas onde as bibliotecas e organizações relacionadas ainda podem fazer contribuições importantes.

Em relação à questão de direitos autorais, o relatório aponta que de acordo com as políticas, contratos e condições locais das bibliotecas, certos dados têm uso restrito. Portanto, os dados podem ter problemas de permissão pouco claros e não testados que os impedem de serem liberados como dados abertos. As questões de direitos variam muito entre os países, dificultando a colaboração na publicação de dados abertos.

Desde o surgimento dos catálogos online, a propriedade de registros de catálogos tornou-se muito complicada devido ao grau de compartilhamento de dados entre as bibliotecas. Frequentemente os registros são copiados e as cópias são modificadas ou aprimoradas para uso por catalogadores locais. Esses registros podem ser posteriormente adicionados aos diretórios de consórcios regionais, nacionais e internacionais. Dessa forma, é complexo atribuir direitos legais de propriedade intelectual entre esses agentes, e a falta de certeza impede o compartilhamento de dados abertos nesta comunidade que é tradicionalmente cautelosa com questões autorais.

⁴ Pode ser acessado por intermédio do link: <https://www.w3.org/2005/Incubator/ld/XGR-ld-vocabdataset-20111025/>.

O estudo apontou a importância da discussão sobre dados conectados e direitos autorais entre os gerentes de bibliotecas, buscando chegar a um acordo com os proprietários sobre direitos e permissões de uso em nível de consórcio de bibliotecas e até mesmo em escala nacional ou internacional. Um norteador para orientar os profissionais nesse sentido pode ser encontrado na seção de Direitos e Licenciamento do Guia de Dados Bibliográficos Abertos para bibliotecas de Ensino Superior do Reino Unido⁵.

Como recomendações, o relatório indica a inserção das bibliotecas na rede de dados, tanto para disponibilizar dados conectados, quanto para usar a rede de dados em serviços da biblioteca. Idealmente, os dados da biblioteca devem ser totalmente integrados com outros recursos na *Web* para criar maior visibilidade para a biblioteca e levar os serviços para os buscadores de informações. Quando a biblioteca conecta-se à rede de dados, ela pode assumir um papel de liderança de acordo com suas atividades tradicionais: gerenciamento de recursos para uso atual e preservação de longo prazo; descrição dos recursos de acordo com as regras acordadas; e, assim, respondendo às necessidades dos buscadores de informações.

Em outro estudo, desenvolvido por Jesus e Castro (2019), foi realizada uma Revisão Sistemática de Literatura (RS) com o objetivo de analisar como a relação entre dados bibliográficos e o *Linked Data* tem sido abordada pelos profissionais da Ciência da Informação. Foram analisados 56 artigos que resultaram na conclusão de que as principais preocupações quando se trata do tema são: “apresentar projetos de adequação de dados bibliográficos ao *Linked Data*; discutir o processo de conversão de dados bibliográficos ao *Linked Data*; e realizar discussões conceituais sobre a adequação” (JESUS; CASTRO, 2019, p.45).

Conclui-se também que há uma similaridade nos 56 artigos estudados em considerar imprescindível a adequação dos dados bibliográficos ao *Linked Data*, tratando essa adequação como um processo natural. Apesar de esses documentos mencionarem muitos desafios a serem superados e alguns estudos concluírem que as bibliotecas ainda estão muito longe de adequar seus dados ao *Linked Data*, há um consenso entre eles de que a adaptação é benéfica para a biblioteca.

⁵ Disponível no link: <http://obd.jisc.ac.uk/rights-and-licensing>

Jesus e Castro (2019) apontaram neste estudo as principais vantagens do uso do Linked Data por bibliotecas, descritas a seguir:

- 1) A possibilidade de interoperar os dados da biblioteca com os dados de fontes externas como os de outras instituições do patrimônio cultural (principalmente bibliotecas, arquivos e museus), fontes comerciais (casas publicadoras, e-commerce etc.) e os sites do ambiente Web;
- 2) Promover uma maior visibilidade dos recursos das bibliotecas no ambiente Web;
- 3) Promover melhores resultados de busca e;
- 4) Prover a descoberta acidental de recursos. Todas essas vantagens refletem a preocupação em melhorar os serviços oferecidos aos usuários (JESUS; CASTRO, 2019, p.48).

Em relação aos desafios identificados, Jesus e Castro (2019) destacam os seguintes:

- 1) Confiar em fontes externas para ligação de dados. A heterogeneidade dos dados disponibilizados na *Web* faz com que as bibliotecas tenham dificuldade em confiar em fontes externas para criar links semânticos. A preocupação em associar os dados das bibliotecas com os dados de fontes externas se reflete ainda na abordagem dos documentos, onde muitos se preocuparam em apresentar fontes confiáveis para a obtenção de dados em *Linked Data*.
- 2) Converter os dados bibliográficos descritos em MARC 21 para o RDF. A conversão dos dados descritos em MARC21 será um processo trabalhoso e custoso, os documentos abordam a criação ou a escolha de ferramentas que possam ser utilizadas para converter os dados bibliográficos de maneira automatizada sem que a consistência desses dados se perca no processo (JESUS; CASTRO, 2019, p. 48).

Os autores esclareceram que a pesquisa não esgota o tema discutido e que, ao contrário, proporciona um referencial teórico que pode colaborar com estudos futuros e desdobramentos da pesquisa.

Outro estudo correlato selecionado foi realizado por Serra e Santarem Segundo (2017, p. 167), cujo objetivo foi “discorrer sobre as possibilidades de aplicação de conceitos da *Web* semântica, por meio do *Linked Data*, no catálogo da biblioteca, favorecendo a troca de informações de forma dinâmica, legível por pessoas e máquinas, com dados estruturados e conectados”.

O estudo buscou identificar quais as possibilidades de aplicação do *Linked Data* nos catálogos das bibliotecas, onde possa favorecer a troca de dados entre os registros de uma mesma coleção, catálogo de outras instituições ou na *Web*, ampliando as possibilidades de exploração de recursos e documentos. Desse modo,

pretendeu-se identificar formas de aplicação do *Linked Data* aos catálogos, que fosse favorável para os usuários ao fornecer descrições detalhadas de recursos bibliográficos, e, também, ganhos aos profissionais da informação ao possibilitar o reuso dos dados por meio de ontologia e compartilhamento de elementos descritivos. O método empregado foi de pesquisa exploratória realizada a partir de um levantamento bibliográfico, com recorte temporal entre os anos de 2001 a 2016 (SERRA; SANTAREM SEGUNDO, 2017).

Os autores observaram que embora as bibliotecas utilizem e façam intercâmbio de dados, os catálogos ainda são identificados como silos de informação onde seus conteúdos ainda se encontram restritos à coleção de uma instituição. A *Web* semântica aplicada nesse contexto agrega elementos descritivos aos conteúdos dos catálogos, interligando dados e propiciando sua reutilização por meio do *Linked Data* (SERRA; SANTAREM SEGUNDO, 2017).

A *Web* semântica sugere uma nova estrutura de conteúdos para os sites da *Web*, criando possibilidades para que agentes computacionais e pessoas realizem suas tarefas de maneira otimizada, fazendo com que os sites não sejam mais páginas isoladas e sim interligadas entre si, agregando contexto na sua utilização. As principais tecnologias empregadas na estruturação da *Web* semântica são o modelo de dados *Resource Description Framework* (RDF), ferramentas para descrição semântica e lógica dos dados como o *RDF Schema*, *OWL* e o *SPARQ* (SERRA; SANTAREM SEGUNDO, 2017).

As iniciativas identificadas no estudo de Serra e Santarem Segundo (2017) sobre o emprego do *Linked Data* nos catálogos de bibliotecas, baseadas nos estudos de Marcun (2011) e Mitchell (2016), são apresentadas a seguir:

1) *Bibframe - Bibliographic Framework Transition Initiative*, criado em 2011 pela Library of Congress como um padrão de estrutura de dados, voltado para o ambiente *Web* e baseado em princípios do *Linked data* e no modelo de dados RDF. A iniciativa surgiu em virtude das atuais limitações identificadas no formato MARC que não é capaz de realizar a criação de relacionamentos entre registros.

2) A reestruturação da lista de cabeçalhos de assunto da *Library of Congress* (LCSH), baseada em padrões semânticos, permitindo compartilhamento da terminologia da instituição.

3) A *National Library of Medicine* (NLM) está desenvolvendo e testando aplicações em dados conectados e modelos RDF, incluindo o vocabulário *Medical Subject Headings* (MeSH).

4) O *Virtual International Authority File* (VIAF), que é um consórcio de cooperação internacional, adotou a disponibilização dos arquivos de autoridades em formato de Dados Abertos Conectados. Os autores destacaram que:

Ao selecionar uma autoridade “Pessoa” em uma fonte aberta e com estrutura de dados interligados como o VIAF, por exemplo, informações adicionais como datas, local de nascimento e/ou atividade, obras publicadas, temas de interesse, relacionamentos com outros autores etc., são incluídos ao registro de forma dinâmica, minimizando o tempo de descrição necessário, agregando contexto à autoridade (SERRA e SANTAREM SEGUNDO, 2017, p. 182).

5) Além do VIAF, outras alternativas de controle de autoridade também estão disponibilizando informações nos padrões de *Linked Data*. São eles: ORCID, *ResearcherID*, *Scopus Author ID*.

6) Como projeto foi identificado o *Linked Data for Libraries* (LD4L) como um projeto colaborativo que visa desenvolver estudos para facilitar a transição dos fluxos de trabalhos na produção de metadados à comunidade de *Linked Data*, contribuindo com estudos sobre o *Bibframe* e criação de ontologias. Participam do projeto as *Universidades de Cornell, Harvard, Stanford, Columbia, Princeton* e a *Library of Congress*.

7) Como pesquisas anteriores foi citada o estudo de Mitchell (2016), realizado em 2004, que visou identificar casos de uso em *Linked Data* por bibliotecas, destacando-se que as principais motivações encontradas são:

- a) Enriquecer a descrição do registro bibliográfico;
- b) Relacionar registros harmonizando diversas fontes, como se fosse uma fonte de referência;
- c) Automatizar o controle de autoridades;
- d) Enriquecer a aplicação utilizada pela instituição;
- e) Publicar dados de forma ampla, apresentando casos de uso potenciais (SERRA; SANTAREM SEGUNDO, 2017, p.182).

Os autores enfatizaram nas considerações finais do estudo que a agregação de *Linked Data* na estrutura dos catálogos representa um avanço no tratamento

descritivo da informação, permitindo que os dados sejam preenchidos com o contexto e que sejam interoperáveis, facilitando assim a reutilização de recursos.

Eles ressaltaram ainda que o uso do *Linked Data* nos catálogos deve ser constantemente estimulado, sobretudo em instituições que têm acervos raros, e também naquelas que elaboram e mantêm catálogos de controle de autoridades, tanto de pessoas como de assuntos (SERRA; SANTAREM SEGUNDO, 2017).

E, por fim, como desafios o estudo de Serra e Santarem Segundo (2017) apontou que os bibliotecários precisam estar atentos ao uso de ontologias e catálogos apropriados para determinar as ferramentas a serem utilizadas, a fim de proteger o escopo da instituição e seus usuários.

Outra revisão sistemática de literatura selecionada foi a realizada por Ali e Warraich (2018) que teve como objetivo “explorar iniciativas de dados conectados (LD) em bibliotecas e centros de informação, juntamente com fatores motivadores para iniciar esses projetos de LD e desafios enfrentados por bibliotecários na implementação da tecnologia LD” (ALI; WARRAICH, 2018, p.925, tradução nossa).

O estudo se concentrou nas seguintes questões de pesquisa:

- 1) Quais são os fatores motivacionais para iniciar um projeto *Linked Data* em bibliotecas e centros de informação?
- 2) Quais são os desafios para implementar uma tecnologia LD em bibliotecas e centros de informação?
- 3) Qual etapa do processo de LD está focada em iniciar esses projetos em bibliotecas e centros de informação? (ALI; WARRAICH, 2018, p. 926, tradução nossa).

O estudo de Ali e Warraich (2018) identificou projetos de LD no Brasil, França, Irã, Irlanda, Japão, Reino Unido e EUA. A tabela 12 relaciona os dados extraídos dos 12 (doze) estudos selecionados para a pesquisa:

TABELA 12: Dados extraídos dos estudos selecionados

Estudo	País	Ano	Fatores motivadores	Desafios	Foco do projeto ou estágio do LD
Charles et al.,	Não se aplica. (O estudo aborda bibliotecas europeias que fazem parte da European Library e Europeana)	2014	Existem muitos benefícios de LD e web semântica.	<p>A tarefa LD exige recursos humanos e recursos tecnológicos.</p> <p>A tarefa LD exige experiência em tecnologia da informação e tecnologia da <i>Web</i>.</p>	Focado em vocabulário aberto vinculado à pesquisa multilíngue.
Cole et al.,	EUA	2013	Existem limitações ao usar MARC com RDF em comparação com o esquema de descrição do objeto de metadados e outros esquemas.	<p>Para implementação do LD em bibliotecas, deve haver diretrizes claras sobre a conversão de dados da biblioteca em dados abertos.</p> <p>Para implantação de dados abertos conectados, há uma necessidade de uma operadora para registros bibliográficos de biblioteca, como o schema.org.</p> <p>Para converter dados bibliográficos em RDF, há muitas opções disponíveis. Vários primeiros experimentos mostram que eles criam seus próprios namespaces para os dados da biblioteca.</p>	Conversão de registro Marc em RDF.

Debruyne et al.,	Irlanda	2016	Os princípios de LD e RDF fornecem oportunidade fácil para modelagem de dados. Esforço colaborativo de arquivista digital e engenheiros de computação fornece uma plataforma para colaboração na forma de um projeto LD.	Problemas durante a criação de triplos RDF. Necessária equipe para curadoria de informações e engenheiros para tarefas técnicas.	Focado no processo de construção de ontologias e arquitetura semântica.
Deliot, C.	Reino Unido	2014	Aumento do compromisso do governo para fornecer acesso a um público mais amplo, convertendo dados do formato específico da biblioteca para o padrão baseado em XML. Conjuntos de dados consistentes em fontes de dados autênticos fornecem informações sobre as publicações do Reino Unido. Os dados a serem publicados no LD eram consistentes, pois todos os registros têm números de classificação e cabeçalhos de assuntos elaborados a partir de vocabulário controlado.	Problema na conversão de MARC. Avançar com a nova tecnologia apresenta problemas. Há possibilidades de duplicação em URIs ao gerar URIs a partir de títulos de nomes. A decisão sobre ontologia era problemática para usar ontologias existentes ou criar novas ontologias.	Focado na transformação de dados bibliográficos MARC 21 em LD e no desenvolvimento de modelo de dados RDF.

Eslami, S. e Vaghefzadeh, M.H.	Irã	2013	Os modelos RFD da IFLA oferecem oportunidade para criar uma versão LD dos arquivos de autoridade da Biblioteca Nacional e Arquivos do IRÃ (NLAI).	A seleção da ontologia apropriada é um grande desafio. Muitas ferramentas são usadas para vincular dados a recursos externos. O NLAI não pode publicar LD e ainda está no estágio inicial. Alguns registros do NLAI estão incompletos ou expressos em língua persa. Isso cria problemas de conexão com fontes externas. A manutenção de links é um desafio para a NLAI. Quando conjuntos de dados interligados são atualizados, pode ocorrer um erro de links inválidos.	Transformação de dados em RDF.
Hanson, E.M	Carolina do Norte (EUA)	2014	Conclusão bem sucedida do projeto com a equipe pequena e pouco tempo motivada para novas iniciativas no futuro.	Serialização RDF e saber como escrever XSLT consumiam tempo e energia para a equipe. Ferramentas de reconciliação sofisticadas são necessárias para o processo LD.	Focado nas várias etapas do projeto LD.
Haslhofer, B. e Isaac, A.	Não se aplica. (O texto ao qual o autor se refere aborda a Europeia. Fazem parte da Europeia 4.000 instituições em toda Europa)	2011	Disponibilizar metadados na <i>Web</i> para terceiros através da tecnologia de dados conectados. Fazer um novo modelo de ESE para EDM. Fornecer facilidade para os provedores de dados Europeia converterem seus metadados para LD.	Problema de mapeamento de dados de elementos semânticos (Europeana - ESE) para modelo de dados (Europeana - EDM). A conversão de URIs ESE em URIs não referenciáveis (EDM) representa um grande desafio.	Focado no processo completo de LD.

Illien, G.	França	2012	Benéfico para outros aplicativos de bibliotecas. Os recursos da BNF podem se conectar com outros recursos de bibliotecas. Outras organizações podem usar metadados de data.bnf.fr. LOD também são úteis para empresas privadas, pois podem usar metadados e criar novos produtos com base de metadados existentes.	Enorme carga financeira.	Metade do processo de LD foi feito.
Rademaker et al.,	Brasil	2015	A tecnologia proposta em relação ao LD pode ser usada para outros arquivos e bibliotecas digitais para fins de migração. Menos dependência da equipe de TI após a conclusão do projeto LD. O projeto LD oferece muitas oportunidades de pesquisa para análise automática no domínio LD.	Enfrenta desafios devido ao modelo de banco de dados relacional existente em um novo modelo aprimorado e implementado.	Focado no desenvolvimento da arquitetura baseada em recursos LD.

Ryan et al.,	Irlanda	2015	A conversão de dados bibliográficos em LD aprimora as capacidades de pesquisa e descoberta. O Logainm, banco de nomes de locais irlandeses, fornece a capacidade de mostrar os em uma natureza hierárquicas e criar relacionamentos entre eles.	Correspondência de grafias do idioma irlandês com conjuntos de dados externos. Correspondência de categorias de conjuntos do banco de dados Logainm com conjuntos de dados externos.	Para criar um novo conjunto de dados LD. Conversão de dados XML em dados de links RDF para conjuntos de dados externos, como DBpedia.
Szekely et al.,	EUA	2013	O LD do <i>Art Museum</i> (SAAM) será útil para o próprio SAAM e sua comunidade, pois o foco é criar links entre dados dos artista SAAM para DBpedia e artista SAAM para a lista de artista do <i>Getty</i> . O objetivo por trás do projeto SAAM LD era desenvolver uma ferramenta para mapear dados estruturados em RDF.	Desafios enfrentados durante a preparação de dados, pois é necessário filtrar e transformar dados para converter e modelar dados RDF. Problemas relacionados ao mapeamento de dados. Outros problemas foram o design de ontologias.	Focado no processo completo de LD.
Nagai et al.,	Japão	2015	Para promover LD e fornecer metadados de bibliotecas como LD.	Ambiguidade encontrada nos dados bibliográficos da biblioteca. Não disponibilidade de dados como uma licença aberta. Vocabulários padrões e metadados necessários para a conversão de dados bibliográficos em LD.	Focado no processo completo de LD.

Os dados extraídos dos estudos selecionados sobre iniciativas de *Linked Data* em bibliotecas e centros de informação também revelaram que existem muitos fatores referentes à motivação e desafios para a implementação do *Linked Data*. Esses dados estão representados no Quadro 5 por meio da matriz SWOT⁶, que é uma ferramenta utilizada no processo de planejamento estratégico.

QUADRO 5: Matriz SWOT dos fatores motivacionais e desafios de Linked Data em bibliotecas e centros de informação

Forças	Fraquezas
Comprometimento governamental para identificar os benefícios do aprimoramento da aplicação do LD. A tecnologia LD pode ser utilizada para diversos serviços digitais. Menos dependências da equipe de TI após a conclusão do projeto LD.	A decisão sobre usar ontologias existentes ou criar novas ontologias é um desafio. A criação de um projeto de ontologia também é desafiadora. Vocabulários e metadados padrão são necessários para a conversão de dados bibliográficos em LD. A equipe deve realizar a serialização RDF, o que consome tempo e energia. A criação de triplas RDF é uma tarefa desafiadora. Necessidade de ferramentas de reconciliação sofisticadas. Problema de mapeamento de dados. A manutenção de vínculos é um desafio. Experiência exigida de tarefa LD. Falta de diretrizes claras e disponíveis para implementar LD em bibliotecas.
Oportunidades	Ameaças
Fácil conversão de dados. Oportunidade de acesso a um público amplo. Oportunidade de criar relacionamento entre dados. Vinculação de conjuntos de dados de biblioteca a conjuntos de dados externos, como DBpedia. Conceber um novo modelo para conjuntos de dados existentes. Oportunidade de modelagem de dados.	Não disponibilidade de licenças abertas.

