

UNIVERSIDADE FEDERAL DE SÃO CARLOS
CENTRO DE EDUCAÇÃO E CIÊNCIAS HUMANAS
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO

DAYANE ONAGA FERREIRA MACHADO

SCHEMA.ORG PARA REPRESENTAÇÃO NOS CATÁLOGOS DIGITAIS: a
descoberta de recursos informacionais na *Web*

São Carlos, SP
2022

DAYANE ONAGA FERREIRA MACHADO

SCHEMA.ORG PARA REPRESENTAÇÃO NOS CATÁLOGOS DIGITAIS: a descoberta de recursos informacionais na *Web*

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Informação da Universidade da Universidade Federal de São Carlos (UFSCar), como parte das exigências para obtenção do título de Mestre em Ciência da Informação.

Área de concentração: Conhecimento, Tecnologia e Inovação.

Linha de pesquisa: Tecnologia, Informação e Representação.

Orientadora: Profa. Dra. Ana Carolina Simionato Arakaki



DADOS E METADADOS
grupo de pesquisa e extensão

São Carlos, SP
2022

Onaga Ferreira Machado, Dayane

Schema.org para representação nos catálogos digitais: a descoberta de recursos informacionais na Web / Dayane Onaga Ferreira Machado -- 2022.
109f.

Dissertação (Mestrado) - Universidade Federal de São Carlos, campus São Carlos, São Carlos

Orientador (a): Ana Carolina Simionato Arakaki

Banca Examinadora: Barbara Coelho Neves, Sandra

Milena Roa Martínez

Bibliografia

1. Catálogo digital. 2. Catálogo de acesso público online.
3. Schema.org. I. Onaga Ferreira Machado, Dayane. II. Título.

Ficha catalográfica desenvolvida pela Secretaria Geral de Informática
(SIn)

DADOS FORNECIDOS PELO AUTOR

Bibliotecário responsável: Ronildo Santos Prado - CRB/8 7325

DAYANE ONAGA FERREIRA MACHADO

SCHEMA.ORG PARA REPRESENTAÇÃO NOS CATÁLOGOS DIGITAIS: a descoberta de recursos informacionais na *Web*

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Informação da Universidade da Universidade Federal de São Carlos (UFSCar), como parte das exigências para obtenção do título de Mestre em Ciência da Informação.

Área de concentração: Conhecimento, Tecnologia e Inovação.

Linha de pesquisa: Tecnologia, Informação e Representação.

BANCA EXAMINADORA

Profa. Dra. Ana Carolina Simionato Arakaki
Universidade Federal de São Carlos (UFSCar)

Profa. Dra. Barbara Coelho Neves
Universidade Federal da Bahia (UFBA)/Universidade Federal de São Carlos
(UFSCar)

Profa. Dra. Sandra Milena Roa Martínez
Universidad del Cauca (Unicauca)



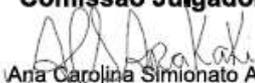
UNIVERSIDADE FEDERAL DE SÃO CARLOS

Centro de Educação e Ciências Humanas
Programa de Pós-Graduação em Ciência da Informação

Folha de Aprovação

Defesa de Dissertação de Mestrado da candidata Dayane Onaga Ferreira Machado, realizada em 12/05/2022.

Comissão Julgadora:


Profa. Dra. Ana Carolina Simionato Arakaki (UFSCar)

Profa. Dra. Barbara Coelho Neves (UFBA)

Profa. Dra. Sandra Milena Roa-Martínez (UNICAUCA)

O Relatório de Defesa assinado pelos membros da Comissão Julgadora encontra-se arquivado junto ao Programa de Pós-Graduação em Ciência da Informação.

AGRADECIMENTOS

Aos meus pais, Nilton e Marcia, por todo apoio, paciência, incentivo e ensinamentos ao longo de toda minha vida.

À minha orientadora, Dra. Ana Carolina Simionato Arakaki, pela confiança depositada em mim, pelo aprendizado e incentivo desde a graduação.

Aos professores do PPGCI pelo empenho em manter a qualidade do ensino em meio a pandemia de Covid 19, um período atípico e extremamente conturbado.

Ao grupo de pesquisa Dados e Metadados pelas discussões enriquecedoras e por despertar e intensificar o interesse pelos assuntos tratados nesta dissertação.

À banca, composta pelas professoras Dra. Barbara Coelho Neves e Dra. Sandra Milena Roa Martínez pela atenção e tempo investidos na leitura da dissertação e pelas contribuições para o desenvolvimento da pesquisa.

Ao meu namorado, Thiago Fernando, pela companhia e apoio ao longo de toda trajetória.

Aos amigos e colegas, que apesar da distância proporcionaram momentos de troca, companheirismo e de descontração, tornando o percurso mais leve.

Por fim, agradeço a Universidade Federal de São Carlos (UFSCar) pelas oportunidades, aprendizados e experiências inestimáveis.

“Suportar não significa sucumbir, mas resistir às incertezas e continuar. Para resistir às incertezas é preciso ter audácia. Repita-se: não confunda audácia com aventura. Audacioso ou audaciosa é aquele ou aquela que planeja, organiza, estrutura e vai.”