Fonte: Ali e Warraich, 2018, tradução nossa.

⁶ SWOT é um acrônimo formado a partir das iniciais das palavras *Strengths*, *Weaknesses*, *Opportunities* e *Threats*.

As conclusões apresentadas pelos autores são as seguintes:

Os dados extraídos dos estudos selecionados após revisão sistemática mostram que a maioria das iniciativas e atividades de LD em centros de informação e bibliotecas ocorreu em países desenvolvidos. As bibliotecas e centros de informação analisados estão em diferentes estágios de implementação do LD. Bibliotecas e profissionais da informação parecem motivados para implementar LD. No entanto, há uma necessidade premente de fazer muito trabalho para implementar as tecnologias de LD de maneira eficaz e enfrentar seus desafios. Os principais desafios identificados são tecnológicos (ALI; WARRAICH, 2018, p.935, tradução nossa).

Com base nos resultados da pesquisa, Ali e Warraich (2018) discorreram algumas considerações que devem ser colocadas em prática: criação de cursos de Pós-graduação *stricto sensu* de LD ministrados pelas escolas de Biblioteconomia; a conscientização do profissional bibliotecário, por meio de seminários, acerca da importância da LD; órgãos como IFLA, OCLC e WWW deverão confeccionar diretrizes para os desafios técnicos de implantação do LD; estudantes de Biblioteconomia e áreas afins de países em fase de desenvolvimento devem voltar às atenções para a tecnologia LD; IFLA, OCLC, *American Library Association*, assim como também outras instituições deverão envidar esforços para a disponibilidade de dados em licença aberta na *Web*; e ainda, os centros de informações e as bibliotecas deverão motivar o uso da tecnologia LD.

Outro estudo selecionado foi desenvolvido por Senso e Arroyo Machado (2018), que realizaram um estudo de caso a partir de um conjunto de registros extraídos da Biblioteca Universitária de Granada, na Espanha. O objetivo do estudo é conhecer os principais problemas que pode se encontrado por qualquer Centro que pretenda converter seus registros bibliográficos em dados conectados sem terem que, para isso, mudarem o sistema de automação da biblioteca.

A principal motivação do estudo foi estabelecer um modelo que ajude as bibliotecas a definirem um *workflow* que facilite o processo de publicação dos registros bibliográficos armazenados em seus catálogos automatizados como dados conectados. A metodologia adotada baseia-se em estudo bibliográfico com o objetivo de determinar o estado da arte do *Linked Data* nas bibliotecas, prestando especial atenção à forma como os principais projetos internacionais têm prosseguido, sendo eles a Biblioteca Nacional da Espanha, a Biblioteca Britânica, a Biblioteca Nacional da França, a Europeia e a Biblioteca do Congresso. Concluída essa fase,

obteve-se uma visão global de como esses projetos foram desenvolvidos e os elementos necessários para elaborar o modelo.

O modelo desenvolvido foi implementado em um pequeno conjunto de dados, de forma a observar as possíveis deficiências que o método possa ter, bem como determinar as principais dificuldades que este processo de transformação acarreta (SENSO; ARROYO MACHADO, 2018). O modelo proposto tem a pretensão de permitir a transformação de registros bibliográficos em *Linked Data*, independentemente do software ou do ambiente em que esses registros se encontram. Este modelo é mostrado na tabela a seguir:

TABELA 13 - Proposta de metodologia para publicação de registros bibliográficos como Linked data

<i>Etapa</i>	<i>Descripción</i>	<i>Tareas</i>
1. <i>Determinar</i>	Identificación y descripción de los datos	a. Identificar y analizar los datos y fuente de datos (software, formato, base de datos...)
		b. Identificar su licencia
		c. Determinar una licencia
2. <i>Limpiar</i>	Almacenamiento y corrección de los datos	a. Data curation
3. <i>Modelar</i>	Desarrollo de un vocabulario para describir los datos en formato RDF	a. Seleccionar los vocabularios
		b. Creación de mapa
		c. Asignar URIs
4. <i>Generar</i>	Generación de los recursos RDF	d. Seleccionar las tecnologías para la generación de RDF
		e. Transformar los datos fuente en RDF
		f. Validarlo
5. <i>Enlazar</i>	Conectar el dataset a otros que lo enriquezca	a. Buscar datasets relevantes
		b. Descubrir relaciones
		c. Enlazar
		d. Verificar los enlaces
6. <i>Publicar</i>	Publicación del dataset	a. Escoger el formato y plataforma
		b. Publicar el dataset
		c. Publicar sus metadatos

Fonte: Senso e Arroyo Machado, 2018.

Segundo os autores, ao utilizar as 6 (seis) etapas do modelo proposto, também se obteve uma visão geral das ferramentas computacionais mais usadas, que podem ser utilizadas no desenvolvimento e implementação de cada etapa. O grupo selecionou ferramentas de código aberto ou gratuitas mais usadas atualmente para contribuir no aumento da possibilidade desse modelo ser colocado em prática.

A tabela a seguir elenca as ferramentas utilizadas em cada um dos processos:

TABELA 14 - Ferramentas mais utilizadas em cada um dos processos

Almacenamiento y gestión de datos				
Nombre	URL	Descripción	Licencia	Plataforma
Apache Hadopp	http://hadoop.apache.org/	Framework de software open-source para el almacenamiento distribuido de conjuntos de datos muy grandes en clusters de ordenadores.	Apache License 2.0	Multiplataforma
Cloudera Distributed Hadoop (CDH)	http://www.cloudera.com	Distribución de Apache orientada al mundo empresarial	Apache License 2.0	Linux
MongoDB	https://www.mongodb.com	Base de datos NoSQL	GNU AGPL 3.0	Multiplataforma

Extracción y limpieza de datos				
Nombre	URL	Descripción	Licencia	Plataforma
Spoon - Pentaho's Data Integration	http://community.pentaho.com/projects/data-integration/	Herramienta open-source para la extracción, transformación, transporte y carga de datos (ETL)	Apache License 2.0	Multiplataforma
Virtuoso Sponger	https://virtuoso.openlinksw.com/dataspace/doc/dav/wiki/Main/VirtSponger	Se trata de un componente middleware de Virtuoso Open-Source (VOS) que permite importar datos en diversos formatos (CSV, RSS, vCard...) y transformarlos en RDF	GNU General Public License 2.0	Multiplataforma
D2RQ	http://d2rq.org	Sistema open-source que permite acceder a bases de datos relacionales como grafos RDF virtuales, pudiendo lanzar consultas SPARQL en bases de datos no RDF, así como exportar la base de datos en RDF	Apache License 2.0	Multiplataforma
OpenRefine	http://openrefine.org	Herramienta ETL (Extraer, Transformar y Cargar) enfocada a la limpieza, transformación, exploración y enlazado de datos procedentes de diversos formatos. Sus funciones se pueden expandir con el uso de extensiones, destacando RDF Refine extension o DBpedia extension	BSD	Multiplataforma
GraphDB Free Edition	http://ontotext.com	Se trata de un repositorio semántico, un sistema de base de datos NoSQL que permite almacenar, consultar y gestionar datos estructurados. Utiliza ontologías para razonar automáticamente sobre los datos.	Licencia libre tipo RDBMS	Multiplataforma

Modelización				
Protégé	http://protege.stanford.edu/	Herramienta open-source que permite la construcción de modelos de dominio y aplicaciones basadas en el conocimiento con ontologías. Cuenta con una versión web y otra de escritorio. Es compatible con la última versión del Lenguaje de ontologías OWL 2 y especificaciones RDF de la World Wide Web Consortium (W3C)	FreeBSD	Multiplataforma
CmapTools Ontology Editor (COE)	http://cmap.ihmc.us/coe/test/v401ReleaseNotes.html#	Versión de CmapTools, herramienta para los mapas conceptuales, orientada a construir, compartir y visualizar ontologías OWL	-	Multiplataforma
OntoWiki	http://ontowiki.net/	Esta herramienta open-source permite la edición del contenido de archivos RDF de una forma muy visual, del mismo modo que un editor WYSIWIG para documentos de texto.	GNU General Public License 2.0	Multiplataforma
OOPS!	http://oops.linkeddata.es/	Se trata de una herramienta online de validación que permite detectar algunos de los errores más comunes que aparecen al desarrollar ontologías	-	Online
W3C RDF Validation Service	https://www.w3.org/RDF/Validator/	Herramienta online de W3C para la validación y visualización de documentos RDF (RDF/XML)	-	Online
Enlazado				
Limes	http://aksw.org/Projects/LIMES.html	Framework que implementa métodos eficientes en tiempo para el descubrimiento de enlaces a gran escala basados en las características de los espacios métricos	GNU General Public License	Multiplataforma
Silk	http://silkframework.org/	Open-source framework para combinar fuentes de datos heterogéneas, permitiendo generar enlaces entre elementos de datos contenidos en distintas fuentes	Apache License 2.0	Multiplataforma
Publicación				
Virtuoso	https://virtuoso.openlinksw.com/	Servidor multiplataforma escalable para el acceso a datos, integración y gestión de bases de datos relacionales, RDF y XML con un servidor de aplicaciones, de servicios Web	Apache License 2.0	Multiplataforma

Fonte: Senso e Arroyo Machado, 2018.

Para determinar a viabilidade das 6 (seis) etapas do modelo proposto, foi realizado um estudo piloto usando um conjunto de registros bibliográficos extraídos do catálogo geral da Biblioteca Universitária de Granada (BUGR). No experimento os autores do estudo utilizaram os dados de autor, título, publicação, assunto e ISBN na área de "Biblioteconomia" e "Documentação", sendo recuperados 1.251 registros para experimentação (SENSO; ARROYO MACHADO, 2018).

Senso e Arroyo Machado (2018) concluem que existem muitos projetos de bibliotecas que converteram seus registros bibliográficos em dados conectados a fim de aproveitar as inúmeras vantagens proporcionadas por este modelo. Muitos fatores determinam o uso de métodos diferentes para obter essa conversão. A fonte de dados original, os programas usados e o produto a ser criado são apenas algumas das muitas restrições que impedem o uso de modelos padronizados para realizar essa migração.

Como produto da pesquisa dos autores e considerando as tarefas associadas às 6 (seis) etapas, concluiu-se também que é possível automatizar totalmente o processo de extração de dados, limpeza e geração de grafos RDF. Além disso, na perspectiva da persistência de implementação futura este trabalho também pode ser automatizado através da introdução de novos dados em uma fase posterior (SENSO; ARROYO MACHADO, 2018).

Os autores destacaram ainda o grande número de ferramentas disponíveis que podem ser utilizadas para realizar quase todas as etapas do método. Essas aplicações são proeminentes na conversão, enriquecimento e limpeza de registros bibliográficos. Por outro lado, Senso e Arroyo Machado (2018) enfocaram que existem poucos *softwares* de biblioteca que publicam os dados catalogados como *Linked Data*:

Em um ambiente ideal, o lógico seria que o próprio programa de automação da biblioteca pudesse publicar os dados catalogados em *Linked Data*. Desta forma, este processo seria completamente transparente e atualizado à medida que o catálogo da biblioteca é atualizado com suas adições, exclusões, etc. Além disso, o ideal seria que no mesmo processo de catalogação fosse possível escolher quais dados do registro com o qual estamos trabalhando têm maior probabilidade de estar conectados a outros conjuntos de dados, já pré-carregados no sistema, e esse vínculo é feito a partir lá. No entanto, atualmente isso não acontece, uma vez que são poucos os sistemas integrados que permitem realizar este tipo de tarefa ou similar. O mercado oferece poucas soluções, entre as quais o DigiBIB espanhol (com variantes para arquivos, DigiArch, e para museus, DigiMus), da Digibis, e o serviço Innovative *Linked Data* da North American Innovative Interfaces Inc.; que é oferecido

como um extra para seus programas Sierra e Polaris. Possivelmente, nos novos requisitos funcionais para este tipo de programas, vários itens deverão ser incluídos para que o *software* seja capaz de realizar essas funções (SENSO; ARROYO MACHADO, 2018, p.4, tradução nossa).

Os principais obstáculos encontrados no estudo supracitado foram os seguintes: a necessidade de entender linguagens de programação e consultar bancos de dados. No entanto, os autores afirmaram que existe uma lacuna entre as ferramentas existentes devido à falta de ferramentas como o CMS (*Content Management System*) para facilitar o processo de publicação (SENSO; ARROYO MACHADO, 2018).

Por fim, Senso e Arroyo Machado (2018, p.18, tradução nossa) destacaram as limitações enfrentadas pelas bibliotecas acerca da implementação de dados catalogados como *Linked Data* e ressaltaram:

Metodologias, como a apresentada neste trabalho, não fariam sentido se as bibliotecas tivessem, por meio de seu próprio programa de automação, mecanismos para publicar automaticamente seus registros bibliográficos em *Linked Data*. Desta forma, os conjuntos de dados oferecidos estariam em dia e não seria necessário um esforço dobrado, como é o caso hoje. Fingir que uma biblioteca, com os problemas econômicos que esse setor arrasta, pode se dar ao luxo de ter dois ambientes diferentes administrados em paralelo, não tem sentido. Até que isso aconteça, a maioria das bibliotecas está condenada a atrasar seu salto para o *Linked Data*. Por isso, e na situação em que nos encontramos hoje, esta metodologia, aplicável a qualquer catálogo, tem razão de ser.

O próximo estudo selecionado é proposto por Vila-Suero e Gómez-Pérez (2013), com o objetivo de fornecer um método detalhado de utilização da ferramenta *Marimba* para publicação de dados conectados de catálogos de bibliotecas no formato MARC 21, juntamente com sua aplicação no projeto datos.bne.es da Biblioteca Nacional da Espanha. A ferramenta *Marimba* foi desenvolvida pelo projeto para transformar as fontes de dados em formato MARC 21 para o vocabulário RDFS/OWL (VILA-SUERO; GÓMEZ-PÉREZ, 2013).

O datos.bne.es é um projeto do Ministério da Cultura da Espanha que, desde 2006, busca uma forma de melhorar o controle de acesso e a interoperabilidade entre os arquivos de acesso das bibliotecas espanholas. Nesse sentido, foi proposta a criação de um documento de autoridade administrado pela Biblioteca Nacional da Espanha, que pode ser utilizado como ferramenta de referência para bibliotecas na Espanha e na América Latina. A razão para estabelecer e manter este sistema de autoridades é evitar registros duplicados, melhorar a qualidade e expansibilidade da

catalogação, economizando custos operacionais (VILA-SUERO; GÓMEZ-PÉREZ, 2013).

Nesse contexto, motivado pelo crescente interesse por LOD e tecnologias semânticas, em 2011 a Biblioteca Nacional da Espanha e o *Ontology Engineering Group* da “*Universidad Politécnica de Madrid*” iniciaram um projeto com o objetivo de transformar a autoridade e os catálogos bibliográficos em RDF, seguindo as melhores práticas do LD (VILA-SUERO; GÓMEZ-PÉREZ, 2013).

Para realizar a transformação, vinculação e publicação do conjunto de dados conectados da BNE, Vila-Suero e Gómez-Pérez (2013) adotaram o método de Villazón-Terrazas et al. (2011), baseado em uma modificação e extensão que consiste nas seguintes atividades: 1) especificação; 2) curadoria de dados; 3) modelagem; 4) geração; 5) vinculação; 6) publicação e 7) exploração.

Um dos principais resultados do projeto *datos.bne.es* destacados pelos autores foi o trabalho conjunto entre programadores e bibliotecários, que resultou numa equipe sólida e em diversos cursos de formação em dados conectados na Biblioteca Nacional de Espanha (VILA-SUERO; GÓMEZ-PÉREZ, 2013).

A curadoria de fontes de dados também é enfatizada na conclusão do estudo. Segundo Vila-Suero e Gómez-Pérez (2013) uma atividade que seja realizada de cunho transversal junto a outras de ordem de especificação, modelagem e geração poderão ter os problemas corrigidos nas fontes de dados, possibilitando a minimização de custos.

Nessa perspectiva, após análise dos estudos selecionados, entende-se que há uma vasta literatura empregada para a resolução de várias questões relacionados à utilização de dados conectados e dados abertos conectados em bibliotecas.

Para dar prosseguimento ao estudo, será apresentado agora o percurso metodológico que a pesquisa articulou para a coleta e análise de dados.

4. METODOLOGIA

O percurso metodológico possibilitou a coleta e o processamento de informações, visando à resolução do problema de pesquisa aqui investigado, representado pela seguinte pergunta: Quais os desafios são encontrados pelas

bibliotecas para a disponibilização dos catálogos online como dados abertos conectados.

Esta pesquisa se caracteriza como exploratória, do ponto de vista do objetivo, pois buscou a aquisição de conhecimentos sobre os desafios das práticas de implementação de dados abertos conectados por bibliotecas. “Pesquisas exploratórias são desenvolvidas com o objetivo de proporcionar visão geral, de tipo aproximativo, acerca de determinado fato” (GIL, 2008, p. 27).

Como método de estudo, do ponto de vista da natureza, esta pesquisa é classificada como aplicada, pois teve o propósito de mapear os desafios encontrados pelas bibliotecas para disponibilizarem seus catálogos online como dados abertos conectados, de forma a gerar conhecimento para que tal prática seja adotada por um número maior de bibliotecas. A “pesquisa aplicada objetiva gerar conhecimentos para aplicação práticas dirigidas à solução de problemas específicos. Envolve verdades e interesses locais” (PRODANOV, 2013, p.51).

Quanto aos procedimentos, a pesquisa caracteriza-se como documental e bibliográfica. Documental tendo em vista que utilizou como instrumento de estudo documentos oficiais, relatórios de pesquisa, relatórios de empresas, tabelas estatísticas, entre outros. E bibliográfica tendo em vista que aplicou como instrumentos de análise materiais publicados, tais como livros, artigos de periódicos, teses e dissertações para alcançar o objetivo proposto.

Do ponto de vista da forma de abordagem do problema, esta pesquisa caracteriza-se como qualitativa, visto que buscou compreender o contexto em que o problema está inserido. Para a análise dos resultados optou-se pelo uso da Análise de Conteúdo que é:

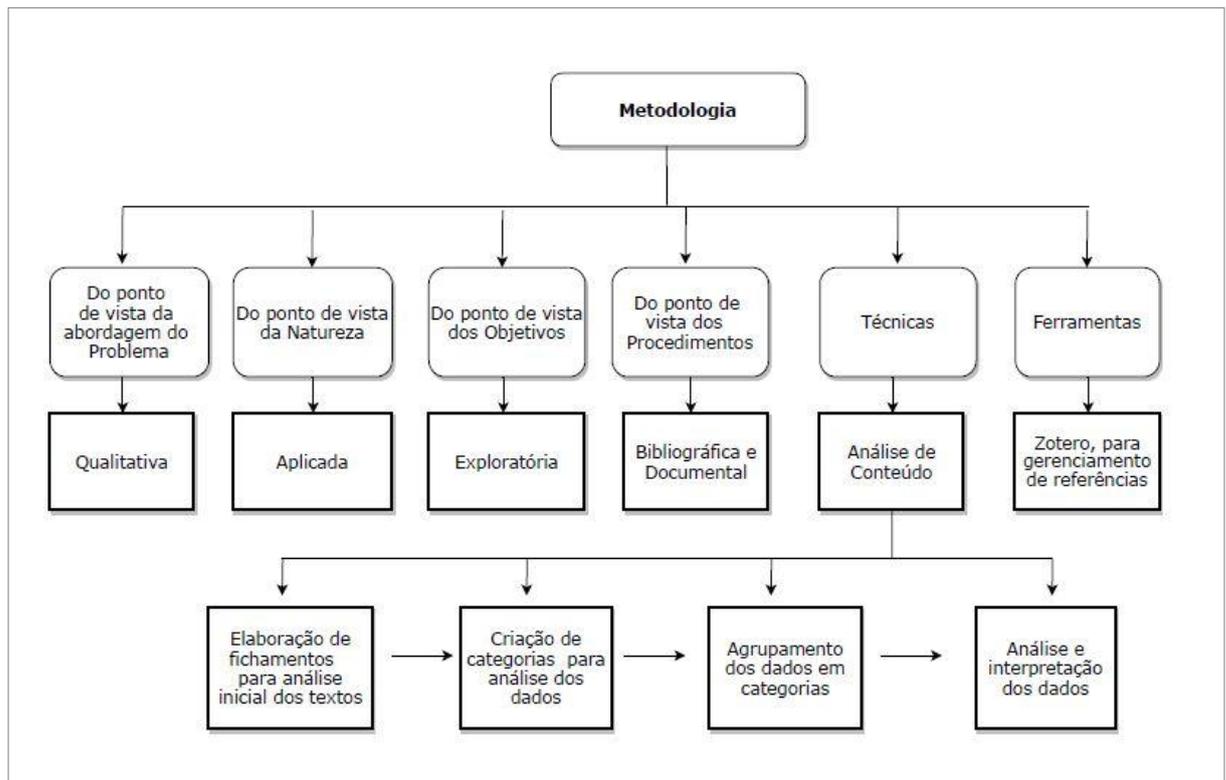
Um conjunto de técnicas de análise das comunicações visando obter, por procedimentos sistemáticos e objetivos de descrição do conteúdo das mensagens, indicadores (quantitativos ou não) que permitam a inferência de conhecimentos relativos às condições de produção/recepção (variáveis inferidas) destas mensagens (BARDIN, 2011, p. 47).