Mario Sergio Cortella

RESUMO

Introdução: Os mecanismos de busca da *Web* fornecem o acesso à multiplicidade de informações disponíveis, com resultados dinâmicos, precisos e visualmente enriquecidos. Em virtude dessas características, os usuários contemporâneos estão habituados à praticidade e dinamismo oferecidos por ambientes abertos e intuitivos como a *Web*. Tendo isso em vista, é necessário repensar o tratamento descritivo nos catálogos, pois, a arquitetura dos atuais *softwares* disponíveis para as bibliotecas e os protocolos de comunicação utilizados não permitem a interoperabilidade com os mecanismos de busca ou com outras fontes de informação externas ao domínio bibliográfico. Em contrapartida, a *Web* oferece um conjunto de tecnologias com enfoque na descoberta e na comunicação entre domínios distintos, uma dessas tecnologias é o vocabulário *Schema.org*. **Objetivo:** O objetivo da pesquisa consiste em analisar o *Schema.org* como uma alternativa para representação de recursos informacionais nos catálogos, identificando as principais características e possíveis benefícios e desafios da aderência do vocabulário no contexto das bibliotecas. **Procedimentos metodológicos:** Esta pesquisa é de natureza teórica, realizada a partir de uma abordagem qualitativa. Do ponto de vista dos objetivos, caracteriza-se como descritiva e em relação aos procedimentos técnicos, como bibliográfica e documental, visando explorar a intersecção entre as temáticas, catálogos digitais e *Schema.org*. **Resultados:** Identificou-se que o vocabulário apresenta limitações para representação em domínios específicos, como no caso das bibliotecas. Contudo, as diretrizes e a estrutura do *Schema.org* são flexíveis, permitindo o desenvolvimento de extensões, as classes, propriedades e relacionamentos podem ser adaptados para representação em contextos distintos. O *Schema.org* é compatível com formatos semânticos, recomendados pelo *World Wide Web Consortium (W3C)* para dados estruturados. Dessa maneira, o vocabulário viabiliza a descoberta de informações em *Rich Snippets* e a integração dos dados ao *Knowledge Graph*. Com essas tecnologias os resultados de busca são apresentados de modo diferenciado, enriquecidos com elementos adicionais. Identificou-se que no universo bibliográfico, o *Schema.org* foi incluído nos projetos da *Online Computer Library Center (OCLC)* e alinhado aos princípios *Linked Data* para estruturação e descoberta dos registros do *WorldCat*. **Conclusão:** Pondera-se que a aderência do *Schema.org* em conjunto com os instrumentos de representação desenvolvidos no âmbito das bibliotecas possa favorecer a exposição dos dados dos catálogos na *Web*. Desse modo, além de se beneficiarem de tecnologias alinhadas ao escopo da *Web Semântica*, as bibliotecas podem contribuir para o desenvolvimento da mesma, a partir do compartilhamento de informações estruturadas, relevantes e fidedignas.

Palavras-Chave: Catálogo digital. Catálogo de acesso público *online*. *Schema.org*.

ABSTRACT

Introduction: The web search engines provide access to the multitude of available information, with dynamic, accurate and visually enriched results. Due to these characteristics, contemporary users are used to the practicality and dynamism offered by open and intuitive environments such as the Web. It is necessary to rethink the descriptive treatment in the catalogs since the architecture of the current software available for libraries and the communication protocols used do not allow interoperability with search engines or with other sources of information external to the bibliographic domain. In contrast, the Web offers a set of technologies focused on discovery and communication between different domains, one of these technologies is the Schema.org vocabulary. **Objective:** The objective of the research is to analyze Schema.org as an alternative for representing informational resources in catalogs, identifying the main characteristics and possible benefits and challenges of vocabulary adherence in the context of libraries. **Methodological procedures:** This study is theoretical in nature, carried out from a qualitative approach. From the point of view of the objectives, it is characterized as descriptive and in relation to the technical procedures, as bibliographic and documentary, aiming to explore the intersection between the themes, digital catalogs and Schema.org. **Results:** It was identified that the vocabulary has limitations for representation in specific domains, as in the case of libraries. However, the guidelines and structure of Schema.org are flexible, allowing the development of extensions, the classes, properties, and relationships can be adapted for representation in different contexts. Schema.org vocabulary can be used with encodings recommended by the World Wide Web Consortium (W3C) for structured data. In this way, it is possible to discover information in Rich Snippets and integrate the data into the Knowledge Graph. With these technologies, search results are enriched with additional elements. It was identified that in the bibliographic universe, Schema.org was included in the Online Computer Library Center projects and aligned with Linked Data principles for structuring and discovering WorldCat records. **Conclusion:** It is considered that the adherence of Schema.org together with the representation instruments developed within the scope of libraries can favor the exposure of catalog data on the Web. Thus, in addition to benefiting from technologies aligned with the scope of the Semantic Web, libraries can contribute to its development by sharing structured, relevant, and reliable information.

Keywords: Digital catalog. Online Public Access Catalog. Schema.org.

LISTA DE FIGURAS

Figura 1 - Hierarquia principal <i>Schema.org</i>	51
Figura 2 – Hierarquia <i>DataTypes</i>	52
Figura 3 - Exemplo de codificação em RDFa para relacionamentos com <i>Role</i>	53
Figura 4 - <i>Role</i> para descrição da função do autor de um prefácio.....	54
Figura 5 – Descrição dos relacionamentos e da função do autor de um posfácio com <i>Role</i>	55
Figura 6 - Resultado enriquecido para <i>Nineteen Eighty-Four</i>	58
Figura 7 - Resultado enriquecido com a opção de empréstimo	59
Figura 8 - <i>Snippet</i> de avaliação para Histórias extraordinárias	60
Figura 9 - Exemplo de estrutura para entidade <i>Book</i> em um <i>feed</i>	61
Figura 10 – Descrição de <i>LibrarySystem</i> em um <i>feed</i>	62
Figura 11 – Informações sobre Edgar Allan Poe em um <i>Knowledge panel</i>	63
Figura 12 - Relacionamentos a partir de <i>CreativeWork</i>	71
Figura 13 - OCLC <i>Model of Works</i>	76
Figura 14 - Representação de Obra como Produto	78
Figura 15 – Resultado de busca no <i>Worldcat</i> para <i>Pride and Prejudice</i>	79
Figura 16 - Edições e formatos de <i>Pride and Prejudice</i> no <i>Worldcat</i>	80
Figura 17 - Código fonte de <i>Pride and Prejudice</i> no <i>Worldcat</i>	81
Figura 18 - Estrutura para <i>Le concerto</i> em JSON-LD	81
Figura 19 - Codificação em JSON-LD para <i>Le concerto</i>	82
Figura 20 - Registro <i>Worldcat</i> recuperado como resultado de busca	83
Figura 21 – Modelo baseado em <i>Schema.org</i> e modelo BIBFRAME	84

LISTA DE QUADROS

Quadro 1 - Plano de trabalho	15
Quadro 2 - Critérios para seleção de estudos sobre a temática <i>Schema.org</i>	16
Quadro 3 - Projetos de aplicação do <i>Schema.org</i> em contextos distintos	18
Quadro 4 – Abordagens de aplicação do <i>Schema.org</i> no universo bibliográfico	23
Quadro 5 - Tipos de metadados, propriedades e funções	37
Quadro 6 - Subclasses de <i>Thing</i>	50
Quadro 7 - Propriedades, tipos esperados e descrição para <i>Book</i>	51
Quadro 8 - Categorias gerais de resultados de busca	57
Quadro 9 - Propriedades de <i>CreativeWork</i> herdadas de <i>Thing</i>	67
Quadro 10 - Propriedades e tipos esperados de <i>CreativeWork</i>	68
Quadro 11 - Grupos FRBR	75

SUMÁRIO

1 INTRODUÇÃO	10
1.1 Objetivos	12
1.2 Justificativa	12
1.3 Procedimentos metodológicos	14
1.4 Plano de trabalho	15
1.5 Estrutura da dissertação	17
2 O SCHEMA.ORG NA REVISÃO DE LITERATURA	18
3 A REPRESENTAÇÃO DE RECURSOS INFORMACIONAIS NO CONTEXTO DAS BIBLIOTECAS	27
4 REPRESENTAÇÃO E METADADOS EM AMBIENTES DIGITAIS.....	36
5 SCHEMA.ORG	46
5.1 Estrutura e organização do <i>Schema.org</i>	49
5.2 Resultados enriquecidos	56
6 A REPRESENTAÇÃO A PARTIR DA CLASSE <i>CREATIVEWORK</i>	65
7 O SCHEMA.ORG NO CONTEXTO DAS BIBLIOTECAS.....	74
8 CONSIDERAÇÕES FINAIS	89
REFERÊNCIAS	92

1 INTRODUÇÃO

Nas bibliotecas, os recursos informacionais são disponibilizados aos usuários por intermédio dos catálogos digitais ou *Online Public Access Catalogs* (OPACs). O tratamento descritivo dos recursos é realizado a partir das técnicas e processos de catalogação que contemplam instrumentos de representação, como códigos, modelos e padrões reconhecidos internacionalmente. Desse modo, a catalogação permite a descrição formal e padronizada dos recursos, visando garantir a qualidade dos dados bibliográficos, dados de autoridade e outros disponíveis nos catálogos. Devido ao tratamento minucioso e sistematizado desses dados, os catálogos são caracterizados como ambientes informacionais confiáveis, pois, fornecem o acesso a informações fidedignas.

O constante avanço das Tecnologias de Informação e Comunicação (TIC) e, sobretudo, o advento da *Web*, ampliaram consideravelmente as possibilidades de compartilhamento, busca, recuperação e acesso à informação. Na *Web*, os mecanismos de busca propiciam aos usuários a obtenção de respostas instantâneas e o acesso à multiplicidade de informações disponíveis.

A praticidade e o dinamismo da *Web* resultaram em alterações no comportamento e nas necessidades informacionais dos usuários, elevando as expectativas em relação aos ambientes digitais. Para Barreto (2002) as TIC modificaram aspectos fundamentais no processo comunicativo, favorecendo a troca de informações. Segundo o autor, “Essas tecnologias intensas modificaram radicalmente a qualificação de tempo e espaço entre as relações do emissor, os estoques e os receptores da informação.” (BARRETO, 2002, p. 73).

Diante dessa realidade, as bibliotecas deparam-se com novos desafios em relação à adequação dos catálogos às necessidades informacionais dos usuários contemporâneos. Os instrumentos de representação que norteiam a estruturação dos dados nos catálogos não permitem a interoperabilidade com os mecanismos de busca ou com outros ambientes externos ao âmbito das bibliotecas, a exemplo do Código de Catalogação Anglo-Americano (AACR2r) e o formato MARC 21. Em vista disso, os catálogos são ambientes restritos, limitados aos sistemas compatíveis.

No contexto atual, os mecanismos de busca são o ponto de partida para recuperação de informações na *Web*. Tendo isso em vista, para que o conteúdo das bibliotecas seja localizado pelos usuários, é necessário repensar a estruturação dos

dados nos catálogos, a partir do alinhamento dos instrumentos de representação com tecnologias que favoreçam o compartilhamento e a interpretação dos mecanismos de busca.

No âmbito da Ciência da Informação, os estudos acerca da representação da informação têm como base o uso estratégico das TIC, visando o aperfeiçoamento do tratamento descritivo para o atendimento das necessidades informacionais emergentes. Nesse aspecto, as melhorias e adequações dos catálogos ao contexto atual são encontradas no escopo das tecnologias semânticas desenvolvidas para atender a proposta da *Web Semântica*. A *Web Semântica* propõe a atribuição de significado aos dados a partir de um conjunto de tecnologias que viabilizam a interpretação eficiente dos agentes de *software*, de modo a promover melhores experiências em relação à busca, acesso e recuperação de informações.

Nesse contexto, as principais empresas de mecanismos de busca desenvolveram o *Schema.org*, um vocabulário genérico e extensível para representação de informações na *Web*. A iniciativa *Schema.org* propõe criar, manter e promover esquemas para dados estruturados, visando favorecer a interpretação dos mecanismos de busca e assim propiciar melhores experiências aos usuários.

O termo vocabulário é compreendido nesta pesquisa conforme a definição do World Wide Web Consortium (2015d, não paginado, tradução nossa) “Vocabulários são utilizados para classificar termos que podem ser empregados em uma aplicação específica, caracterizando possíveis relacionamentos, e definindo possíveis restrições sobre o uso desses termos [...]”. Os vocabulários consistem em “[...] blocos de construção básicos para técnicas de inferência na *Web Semântica*.” (WORLD WIDE WEB CONSORTIUM, 2015d, não paginado, tradução nossa).