Esse estudo foi construído com base nas seguintes etapas: 1) levantamento bibliográfico a partir de diferentes bases dados, utilizando os descritores de busca; 2) criação e elaboração das categorias com foco em compilar elementos com base teórica para categorização dos dados; 3) categorização dos dados com base nas

categorias previamente estabelecidas e seguida pela etapa 4) de análise e interpretação dos dados.

Cada uma dessas etapas são apresentadas por meio da próxima figura, representada por um diagrama para ilustrar de forma simplificada o percurso metodológico adotado para realização dessa pesquisa:

FIGURA 10 – Diagrama da metodologia adotada



Fonte: elaborado pela autora, 2021.

Nessa perspectiva, de discorrer os fundamentos que embasaram os tipos e abordagens metodológicas da pesquisa, será agora enfatizado como se deu o levantamento bibliográfico.

4.1 LEVANTAMENTO BIBLIOGRÁFICO

Etapa 1: para obtenção dos dados dispostos nesse estudo, foram usados as seguintes bases de dados para pesquisa documental: os sites institucionais da *Online Computer Library Center (OCLC)*, *World Wide Web (W3C)*, *International Federation of Library Associations (IFLA)*, *American Library Association (ALA)*, da *LIBER* e da *Library of Congress*.

Já para a pesquisa bibliográfica foram utilizados instrumentos bibliográficos para coleta dos dados, tais como livros, artigos de periódicos científicos, teses e dissertações para realização do levantamento dos dados. As fontes bibliográficas foram pesquisadas no Portal da CAPES; nas bases de dados *Advanced Technologies e Aerospace Database*, *Library e Information Science Abstracts*, *Scopus* (Elsevier), *Directory of Open Access Journals (DOAJ)*, *OneFile* (GALE), *Social Sciences Citation Index (Web of Science)*, *Emerald Insight*, na base de dados em Ciência da Informação (BRAPCI), no catálogo online da Rede Pergamum, no catálogo do Sistema de Bibliotecas da UFMG e nas bibliotecas digitais das universidades brasileiras que têm Programa de Pós-Graduação em Ciência da Informação (PPG em CI).

A pesquisa bibliográfica referente ao Capítulo 2, que apresenta o embasamento teórico deste estudo, foi realizada a respeito da temática que discorre sobre os princípios e teorias da catalogação, catálogos, registros bibliográficos, dados abertos conectados, web semântica e fonte de dados conectados. Os textos foram selecionados de forma a delimitar o tema de pesquisa e indicar os referenciais teóricos adequados.

Para o Capítulo 3, que apresenta a revisão de literatura e compõe também os textos selecionados para a categorização dos resultados deste estudo, a pesquisa bibliográfica foi realizada a respeito da temática que aborda as iniciativas, estudo de casos e revisão sistemática de literatura sobre dados abertos conectados no âmbito de bibliotecas. Os descritores de buscas utilizados na pesquisa foram aplicadas da seguinte maneira: ("*library linked data*" OR "*library linked open data*") and ("*initiative*" OR "*case study*" OR "*systematic literature review*").

Foram selecionados, dentre os resultados obtidos, os materiais bibliográficos e documentais que atenderam aos critérios de inclusão e exclusão.

Foram utilizados os seguintes critérios de inclusão:

- a) Estudos publicados entre 2011 e 2021;
- b) Estudos com os termos *Linked Data* ou *Linked Open Data*;
- c) Estudos focados em iniciativas de *Linked Data* em catálogos ou registros bibliográficos de bibliotecas.

Nesse sentido, foram utilizados os seguintes critérios de exclusão:

- a) Estudos de *Linked Data* ou *Linked Open Data* não relevantes para bibliotecas;
- b) Iniciativas de *Linked Data* ou *Linked Open Data* não direcionados aos catálogos (registros) bibliográficos de bibliotecas;
- c) Iniciativas de *Linked Data* ou *Linked Open Data* de biblioteca que não descrevem os desafios de implementação.

Um total de 18 (dezoito) registros foram recuperado por esta estratégia de busca na pesquisa bibliográfica. Destes registros de texto completo, 5 (cinco) atenderam aos critérios de inclusão e foram selecionados. A pesquisa documental realizada nos sites institucionais da OCLC, da W3C, da IFLA, da ALA, da LIBER e da *Library of Congress* resultou em um total de 3 (três) registros que atenderam aos critérios de inclusão e exclusão. Dessa forma 08 (oito) textos foram selecionados e analisados por intermédio das categorias para análise dos dados, que serão percorridas na próxima subseção, exemplificando todo o processo.

A pesquisa bibliográfica para estruturação do presente Capítulo, que discorre sobre a metodologia adotada nesta pesquisa, foi desenvolvida em torno da temática que discorre sobre metodologia do trabalho científico, técnica de análise de conteúdo, dados abertos conectados, web semântica, registros bibliográficos e inovação organizacional.

4.2 CRIAÇÃO DAS CATEGORIAS

Etapa 2: a categorização consistiu em criar as categorias para reunir um grupo de elementos levando em consideração a orientação teórica e os objetivos da pesquisa. Bardin (2011) indica a possibilidade de uma categorização com o uso de categorias, a priori, sugeridas pelo referencial teórico e com categorias, a posteriori, elaboradas após a análise do material.

A fim de tornar a análise de conteúdo consistente com os objetivos propostos neste estudo, foram criadas 04 (quatro) categorias com base nos objetivos gerais e objetivos específicos deste estudo. Para cada categoria é feita a subdivisão desta em subcategorias e sua respectiva descrição, acompanhada pelo referencial teórico que a fundamenta.

As 04 (quatro) categorias criadas foram:

1. **Categoria Desafios** encontrados na implementação de práticas de dados abertos conectados em bibliotecas para a disponibilização de seus catálogos online na *Web*;
2. **Categoria Barreiras** encontradas na implementação de práticas de dados abertos conectados em bibliotecas para a disponibilização de seus catálogos online na *Web*;
3. **Categoria Potencialidades** dos registros bibliográficos para uso em formato de dados abertos para conectar os catálogos das bibliotecas;
4. **Categoria Requisitos** para publicação de dados abertos conectados.

A primeira categoria corresponde ao objetivo geral da pesquisa, que são os *“Desafios encontrados na implementação de práticas de dados abertos conectados em bibliotecas para a disponibilização de seus catálogos online na Web”*.

Para elencar as subcategorias da categoria “Desafios [...]” utilizou-se do referencial teórico de Lóscio, Burle e Calegari (2017), autores das Boas Práticas para Dados na *Web* da W3C, que conceituam e descrevem os principais desafios para publicar dados na *Web*. Dessa forma fazem partes desta categoria às seguintes subcategorias indicadas no quadro 6:

QUADRO 6 - Categoria Desafios encontrados na implementação de práticas de dados abertos conectados em bibliotecas para a disponibilização de seus catálogos online na *Web*

Categoria	Subcategorias	Descrição	Referencial teórico
Desafios	Metadados	Como eu forneço metadados para pessoas e máquinas? Fornecer metadados, metadados descritivos e metadados estruturais.	Lóscio, Burle e Calegari, (2017).
	Licença de dados	Como forneço e restrinjo o acesso? Fornecer informações sobre a licença de dados.	
	Proveniência e Qualidade	Como posso aumentar a confiança? Fornecer informações de proveniência dos dados e fornecer informações de qualidade dos dados.	
	Versionamento de dados	Como posso rastrear versões e histórico de versões? Fornecer indicador de versão e fornecer o histórico de versão.	
	Identificadores de dados	Como posso identificar conjuntos de dados e distribuições? Usar URIs persistentes como identificadores de conjuntos de dados.	

		Usar URIs persistentes como identificadores dentro de conjuntos de dados. Atribuir URIs para as versões dos conjuntos de dados e séries.
	Formato de dados	Quais formatos de dados devo usar? Usar formatos de dados padronizados e legíveis por máquinas. Usar representações de dados que sejam independentes de localidade (local e neutral). Fornecer dados em vários formatos.
	Vocabulário de dados	Como melhorar a interoperabilidade do dado? Reutilizar vocabulários, dando preferência aos padronizados. Escolher o nível de formalização adequado.
	Acesso a dados	Como posso fornecer acesso ao dado? Fornecer <i>download</i> em massa (<i>bulk download</i>). Fornecer subconjuntos para grandes conjuntos de dados. Usar negociação de conteúdo para servir os dados disponíveis em vários formatos. Fornecer acesso em tempo real. Fornecer dados atualizados. Fornecer uma explicação para os dados que não estão disponíveis. Tornar os dados disponíveis por meio de uma API. Usar padrões <i>Web</i> como base para construção de APIs. Fornecer documentação completa para as APIs. Evitar alterações que afetem o funcionamento de sua API.
	Preservação de dados	Como os dados podem ser arquivados? Preservar identificadores. Avaliar a cobertura do conjunto de dados.
	Enriquecimento de dados	Como posso engajar usuários? Coletar <i>feedback</i> dos consumidores de dados. Compartilhar o <i>feedback</i> disponível.
	Republicação de dados.	Como posso reutilizar dados responsabilmente? Fornecer <i>feedback</i> para o publicador original. Obedecer aos termos de licença. Citar a publicação original do conjunto de dados.

Fonte: elaborado pela autora, 2021.

Percebeu-se, como indicado pelo documento que define as Boas Práticas para publicação de Dados na *Web*, que as orientações foram elaboradas para “atender à demanda das equipes de gestão de informações, desenvolvedores, além de grupos mais amplos, tais como cientistas que estejam interessados em compartilhar e reutilizar dados na *Web*” (Lóscio, Burle e Calegari, 2017, não paginado). Os desafios elencados nessa categoria são de caráter tecnológico.

A segunda categoria criada levou em consideração o primeiro objetivo específico da pesquisa referente às *“Barreiras encontradas na implementação de práticas de dados abertos conectados em bibliotecas para a disponibilização de seus catálogos online na Web”*.

Para elencar as subcategorias dessa classe utilizou-se do referencial teórico do Manual de Oslo: diretrizes para coleta e interpretação de dados sobre inovação (2005), publicado pela OCDE (Organização para Cooperação e Desenvolvimento Econômico).

O Manual de Oslo foi publicado pela primeira vez em 1992 e é um guia de referência internacional para a coleta e uso de dados de inovação que fornece o entendimento e as características da inovação no nível institucional, bem como os fatores internos e sistêmicos que a afetam, o que é um pré-requisito para explorar e analisar políticas que incentivem a inovação (OCDE, 2005). As subcategorias criadas estão indicadas no quadro 7:

QUADRO 7 - Categoria Barreiras encontradas na implementação de práticas de dados abertos conectados em bibliotecas para a disponibilização de seus catálogos online na *Web*

Categoria	Subcategorias	Descrição	Referencial teórico
Barreiras	Fatores relativos ao custo	Descrição não declarada pelo autor.	OCDE (Manual de Oslo, 2005).
	Fatores relativos aos conhecimentos	Descrição não declarada pelo autor.	
	<ul style="list-style-type: none"> ▪ Carência de pessoal qualificado ▪ Inflexibilidades organizacionais no interior da empresa ▪ Atitude do pessoal com relação a mudanças ▪ Atitude da gerência com relação a mudanças 	Descrição não declarada pelo autor.	
	Fatores de mercado.	Descrição não declarada pelo autor.	

Fonte: elaborado pela autora, 2021.

A terceira categoria criada levou em consideração o segundo objetivo específico da pesquisa, que compreende as “*Potencialidades dos registros bibliográficos para uso em formato de dados abertos para conectar os catálogos das bibliotecas*”. Para elencar a subcategoria dessa classe utilizou-se do referencial teórico de Baker et al. (2011), autores do *Library Linked Data Incubator Group Final Report*, publicado pelo *W3C Incubator Group Report*. A subcategoria criada foi “*Conjunto padronizado de dados*”, conforme demonstrado no quadro 8:

QUADRO 8 - Categoria Potencialidades dos registros bibliográficos para uso em formato de dados abertos para conectar os catálogos das bibliotecas

Categoria	Subcategoria	Descrição	Referencial teórico
Potencialidades dos registros bibliográficos	Conjunto padronizado de dados	Conjunto de dados que descrevem os recursos relacionados à biblioteca	Baker et al., (2011)

Fonte: elaborado pela autora, 2021.

A quarta categoria criada levou em consideração o terceiro objetivo específico da pesquisa que são os “*Requisitos para publicação de dados abertos conectados*”. Para elencar as subcategorias dessa classe utilizou-se do referencial teórico de Derilinx; Lóscio e Archer (2015), autores de *Data on the Web Best Practices Use Cases & Requirements* do W3C, que conceituaram e descreveram os principais requisitos para publicar dados na *Web*.

A categoria e suas respectivas subcategorias estão representadas no quadro 9:

QUADRO 9 - Categoria Requisitos para publicação de dados abertos conectados

Categoria	Subcategorias	Descrição	Referencial teórico
Requisitos para publicação	Requisitos para acesso a dados	Os dados devem estar disponíveis para download em massa. O nível de acesso dos dados deve ser fornecido juntamente com as condições de acesso, por exemplo, aberto, restrito ou fechado. Quando os dados são produzidos em tempo real, eles devem estar disponíveis na <i>Web</i> em tempo real. Os dados devem estar disponíveis de forma atualizada e o ciclo de atualização deve ser	Derilinx; Lóscio e Archer (2015).

	explicitado. Se os dados estiverem disponíveis por meio de uma API, a API deve ser documentada.	
Requisitos para enriquecimento de dados	Deve ser possível realizar algumas tarefas de enriquecimento de dados a fim de agregar valor aos dados, portanto, fornecendo mais valor para os aplicativos e serviços do usuário.	
Requisitos para formatos de dados	Informações sobre os parâmetros de local (formatos de data e número, idioma) devem ser disponibilizadas. Os dados devem estar disponíveis em um formato legível por máquina que seja adequado para seu uso pretendido ou potencial. Os dados devem estar disponíveis em vários formatos. Os dados devem estar disponíveis em um formato aberto. Os dados devem estar disponíveis em um formato padronizado. Por meio da padronização, a interoperabilidade também é esperada.	
Requisitos para Identificação de Dados	Cada recurso de dados deve ser associado a um identificador único.	
Requisitos para Seleção de Dados	Os dados devem ser designados se forem irreproduzíveis. As etapas preliminares do ciclo de vida dos dados não devem infringir os direitos de propriedade intelectual do indivíduo. Os dados devem ser identificados por um estágio de ciclo de vida designado.	
Requisitos para vocabulários de dados	Os vocabulários devem ser claramente documentados. Os vocabulários devem ser compartilhados de forma aberta. Os vocabulários de referência existentes devem ser reutilizados sempre que possível. Os vocabulários devem incluir	

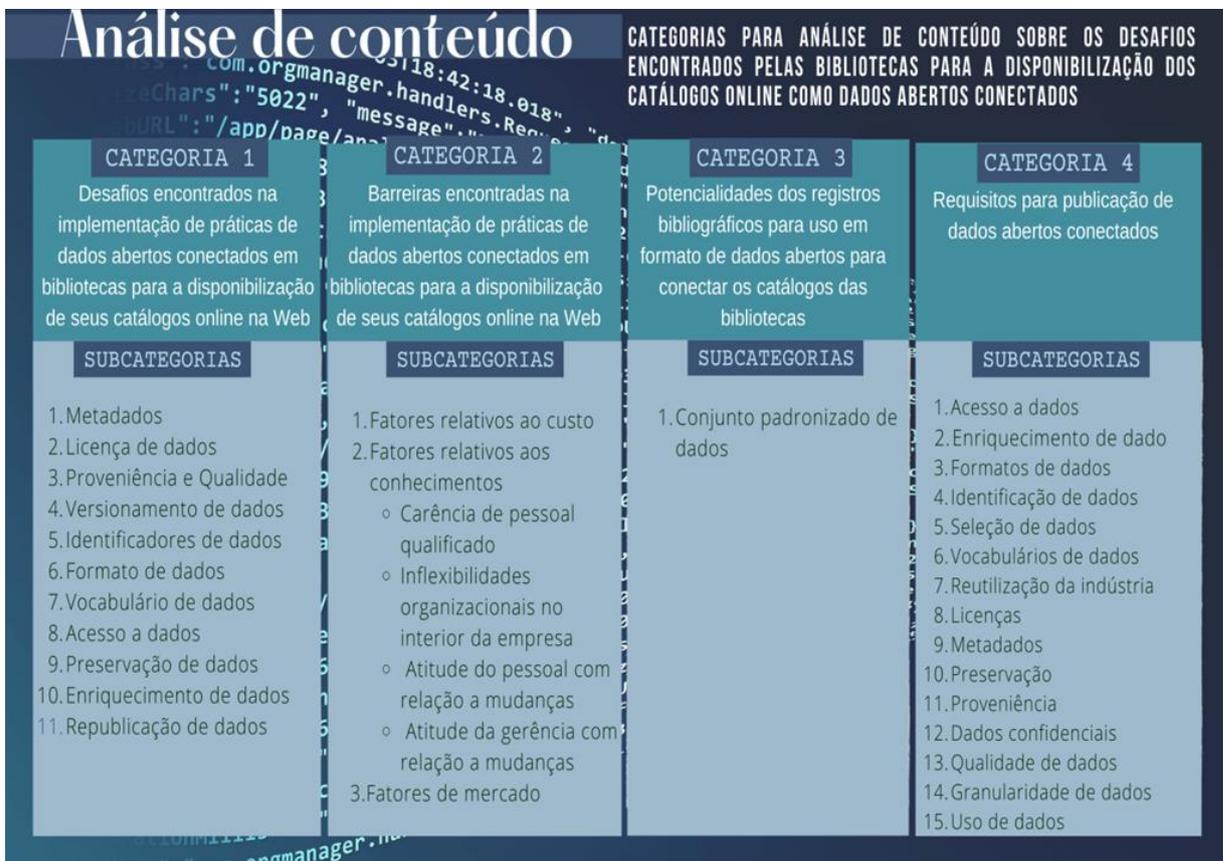
		informações de versão.	
	Requisitos para reutilização da indústria	Acordos de nível de serviço (SLAs) para a reutilização dos dados pelo setor devem estar disponíveis se solicitados (por meio de um ponto de contato definido). Um SLA é um tipo de metadados, portanto, todos os requisitos de metadados também se aplicam aqui.	
	Requisitos para licenças	Os dados devem ser associados a uma licença. Os termos de responsabilidade associados ao uso de Dados na <i>Web</i> devem ser claramente delineados.	
	Requisitos para metadados	As informações do contexto de produção devem ser associadas aos dados, se relevantes, por exemplo, descrições de serviço/processo. <i>Data production Context</i> é um tipo de metadados, portanto, todos os requisitos de metadados também se aplicam aqui. <i>Geographical Context</i> (países, regiões, cidades, etc.) deve ser referido de forma consistente. <i>Geographical Context</i> é um tipo de metadados, portanto, todos os requisitos de metadados também se aplicam aqui. Os metadados devem estar disponíveis. O vocabulário de metadados, ou valores, se o vocabulário não for padronizado, deve ser bem documentado. Os metadados devem ser legíveis por máquina. Os metadados devem ser padronizados. Por meio da padronização, a interoperabilidade também é esperada.	
	Requisitos para preservação	Um identificador para um determinado recurso deve ser resolvido na <i>Web</i> e associado, em um futuro próximo, a um único recurso ou a informações sobre o motivo pelo qual o recurso não está mais disponível.	

Requisitos para Proveniência	Se houver versões diferentes dos dados, o controle de versão dos dados deve ser fornecido. As informações sobre a proveniência dos dados devem estar disponíveis. Os dados de proveniência são um tipo de metadados, portanto, todos os requisitos de metadados também se aplicam aqui.
Requisitos para dados confidenciais	Os dados não devem infringir o direito de uma pessoa à privacidade. Os dados não devem infringir a segurança de uma organização (governo local, governo nacional, negócios).
Requisitos para Qualidade de Dados	Os editores devem indicar se os dados estão parcialmente ausentes ou se o conjunto de dados está incompleto. Os dados devem ser comparáveis com outros conjuntos de dados. Os dados devem ser completos. Os dados devem ser associados a um conjunto de métricas de qualidade documentadas, objetivas e, se disponíveis, padronizadas. Este conjunto de métricas de qualidade pode incluir métricas definidas pelo usuário ou específicas do domínio. Opiniões subjetivas de qualidade sobre os dados devem ser apoiadas.
Requisitos para granularidade de dados	Os dados disponíveis em diferentes níveis de granularidade devem ser acessíveis e modelados de maneira comum.
Requisitos para Uso de Dados	Deve ser possível citar dados na <i>web</i> . Deve ser possível rastrear o uso de dados. Os consumidores de dados devem ter uma maneira de compartilhar <i>feedback</i> e dados de classificação.

Fonte: elaborado pela autora, 2021.

O diagrama apresentado abaixo apresenta uma síntese das categorias criadas para análise de conteúdo desenvolvida para este estudo:

FIGURA 11 – Diagrama das categorias criadas para análise de conteúdo



Fonte: elaborado pela autora, 2021.

Dessa forma, cabe agora apresentar como se deu a etapa do processo de agrupamento dos dados em categorias, que será colocado na subseção a seguir.

4.3 AGRUPAMENTO DOS DADOS EM CATEGORIAS

Etapa 3: nesta etapa os dados coletados foram armazenados em uma planilha *Excel* e correlacionados às categorias e subcategorias decorrentes dos objetivos da pesquisa. Essa categorização serviu de base para distribuir os dados coletados, possibilitando a geração dos resultados e a discussão sobre eles. Feito isso, foi seguido para a análise e interpretação dos dados, etapa muito significativa para a pesquisa e que será discorrida na próxima subseção.

4.4 ANÁLISE E INTERPRETAÇÃO DOS DADOS

Etapa 4: nesta etapa os dados foram analisados a partir do referencial teórico pertinente à investigação, pois eles deram o embasamento e as perspectivas significativas para o estudo. O significado da interpretação foi dado pela relação entre os dados obtidos e o embasamento teórico. As interpretações a que levam as inferências foram realizadas no sentido de buscar o que está por trás da realidade aparente, isto é, o que em profundidade significa certos enunciados aparentemente superficiais.

Apresentado o percurso metodológico da pesquisa, será seguido agora para os resultados obtidos na construção desta dissertação, que será explanado a seguir.

5. RESULTADOS

Neste Capítulo apresenta-se a categorização dos resultados alcançados a partir da revisão de literatura e relacionados às categorias e subcategorias decorrentes do objetivo geral e dos objetivos específicos da pesquisa, assim como a discussão sobre eles. Ressalta-se que são 4 (quatro) as categorias analisadas neste estudo:

1. Categoria Desafios encontrados na implementação de práticas de dados abertos conectados em bibliotecas para a disponibilização de seus catálogos online na Web;
2. Categoria Barreiras encontradas na implementação de práticas de dados abertos conectados em bibliotecas para a disponibilização de seus catálogos online na Web;
3. Categoria Potencialidades dos registros bibliográficos para uso em formato de dados abertos para conectar os catálogos das bibliotecas;
4. Categoria Requisitos para publicação de dados abertos conectados.

Os resultados extraídos dos 8(oito) textos selecionados para este estudo foram atribuídos a cada categoria criada para análise de conteúdo e são apresentados a seguir, por intermédio dos quadros 10, 11, 12 e 13:

QUADRO 10 – Resultados atribuídos à categoria Desafios encontrados na implementação de práticas de dados abertos conectados em bibliotecas para a disponibilização de seus catálogos online na *Web*

Categoria	Subcategorias	Resultados
Desafios	Metadados	<ul style="list-style-type: none"> ▪ Problemas durante a criação de triplos RDF (ALI; WARRAICH, 2018, p.929, tradução nossa). ▪ Problema na conversão de dados bibliográficos no formato MARC 21 em LD e no desenvolvimento de modelo de dados RDF (ALI; WARRAICH, 2018, p.930, tradução nossa). ▪ Desafios enfrentados durante a preparação de dados, pois é necessário filtrar e transformar dados para converter e modelar dados RDF (ALI; WARRAICH, 2018, p.932, tradução nossa). ▪ Compreender como os dados são estruturados antes de usá-los (SMITH-YOSHIMURA, 2018, não paginado, tradução nossa).
	Licença de dados	<ul style="list-style-type: none"> ▪ Verificação sobre quem é o proprietário dos dados (SMITH-YOSHIMURA, 2018, não paginado, tradução nossa). ▪ Não disponibilidade de dados com licença aberta (ALI; WARRAICH, 2018, p.932, tradução nossa). ▪ Certos dados têm uso restrito. Portanto, os dados podem ter problemas de permissão pouco claros e não testados que os impedem de serem liberados como dados abertos. As questões de direitos variam muito entre os países, dificultando a colaboração na publicação de dados abertos (BAKER et al., 2011).
	Proveniência e Qualidade	<ul style="list-style-type: none"> ▪ Inconsistência em dados antigos (SMITH-YOSHIMURA, 2018, não paginado, tradução nossa). ▪ Combinar, eliminar a ambiguidade e alinhar os dados de origem aos recursos de dados conectados (SMITH-YOSHIMURA, 2018, não paginado, tradução nossa). ▪ O que é publicado na Internet como dados conectados nem sempre é reutilizável ou carece de URIs (SMITH-YOSHIMURA, 2018, não paginado, tradução nossa). ▪ Confiabilidade do serviço (SMITH-YOSHIMURA, 2018, não paginado, tradução nossa). ▪ Conjunto de dados desatualizados (SMITH-YOSHIMURA, 2018, não paginado, tradução nossa). ▪ Falta de controle de autoridade (SMITH-YOSHIMURA, 2018, não paginado, tradução nossa).

Versionamento de dados	<ul style="list-style-type: none"> Para esta subcategoria, não há resultados.
Identificadores de dados	<ul style="list-style-type: none"> Estabelecimento de links (SMITH-YOSHIMURA, 2018, não paginado, tradução nossa).
Formato de dados	<ul style="list-style-type: none"> Para implementação do LD em bibliotecas, deve haver diretrizes claras sobre a conversão de dados da biblioteca em dados abertos (ALI; WARRAICH, 2018, p.929, tradução nossa). Tamanho dos RDF dumps (despejos RDF) (SMITH-YOSHIMURA, 2018, não paginado, tradução nossa). Volatilidade de formatos de dados (SMITH-YOSHIMURA, 2018, não paginado, tradução nossa).
Vocabulário de dados	<ul style="list-style-type: none"> Seleção de ontologias apropriadas para representar os dados (SMITH-YOSHIMURA, 2018, não paginado, tradução nossa). A desambiguação de termos em diferentes idiomas é difícil (SMITH-YOSHIMURA, 2018, não paginado, tradução nossa). Mapeamento de vocabulário (SMITH-YOSHIMURA, 2018, não paginado, tradução nossa). A decisão sobre usar ontologias existentes ou criar novas ontologias é um desafio. A criação de um projeto de ontologia também é desafiadora (ALI; WARRAICH, 2018, p.933, tradução nossa).
Acesso a dados	<ul style="list-style-type: none"> Endpoints instáveis (SMITH-YOSHIMURA, 2018, não paginado, tradução nossa).
Preservação de dados	<ul style="list-style-type: none"> A manutenção de links é um desafio. Quando conjuntos de dados conectados são atualizados pode ocorrer erros gerando links inválidos (ALI; WARRAICH, 2018, p.930, tradução nossa).
Enriquecimento de dados	<ul style="list-style-type: none"> Para esta subcategoria, não há resultados.
Republicação de dados.	<ul style="list-style-type: none"> Para esta subcategoria, não há resultados.

Fonte: elaborado pela autora, 2021.

A categoria Desafios encontrados na implementação de práticas de dados abertos conectados em bibliotecas para a disponibilização de seus catálogos online na *Web* buscou respostas para o objetivo geral deste estudo. É composta por 11 (onze) subcategorias e foi criada a partir do referencial teórico de Lóscio, Burle e

Calegari, (2017). A aplicação das subcategorias mostrou-se relevante, pois permitiu a categorização de resultados para 08 (oito) subcategorias dentre as 11 estabelecidas, conforme apresentada no Quadro 11.

Foi possível mapear desafios para as subcategorias formato de Metadados, Licenças, Proveniência e Qualidade, Identificadores, Formatos, Vocabulários, Acesso e Preservação de dados. Foi notado que a *Web* é um ambiente diferente daqueles em que às bibliotecas estão habituadas a utilizar.

Para a subcategoria metadados os resultados apontam desafios relacionados ao formato atual dos dados de biblioteca, aliado a carência de *softwares* de automação capazes de publicar dados dos catálogos em *Linked Data*, o que acarreta na necessidade de conversão dos usuais registros bibliográficos em um formato compatível com a *Web*. Isso ocasiona dificuldades na conversão dos dados bibliográficos no formato MARC 21 em *Linked Data* e na modelagem dos dados em RDF.

Nesse contexto, é provável que o *Bibframe* passe a ter uma maior aceitação entre a comunidade bibliotecária para a adaptação dos dados das bibliotecas aos requisitos requeridos para publicação de dados conectados na *Web* (STEELE, 2018). O que se deve ao fato do modelo estar sendo desenvolvido pela Biblioteca do Congresso, apesar de não ser o único modelo de dados existente para descrição de recursos na *Web*. Nos estudos de Smith-Yoshimura (2018), que investigou 143 instituições em 23 países, o *schema.org*, no ano de 2018, foi o modelo mais usado entre os projetos analisados. A autora salienta que houve um aumento na publicação de dados conectados em *schema.org* e *Bibframe*, refletido por uma diminuição em SKOS, comparando as duas pesquisas realizadas por ela, uma em 2014 e outra em 2018.

Foi possível observar também dificuldades atribuídas à licença de dados, considerando que a verificação sobre quem é o proprietário nem sempre é clara. Nesse contexto, o indicado é o uso de licença aberta para publicação dos dados, colocando-os em domínio público ou usando a licença *Creative Commons* (CC). Isso é importante porque permite que os consumidores de dados usem, reutilizem, vinculem e realoquem os dados livremente para melhor atender às suas necessidades específicas (BE PART OF THE WEB, 2011).

Quanto à subcategoria proveniência e qualidade dos dados, os resultados apontam a existência de conjunto de dados desatualizados, inconsistentes e a ausência de controle de autoridades. Fornecer informações sobre a proveniência dos dados aumenta a confiança entre editores e consumidores de dados. A proveniência dos dados aparenta ser um dos maiores desafios encontrados no âmbito do *Linked Data*. O controle de qualidade deve ser executado quando o URI é criado e reiterado ao longo do tempo (BE PART OF THE WEB, 2011). “Deve-se [...] investir em treinamento de criação, derivação e publicação de URIs, bem como criação de links e uso de links em ambientes de descoberta)” (BE PART OF THE WEB, 2011, não paginado, tradução nossa).

Em relação à subcategoria Identificadores de dados, os resultados alcançados apontam que no estabelecimento e na manutenção de links percebe-se a preocupação com a geração de links inválidos. Isso pode ocorrer na atualização de um conjunto de dados conectados, gerando o desafio de atuar para que isso não aconteça.

Os resultados alcançados para a subcategoria vocabulários de dados apontam que a escolha de qual vocabulário ou “*datasets*” utilizar para o enriquecimento dos dados também é um desafio para publicação de dados na *Web*, pois é complexo lidar com a variedade de vocabulários de dados estruturados existentes ao planejar a criação, publicação e gerenciamento de dados conectados (BE PART OF THE WEB, 2011).

A reutilização de vocabulários é importante, preferencialmente os padronizados, pois favorece a compreensibilidade, o que significa que os seres humanos compreenderão melhor a estrutura e o significado dos dados, bem como a natureza dos metadados e conjuntos de dados; a facilidade de processamento, onde máquinas ou agentes de *software* serão capazes de processar e manipular dados automaticamente e a interoperabilidade, já que facilita o consenso entre editores e consumidores de dados (LÓSCIO; BURLE; CALEGARI, 2017).

Para a subcategoria acesso aos dados, o desafio atribuído foi *endpoints* instáveis. Sempre que os dados não estiverem disponíveis é necessário fornecer uma explicação sobre como os dados podem ser acessados e quem pode fazê-lo. Adotar tal prática fornece meios para que os publicadores identifiquem claramente lacunas de conhecimento, isto fornece uma explicação contextual para comunidades

de consumidores e incentiva o uso dos dados que estão disponíveis (LÓSCIO; BURLE; CALEGARI, 2017).

Quanto à subcategoria preservação de dados, os resultados indicam que a manutenção de links é um desafio. Quando conjuntos de dados conectados são atualizados podem ocorrer erros gerando links inválidos (ALI; WARRAICH, 2018).

E, finalmente, para a última subcategoria explorada, que trata do enriquecimento de dados, os resultados alcançados indicam que tornar os dados 'vinculáveis' permite ao usuário enriquecê-los com a ajuda recursos externos. Este enriquecimento pode ser realizado adicionando dados faltantes ou novas informações. Links também podem ser usados para encontrar discrepâncias nos dados e, assim, corrigir erros. Incentive os usuários a dizerem o que eles estão fazendo com seus dados para coletar essas informações (FROSTERUS et al., 2021).

Os resultados mapeados nessa categoria apontam desafios, de caráter tecnológico, enfrentados pelas bibliotecas ao publicarem dados conectados ou dados abertos conectados na *Web*. Estão relacionados às questões que envolvem o uso de tecnologia. O termo “tecnológico”, na definição do dicionário Michaelis (2021, não paginado) significa “relativo à tecnologia: desenvolvimento tecnológico”.

A partir do conjunto de desafios levantados percebe-se que abertura e a flexibilidade da *Web* trazem novos desafios aos editores e consumidores de dados para representar, descrever e fornecer dados de uma maneira fácil de entender e encontrar. Nesse caso, diretrizes devem ser fornecidas aos editores para que o gerenciamento de dados seja mais consistente. Independentemente da tecnologia usada, essas diretrizes podem ajudar a promover a reutilização de dados e criar confiança nos dados entre os desenvolvedores (LÓSCIO; BURLE; CALEGARI, 2017).

Desse modo, a observância dos requisitos mapeados neste estudo, que serão apresentados no decorrer da apresentação das demais categorias, deve ser adotada por serem consideradas Boas Práticas para Dados na *Web*, para que os desafios identificados nessa pesquisa sejam minimizados pelas instituições que pretendem publicar dados na rede (LÓSCIO; BURLE; CALEGARI, 2017).

Das subcategorias elencadas, 03 (três) não tiveram resultados mapeados dentre os textos selecionados. No entanto, elas não devem ser descartadas das

demais subcategorias, por fazerem parte do conjunto de desafios baseados em estudos de casos definidos por Lóscio, Burle e Calegari, (2017).

A subcategoria “Versionamento de dados” se refere à necessidade de indicar um número de versão ou data para cada conjunto de dado publicado na *Web* que pode mudar no decorrer do tempo. Alguns conjuntos de dados são atualizados regularmente, enquanto outros são alterados à medida que melhorias são implantadas na coleta de dados e atualizações. No entanto, não há consenso sobre em qual momento, por conta das alterações em um conjunto de dados, o mesmo deve ser considerado como um conjunto de dados completamente diferente e não somente uma nova versão (LÓSCIO; BURLE; CALEGARI, 2017).

As informações sobre a versão tornam a revisão de um conjunto de dados identificável de forma única. Esta unicidade pode ser utilizada pelos consumidores de dados para determinar especificamente com qual versão de um conjunto de dados estão trabalhando. O bom versionamento dos dados possibilita aos consumidores entender se uma nova versão de um conjunto de dados está disponível. O versionamento explícito leva em conta a repetitividade na pesquisa, permite comparações e evita confusão. A utilização de números únicos de versão com uma abordagem padronizada também pode contribuir na expectativa do consumidor, acerca da diferença entre versões (LÓSCIO; BURLE; CALEGARI, não paginado, 2017).

A subcategoria “Enriquecimento de dados” infere que o enriquecimento pode melhorar muito a facilidade de processamento, especialmente no caso de dados não estruturados. Os conjuntos de dados com dados ausentes serão melhorados preenchendo esses valores. Se atributos relevantes forem adicionados, a estrutura do conjunto de dados será validada e sua utilidade será aprimorada, desde que esse incremento não distorça os resultados da análise, significância ou poder estatístico (LÓSCIO; BURLE; CALEGARI, 2017).

E, por fim, a subcategoria “Republicação de dados”, que é uma das formas de publicar dados na *Web*, bastando para isso somente republicá-los. Pode assumir a forma de combinação de dados existentes com outros conjuntos de dados, criar aplicativos ou visualizações da *web* ou reempacotar os dados em um novo formato (LÓSCIO; BURLE; CALEGARI, 2017). Para a republicação de dados devem-se observar as Boas Práticas, a saber:

Informar ao publicador original quando você está promovendo o reuso de seus dados. Informe-o caso encontre um erro, ou tenha sugestões ou elogios a tecer. Os publicadores geralmente querem saber se os dados que publicam têm sido úteis. Ademais, eles podem ser obrigados a relatar estatísticas de uso a fim de alocar recursos para atividades de publicação

de dados. Ao informar como dos dados publicados foram usados, você estará ajudando os publicadores originais a justificar a aplicação de recursos no lançamento de dados. Fornecer feedback recompensa os publicadores por seus esforços e os ajuda diretamente a melhorar seu conjunto de dados para futuros usuários. Uma melhor comunicação facilitará para os publicadores originais determinar como os dados que postaram estão sendo usados, o que por sua vez os ajuda a justificar a publicação dos dados. Eles também terão clareza sobre quais medidas podem adotar para melhorar seus dados. Isto conduz a mais e melhores dados para todos (LÓSCIO; BURLE; CALEGARI, não paginado, 2017).