Nessa perspectiva, o objetivo da pesquisa consistiu em analisar o vocabulário *Schema.org* como uma alternativa para representação de recursos informacionais nos catálogos digitais. O desenvolvimento da pesquisa foi norteado a partir do seguinte questionamento: como o *Schema.org* pode contribuir para adequação dos catálogos digitais?

Desse modo, buscou-se apresentar as principais características e possíveis benefícios e desafios da aderência do *Schema.org* no contexto das bibliotecas.

1.1 Objetivos

O objetivo geral da pesquisa, consistiu em analisar o vocabulário *Schema.org* como uma alternativa para representação de recursos informacionais nos catálogos digitais. Nessa perspectiva, foram definidos os seguintes objetivos específicos:

- Identificar os principais aspectos da representação de recursos informacionais no contexto das bibliotecas nos catálogos digitais;
- Descrever o *Schema.org* a partir da estrutura e metadados do vocabulário para representação;
- Investigar a aplicação do *Schema.org* para estruturação dos dados dos catálogos digitais;
- Apresentar os possíveis benefícios e desafios da aderência do *Schema.org* no contexto das bibliotecas.

1.2 Justificativa

A predominância da *Web* como principal fonte de informação na sociedade contemporânea, resultou no aumento expressivo de dados, gerando novos desafios para recuperação e identificação de informações relevantes e fidedignas.

Diante dessa realidade, os catálogos digitais destacam-se como fontes confiáveis, cuidadosamente estruturadas pelas instituições visando garantir a recuperação e o acesso a informações de qualidade. Entretanto, os instrumentos de representação que norteiam o tratamento descritivo nos atuais catálogos, não foram projetados para os ambientes *online*. Por essa razão, os dados dos catálogos não são compartilhados na *Web*, reduzindo consideravelmente o alcance do conteúdo das bibliotecas.

Nesse contexto, as áreas de Biblioteconomia e Ciência da Informação (CI) concentram-se na análise de alternativas e no desenvolvimento de estratégias para revitalização dos catálogos, com o propósito de aperfeiçoar o tratamento descritivo de recursos informacionais em consonância com as tecnologias emergentes.

Desse modo, o tema de pesquisa, justifica-se, a partir dos possíveis benefícios da aderência do *Schema.org* nos catálogos digitais. O *Schema.org* é apresentado pela literatura como um vocabulário amplo, pragmático e extensível, alinhado ao escopo

da *Web Semântica*. A aplicação do *Schema.org* na descrição de recursos informacionais, atribui valor semântico aos dados, favorecendo o processamento e a interpretação dos mecanismos de busca da *Web*.

Tendo em vista a necessidade de adequação dos catálogos ao cenário atual, considera-se o *Schema.org* como uma alternativa para o enriquecimento semântico dos dados bibliográficos e para o compartilhamento dos dados na *Web*. A aderência do vocabulário pelas bibliotecas poderá beneficiar à sociedade em relação à disseminação e ao acesso a informações relevantes e confiáveis, promovendo o melhor aproveitamento dos recursos disponíveis. Além disso, a exposição dos dados dos catálogos na *Web* poderá favorecer a descoberta de informações, atraindo usuários potenciais.

Ao iniciar pelo embasamento teórico da proposta, observou-se que não há um delineamento teórico ou pragmático de publicações sobre essa temática. O *Schema.org* ainda é pouco explorado pela Ciência da Informação, principalmente em âmbito nacional. Sendo assim, no aspecto científico, a pesquisa pode proporcionar uma nova perspectiva para área em relação à representação de recursos informacionais.

Ademais, a temática da pesquisa está em conformidade com a linha de pesquisa 2, intitulada “Tecnologia, Informação e Representação” do Programa de Pós-Graduação em Ciência da Informação da Universidade Federal de São Carlos (UFSCar). A linha 2 defende a realização de estudos que promovam o uso estratégico das TIC para representação, tendo em vista que “[...] os processos de representação documental consagram-se como elo fundamental para a garantia de recuperação da informação.” (PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO, 2021, não paginado).

Nessa perspectiva, buscou-se repensar o tratamento descritivo nos catálogos digitais, a partir do aprofundamento sobre o *Schema.org*, visando identificar os possíveis benefícios do vocabulário para a adequação e compartilhamento dos dados dos catálogos.

1.3 Procedimentos metodológicos

Esta pesquisa é de natureza teórica sendo realizada a partir de uma abordagem qualitativa. Em relação aos objetivos, é caracterizada como descritiva, e quanto aos procedimentos técnicos, como pesquisa bibliográfica e documental, com o propósito de explorar a intersecção entre as temáticas: catálogos digitais e *Schema.org*.

A pesquisa descritiva consiste na observação, registro, análise e ordenação dos dados sem que o pesquisador interfira nos mesmos (PRODANOV; FREITAS, 2013). Segundo Gil (2008, p. 28) “As pesquisas descritivas são, juntamente com as exploratórias, as que habitualmente realizam os pesquisadores sociais preocupados com a atuação prática.”

Do ponto de vista dos procedimentos técnicos, a pesquisa bibliográfica é elaborada, a partir de documentos publicados, como livros, revistas, publicações em periódicos e artigos científicos, jornais, boletins, monografias, dissertações, teses, material cartográfico e *internet*, com o intuito de possibilitar que o pesquisador tenha contato com todo material já escrito sobre o assunto pesquisado (PRODANOV; FREITAS, 2013). Gil (2008, p. 50) destaca que a pesquisa bibliográfica apresenta como principal vantagem “[...] o fato de permitir ao investigador a cobertura de uma gama de fenômenos muito mais ampla do que aquela que poderia pesquisar diretamente.”