O quadro 11 lista os resultados identificados para a 2ª categoria analisada nesta pesquisa, a categoria *Barreiras* [...]:

QUADRO 11 – Resultados atribuídos à categoria Barreiras encontrados na implementação de práticas de dados abertos conectados em bibliotecas para a disponibilização de seus catálogos online na Web

Categoria	Subcategorias	Resultados
Barreiras	Fatores relativos ao custo	<ul style="list-style-type: none"> ▪ Enorme carga financeira (ALI; WARRAICH, 2018, p.931, tradução nossa). ▪ Fingir que uma biblioteca, com os problemas econômicos que esse setor arrasta, pode se dar ao luxo de ter dois ambientes diferentes administrados em paralelo, não tem sentido. Até que isso aconteça, a maioria das bibliotecas está condenada a atrasar seu salto para o Linked Data. A maioria dos projetos internacionais tem se concentrado na área de grandes bibliotecas (SENSO; ARROYO MACHADO, 2018, p.18, tradução nossa). ▪ Os custos de mão de obra constituem a maior parte dos recursos necessários na implementação de projetos de dados conectados (FROSTERUS; et. al. 2021, p.5, tradução nossa). ▪ Falta de recursos (SMITH-YOSHIMURA, 2018, não paginado, tradução nossa).
	Fatores relativos aos conhecimentos	
	Carência de pessoal qualificado	<ul style="list-style-type: none"> ▪ A despesa mais notável relacionada à publicação de dados conectados é o trabalho humano (FROSTERUS et al., 2021, p.3, tradução nossa). ▪ A equipe precisa ter o conhecimento para lidar com várias tarefas que incluem escolher um conjunto de dados, modelar os dados, descrevê-los usando vocabulários padrão, decidir como apresentar os dados usando URIs, converter os dados, fornecer acesso à máquina aos dados, escolher uma licença de publicação, anunciar os conjuntos de dados publicados, comunicar as partes interessadas e reconhecer o contrato social (ou seja, manter os dados disponíveis e atualizados uma vez

		<p>publicados) (FROSTERUS et al., 2021, p.5, tradução nossa).</p> <ul style="list-style-type: none"> ▪ O trabalho com Linked Data exige pessoal com experiência em tecnologia da informação e tecnologia da <i>Web</i>. Exige equipe para curadoria de informações e engenheiros para tarefas técnicas (ALI; WARRAICH, 2018, p.929, tradução nossa). ▪ Necessidade de pessoal com capacitação para entender linguagens de programação e consultar bancos de dados (SENSO; ARROYO MACHADO, 2018, p.16). ▪ A maioria das bibliotecas carece de potencial (econômico, tecnológico e/ou humano) para realizar uma tarefa tão especializada que penetra tão claramente no ambiente computacional como a implementação de projetos de dados conectados (SENSO; ARROYO MACHADO, 2018, p. tradução nossa). ▪ Curva de aprendizado íngreme para a equipe (SMITH-YOSHIMURA, 2018, não paginado, tradução nossa). ▪ Pouca documentação ou orientação sobre como construir sistemas (SMITH-YOSHIMURA, 2018, não paginado, tradução nossa).
	<ul style="list-style-type: none"> ▪ Inflexibilidades organizacionais no interior da empresa 	<ul style="list-style-type: none"> ▪ Para esta subcategoria, não há resultados.
	<ul style="list-style-type: none"> ▪ Atitude do pessoal com relação a mudanças 	<ul style="list-style-type: none"> ▪ Dificuldade por parte das bibliotecas em confiar em fontes externas para ligação de dados (JESUS; CASTRO, 2019).
	<ul style="list-style-type: none"> ▪ Atitude da gerência com relação a mudanças 	<ul style="list-style-type: none"> ▪ Para esta subcategoria, não há resultados.
	Fatores de mercado.	<ul style="list-style-type: none"> ▪ Dependência de fornecedores de sistemas, empresas, consultores ou desenvolvedores externos. Necessidade de ferramentas sofisticadas de reconciliação para o processo LD (ALI; WARRAICH, 2018, p.932, tradução nossa). ▪ Dependência em relação ao fornecedor de <i>software</i> de biblioteca, que precisa estar disposto a adaptar seus produtos às necessidades das bibliotecas. As bibliotecas não contam com uma infraestrutura própria quando desejam adotar o <i>Linked Data</i>

		<p>(BAKER et al., 2011).</p> <ul style="list-style-type: none"> ▪ Não existe uma ferramenta que sirva para todos. Uma grande variedade de ferramentas, tanto comerciais quanto abertas, gerais ou especializadas, são usadas junto com ferramentas desenvolvidas localmente (FROSTERUS et al., 2021, p.3, tradução nossa). A conversão dos dados descritos em MARC21 depende da escolha de ferramentas que possam ser utilizadas para converter os dados bibliográficos de maneira automatizada sem que a consistência desses dados se perca no processo (JESUS; CASTRO, 2019, p.48). ▪ A maioria dos projetos de dados conectados no ambiente de bibliotecas são realizados por grandes instituições que, em muitas ocasiões, necessitaram da ajuda de terceiros (empresa ou universidade) para concluir com sucesso as suas implementações (SENSO; ARROYO MACHADO, 2018, p.3, tradução nossa). ▪ Falta de ferramentas prontas para uso (SMITH-YOSHIMURA, 2018, não paginado, tradução nossa). ▪ <i>Software</i> imaturo (SMITH-YOSHIMURA, 2018, não paginado, tradução nossa).
--	--	--

Fonte: elaborado pela autora, 2021.

Durante a análise da categoria Barreiras encontradas na implementação de práticas de dados abertos conectados em bibliotecas para a disponibilização de seus catálogos online na *Web*, buscaram-se elementos para discutir o 1º objetivo específico deste estudo. A categoria barreiras é composta por 03 (três) subcategorias e foi criada a partir do referencial teórico da OCDE (2005). A aplicação das subcategorias mostrou-se relevante, pois permitiu a categorização de resultados para 03 (três) subcategorias dentre as 03 estabelecidas, conforme categorização apresentada no Quadro 12.

As barreiras mapeadas neste estudo incluem questões sistêmicas e as subcategorias criadas são: fatores relativos ao custo; fatores relativos aos conhecimentos, subdividida em 4 (quatro) tópicos: carência de pessoal qualificado, atitude do pessoal com relação a mudanças, inflexibilidades organizacionais no interior da empresa e atitude da gerência com relação a mudanças. Para os 2 (dois)

últimos tópicos não foram obtidos resultados a partir dos textos selecionados para estudo nesta pesquisa. E, por último, a subcategoria fatores de mercado.

Vê-se, pelos resultados da pesquisa que a absorção de novas tecnologias exige mudanças organizacionais, principalmente em países em desenvolvimento que geralmente absorvem as tecnologias dos países desenvolvidos. Nesse sentido, o comportamento das organizações torna-se importante para explicar a diferença de desempenho entre elas e a qualidade dos serviços oferecidos (OCDE, 2005).

Dentre as barreiras sistêmicas analisadas, foi possível observar para a primeira subcategoria, que aborda fatores relacionados ao custo, enorme carga financeira relacionada à carência de pessoas capacitadas e que tenham habilidades com as tecnologias da informação, já que o trabalho humano é considerado a despesa mais alta nessa categoria. No entanto, nota-se que aí surge uma oportunidade de aperfeiçoamento da equipe profissional que já trabalha nas instituições.

Para a segunda subcategoria analisada, que aborda fatores relativos ao conhecimento, foram identificadas barreiras referentes à atitude do pessoal com relação a mudanças e dificuldades por parte das bibliotecas em confiar em fontes externas para ligação de dados. Os bibliotecários tradicionalmente preferem ater-se a modelos e recursos sobre os quais tenham controle total, em vez de usar ferramentas centradas no usuário e publicamente editáveis (SINGER, 2009). Entretanto, como afirma Serra (2019, p.70), as bibliotecas “precisam reconhecer-se como potenciais provedoras de dados confiáveis” para fornecê-los ao ambiente *Web*.

Identificou-se também que o desenvolvimento de projetos com *Linked Data* demanda o envolvimento de pessoal com experiência em tecnologia da informação e tecnologia da *Web*, sendo este aspecto uma das barreiras mais notáveis a serem superadas por bibliotecas, visto que “a maioria das bibliotecas carece de potencial (econômico, tecnológico e/ou humano) para realizar uma tarefa tão especializada que penetra tão claramente no ambiente computacional como a implementação de projetos de dados conectados” (SENSO; ARROYO MACHADO, 2018, p. tradução nossa).

Ademais, fatores de mercado também foram identificados como barreiras, correspondendo a terceira e última subcategoria analisada. Há uma dependência por parte das bibliotecas com relação a fornecedores e desenvolvedores de

sistemas. Soma-se, ainda, a escassez de sistemas integrados que permitem publicar os dados do catálogo em *Linked Data*. O mercado oferece poucas soluções, entre as quais o DigiBIB da Digibis, em espanhol (com variantes para arquivos, DigiArch, e para museus, DigiMus), e o serviço *Innovative Linked Data* da *North American Innovative Interfaces Inc.*; oferecido como um extra para seus programas Sierra e Polaris. (SENSO; ARROYO MACHADO, 2018).

Outra opção é o *software* Alma, da ExLibris, uma das grandes fornecedoras de softwares para bibliotecas. Ele possibilita que todos os registros bibliográficos possam ser visualizados e exportados como *Bibframe* e acessados por meio de um URI exclusivo (SIMIONATO et al., 2019).

Os resultados mapeados nessa categoria apontam que as barreiras identificadas possuem caráter sistêmico, pois se referem a questões inerentes ao ambiente organizacional no qual as bibliotecas estão inseridas. O termo “sistêmico”, na definição do dicionário Michaelis (2021, não paginado) significa “próprio ou relacionado com um sistema”. Os vieses tecnológicos e sistêmicos estão intimamente relacionados quanto se trata de inovação organizacional.

Salienta-se que a mudança organizacional é uma resposta à mudança da tecnologia e condição necessária para a inovação tecnológica. A inovação organizacional tende a melhorar a qualidade e eficiência do trabalho e fortalecer a troca de informações, aprimorando a capacidade das organizações de aprender e usar conhecimento e tecnologia (OCDE, 2015; LAM, 2005).

Desse modo percebeu-se a necessidade de desenvolver nessas instituições onde os estudos foram realizados grupos de trabalho sobre inovação, com o intuito de minimizar a resistência às mudanças, visto que a transformação organizacional é uma resposta à mudança da tecnologia e condição necessária para a inovação tecnológica.

O quadro 12 lista os resultados identificados para a 3ª categoria analisada nesta pesquisa, a categoria *Potencialidades [...]*:

QUADRO 12 – Resultados atribuídos à categoria Potencialidades dos registros bibliográficos para uso em formato de dados abertos para conectar os catálogos das bibliotecas

Categoria	Subcategoria	Resultados
-----------	--------------	------------

<p>Potencialidades dos registros bibliográficos</p>	<p>Conjunto padronizado de dados</p>	<ul style="list-style-type: none"> ▪ Os dados bibliográficos podem contribuir para o <i>Linked Data</i> ao fornecer dados tradicionalmente consistentes e confiáveis (JESUS; CASTRO, 2019, p.46). ▪ A catalogação é um processo complexo e árduo, que exige muito esforço e profissionalismo dos responsáveis pela sua execução. O produto desse trabalho é o registro bibliográfico composto por inúmeros dados (autor, título, local, editora, data...) que, conectados a dados diferentes de outros conjuntos de dados, poderiam fornecer informações individuais sobre cada um deles, oferecendo ao utilizador uma infinidade de novas informações (onde nasceu aquele autor, a história local daquele lugar, dados contemporâneos...) e, para a biblioteca, a possibilidade de expandir os seus horizontes para além do catálogo (SENSO; ARROYO MACHADO, 2018, p.2, tradução nossa). ▪ Os documentos e metadados que as bibliotecas têm nas mãos atingem um elevado nível de estrutura, sendo um terreno ideal para projetos de <i>Linked Data</i> (SENSO; ARROYO MACHADO, 2018, p.3, tradução nossa). ▪ A conexão de registros bibliográficos com outros conjuntos de dados para interligar informações proporciona, tanto aos usuários quanto às bibliotecas, uma melhoria na visibilidade (dos dados e da instituição que os oferece), permite reutilizar o dados de registros publicados e adicionar novos, estabelece vínculos com outros serviços além de facilitar a modelagem de "coisas de interesse" relacionadas a um recurso bibliográfico, como pessoas, lugares e eventos (SENSO; ARROYO MACHADO, 2018, p.3, tradução nossa). ▪ As bibliotecas têm como prática o desenvolvimento de trabalhos de padronização da representação descritiva, com esforços canalizados nos catálogos e na construção de listas de autoridades e terminologias. O uso de instrumentos como vocabulários controlados e tesouros é uma prática recorrente na Biblioteconomia. Agregar estrutura a estes instrumentos representa um avanço no tratamento descritivo, permitindo que dados sejam preenchidos com contexto e que sejam interoperáveis, favorecendo a reutilização de recursos (SERRA; SANTAREM SEGUNDÓ, 2017, p.181). ▪ Os catálogos das bibliotecas são construídos por um conjunto de práticas, regras, modelos de dados e métodos em evolução e altamente especializados (VILA-SUERO; GÓMEZ-PÉREZ, 2013).
--	--------------------------------------	---

Fonte: elaborado pela autora, 2021.

A categoria Potencialidades dos registros bibliográficos para uso em formato de dados abertos para conectar os catálogos das bibliotecas buscou respostas para o 2º objetivo específico deste estudo. Esta categoria é composta por 01 (uma) subcategoria e foi criada a partir do referencial teórico de Baker et al. (2011). A aplicação da subcategoria mostrou-se relevante, pois permitiu a categorização de resultados consistentes para compor a análise de dados.

A partir dos resultados mapeados confirma-se que os registros bibliográficos possuem grande potencial para serem disponibilizados como dados abertos conectados quando são constituídos por um conjunto padronizado de dados, estruturados a partir de normas e padrões pré-estabelecidos e originados a partir de lista de autoridades e vocabulários controlados. São, portanto, como afirmam Jesus e Castro (2019), fontes de informação consistentes e confiáveis.

Os metadados dos registros bibliográficos das bibliotecas são adequados como fonte inicial de dados conectados, devido aos arquivos de autoridades que suportam sua terminologia controlada. Os registros dos arquivos de autoridade podem ser publicados usando URIs estáveis e duráveis e os dados desses registros são valiosos para uma correlação mais ampla (BE PART OF THE WEB, 2011).

Ao vincular registros bibliográficos a outros conjuntos de dados, as bibliotecas agregam estrutura ao registro, permitindo que os dados sejam preenchidos com contexto e fornecem aos usuários uma visibilidade aprimorada dos dados, pois possibilita a reutilização dos registros publicados para relacionamentos com outros recursos relacionados ao registro. Uma variedade de novas informações pode ser disponibilizada, como, por exemplo, a biografia do autor, outras obras publicadas por ele, informações geográficas sobre o local pesquisado, dentre outras (SENSO; ARROYO MACHADO, 2018, SERRA; SANTAREM SEGUNDO, 2017).

Para a biblioteca é uma grande oportunidade de disseminar informação para além do catálogo, fornecendo para a *Web* informações de qualidade, disponibilizando conjuntos padronizados de dados que podem ser reutilizados e compartilhando instrumentos consagrados na área como os catálogos de autoridades e vocabulários controlados (BAKER, 2011; BYRNE; GODDARD, 2010; SERRA, 2019).

O quadro 13 lista os resultados identificados para a 4ª categoria analisada nesta pesquisa, a categoria Requisitos [...]:

QUADRO 13 – Resultados atribuídos à categoria Requisitos para publicação de dados abertos conectados

Categoria	Subcategorias	Resultados
Requisitos Para publicação de dados abertos conectados	Requisitos para acesso a dados	<ul style="list-style-type: none"> ▪ Disponibilize arquivo em massa para download. É muito fácil de configurar e bastante atraente para usuários que desejam acessar os dados. Ter um ponto de extremidade SPARQL15 para consultar os dados também é uma ótima ideia. SPARQL permite que o usuário extraia informações por meio de consultas e, na verdade, fornece um API “programável” para os dados (FROSTERUS et al., 2021, p.15, tradução nossa).
	Requisitos para enriquecimento de dados	<ul style="list-style-type: none"> ▪ Além da ligação entre dados locais e externos, existem também procedimentos específicos, onde a vinculação é usada para fins de refinamento de dados. A reconciliação, por exemplo, envolve a substituição de vocabulários controlados bem conhecidos, como a Lista de Cabeçalhos de Assuntos da Biblioteca do Congresso (LCSH) ou Arte e Arquitetura <i>Thesaurus</i> (AAT). Mais enriquecimento também é alcançado pela adoção de entidades nomeadas NER (<i>Named Entity Recognition</i>) e pela adição de seus URIs para o conjunto de dados. Ambos os processos podem ser executados com extensões <i>Open Refine</i>. Também deve ser observado que o número de ferramentas de código aberto, especialmente no caso de NER, é bastante limitado (FROSTERUS et al., 2021, p.7, tradução nossa).
	Requisitos para formatos de dados	<ul style="list-style-type: none"> ▪ Para esta subcategoria, não há resultados.
	Requisitos para Identificação de Dados	<ul style="list-style-type: none"> ▪ No projeto da Biblioteca Nacional da Espanha foram distinguidos dois tipos de URIs. Os de vocabulário, que identificam os componentes terminológicos (classes RDF e propriedades RDF) para descrever as entidades e seus relacionamentos e atributos no conjunto de dados RDF e URIs de dados, que identificam os recursos (também chamados de instâncias ou indivíduos) que estamos publicando (VILA-SUERO; GÓMEZ-PÉREZ, 2013).
	Requisitos para Seleção de Dados	<ul style="list-style-type: none"> ▪ Para esta subcategoria, não há resultados.

Requisitos para vocabulários de dados		<p>Vocabulários que atendem aos requisitos:</p> <ul style="list-style-type: none"> ▪ <i>GeoNames</i>, VIAF, ISNI, <i>Wikidata</i>, Agrovoc, MeSH, Lista de Cabeçalhos de Assuntos da Biblioteca do Congresso (LCSH), Arte e Arquitetura <i>Thesaurus</i> (AAT), Eurovoc, ULAN, (FROSTERUS et al., 2021). ▪ id.loc.gov, VIAF, Dbpedia, <i>GeoNames</i>, <i>Wikidata</i>, <i>WordCat.org</i>, Vocabulários Getty, FAST ISNI (SMITH-YOSHIMURA, 2018).
Requisitos para reutilização da indústria		<ul style="list-style-type: none"> ▪ Para esta subcategoria, não há resultados.
Requisitos para licenças		<ul style="list-style-type: none"> ▪ Ao disponibilizar os dados é preciso decidir sobre o licenciamento. Dados abertos implica o uso de uma licença aberta e existem algumas opções específicas para escolha. A mais simples é ir para um Creative Commons (CC) licença. CC018 é a mais permitida das licenças CC, mas existem outras opções. Também é uma boa prática incluir as informações da licença no conteúdo dos dados (FROSTERUS et al., 2021, p.16, tradução nossa). ▪ Quem possui os dados? No caso de autoridade MARC21 e registros bibliográficos, muitos deles importados de outras bibliotecas ou fornecedores, certifique-se de que os direitos relacionados foram esclarecidos e liberados, e não há restrições que possam dificultar a publicação do conjunto de dados como LOD (FROSTERUS et al., 2021, p.6, tradução nossa). ▪ Várias organizações importantes, como a Europeana, CENL (Conferência de Bibliotecários Nacionais Européias) ou a Biblioteca de Harvard apoiam as licenças abertas. Mais especificamente, a licença de domínio público Creative Commons também conhecido como CC0. Desde então da British Library e datos.bne.es, entre outros, lançaram conjuntos de dados conectados abertos sob a licença CC0 (VILA-SUERO; GÓMEZ-PÉREZ, 2013). ▪ O estudo aponta sobre a importância da discussão sobre dados conectados e direitos autorais entre os gerentes de bibliotecas buscando chegar a um acordo com os proprietários sobre direitos e permissões de uso em nível de consórcio de bibliotecas e até mesmo em escala nacional ou internacional (BAKER et al., 2011).
Requisitos para metadados		<ul style="list-style-type: none"> ▪ Inclua documentação sobre os dados publicados, o que pode ser feito de forma formal e processável por máquina usando o Vocabulário de conjuntos de dados interligados (VoID - Vocabulary of Interlinked Datasets), da W3C (FROSTERUS et al., 2021, p.16, tradução

		<p>nossa).</p> <p>Vocabulários de metadados existentes:</p> <ul style="list-style-type: none"> ▪ SKOS, Schema.org, FOAF, Dublin Core (FROSTERUS; et. al. 2021). ▪ <i>Schema.org</i>, SKOS, Dublin Core Terms, FOAF, DCMI Metadata Terms, RDF <i>Schema</i>, <i>BibFrame</i>, Vocabulário Local (SMITH-YOSHIMURA, 2018).
	Requisitos para preservação	<ul style="list-style-type: none"> ▪ É necessário ter um plano de manutenção e programação para cada conjunto de dados LOD que garanta que os dados sejam mantidos atualizados e de alta qualidade (FROSTERUS et al., 2021).
	Requisitos para Proveniência	<ul style="list-style-type: none"> ▪ No projeto da Biblioteca Nacional da Espanha, foram definidos os seguintes aspectos de proveniência: o criador e editor dos dados; e informações temporais, por exemplo, criação de dados e data de recuperação (VILA-SUERO; GÓMEZ-PÉREZ, 2013). ▪ Em relação à descrição da proveniência dos recursos, vários vocabulários estão disponíveis, por exemplo, vocabulário do modelo de proveniência aberta OPMV, PROV e PROV- O, que estão sendo padronizados pelo W3C. Em datos.bne.es, descrevemos as informações de metadados usando os vocabulários OPMV e os termos de metadados Dublin Core (VILA-SUERO; GÓMEZ-PÉREZ, 2013).
	Requisitos para dados confidenciais	<ul style="list-style-type: none"> ▪ Conjuntos de dados com problemas de privacidade devem ser anonimizado ou excluído do processo de publicação de dados conectados (FROSTERUS et al., 2021).
	Requisitos para Qualidade de Dados	<ul style="list-style-type: none"> ▪ Na seleção do conjunto de dados é necessário verificar se o conjunto de dados é confiável, completo e consistente (FROSTERUS et al., 2021). ▪ O objetivo da curadoria das fontes de dados no projeto datos.bne.es é avaliar e garantir a qualidade das fontes de dados e do LD publicado. Como a qualidade dos dados nas fontes de dados originais tem um impacto direto na qualidade do RDF gerado, a curadoria de dados é uma atividade crucial nos estágios iniciais do processo de geração de LLD. A tarefa de curadoria da fonte de dados é decomposta em três sub tarefas: identificar os problemas de dados; relatar os problemas de dados; e corrigir os problemas de dados (VILA-SUERO; GÓMEZ-PÉREZ, 2013).
	Requisitos para	<ul style="list-style-type: none"> ▪ Para esta subcategoria, não há resultados.

	granularidade de dados	
	Requisitos para Uso de Dados	<ul style="list-style-type: none"> ▪ Uma vez que os dados são abertos, é difícil rastrear quem faz uso deles. É possível rastrear downloads, uso de API e outras estatísticas da web, mas isso não dá uma ideia de como os dados são usados. Incentivar os usuários a dizerem como eles utilizaram dados é uma alternativa, para que esses feedbacks possibilitem a elaboração de um relatório para um melhor entendimento quanto ao uso dos dados. Procurar links para os identificadores criados também é uma opção e pode mostrar quem está usando esses dados, mas medir o sucesso ainda é um desafio (FROSTERUS et al., 2021).