Ainda do ponto de vista dos procedimentos técnicos, a pesquisa pode ser caracterizada como documental. Segundo Gil (2008) a única diferença entre a pesquisa bibliográfica e a documental, está na natureza das fontes. Enquanto a pesquisa bibliográfica é pautada principalmente nas contribuições de vários autores sobre um tema específico, a pesquisa documental baseia-se em documentos que não receberam ainda um tratamento analítico. Segundo Prodanov e Freitas (2013) esse tipo de pesquisa é empregada para organizar informações que se encontram dispersas, valorizando-as como importantes fontes de consulta, sendo constituída por documentos oficiais, reportagens de jornal, cartas, contratos, diários, filmes, fotografias, gravações, etc. Nesse viés, além de fontes bibliográficas, foram consultadas fontes oficiais do *Schema.org*, *Google* e *World Wide Web Consortium* (W3C).

Para identificar e selecionar documentos relevantes para pesquisa, foi realizada uma revisão de literatura sobre as temáticas: catálogos digitais e *Schema.org*.

Segundo Grant e Booth (2009) a revisão de literatura consiste na identificação e seleção de documentos potenciais para inclusão, exigindo ou não uma pesquisa formal na literatura, dessa maneira, os documentos podem ser sintetizados em forma textual, tabular ou gráfica para análise de sua contribuição, ou valor.

O desenvolvimento da pesquisa foi norteado a partir da observação sistemática, onde “[...] o pesquisador, antes da coleta de dados, elabora um plano específico para a organização e o registro das informações.” (PRODANOV; FREITAS, 2013, p. 104). Desse modo, foram estabelecidas cinco etapas para realização da pesquisa, descritas no plano de trabalho na próxima subseção.

1.4 Plano de trabalho

Com base nos objetivos e nos procedimentos metodológicos, foi elaborado um plano de trabalho para execução da pesquisa, o Quadro 1 apresenta as etapas e atividades estabelecidas.

Quadro 1 - Plano de trabalho

Etapas	Atividades
1ª etapa: Revisão de literatura e seleção dos documentos.	Identificação da literatura, visando o embasamento teórico sobre as temáticas selecionadas para pesquisa, a partir de estudos publicados em nível nacional e internacional.
2ª etapa: Leitura e interpretação dos documentos selecionados.	Leitura, interpretação e análise das informações, com o objetivo de discutir sobre os posicionamentos identificados na literatura.
3ª etapa: Análise e estabelecimento das principais características extraídas da literatura.	Análise das principais características extraídas da literatura para elucidar o problema de pesquisa.
4ª etapa: Sistematização do estudo exploratório.	Identificação e estabelecimento da conexão entre as temáticas tratadas.
5ª etapa: Elaboração da redação final e divulgação da pesquisa.	Elaboração do relatório final a partir da sistematização das informações para divulgação da pesquisa.

Fonte: Elaboração da autora.

Na 1ª etapa foi realizada a revisão de literatura em nível nacional e internacional, visando identificar e selecionar os documentos relevantes para o embasamento e direcionamento da pesquisa. As buscas foram realizadas via Portal de Periódicos Capes, nas bases de dados: Base de Dados Referenciais de Artigos de Periódicos em Ciência da Informação (BRAPCI), Biblioteca Digital Brasileira de Teses e Dissertações (BDTD), Google Acadêmico, *Information Science & Technology Abstracts* (ISTA), *Library, Information Science & Technology Abstracts* (LISTA), *Scientific Electronic Library Online* (SciELO), *Scopus* e *Web of Science* (WoS).

Como forma de abranger os principais assuntos tratados na pesquisa e identificar a conexão entre os mesmos, foram selecionados os seguintes termos: *Schema.org*, catálogo digital, catálogo bibliográfico, catálogo de acesso público online, biblioteca, dados bibliográficos e suas respectivas traduções para o inglês: *digital catalog*, *bibliographic catalog*, *online public access catalog*, *library* e *bibliographic data* e para o espanhol: *catálogo de acceso público en línea* e *datos bibliográficos*.

No que tange a temática sobre os catálogos digitais, as buscas foram realizadas a partir dos idiomas determinados e sem recorte temporal, visando recuperar tanto estudos publicados recentemente, como estudos clássicos sobre os procedimentos de catalogação.

Para garantir a recuperação precisa de estudos sobre a temática “*Schema.org*”, determinou-se o período de 2011 a 2021, pois o ano inicial corresponde ao lançamento oficial do vocabulário e o ano final ao período estabelecido para execução da pesquisa. A relevância dos estudos recuperados foi determinada a partir dos critérios de inclusão e exclusão, descritos no Quadro 2.

Quadro 2 - Critérios para seleção de estudos sobre a temática *Schema.org*

Critérios para inclusão	Critérios para exclusão
Estudos que abordam a estrutura do vocabulário, organização e diretrizes para descrição.	Estudos anteriores ao período estabelecido.
Estudos que apontam possíveis vantagens e desvantagens do <i>Schema.org</i> e/ou apresentam os impactos de implementação.	Estudos que não correspondem aos idiomas definidos na estratégia de busca.
Estudos que apresentam exemplos sobre a codificação dos dados com <i>Schema.org</i> .	Estudos indisponíveis para acesso completo de modo gratuito na <i>Web</i> , ou via Portal de Periódicos Capes.
Estudos que apontam lacunas e inconsistências no vocabulário e trazem propostas de extensão.	Estudos recuperados por apenas incluírem o termo “ <i>schema.org</i> .”

Fonte: Elaboração da autora.

A 2ª etapa consistiu na leitura e interpretação dos estudos, permitindo atingir o primeiro e segundo objetivos específicos. Na 3ª etapa os principais aspectos identificados foram analisados, permitindo elucidar o problema de pesquisa. Na 4ª etapa, identificou-se a conexão entre as temáticas, oportunizando a discussão das características fundamentais extraídas da literatura, de modo a contemplar o terceiro e o quarto objetivos específicos. Na 5ª etapa as informações foram sistematizadas para elaboração do relatório final e divulgação da pesquisa à comunidade acadêmica.

1.5 Estrutura da dissertação

Compõem esta dissertação a seção de Introdução e respectivas subseções que incluem: objetivos, justificativa, procedimentos metodológicos, plano de trabalho e estrutura da dissertação, assim como as seguintes seções:

A **Seção 2** – O *SCHEMA.ORG* NA REVISÃO DE LITERATURA descreve sinteticamente alguns dos projetos de aplicação do *Schema.org* identificados na revisão de literatura.

A **Seção 3** – A REPRESENTAÇÃO DE RECURSOS INFORMACIONAIS NO CONTEXTO DAS BIBLIOTECAS versa sobre os principais aspectos da representação de recursos informacionais nos catálogos digitais, abordando os códigos, princípios, modelos e padrões que compreendem o tratamento descritivo no contexto das bibliotecas.

A **Seção 4** – REPRESENTAÇÃO E METADADOS EM AMBIENTES DIGITAIS aborda a relevância e a função dos metadados no processo de representação em ambientes digitais.

A **Seção 5** – *SCHEMA.ORG* discorre sobre as principais características do vocabulário, como a estrutura, organização e diretrizes para descrição de recursos informacionais. Aborda ainda os benefícios da estruturação dos dados conforme as diretrizes do vocabulário.

A **Seção 6** – A REPRESENTAÇÃO A PARTIR DA CLASSE *CREATIVEWORK* apresenta a classe *CreativeWork* do *Schema.org* com enfoque na descrição de recursos bibliográficos.

A **Seção 7** – O *SCHEMA.ORG* NO CONTEXTO DAS BIBLIOTECAS discorre sobre a aderência do *Schema.org* para representação do conteúdo dos catálogos digitais.

A **Seção 8** – CONSIDERAÇÕES FINAIS aponta os principais aspectos tratados ao longo do texto, apresentando observações e reflexões acerca dos resultados da pesquisa.

2 O SCHEMA.ORG NA REVISÃO DE LITERATURA

O *Schema.org* é um vocabulário genérico, desenvolvido para descrição de recursos heterogêneos, segundo Wallis (2022, p. 40, tradução nossa) o vocabulário é significativamente abrangente, “[...] na amplitude do mundo, a maioria das coisas possui um tipo em *Schema.org*, que pode ser utilizado.” Na revisão de literatura foram identificados projetos de aplicação do vocabulário em domínios distintos. Tendo em vista o enfoque da presente pesquisa nessa temática, considerou-se relevante apresentar a pluralidade de abordagens identificadas.

Os estudos foram selecionados e apresentados sinteticamente com base nos critérios de inclusão estabelecidos na seção 1.4, no Quadro 2 e organizados por semelhança, nas categorias gerais: conjuntos de dados de pesquisa, *e-commerce*, *internet of things* (IoT), redes sociais, repositórios e turismo. No Quadro 3 é possível visualizar os autores, ano, título e descrição de cada estudo nas categorias estabelecidas.

Cabe ressaltar que o Quadro 3 não expõe a totalidade de publicações que se enquadram nas categorias, o objetivo consistiu em apresentar e discutir sobre algumas das possibilidades de aplicação do *Schema.org*, demonstrando a flexibilidade do mesmo para representação em contextos distintos.

Quadro 3 - Projetos de aplicação do *Schema.org* em contextos distintos

Categoria	Autor	Título	Descrição
CONJUNTOS DE DADOS DE PESQUISA	ALRASHED <i>et al.</i> (2021)	<i>Dataset or not? A study on the veracity of semantic markup for dataset pages.</i>	Analisa a eficiência do <i>Schema.org</i> para descrição de conjuntos de dados de pesquisa com enfoque no <i>Dataset Search da Google</i> . Identifica problemas em relação a veracidade das marcações.
	OUCHI; SIMIONATO (2018)	Descrição de conjuntos de dados na <i>Web</i> com <i>Schema.org</i> .	Analisa o <i>Schema.org</i> como uma alternativa para tornar acessíveis os conjuntos de dados de pesquisa na <i>Web</i> .
	SILVA (2020)	Publicação de Dados de Pesquisa Científica: proposta de estruturação semântica de cadernos abertos de pesquisa frente às dimensões da <i>e-Science</i> .	Analisa os vocabulários <i>Schema.org</i> , <i>DC Terms</i> , <i>Simple Knowledge Organization System</i> (SKOS) e <i>RDA Element sets</i> para descrição dos conjuntos de dados de cadernos de pesquisa.
E-COMMERCE	HEPP (2015)	<i>The Web of Data for e-commerce: Schema.org and GoodRelations for researchers and practitioners.</i>	Analisa e fornece diretrizes para aplicação do <i>Schema.org</i> no setor de <i>e-commerce</i> a partir do modelo <i>GoodRelations</i> .