Fonte: elaborado pela autora, 2021.

Os requisitos para publicação de dados abertos conectados verificados na coleta de dados visam responder ao 3º e último objetivo específico definido nesta pesquisa que é identificar na literatura os requisitos para publicação de dados abertos conectados na *Web*. As 15 (quinze) subcategorias dessa categoria foram criadas a partir do referencial teórico de Derilinx; Lóscio e Archer (2015). Os autores estabeleceram um conjunto de requisitos baseados nos desafios enfrentados por editores e consumidores de dados abertos conectados, mostrando que alguma orientação é necessária para publicação de dados na *Web*.

A aplicação das subcategorias mostrou-se relevante, pois permitiu a categorização de resultados para 11 (onze) requisitos dentre os 15 estabelecidos, conforme resultados apresentados no Quadro 13. Ademais, constatou-se que os requisitos definidos por Derilinx; Lóscio e Archer (2015) são aplicáveis e, também, conforme visto no referencial teórico deste estudo, que requisitos são requeridos no contexto da *Web* de Dados, pois ela exige um modelo padrão para a representação e conexão de dados, favorecendo assim a interoperabilidade entre eles (ISOTANI; BITTENCOURT, 2015).

Os resultados alcançados para a primeira subcategoria analisada que trata dos requisitos para acesso a dados apontam sobre a necessidade dos editores disponibilizarem arquivo em massa para download.

Permitir que os consumidores acessem o conjunto de dados completos em uma única solicitação é uma forma de facilitar o acesso aos dados. É importante porque quando os dados estiverem distribuídos na *Web* através de muitos URIs será possível organizá-los logicamente como um pacote. O acesso em massa garante

uma forma consistente de tratar os dados como um conjunto de dados. Acessá-los individualmente ao longo de muitas consultas pode ser difícil porque se forem usados para remontar um conjunto de dados completos pode levar a uma manipulação inconsistente dos dados. A disponibilização de arquivos em massa possibilita a realização de transferências de arquivos grandes por meio de protocolos de transferência de arquivos, do inglês *file-transfer protocols* (FTP) (LÓSCIO; BURLE; CALEGARI, 2017).

Outro requisito dessa subcategoria encontrado nos estudos analisados diz respeito ao fornecimento de API aos consumidores para acesso aos dados. Oferecer aos consumidores essa interface possibilita maior flexibilidade e facilidade de processamento dos dados, permitindo que sejam usados em tempo real, assim como o uso de filtros para acesso aos dados. Uma API é indicada, principalmente, para grandes conjuntos de dados que tenham uma frequência de atualização constante (LÓSCIO; BURLE; CALEGARI, 2017).

Para a segunda subcategoria aplicada, correspondentes aos requisitos para enriquecimento de dados, observaram-se resultados relacionados à vinculação, refinamento e reconciliação de dados.

O enriquecimento de dados melhora a facilidade de processamento, especialmente no caso de dados não estruturados. Em alguns casos, dados inexistentes podem ser adicionados e novas atribuições e mensurações podem ser acrescentadas a partir de dados brutos pré-existentes. Conjuntos de dados também podem ser enriquecidos por meio de coletas adicionais da mesma maneira que os dados originais são coletados, ou combinando dados originais com outros conjuntos de dados. Se feito corretamente e com ética, publicar um conjunto de dados mais completo pode aumentar a confiança nos dados entre editores e consumidores. Agregar valor a eles é importante porque economiza tempo dos usuários e encoraja mais tipos de reutilização (LÓSCIO; BURLE; CALEGARI, 2017).

Para a terceira subcategoria aplicada, que são os requisitos para formatos de dados, não se obteve nenhum resultado a partir dos textos selecionados para estudo. No entanto é importante ressaltar a necessidade do uso de formatos padronizados de dados legíveis por máquinas, pois, na medida em que os dados passam a ser mais ubíquos e os conjuntos de dados maiores e mais complexos, o processamento por computadores torna-se cada dia mais relevante.

Disponibilizar dados em um formato que não seja legível por máquinas impõe severas limitações para a utilidade. Os dados tornam-se úteis ao serem processados e transformados em informação. Observe que há uma importante distinção entre formatos que podem ser lidos e editados por pessoas utilizando um computador e formatos que são legíveis por máquinas. O último termo implica que os dados sejam prontamente extraídos, transformados e processados por um computador (LÓSCIO; BURLE; CALEGARI, 2017).

Para a quarta subcategoria, requisitos para identificação de dados, foram encontrados resultados que apontam sobre a necessidade de distinção entre dois tipos de URIs: os de vocabulário e os de dados. “Usar URIs persistentes como identificadores de conjuntos de dados [...] são pré-condições essenciais para o gerenciamento e reutilização dos dados de forma adequada” (LÓSCIO; BURLE; CALEGARI, 2017, não paginado).

A quinta subcategoria, requisitos para seleção de dados, não retornou nenhum resultado a partir dos textos selecionados para estudo.

Para a sexta subcategoria, requisitos para vocabulários de dados, obteve-se como resultados alguns vocabulários que podem ser usados para codificar dados e metadados: *GeoNames*, *VIAF*, *ISNI*, *Wikidata*, *Agrovoc*, *MeSH*, *Lista de Cabeçalhos de Assuntos da Biblioteca do Congresso (LCSH)*, *Arte e Arquitetura Thesaurus (AAT)*, *Eurovoc*, *ULAN*, *Dbpedia*, e *FAST*.

A utilização de vocabulários já em uso por outros estimula e facilita o consenso entre comunidades, aumenta a interoperabilidade e reduz as redundâncias, incentivando assim a reutilização de seus próprios dados.

Particularmente, a aplicação de vocabulários compartilhados para metadados (especialmente os metadados estruturais, de procedência, de qualidade e de versionamento) auxilia o processo de comparação e o processamento automático, tanto dos dados quanto dos metadados. Além disso, referir-se a códigos e terminologia padronizada ajuda a evitar ambiguidades e conflitos entre elementos e valores semelhantes (LÓSCIO; BURLE; CALEGARI, 2017).

A sétima subcategoria, requisitos para reutilização da indústria, não retornou nenhum resultado a partir dos textos selecionados para estudo.

Para a oitava subcategoria, requisitos para licenças, as repostas obtidas indicam que ao disponibilizar os dados é preciso decidir sobre o licenciamento e

incluir tais informações no conteúdo dos dados. Dados abertos implicam em uso de uma licença aberta e existem algumas opções específicas para a escolha. A tendência entre bibliotecas é o uso da licença de domínio público *Creative Commons* também conhecida como CC0. Fornecer informações de permissão de uso é essencial para que os consumidores de dados avaliem a utilidade dos dados. De posse dessa informação, as pessoas serão capazes de entender a licença de dados que descreve quaisquer restrições ao uso e o agente de *software* será capaz de detectar automaticamente a licença de dados de uma distribuição (LÓSCIO; BURLE; CALEGARI, 2017).

Para a nona subcategoria, requisitos para metadados, os resultados apontam sobre a necessidade de inclusão da documentação sobre os dados publicados. Para Lóscio; Burle e Calegari (2017), fornecer metadados é um requisito básico para publicar dados na Internet, porque editores e consumidores de dados podem não se conhecer. Desse modo, é necessário fornecer informações que ajudem as pessoas e os aplicativos de computador a compreenderem os dados e a descrição de outros aspectos importantes do conjunto de dados da distribuição. Dentre os vocabulários de metadados citados pelos autores estão o Schema.org, SKOS, Dublin Core Terms, FOAF, DCMI Metadata Terms, RDF Schema, e o *Bibframe*. Com RDF podem ser usados vários vocabulários de metadados simultaneamente.

Para a décima subcategoria, requisitos para preservação, os resultados apontam para a importância da elaboração do plano de manutenção e programação para cada conjunto de dados LOD que garanta que os dados sejam mantidos atualizados e de alta qualidade. Segundo Lóscio; Burle e Calegari (2017), ao remover dados da *Web* se faz necessário preservar o identificador e fornecer informações sobre o recurso arquivado. Dessa forma “O URI de um recurso irá sempre se referenciar para o recurso ou redirecionar para informações a respeito” (LÓSCIO; BURLE; CALEGARI, 2017, não paginado).

Para a décima primeira subcategoria, requisitos para proveniência, os resultados alcançados indicam que oferecer informações sobre o criador e editor dos dados, assim como de informações temporais, por exemplo, criação de dados e data de recuperação atendem a esse requisito. Para Lóscio; Burle e Calegari (2017), oferecer informações sobre a proveniência dos dados é uma das maneiras pelas quais os consumidores julgam a qualidade do conjunto de dados. Compreender sua

origem e histórico ajuda a determinar se os dados são confiáveis e fornece contexto para a interpretação do que é relevante.

Para a décima segunda subcategoria, requisitos para dados confidenciais, os resultados apontam que conjuntos de dados com problemas de privacidade devem ser anonimizado ou excluído do processo de publicação de dados conectados (FROSTERUS et al., 2021).

Para a décima terceira subcategoria, requisitos para qualidade de dados, os resultados apontam para a importância de se verificar no momento da seleção do conjunto de dados se ele é confiável, completo e consistente. A curadoria das fontes de dados desempenha esse papel, pois se destina a identificar os problemas de dados, relatá-los e corrigi-los. De acordo com Lóscio, Burle e Calegari (2017), a qualidade dos dados pode afetar gravemente o uso de dados de aplicativos específicos, incluindo os que são muito diferentes do propósito original de criação. O registro da qualidade dos dados simplifica muito o processo de seleção do conjunto de dados e aumenta as chances de reutilização dos dados. Independentemente das peculiaridades em um determinado campo, a qualidade dos dados deve ser documentada e os problemas de qualidade conhecidos devem ser claramente indicados nos metadados.

A décima quarta subcategoria, requisitos para granularidade de dados, não retornou nenhum resultado a partir dos textos selecionados para estudo.

E finalmente, para a décima quinta e última subcategoria, requisitos para uso de dados, compreende-se que a partir dos resultados alcançados, que incentivar os usuários a dizerem como eles utilizaram os dados é a melhor maneira para se entender como os dados estão sendo usados.

Receber *feedback* ajuda os editores a entender as necessidades de seus consumidores de dados e a melhorar a qualidade dos dados publicados. Além disso, também gera confiança porque mostra aos consumidores que os editores se preocupam em atender às suas necessidades. O ideal é especificar claramente o mecanismo de *feedback* para evitar que os consumidores de dados procurem por uma maneiras de fazê-lo (LÓSCIO; BURLE; CALEGARI, 2017).

A observação de todos os requisitos elencados nessa categoria fornece orientação às partes interessadas na publicação e consumo de dados conectados

melhorando a consistência na forma como os dados são publicados, gerenciados, referenciados e usados na *Web* (DERILINX; LÓSCIO; ARCHER, 2015).

Esta série de requisitos estabelecidos em subcategorias é derivada de casos de uso com o objetivo de orientar o desenvolvimento de boas práticas de publicação de dados na *Web*, independentes de domínios e aplicativos. No entanto, eles podem ser expandidos ou complementados por outros documentos de boas práticas ou padrões para situações mais específicas, como no caso das bibliotecas. Os benefícios da adoção dos requisitos são muitos, entre eles foram identificados os seguintes: facilidade de compreensão, processamento, descoberta, reutilização, confiabilidade, conexão, acesso e interoperabilidade (LÓSCIO; BURLE; CALEGARI, 2017).

Apesar de três das subcategorias não terem retornado resultados nos textos selecionados, elas não devem ser descartadas das demais subcategorias, pois fazem parte do conjunto de requisitos definidos para as melhores práticas de publicação de dados conectados na *Web*. Dentre elas encontram-se:

A subcategoria Requisitos para Seleção de Dados, que, conforme definido em sua descrição, tem a premissa de que os dados devem ser cuidadosamente gerenciados para melhorar a eficiência da coleta e uso, assim como da segurança, garantindo a confidencialidade dos dados pessoais e promovendo a adoção de boas práticas no processo de coleta, processamento e arquivamento (DERILINX; LÓSCIO; ARCHER, 2015).

A subcategoria Requisitos para reutilização da indústria estabelece que os dados abertos precisam de padrões mais fortes para terem aceitação industrial em nível de metadados e de conjunto de dados. O uso industrial tem requisitos diferentes daqueles desenvolvidos pela sociedade civil, então é importante pensar como os dados abertos podem ser usados pela indústria no momento da publicação (DERILINX; LÓSCIO; ARCHER, 2015).

A subcategoria Requisitos para granularidade de dados estabelece que os dados disponíveis em diferentes níveis de granularidade devem ser acessíveis e modelados de maneira comum. Granularidade diz respeito ao nível de detalhe em que as unidades de dados são mantidas. Quanto maior o nível de detalhe dos dados, menor o nível de granularidade, e quanto maior o nível de granularidade, menor será o nível de detalhes dos dados (DERILINX; LÓSCIO; ARCHER, 2015).

Apresentado os resultados, o próximo capítulo apresentará as considerações finais da pesquisa, bem como a indicação das limitações desse estudo e sugestão para trabalhos futuros.

6. CONSIDERAÇÕES FINAIS

Comparado com a prática atual de criação e disponibilização de registros bibliográficos de biblioteca, o método de dados conectados tem vantagens significativas e, ao mesmo tempo, fornece uma extensão natural para o modelo de compartilhamento colaborativo que as bibliotecas tradicionalmente adotam. Dados conectados, especialmente dados abertos conectados, são compartilháveis, extensíveis e fáceis de reutilizar; oferecem suporte a funções multilíngues para dados e serviços do usuário, que são recursos inerentes ao padrão *Linked Data*. Os recursos podem ser descritos em colaboração com outras bibliotecas e conectados a dados fornecidos por outras comunidades ou até mesmo outros indivíduos.

Em contrapartida, as bibliotecas podem fornecer conjuntos de dados padronizados, reutilizáveis e compartilhar ferramentas consagradas no desenvolvimento dos procedimentos técnicos do bibliotecário, como os catálogos de autoridades e os vocabulários controlados.

No entanto, para que essa junção ocorra alcançando interoperabilidade com outros sistemas e conjuntos de dados na rede, as bibliotecas precisam adaptar seus metadados para uso no novo ambiente. Dentre os modelos de dados existentes que atendem aos requisitos do *Linked Data* estão o *Schema.org*, *SKOS*, *Dublin Core Terms*, *FOAF*, *DCMI Metadata Terms*, *RDF Schema*, e o *Bibframe*. Dentre esses, é provável que o *Bibframe* passe a ter uma maior aceitação entre a comunidade bibliotecária pelo fato de estar sendo desenvolvido pela Biblioteca do Congresso.

Notou-se que tecnologia não é mais o principal obstáculo para publicar dados conectados a partir de catálogos de bibliotecas, embora alterar o modelo atual de metadados adotado não seja uma tarefa fácil, pois se viu que são muitos os desafios de carácter tecnológico identificados. Observou-se, portanto, que os desafios existem não por falta de requisitos, mas sim pela complexidade que envolve o processo de publicação de dados abertos conectados a partir de catálogos online, aliada à

carência de pessoal qualificado para desempenhar a função, da escassez de recursos financeiros, da necessidade de conscientização de pessoal sobre os benefícios advindos da adoção dos dados abertos conectados pelas bibliotecas e pela dependência do mercado externo para aquisição de softwares para biblioteca.

A necessidade de pessoal qualificado para implementação de projetos de dados abertos conectados está relacionada à subcategoria criada denominada “*Carência de pessoal qualificado*” da categoria “*Barreiras encontradas na implementação de práticas de dados abertos conectados em bibliotecas para a disponibilização de seus catálogos online na Web*” e possui caráter sistêmico.

Nesse contexto, como aponta Santarem Segundo (2018), é importante observar que existe uma estreita relação entre a pesquisa realizada em Ciência da Informação e a Ciência da Computação, podendo ocorrer, por vezes, conflitos. Mas há diferenças evidentes na forma como esses dois campos contribuem diretamente para as pesquisas. O autor afirma que “estudos relacionados a Dados, *Web Semântica*, Ontologias e Dados Ligados precisam ser compartilhados entre pesquisadores das duas áreas” (SANTAREM SEGUNDO, p. 119, 2018). No estudo de Ali e Warraich (2018) essa barreira foi identificada quando eles apontam que o trabalho com *Linked Data* exige pessoal com experiência em tecnologia da informação e tecnologia da *Web*. Em Senso e Arroyo Machado (2018), a barreira é citada com uma necessidade de entender linguagens de programação e consultar bancos de dados. Vê-se, portanto, que o envolvimento desses 2 (dois) profissionais é necessária para a implementação de projetos de dados conectados em bibliotecas.

Outra constatação importante verificada nesta pesquisa é que a maioria dos projetos avaliados nos estudos analisados é realizada por grandes instituições como a Biblioteca Nacional da Espanha, a Biblioteca Britânica, a Biblioteca Nacional da França, a Europeana, a Biblioteca do Congresso, entre outras. Considerando a estrutura organizacional da maioria das bibliotecas, em especial as localizadas no Brasil, que carecem de potencial econômico, tecnológico e/ou humano, percebeu-se o quanto de mobilização e investimento em pessoal e tecnologia será necessário para que a maioria das bibliotecas brasileiras possa realizar uma tarefa tão especializada que penetra tão claramente no ambiente computacional.

Observou-se, ainda, que entre as barreiras identificadas no desenvolvimento de dados abertos conectados em bibliotecas está a necessidade de aumentar a

conscientização da comunidade bibliotecária (BYRNE; GODDARD, 2010); (GALVÃO; LOPES, 2012); (SINGER; 2009), fator este relacionado à subcategoria criada “*Atitude do pessoal com relação a mudanças*” da categoria “*Barreiras*” deste estudo. Byrne e Goddard (2010) apontaram que é pequena a publicação de estudos sobre dados abertos conectados na literatura da biblioteconomia, o que pode ser observado nos anais de conferências de bibliotecas e nos principais bancos de dados de artigos da área.

Sobre as limitações encontradas no desenvolvimento desta pesquisa, destaca-se a dificuldade de observar um método claro e uniforme de conversão de registros bibliográficos em dados conectados. Viu-se, ainda, que uma série de variáveis pode alterar o andamento de um determinado projeto, como a qualidade da fonte de dados, o licenciamento, o vocabulário e as ontologias utilizadas; assim como também os conjuntos de dados utilizados para enriquecê-los por meio de links, os métodos de publicação e as técnicas utilizadas no processo, entre outras variáveis. O que significa que pode haver vários caminhos para escolher, o que impede o estabelecimento de uma única metodologia.