	KEJRIWAL <i>et al.</i> (2021)	<i>Empirical best practices on using product-specific Schema.org.</i>	Apresenta o projeto <i>Web Data Commons</i> (WDC). Realiza um estudo empírico sobre o <i>Schema.org</i> para descrição dos produtos de WDC.
	PEETERS <i>et al.</i> (2020)	<i>Using Schema.org annotations for training and maintaining product matchers.</i>	Aborda a aplicação do <i>Schema.org</i> para descrição de produtos no setor de <i>e-commerce</i> e apresenta estratégias para manutenção e atualização dos dados dos produtos.
INTERNET OF THINGS (IoT)	NOURA <i>et al.</i> (2019)	<i>Automatic knowledge extraction to build Semantic Web of things applications.</i>	Identifica as principais aplicações tratadas no contexto de IoT. Aborda as extensões em <i>Schema.org</i> para IoT nas aplicações: <i>smart home, smart city e smart weather.</i>
	THULUVA; ANICIC; RUDOLPH (2018)	<i>IoT semantic interoperability with device description shapes.</i>	Propõe aprimoramentos nas extensões em <i>Schema.org</i> para IoT.
	THULUVA <i>et al.</i> (2020)	<i>Semantic Node-RED for rapid development of interoperable industrial IoT applications.</i>	Propõe uma extensão baseada em <i>iot.Schema.org</i> para ferramenta <i>Node-RED.</i>
REDES SOCIAIS	GARCÍA; GIL (2019)	<i>Social media copyright management using Semantic Web and Blockchain.</i>	Analisa o <i>Schema.org</i> para adaptação da <i>Copyright Ontology</i> , uma ontologia para gerenciamento de direitos autorais. A análise é realizada no contexto das redes sociais para fins jornalísticos.
	TOMOYSE; SANTOS; SIMIONATO (2019)	<i>Schema.org para recuperação da informação em redes sociais.</i>	Analisa o <i>Schema.org</i> para descrição de recursos informacionais em redes sociais. Os benefícios do <i>Schema.org</i> são identificados a partir da aplicação na rede social <i>Pinterest.</i>
	JONES <i>et al.</i> (2021)	<i>It's All About The Cards: Sharing on Social Media Probably Encouraged HTML Metadata Growth.</i>	Avalia o uso de metadados em artigos de notícias em relação aos formatos adotados em redes sociais, como <i>Open Graph Protocol, Standard HTML, Twitter Cards, Schema.org e Facebook Tracking.</i>
REPOSITÓRIOS	MIXTER; OBRIEN; ARLITSCH (2014)	<i>Describing theses and dissertations using Schema.org.</i>	Trata do desenvolvimento de um vocabulário de extensão baseado em <i>Schema.org</i> para descrição de teses e dissertações.
	NEVADO-CHINÉ; ALCARAZ-MARTÍNEZ; NAVALÓN (2021)	<i>Análisis de la implementación Schema.org en el repositorio RODERIC e impacto en el posicionamiento en Google y Google Scholar</i>	Discorre sobre a implementação do <i>Schema.org</i> no repositório RODERIC da Universidade de Valencia. Analisa-se o impacto da implementação a partir do número de acessos e <i>downloads</i> dos recursos.

	ROA-MARTÍNEZ; VIDOTTI; PASTOR- SÁNCHEZ (2018)	<i>Marcado semántico enriquecido para programas de posgrado en Latinoamérica.</i>	Trata do <i>Schema.org</i> para atribuição de significado aos metadados de programas de pós-graduação na América Latina.
TURISMO	KARLE <i>et al.</i> (2017)	<i>Extending the Schema.org vocabulary for more expressive accommodation annotations</i>	Versa sobre o desenvolvimento de extensões para o <i>Schema.org</i> , visando aperfeiçoar a descrição na área de turismo, no setor de acomodações (hotéis).
	LOHVYNENKO; NEDBAI (2019)	<i>Usage of Semantic Web in Austrian regional tourism organizations.</i>	Analisa a aplicação de formatos da <i>Web Semântica</i> e ontologias em <i>sites</i> austríacos da área de turismo. Identifica que esses <i>sites</i> utilizam os formatos recomendados pelos mecanismos de busca, tais como: <i>Schema.org</i> , Microdados e JSON-LD.
	SIMSEK; KARLE; FENSEL (2018)	<i>Machine Readable Web APIs with Schema.org action annotations.</i>	Analisa a aplicação do <i>Schema.org</i> em APIs da <i>Web</i> para serviços de turismo, como a realização de reservas em hotéis ou compra de passagens aéreas.

Fonte: Elaboração da autora.

Há estudos que apontam as lacunas e deficiências identificadas no *Schema.org* e propõem modelos, extensões específicas para as classes, propriedades e relacionamentos do vocabulário, assim como ferramentas para verificar a qualidade dos dados estruturados. Dentre os estudos selecionados, Alrashed *et al.* (2021) na categoria conjuntos de dados e dados de pesquisa, apontam a existência de páginas na *Web* codificadas em *Schema.org* de maneira equivocada. Em vista disso, é proposto um modelo de classificação para verificação da confiabilidade das páginas que utilizam a marcação *Schema.org/Dataset*.

Na categoria *e-commerce*, Hepp (2015) versa sobre a ontologia *GoodRelations*, integrada oficialmente ao *Schema.org*. O autor apresenta um tutorial para aplicação dos esquemas da *GoodRelations* e *Schema.org* para descrição no domínio de *e-commerce*. Na mesma categoria, Peeters *et al.* (2020) discorrem sobre a manutenção e atualização dos dados de *e-commerce*, a partir dos metadados provenientes do *Schema.org*, os autores apresentam estratégias para o aprimoramento de redes neurais para otimizar o processamento desses dados. Ainda em *e-commerce*, Kejriwal *et al.* (2021) discorrem sobre a elaboração de um conjunto de melhores práticas para descrição de produtos no setor comercial utilizando o *Schema.org*.

Na categoria IoT, Noura *et al.* (2019) apresentam o desenvolvimento de extensões em *Schema.org* com enfoque nas tecnologias: *smart home*, *smart city* e

smarth weather. Ainda no contexto de IoT, Thuluva, Anicic e Rudolph (2018) propõem extensões para ‘*iot.Schema.org*’ a partir do formato RDF (RDF *Shapes*) visando a interoperabilidade semântica entre dispositivos. Thuluva *et al.* (2020) abordam a extensão *iot.Schema.org* para a ferramenta *Node-Red* utilizada para aplicações complexas em IoT.

Na categoria redes sociais, García e Gil (2019) apresentam o desenvolvimento de um modelo baseado em *Schema.org* para representação dos principais conceitos de direitos autorais, a partir da adaptação da *Copyright Ontology* no contexto de mídias sociais para fins jornalísticos.

Mixer, O'Brien e Arlitsch (2014) abordam o desenvolvimento de extensões para descrição de teses e dissertações baseadas nos metadados do *Schema.org*, na categoria repositórios. Nesse estudo, as extensões foram testadas no repositório institucional da *Montana State University ScholarWorks*, segundo os autores, os metadados do *Schema.org* foram complementados com vocabulários de domínios específicos, permitindo a descrição detalhada de teses e dissertações. Na mesma categoria, Roa-Martínez, Vidotti e Pastor-Sánchez (2018) abordam o *Schema.org* para atribuição de significado ao conteúdo dos programas de pós-graduação das universidades da América Latina. A partir das diretrizes do *Schema.org*, os autores propõem a criação de uma nova entidade denominada “*ProgramaPosgrado*”.