Entretanto, na maioria dos casos é possível encontrar um denominador comum, que é determinado por aquelas etapas que se repetem, as quais se referem ao planejamento do processo de publicação; escolha e curadoria do conjunto de dados; identificação de recursos para vinculação; aplicação de um modelo de dados: conjuntos de elementos e vocabulários de valores; conversão, disponibilização e atualização dos dados. Podemos entender que essas etapas fazem parte do ciclo de vida comum à maioria dos projetos de dados conectados e dados abertos conectados que têm como objetivo tanto a publicação de dados, como seu posterior enriquecimento.

Outra limitação encontrada refere-se à identificação de políticas de curadoria que estabeleçam como é feita a seleção dos registros de um catálogo de biblioteca para publicá-los como dados abertos conectados. As políticas de bibliotecas, de um modo geral, são definidas tendo em vista a adequação do acervo aos objetivos da biblioteca e das instituições à qual estão inseridas, assim como também é alinhada ao perfil do usuário a que atendem. Dessa forma, para se chegar às delimitações das políticas de curadoria, é necessário realizar pesquisas que foquem em estudos de casos sobre o assunto.

Por fim, à abordagem da temática que trata sobre inovação organizacional no ambiente das bibliotecas, correlacionada à prática de uso dos dados abertos conectados, também constitui uma limitação nesta pesquisa. Embora a temática tenha sido abordada neste estudo, são necessárias outras pesquisas que se proponham a fazer um estudo aprofundado do tema.

Como estudos futuros, sugere-se o desenvolvimento de pesquisas que explorem as limitações indicadas nessa pesquisa, assim como estudos que entrevistem bibliotecários e gestores de bibliotecas no Brasil para apurar qual é a percepção que eles têm sobre dados abertos conectados no campo da biblioteconomia.

Para além das limitações apresentadas acima, que se relacionam aos objetivos e ao desenvolvimento desse estudo, outro desafio, de cunho pessoal, surgiu no desenvolver da pesquisa, relacionado à formação acadêmica da proponente, que foi a interlocução estabelecida com áreas que trouxeram fundamentação teórica e metodológica para o aprofundamento dos estudos sobre *Web* semântica e dados abertos conectados, em especial na categorização dos resultados nas respectivas subcategorias estabelecidas. No entanto, os resultados foram consistentes com o embasamento teórico usado ao longo do processo, o que possibilitou a realização das etapas de categorização e a obtenção das respostas ao questionamento que norteou este estudo, a saber: *Quais os desafios são encontrados pelas bibliotecas para a disponibilização dos catálogos online como dados abertos conectados?* Tanto a questão da pesquisa quanto os objetivos delineados neste estudo foram atendidos e são descritos no Capítulo 7.

Finalmente, no âmbito da Ciência da Informação, considera-se que esta pesquisa contribuiu com os estudos no campo da Gestão e Organização do Conhecimento, ao atingir os objetivos traçados, mapeando os principais desafios encontrados na implementação de práticas de dados abertos conectados em bibliotecas para a disponibilização de seus catálogos online na *Web*. Estes desafios podem ser benéficos para traçar diretrizes e transformá-los em oportunidades.

REFERÊNCIAS

- ALI, Irfan; WARRAICH, Nosheen Fatima. Linked data initiatives in libraries and information centres: a systematic review. 2018. **The Electronic Library**, vol. 36, n. 5, p. 925-937. Disponível em: <https://www.emerald.com/insight/content/doi/10.1108/EL-04-2018-0075/full/html>. Acesso em 21 fev. 2021.
- ARAKAKI, F. A. et al., Bibframe: tendência para a representação bibliográfica na web. **Revista Brasileira de Biblioteconomia e Documentação**, v. 13, p. 2231-2249, 2017. Disponível em: <https://rbbd.febab.org.br/rbbd/article/view/995/1030>. Acesso em: 30 mar. 2021.
- ARAKAKI, F. A. **Linked Data**: ligação de dados bibliográficos. 2016. 144 f. Dissertação (Mestrado em Ciência da Informação) – Universidade Estadual Paulista Júlio de Mesquita Filho, Marília/SP, 2016. Disponível em: <https://repositorio.unesp.br/handle/11449/147979>. Acesso em: 16 jul. 2020.
- ARAÚJO, Carlos Ávila. **Infodemia**: origem, características e consequências. 2021. Disponível em: <https://www.youtube.com/watch?v=u6LWa4kHiew&t=1294s>. Acesso em: 24 ago. 2021.
- ASSUMPÇÃO, Fabrício Silva. **O que é FRBR?** 2012. Disponível em: <https://fabricioassumpcao.com/2012/07/o-que-e-frbr.html#:~:text=O%20Functional%20Requirements%20for%20Bibliographic,FRBR%20foi%20publicado%20em%201998>. Acesso em: 18 mar. 2021.
- ASSUMPÇÃO, Fabrício Silva. **Padrões bibliográficos. 2012.** Disponível em: <https://fabricioassumpcao.com/2012/03/padrees-bibliograficos.html>. Acesso em: 04 abril 2021.
- ASSUMPÇÃO, Fabrício Silva; SANTOS, Plácida Leopoldina Ventura Amorim da Costa; ZAFALON, Regina Zafalon. O controle de autoridade no domínio bibliográfico: os catálogos digitais. **Biblios: Journal of Librarianship and Information Science**, n. 68, p. 21–33, 12 jan. 2018. Disponível em: http://www.scielo.org.pe/scielo.php?script=sci_arttext&pid=S1562-47302017000300002. Acesso em: 18 mar. 2021.
- ÁVILA BARRIENTOS, Eder. **Los Datos enlazados y su uso en bibliotecas**. México: UNAM. Instituto de Investigaciones Bibliotecológicas y de la Información, 2020. Disponível em: http://ru.iibi.unam.mx/jspui/handle/IIBI_UNAM/56. Acesso em: 18 mar. 2021.
- ÁVILA-BARRIENTOS, Eder. BIBFRAME y el futuro de la descripción, vinculación y recuperación de los recursos de información. **Transinformação** [online]. 2020, v. 32. Disponível em: <<https://doi.org/10.1590/1678-9865202032e190069>>. Acesso em: 26 ago. 2021.
- BAKER, Thomas; et al., (2011). **Library Linked Data Incubator Group Final Report**. W3C Incubator Group Report. Disponível em: <http://www.w3.org/2005/Incubator/ld/XGR-ld-20111025/>. Acesso em: 21 abril 2021.
- BAPTISTA, D. M. A catalogação como atividade profissional especializada e objeto de ensino universitário. **Informação & Informação**, v. 11, n. 1, p. 63, 15 jul. 2006. Disponível em: <https://repositorio.unb.br/handle/10482/893>. Acesso em: 18 mar. 2021.

BARBOSA, A. P. **Novos rumos da catalogação**. Rio de Janeiro: BNG/Brasilart, 1978.

BARDIN, Laurence. **Análise de conteúdo**. São Paulo: Edições 70, 2011. 279 p.

BE PART OF THE WEB. **Report of the Stanford Linked Data Workshop**, 27 June. 1 July 2011. Disponível em: <https://www.clir.org/pubs/reports/pub152/>. Acesso em: 05 nov. 2016.

BERNERS-LEE, Tim. **Design Issues**. [S.l.]: W3C, 2006. Disponível em: <http://www.w3.org/DesignIssues/>. Acesso em: 26 de maio de 2020.

BERNERS-LEE, Tim; HENDER, James; LASSILA, Ora. The Semantic Web: a new form of Web content that is meaningful to computers will unleash a revolution of new possibilities. **Scientific American**, New York, v. 284, n. 5, Maio 2011.

BERNERS-LEE, Tim; HENDLER, James. (2001). Publishing on the Semantic Web. **Nature**. 410, april: 1023-1024. Disponível em: <https://github.com/AKSW/semann/blob/master/pdf/T.B.%20Lee%20-%20Publishing%20on%20the%20semantic%20web.pdf>. Acesso em: 12 fev. 2021.

BIZER, C. **The DBpedia Data Provision Architecture**. DBPedia. [Online], 9 nov. 2009. Disponível em: <https://www.w3.org/TR/trig>. Acesso em: 12 de jan. 2021.

BIZER, C.; HEATH, T.; BERNERS-LEE, T. Linked Data – The story so far. [ed.] Tim Heath, M. Hepp and Christian Bizer. **International Journal on Semantic Web and Information System**, Special Issue on Linked Data, 2006.

BYRNE, Gillian; GODDARD, Lisa. The strongest link: libraries and linked data. **D-Lib Magazine**, v.16, n.11/12, 2010. Disponível em: <http://www.dlib.org/dlib/november10/byrne/11byrne.html>. Acesso em: 10 jan. 2021.

CHEN, Peter Pin-Shan. The Entity-Relationship Model: Toward a Unified View of Data. **ACM Transactions on Database Systems**, Vol. 1, No. 1. March 1976, Pages 9-36. Disponível em: <http://www.inf.unibz.it/~nutt/IDBs1011/IDBPapers/chen-ER-TODS-76.pdf>. Acesso em: 18 mar. 2021.

COMMITTEE ON CATALOGING, DESCRIPTION & ACCESS (CCDA). 2000. **Task Force on Metadata**: Final Report. Association for Library Collections & Technical Services (ALCTS) Committee on Cataloging: Description & Access (CCDA), June 16, 2000. Disponível em: <https://www.libraries.psu.edu/tas/jca/ccda/tf-meta6.html>. Acesso em: 12 fev. 2021.

CONEGLIAN, C. S.; SANTAREM SEGUNDO, J. E. Europeana no linked open data: conceitos de web semântica na dimensão aplicada das humanidades digitais. **Pesquisa Brasileira em Ciência da Informação e Biblioteconomia**, v. 12, n. 2, 2017. DOI: 10.22478/ufpb.1981-0695.2017v12n2.36529. Acesso em: 20 jan. 2021.

CONFLUENCE. **Linked Data for Libraries (LD4L) Gateway**. 2020. Disponível em: <https://wiki.lyrasis.org/display/ld4lGW/Linked+Data+for+Libraries+%28LD4L%29+Gateway> . Acesso em: 08 abril 2021.

COYLE, Karen. Changing the nature of library data. **Library Technology Reports**, ISSN 0024-2586. Vol. 46, no 1 (Jan.), 2010, p.14-31. Disponível em: <https://journals.ala.org/index.php/ltr/article/view/4629/5475>. Acesso em: 09 fev. 2021.

COYLE, Karen. **FRBR, before and after: a look at our bibliographic models**. Chicago: Ala editions, 2016.

COYLE, Karen. Library Data in a Modern Context. **Library Technology Reports**. January 2010. Disponível em: <https://journals.ala.org/index.php/ltr/article/view/4630/5478>. Acesso em: 09 fev. 2021. (2010 a).

COYLE, Karen. Library Data in the Web World. **Library Technology Reports**. February/March 2010. Disponível em: <https://www.journals.ala.org/index.php/ltr/article/view/4673>. Acesso em: 09 fev. 2021. (2010 b).

CUTTER, Charles A. **Rules for a dictionary catalog**. 4a. ed., rewritten Washington D.C.: Government Printing office, 1904.

DERILINX; LÓSCIO; ARCHER. **Data on the Web Best Practices Use Cases & Requirements do W3C**. 2015. Disponível em: <https://www.w3.org/TR/dwbp-ucr/>. Acesso em: 09 fev. 2021.

DIRAOL. **Gráfico descrevendo as 5 estrelas dos dados abertos**. 2013. Disponível em: https://pt.wikipedia.org/wiki/Dados_abertos#/media/Ficheiro:Open_data_5stars.svg. Acesso em: 15 mar. 2021.

FAO (Food and Agricultural Organization of the United Nations). 2021. **AGROVOC**. Disponível em: <http://aims.fao.org/agrovoc>. Acesso em: 08 abril 2021.

FIUZA, Marysia Malheiros. A catalogação bibliográfica até o advento das novas tecnologias. **Revista da Escola de Biblioteconomia da UFMG**, Belo Horizonte, v. 1, n. 16, p. 43-53, mar. 1987.

FROSTERUS, Matias et al., **Best Practices for Library Linked Open Data (LOD) Publication**. LIBER Linked Open Data (LOD) Working Group, February 2021. Disponível em: <https://libereurope.eu/wp-content/uploads/2021/02/LOD-Guidelines-FINAL-Feb-2021.pdf>. Acesso em: 28 abril 2021.

FUJITA, MSL. Introdução o contexto da indexação para a catalogação de livros: uma introdução. In: Fujita, MSL, org., et al., **A indexação de livros: a percepção de catalogadores e usuários de bibliotecas universitárias**. Um estudo de observação do contexto sociocognitivo com protocolos verbais [online]. São Paulo: Editora UNESP; São Paulo: Cultura Acadêmica, 2009. 149 p. ISBN 978- 85-7983-015-0.

FUSCO, Elvis. **Aplicação dos FRBR na modelagem de catálogos bibliográficos digitais**. São Paulo: Cultura Acadêmica, 2011. (Coleção PROPG Digital - UNESP). ISBN 9788579832109. Disponível em: <http://hdl.handle.net/11449/109186>. Acesso em: 18 mar.2021.

FUSCO, Elvis. **Modelos conceituais de dados como parte do processo da catalogação: perspectiva de uso dos FRBR no desenvolvimento de catálogos bibliográficos digitais**. 2010.

GALVÃO, Rosa Maria Tavares; CORDEIRO, Maria Inês Novos princípios, modelos e normas para o futuro dos serviços de informação bibliográfica. **Actas do Congresso Nacional de Bibliotecários, Arquivistas e Documentalistas**, n.10, 2010. Disponível em: <https://www.bad.pt/publicacoes/index.php/congressosbad/article/view/197>. Acesso em: 07 fev. 2021.

GALVÃO, Rosa Maria; LOPES, Maria Margarida. Da ficha bibliográfica aos dados relacionados: uma revolução em curso. **Actas do Congresso Nacional de Bibliotecários, Arquivistas e Documentalistas**, n.11, 2012. Disponível em: <https://www.bad.pt/publicacoes/index.php/congressosbad/article/view/406>. Acesso em: 27 jan. 2021.

GONZÁLEZ, P. U. A Strategy for Integrating Printed Catalog Cards from Three Cuban Libraries into the Open Linked Data Space. **Library Trends**, v. 67, n. 4, p. 713–728, 2019.

GORMAN, Michael. Seymour Lubetzky, man of principles. In: CONNELL, Tschera Harkness; MAXWELL, Robert L. **The Future of Cataloging**: insights from the Lubetzky Symposium. Chicago: ALA, 2000. p. 12-21.

GUIA PRÁTICO PARA PUBLICAÇÃO DE DADOS ABERTOS CONECTADOS NA WEB. Curitiba: Editora Appris, 2018.

HALLO, María; et al., Current state of Linked Data in digital libraries. **Journal of information science**, April 2016, Vol. 42(2), pp.117-127. Disponível em: <https://journals-sagepub-com.ez27.periodicos.capes.gov.br/doi/full/10.1177/0165551515594729>. Acesso em: 15 abril 2021.

HARPRING, Patricia; BACA, Murtha. **Introdução aos vocabulários controlados**: terminologia para arte, arquitetura e outras obras culturais. São Paulo: Pinacoteca de São Paulo, 2016. Disponível em: <https://www.sisemsp.org.br/wp-content/uploads/2013/12/Vocabularios%20Controlados%20-%20Digital.pdf>. Acesso em: 07 fev. 2021

HENDLER, James, BERNERS-LEE, Tim e MILLER, Eric. Integrating Applications on the Semantic Web. **Journal of the Institute of Electrical Engineers of Japan**, Vol. 122 (10), outubro de 2002, p. 676-680. Disponível em: <https://www.w3.org/2002/07/swint>. Acesso em: 07 abril 2021.

HITZLER, P.; KRÖTZSCH, M.; RUDOLPH, S. **Foundations of Semantic Web technologies**. Boca Raton: CRC Press, 2009.

HODGE, Gail. 2000. **Systems of Knowledge Organization for Digital Libraries**: Beyond Traditional Authority Files. Washington, DC: Council on Library and Information Resources. Disponível em: <http://www.clir.org/pubs/reports/pub91/contents.html>. Acesso em: 15 abril 2021.

IFLA - INTERNATIONAL FEDERATION OF LIBRARY ASSOCIATIONS AND INSTITUTIONS. **Declaração de Princípios Internacionais de Catalogação**. [S.l.]: IFLA, 2016.

IFLA - INTERNATIONAL FEDERATION OF LIBRARY ASSOCIATIONS AND INSTITUTIONS. **Functional Requirements for Bibliographic Records**. [S.l.]: IFLA, 2009.

IFLA - INTERNATIONAL FEDERATION OF LIBRARY ASSOCIATIONS AND INSTITUTIONS. **Functional Requirements for Bibliographic Records**. [S.l.]: IFLA, 1998.

INTERNATIONAL FEDERATION OF LIBRARY ASSOCIATIONS AND INSTITUTIONS. **IFLA Library Reference Model**: a conceitual model for bibliographic information. Netherlands: IFLA, 2017.

ISOTANI, S.; BITTENCOURT, Ig Ibert. **Dados Abertos Conectados**: em Busca da Web do Conhecimento. [s.l.] Novatec, 2015. Disponível em: <https://ceweb.br/publicacao/livro-dados-abertos/>. Acesso em: 03 fev. 2021.

JESUS, Ananda Fernanda de; CASTRO, Fabiano Ferreira de. Dados bibliográficos para o linked data: uma revisão sistemática de literatura. **Brazilian Journal of Information Studies: Research Trends**. 13:1 (2019) p.45-55. Disponível em: <https://revistas.marilia.unesp.br/index.php/bjis/article/view/8297>. Acesso em: 03 fev. 2021.

KROEGER, Angela. The road to Bibframe: the evolution of the idea of bibliographic transition into a post-MARC future. **Cataloguing & Classification Quarterly**, 51 (8), 873-890, 2013. <https://doi.org/10.1080/01639374.2013.823584>. Acesso em: 20 set.2021.

LAM, Alice. 2005. Organizational innovation. In: FAGERBERG, J; MOWERY, D. C; NELSON, R. R. **The Oxford Handbook of Innovation**. Oxford: Oxford University Press.

LAUFER, Carlos. **Guia de web semântica**. São Paulo: CeWeb.br, 2015. Disponível em: https://nic.br/media/docs/publicacoes/13/Guia_Web_Semantica.pdf. Acesso em: 02 fev. 2021.

LEVACOV, M. Tornando a informação disponível: o acesso expandido e a reinvenção da biblioteca. In: MARCONDES, C. H. et. al. (Org.) **Bibliotecas digitais: saberes e práticas**. Salvador: EDUFBA; Brasília: IBICT, 2005. p. 293-307. Disponível em:

LIBRARY OF CONGRESS. **Linked Data Service**. 2021. Disponível em: <https://id.loc.gov/>. Acesso em 08 abril 2021.

LIBRARY OF CONGRESS. **Overview of the BIBFRAME 2.0 Model**. 2016. Disponível em: <https://www.loc.gov/bibframe/docs/bibframe2-model.html>. Acesso em: 20 mar.2021.

LÓSCIO, Bernadette Farias; BURLE, Caroline; CALEGARI, Newton J. **Boas Práticas para Dados na Web**. W3C. 2017. Disponível em: <https://www.w3.org/Translations/DWBP-pt-BR/>. Acesso em: 02 fev. 2021.

LUBETZKY, Seymour. **Cataloging Rules and Principles**: critique of the A.L.A. Rules for entry and a proposed desing for their revision. Washington: Library of Congress, 1953.

LUZ, Charley dos Santos; CONEGLIAN, Caio Saraiva; SANTAREM SEGUNDO, José Eduardo. Tecnologias da web semântica para a recuperação da informação no wikidata. **Revista Digital de Biblioteconomia & Ciência da Informação**, Campinas, v. 17, p. 1-20, 2019. Disponível em: <https://periodicos.sbu.unicamp.br/ojs/index.php/rdbci/article/view/8651791>. Acesso em: 14 fev. 2021.

MACHADO, Ana Maria Nogueira. **Informação e controle bibliográfico**: um olhar sobre a cibernética. São Paulo: Editora UNESP, 2003.

MACHADO, Raildo de Sousa; ZAFALON, Zaira Regina. **Catálogo**: dos princípios e teorias ao RDA e IFLA LRM. João Pessoa: Ed. UFPB, 2020. ISBN 978-65-5942-000-1. Disponível em: <http://www.editora.ufpb.br/sistema/press5/index.php/UFPB/catalog/book/336>. Acesso em 16 fev. 2021.

MAGALHÃES, Regis Pires. **Um ambiente para processamento de consultas federadas em Linked Data Mashups**. Dissertação (Mestrado em Ciência da Computação) —

Departamento de Ciência da Computação, Mestrado e Doutorado em Ciência da Computação, Universidade Federal do Ceará, Fortaleza/CE, 2012. Disponível em: http://www.repositorio.ufc.br/bitstream/riufc/18554/1/2012_dis_rpmagalhaes.pdf. Acesso em: 15 abril 2021.

MANUAL DOS DADOS ABERTOS: desenvolvedores. São Paulo: Comitê Gestor da Internet no Brasil, 2011. Disponível em: https://www.w3c.br/pub/Materiais/PublicacoesW3C/manual_dados_abertos_desenvolvedores_web.pdf. Acesso em: 15 mar. 2021.

MARCONDES, Carlos Henrique. **Dados abertos interligados**: publicação, recuperação e integração de acervos de arquivos, bibliotecas e museus na web. Marília: Oficina Universitária; São Paulo: Cultura Acadêmica, 2021. Disponível em: https://ebooks.marilia.unesp.br/index.php/lab_editorial/catalog/view/224/1390/2512-1. Acesso em: 30 mar. 2021.

MARCONDES, Carlos Henrique. Interoperabilidade entre acervos digitais de arquivos, bibliotecas e museus: potencialidades das tecnologias de dados abertos interligados. **Perspectivas em Ciência da Informação**, v. 21, n. 2, p. 61–83, jun. 2016. Disponível em: <https://www.scielo.br/j/pci/a/8svGtzqw5HZCrfrPJbRypsb/?lang=pt&format=pdf>. Acesso em 16 fev. 2021.

MARTINHO, N. O.; FUJITA, M. S. L. La catalogación de materias: apuntes históricos sobre su normalización. **Scire**, Zaragoza, v. 16, p. 61-70, 2011.

MAZZOCCHI, Fulvio. 2018. Knowledge organization system (KOS): an introductory critical account. **Knowledge Organization** 45, no. 1: 54-78. Also available in Hjørland, Birger, ed. **ISKO Encyclopedia of Knowledge Organization**. Disponível em: <http://www.isko.org/cyclo/kos>. Acesso em 15 abril 2021.

MEY, Eliane Serrão Alves. **Introdução à catalogação**. Brasília: 1995.

MEY, Eliane Serrão Alves; SILVEIRA, Naira Christofolletti. **Catalogação no plural**. Brasília, DF: Briquet de Lemos / Livros, 2009.

MICHAELIS. **Dicionário Brasileiro da Língua Portuguesa**. Melhoramentos: 2021. Disponível em: <https://michaelis.uol.com.br/moderno-portugues/busca/portugues-brasileiro/desafio/>. Acesso em: 15 abril 2021.

MILES, Alistair; SEAN Bechhofer, eds. 2008. **SKOS, Simple Knowledge Organization System Reference**. W3C Working Draft 25 January 2008. Disponível em: <http://www.w3.org/TR/2008/WD-skos-reference-20080125/>. Acesso em: 15 abril 2021.

MODESTO, Fernando. **Panorama da Catalogação no Brasil**: da década de 30 aos primeiros anos do Século 21. Disponível em: https://www.researchgate.net/publication/228806592_PANORAMA_DA_CATALOGACAO_NO_BRASIL_Da_decada_de_1930_aos_primeiros_anos_do_Seculo_XXI/citation/download. Acesso em: 07 fev. 2021.

OCDE - ORGANIZAÇÃO PARA A COOPERAÇÃO E DESENVOLVIMENTO ECONÔMICO. **Manual de Oslo**. Diretrizes para Coleta e Interpretação de Dados Sobre Inovação. 3. ed. Publicação conjunta da OCDE e Eurostat. Versão Brasileira: Finep, 2005. Disponível em: <http://www.finep.gov.br/images/apoio-e-financiamento/manualoslo.pdf>. Acesso em: 22 Maio 2021.

OCLC. **Web Services**. 2019. Disponível em: <https://www.oclc.org/developer/develop/web-services.en.html>. Acesso em: 09 abril 2021.

OMS - ORGANIZACIÓN PANAMERICANA DE LA SALUD. 2020. **Entenda a infodemia e a desinformação na luta contra a covid-19**. Disponível em: https://iris.paho.org/bitstream/handle/10665.2/52054/Factsheet-Infodemic_por.pdf?sequence=14. Acesso em: 28 ago.2021.

OPEN KNOWLEDGE FOUNDATION. **Open Data Handbook**. 2010. Disponível em: <http://opendatahandbook.org/guide/en/>. Acesso em: 26 de Maio de 2020.

ORCID. 2021. Disponível em: <https://info.orcid.org/documentation/>. Acesso em: 08 de abril 2021.

ORTEGA, C. D. Do princípio monográfico à unidade documentária: exploração dos fundamentos da catalogação. **Liinc em Revista**, v. 7, n. 1, p. 43-60, 2011. Disponível em: <http://repositorios.questoesemrede.uff.br/repositorios/handle/123456789/983>. Acesso em: 15 jan. 2021.

PACHECO, K. L.; ORTEGA, C. D. Origem do modelo frbr. **Biblios** (Peru), n. 60, p. 63-75, 2015. Disponível em: <https://brapci.inf.br/index.php/res/v/69681>. Acesso em: 15 jan. 2021.

PASSINI MORENO, F.; MÁRDERO ARELLANO, M. Á. Requisitos funcionais para registros bibliográficos - FRBR: uma apresentação. **RDBCI: Revista Digital de Biblioteconomia e Ciência da Informação**, v. 3, n. 1, p. 20, 30 set. 2005.

PATRÍCIO, Helena Simões. A Europeana e a agregação de metadados na web: análise dos esquemas ESE/EDM e da aplicação de standards da web semântica a dados de bibliotecas. **Actas do Congresso Nacional de Bibliotecários, Arquivistas e Documentalistas**, n.11, 2012. Disponível em: <https://www.bad.pt/publicacoes/index.php/congressosbad/article/view/458/pdf>. Acesso em: 27 jan. 2021.

POZZI, Brisa; FUJITA, Mariângela Spotti Lopes. Do catálogo impresso ao catálogo on-line: algumas considerações e desafios ao bibliotecário. **Revista ACB**, v. 17, n. 1, p. 59-75, 2012. Disponível em: <http://hdl.handle.net/11449/114976>. Acesso em: 08 abril 2021.

PRODANOV, Cléber Cristiano; FREITAS, Ernani Cesar de. **Metodologia do trabalho científico** [recurso eletrônico]: métodos e técnicas da pesquisa e do trabalho acadêmico. Novo Hamburgo: Feevale, 2013.

RAMALHO, Rogério Aparecido Sá. Bibframe: modelo de dados interligados para bibliotecas. **Informação & Informação**, [S.l.], v. 21, n. 2, p. 292–306, dez. 2016. Disponível em: <http://www.uel.br/revistas/uel/index.php/informacao/article/view/26425>. Acesso em: 20 jan. 2021.

RANGANATHAN, S. R. **As cinco leis da Biblioteconomia**. Brasília: Briquet de Lemos Livros, 2009. 336 p.

RIO-BRANCO, Luciana Beatriz Piovezan. **Interoperabilidade semântica entre linguagens de indexação para bibliotecas universitárias**. Marília, 2020. Disponível em: <https://repositorio.unesp.br/handle/11449/194243?show=full>. Acesso em: 20 jan. 2021.

SANTAREM SEGUNDO, José Eduardo. Web semântica: fluxo para publicação de dados abertos e ligados. **Informação em Pauta**, Fortaleza, v. 3, número especial, p. 117-140, nov. 2018. DOI: <https://doi.org/10.32810/2525-3468.ip.v3iEspecial.2018.39721.117-140>. Disponível em: <http://www.periodicos.ufc.br/informacaoempauta/article/view/39721/pdf>. Acesso em: 20 Maio 2021.

SANTOS NETO, Antonio Laurindo dos, et. al. Tecnologias de dados abertos para interligar bibliotecas, arquivos e museus: um caso machadiano. **Transinformação**, Campinas, v. 25, n. 1, p. 81-87, apr. 2013. Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0103-37862013000100008&lng=en&nrm=iso. Acesso em: 22 jan. 2021.

SANTOS, Plácida L. V Amorim. Prefácio. In: FUSCO, Elvis. **Aplicação dos FRBR na modelagem de catálogos bibliográficos digitais**. São Paulo: Cultura Acadêmica, 2011, p. 9-12, (Coleção PROP Digital - UNESP). ISBN 9788579832109. Disponível em: <http://hdl.handle.net/11449/109186>. Acesso em: 18 mar.2021.

SAYÃO, Luis Fernando; MARCONDES, Carlos Henrique. O desafio da interoperabilidade e as novas perspectivas para as bibliotecas digitais. **Transinformação**. 2008, v. 20, n. 2, pp. 133-148. Disponível em: <https://www.scielo.br/j/tinf/a/LSxTfhK6NfX54t4ypBK87kM/?lang=pt#>. Acesso em: 30 ago.2021.

SENSO, J. A.; ARROYO MACHADO, W. q **Revista española de Documentación Científica**, v. 41, n. 4, 21 nov. 2018. Disponível em: file:///C:/Users/TEMP.lg.019/Downloads/La_publicacion_en_Linked_Data_de_registros_bibliog.pdf. Acesso em: 16 fev. 2021.

SERRA, L. G.; SANTAREM SEGUNDO, J. E. O catálogo da biblioteca e o linked data. **Em Questão**, v. 23, n. 2, p. 167, 26 abr. 2017. Disponível em: <https://seer.ufrgs.br/index.php/EmQuestao/article/view/67162>. Acesso em 07 fev. 2021.

SERRA, Liliana Giusti. **A Web Semântica na gestão de livros digitais licenciados: uma proposta de modelo**. Marília, 2019. 152 f. Disponível em: <https://repositorio.unesp.br/handle/11449/183526>. Acesso em 18 mar. 2021.

SHIEH, J.; REESE, T. The importance of identifiers in the new Web environment and using the Uniform Resource Identifier (URI) in subfield zero (\$0): a small step that is actually a big step. **Journal of Library Metadata**, v. 15, p. 208-226, 2015.

SIMIONATO, Ana Carolina; ANDRADE, M. C.; ARAKAKI, F. A.; GONÇALEZ, P. R. V. A. BIBFRAME: panorama de iniciativas em bibliotecas. In: I Encontro de RDA no Brasil, 2019, Florianópolis. **Anais do I Encontro de RDA no Brasil**. Florianópolis: Editora da UDESC, 2019. p. 31-38. Disponível em: <https://rdanobrasil.org/wp-content/uploads/2019/12/2019-bibframe-panorama-de-iniciativas-em-bibliotecas-texto.pdf>. Acesso em: 20 Jan 2021.

SIMÓN-MARTÍN, J.; ARIAS-COELLO, A.; SIMÓN-BLAS, C. The impact of the economic crisis on Spanish university libraries. **Revista española de Documentación Científica**, v. 39, n. 3, p.142, 30 set. 2016.

SINGER, Ross. 2009. Linked Library Data Now! **Journal of Electronic Resources Librarianship**. 21. n.2: 114-126. <http://dx.doi.org/10.1080/19411260903035809>. Acesso em: 09 fev. 2021.

SMITH-YOSHIMURA, Karen. Analysis of International Linked Data Survey for Implementers. **Code4Lib Journal**. 2018. Disponível em: <https://journal.code4lib.org/articles/13867>. Acesso em: 09 fev. 2021.

SMITH-YOSHIMURA, Karen. Analysis of International Linked Data Survey for Implementers. **D-Lib Magazine**. 2016. Disponível em: <http://www.dlib.org/dlib/july16/smith-yoshimura/07smith-yoshimura.html>. Acesso em: 09 fev. 2021.

STEELE, T. What comes next: understanding BIBFRAME. **Library Hi Tech**, v. 33, n. 3, p. 1-13, 2018. Doi: <https://doi.org/10.1108/LHT-06-2018-0085>. Acesso em: 27 ago. 2021.

SULÉ, A.; et al. Aplicación del modelo de datos RDF en las colecciones digitales de bibliotecas, archivos y museos de España. **Revista española de Documentación Científica**, v. 39, n. 1, p. e121, 30 mar. 2016.

SVENONIUS, E.; MCGARRY, D. (eds.). **Seymour Lubetzky: writings on the classical art of cataloging**. Englewoog: Libraries Unlimited, 2001.

TAYLOR, A. G. **Introduction to cataloging and classification**. 10th ed. Westport: Libraries Unlimited, 2006.

TILLETT, Barbara. **O que é FRBR?**: um modelo conceitual para o universo bibliográfico. Tradução Lidia Alvarenga e Renato Rocha Souza. 2006. Disponível em: <https://www.loc.gov/catdir/cpsol/o-que-e-frbr.pdf>. Acesso em: 23 mar. 2021.

TORINO, E.; VIDOTTI, S. A. B. G.; VECHIATO, F. L. Contribuições do atributo metadados para a encontrabilidade da Informação. **Em Questão**, v. 26, n. 2, p. 437–457, 23 abr. 2020. Disponível em: <https://seer.ufrgs.br/EmQuestao/article/view/93072/0>. Acesso em: 23 mar. 2021.

TORRE-BASTIDA, Ana-Isabel; GONZÁLEZ-RODRÍGUEZ, Marta; VILLAR-RODRÍGUEZ, Esther (2015). Datos abiertos enlazados (LOD) y su implantación en bibliotecas: iniciativas y tecnologías". **El profesional de la información**, v. 24, n. 2, pp. 113-120. Disponível em: https://www.researchgate.net/publication/277930183_Datos_abiertos_enlazados_LOD_y_su_implantacion_en_bibliotecas_iniciativas_y_tecnologias. Acesso em: 23 mar. 2021.

UCL - University College London. MRC Centre of Epidemiology for Child Health. **Data management**. Disponível em: <http://www.ucl.ac.uk/ich/research-ich/mrc-cech/data>. Acessado em: 15 mar. 2021.

VILA-SUERO, Daniel ; GÓMEZ-PÉREZ, Asunción. Datos.bne.es and MARiMbA: an insight into library linked data. **Library Hi Tech**, 2013, Vol.31(4), pp.575-601. Disponível em: <https://www.emerald.ez27.periodicos.capes.gov.br/insight/content/doi/10.1108/LHT-03-2013-0031/full/html>. Acesso em: 21 abril 2021.

W3C LIBRARY LINKED DATA INCUBATOR GROUP. **Datasets, Value Vocabularies, and Metadata Element Sets**. W3C, 2011. Disponível em: https://www.w3.org/2005/Incubator/lld/wiki/Vocabulary_and_Dataset. Acesso em: 21 jan. 2021.

W3C WORKING GROUP NOTE. **RDF 1.1 Primer**. Disponível em: <https://www.w3.org/TR/rdf11-primer/>. Acesso em: 02 fev. 2021.

W3C. **Extensible Markup Language**. 2015. Disponível em: <https://www.w3.org/XML/> . Acesso em: 08 mar. 2021.

W3C. **Linked Data**. 2015. Disponível em: <https://www.w3.org/standards/semanticweb/data.html>. Acesso em: 09 mar.2021.

W3C. **OWL: web ontology language (OWL)**. 2012. Disponível em: <https://www.w3.org/OWL/> . Acesso em: 08 mar. 2021.

W3C. **Semantic Web**. 2015. Disponível em: <https://www.w3.org/standards/semanticweb/>. Acesso em: 15 fev. 2021.

W3C. **SPARQL query language for RDF**. 2008. Disponível em: <https://www.w3.org/TR/rdf-sparql-query/> . Acesso em: 08 mar. 2021.

W3C. **W3C Semantic Web Activity**. 2013. Disponível em: <https://www.w3.org/2001/sw/>. Acesso em: 15 fev. 2021.

XAVIER, A.; HERNANDEZ, F. OAI-PMH and Linked Open Data in the context of Hispana and Europeana : some historical reflections. **JLIS : Italian Journal of Library, Archives and Information Science = Rivista italiana di biblioteconomia, archivistica e scienza dell'informazione** : 11, 1, 2020, n. 11, 2020.

ZENG, Marcia Lei. Interoperability. **Knowledge Organization**, Baden-Baden, v. 46, n. 2, p. 122-146, 2019a. Disponível em: <https://www.isko.org/cyclo/interoperability>. Acesso em: 12 abril 2021.

ZENG, Marcia Lei. Semantic enrichment for enhancing LAM data and supporting digital humanities. Review article. **El profesional de la información**, v. 28, n. 1, e280103, 2019b. Disponível em: https://www.researchgate.net/publication/330348622_Semantic_enrichment_for_enhancing_LAM_data_and_supporting_digital_humanities_Review_article. Acesso em: 04 de abril de 2021.

ŽUMER, Maja. 2018. IFLA Library Reference Model (IFLA LRM): Harmonisation of the FRBR Family. **Knowledge Organization** 45, no. 4: 310-8. Also available in Hjørland, Birger and Gnoli, Claudio eds., **ISKO** Encyclopedia of Knowledge Organization. Disponível em: <http://http://www.isko.org/cyclo/lrm>. Acesso em: 04 abril 2021.

APÊNDICE A

TEXTOS SELECIONADOS PARA ANÁLISE NESTE ESTUDO

Referência	Tipologia	Ano
FROSTERUS, Matias, et al., Best Practices for Library Linked Open Data (LOD) Publication . LIBER Linked Open Data (LOD) Working Group, February 2021. Disponível em: https://libereurope.eu/wp-content/uploads/2021/02/LOD-Guidelines-FINAL-Feb-2021.pdf . Acesso em: 28 abril 2021.	Documental	2021
JESUS, Ananda Fernanda de; CASTRO, Fabiano Ferreira de. Dados bibliográficos para o linked data: uma revisão sistemática de literatura. Brazilian Journal of Information Studies: Research Trends 13:1 (2019) p.45-p.55. Disponível em: https://revistas.marilia.unesp.br/index.php/bjis/article/view/8297 . Acesso em 21 abril 2021.	Bibliográfico	2019
SMITH-YOSHIMURA, Karen. Analysis of International Linked Data Survey for Implementers. Code4Lib Journal . 2018. Disponível em: https://journal.code4lib.org/articles/13867 . Acesso em 21 abril 2021.	Documental	2018
ALI, Irfan; WARRAICH, Nosheen Fatima (2018). <i>Linked data</i> initiatives in libraries and information centres: a systematic review. The Electronic Library , vol. 36, n. 5, p. 925-937. Disponível em: https://www.emerald.com/insight/content/doi/10.1108/EL-04-2018-0075/full/html . Acesso em 21 abril 2021.	Bibliográfico	2018
SENSO, Jose A.; ARROYO MACHADO, Wenceslau. La publicación en Linked Data de registros bibliográficos: modelo e implementación. Revista española de Documentación Científica , v. 41, n. 4, 21 nov. 2018. Disponível em: http://redc.revistas.csic.es/index.php/redc/article/view/1023/1637 . doi/10.1108/LHT-03-2013-0031/full/html. Acesso em 21 abril 2021.	Bibliográfico	2018
SERRA, L. G.; SANTAREM SEGUNDO, J. E. O catálogo da biblioteca e o linked data. Em Questão , v. 23, n. 2, p. 167, 26 abr. 2017. Disponível em: https://seer.ufrgs.br/index.php/EmQuestao/article/view/67162 .	Bibliográfico	2017

Acesso em 07 fev. 2021.		
VILA-SUERO, Daniel; GÓMEZ-PÉREZ, Asunción. Datos.bne.es and MARiMbA: an insight into library linked data. Library Hi Tech , 2013, Vol.31(4), pp.575-601. Disponível em: https://www-emerald.ez27.periodicos.capes.gov.br/insight/content/doi/10.1108/LHT-03-2013-0031/full/html . Acesso em 21 abril 2021.	Bibliográfico	2013
BAKER, Thomas; et.al. (2011). Library Linked Data Incubator Group Final Report . W3C Incubator Group Report. Disponível em: http://www.w3.org/2005/Incubator/ld/XGR-ld-20111025/ . Acesso em 21 abril 2021.	Documental	2011

Fonte: elaborado pela autora, 2021.