Na categoria turismo, Karle *et al.* (2017) abordam o desenvolvimento de extensões em *Schema.org* para o aprimoramento da descrição de acomodações em *sites* de hotéis, visando tornar esses dados mais expressivos e legíveis por máquina. Na mesma categoria Simsek, Karle e Fensel (2018) propõem um método baseado em regras pré-definidas para validar a consistência das anotações do *Schema.org* em conjuntos de dados de domínios específicos, nessa abordagem foram utilizados dados do domínio de turismo.

Ainda há estudos que discorrem sobre as características e possíveis benefícios e impactos da adoção do *Schema.org*, ou abordam o mapeamento do mesmo com vocabulários especializados, ou analisam o aumento da utilização no decorrer dos anos.

Dentre esses estudos, na categoria conjuntos de dados de pesquisa, Ouchi e Simionato (2018) discorrem sobre as características do *Schema.org* e analisam exemplos de aplicação para publicação de dados de pesquisa. Ainda nessa categoria, Silva (2020) analisa as correspondências entre os metadados do *Schema.org* e *DC*

Terms, SKOS e RDA *Element Sets*, visando a realização do mapeamento das propriedades dos vocabulários.

Na categoria redes sociais, Jones *et al.* (2021) avaliam o crescimento do uso de metadados para fins de publicidade nas redes sociais, compara-se a utilização dos formatos *Open Graph Protocol*, *Standard HTML*, *Twitter Cards*, *Schema.org* e *Facebook Tracking*. Na mesma categoria, Tomoyose, Santos e Arakaki (2019) discorrem sobre as principais características e benefícios do *Schema.org* para redes sociais com enfoque no *Pinterest*.

Na categoria sobre repositórios, os autores Nevado-Chiné, Alcaraz-Martínez e Navalón (2021) analisam o impacto da aderência do *Schema.org* no repositório RODERIC para SEO em relação ao posicionamento, número de acessos e *downloads* dos recursos disponíveis.

Na categoria turismo, Lohvynenko e Nedbal (2019) analisam as ontologias utilizadas para estruturação dos dados na área de Turismo e identificam o *Schema.org* como ontologia mais utilizada neste domínio.

De modo geral, os estudos visam estruturar os dados conforme o escopo da *Web Semântica*, nesse sentido, o *Schema.org* é analisado como uma alternativa para atribuição de significado aos dados. Observou-se que independentemente do domínio, os estudos compartilham do objetivo de tornar os dados mais detectáveis na *Web* e otimizar a interpretação e a recuperação das informações via mecanismos de busca. Devido ao caráter genérico do *Schema.org*, em muitos casos é necessário o desenvolvimento de extensões e a realização de adaptações, visando atingir o nível de granularidade exigido em determinados contextos.

À medida que diferentes domínios aderem ao *Schema.org*, o vocabulário é ampliado e aperfeiçoado para o atendimento de realidades distintas, incluindo o universo bibliográfico, que contempla centros de informação, como bibliotecas, arquivos e museus.

Nesse contexto, com base nos mesmos critérios de seleção explicitados no Quadro 2, foram selecionados e brevemente apresentados no Quadro 4 alguns dos estudos que abordam o *Schema.org* no universo bibliográfico.

Quadro 4 – Abordagens de aplicação do *Schema.org* no universo bibliográfico

Autor	Título	Descrição
AGENJO-BULLÓN; HERNÁNDEZ-CARRASCAL (2017)	<i>Avances de Bibframe en 2016: perspectivas del nuevo modelo bibliográfico.</i>	Analisa o progresso do <i>Bibliographic Framework Initiative</i> (BIBFRAME) e a transição entre o <i>Anglo-American Cataloguing Rules</i> (AACR2) e o <i>International Standard Bibliographic Description</i> (ISBD) para <i>Resource Description and Access</i> (RDA) e dos <i>Functional Requirements for Bibliographic Records</i> (FRBR) para <i>Library Reference Model</i> (LRM). Aborda a utilização do <i>Schema.org</i> pela <i>Online Computer Library Center</i> (OCLC) em relação a vinculação de dados bibliográficos.
ARAKAKI (2019)	Metadados administrativos e a proveniência dos dados: modelo baseado na família PROV.	Analisa a aplicação da ontologia PROV-O para a representação de registros bibliográficos em relação à proveniência. Apresenta o <i>crosswalk</i> entre o PROV-O, MARC 21, <i>Dublin Core</i> , PREMIS, BIBFRAME e <i>Schema.org</i> , com o intuito de verificar a compatibilidade dos formatos no que tange à proveniência.
CLARK; ROSSMANN (2017)	<i>The Open SESMO (Search Engine & Social Media Optimization) Project: Linked and Structured Data for Library Subscription Databases to Enable Webscale Discovery in Search Engines.</i>	Discorre sobre o projeto de banco de dados “Open SESMO” da <i>Montana State University</i> (MSU). Analisa o <i>Schema.org</i> e modelos <i>Linked Data</i> para otimização de bases de dados assinadas pelas bibliotecas.
FONS; PENKA; WALLIS (2012)	<i>OCLC’s Liked Data Initiative: using Schema.org to make library relevant on the Web.</i>	Versa sobre o progresso da OCLC em relação à implementação dos princípios <i>Linked Data</i> nos dados do <i>Worldcat</i> . Apresenta os benefícios da aplicação do <i>Schema.org</i> para exposição dos dados do <i>Worldcat</i> na <i>Web</i> .
FOX (2014)	<i>Digital libraries: the systems analysis perspective.</i>	Analisa a otimização do mecanismo de pesquisa (SEO) considerando os riscos e benefícios da exposição do conteúdo das bibliotecas, a partir da aplicação de microdados e <i>Schema.org</i> .
FREIRE; CHARLES; ISAAC (2018)	<i>Evaluation of Schema.org for aggregation of cultural heritage metadata.</i>	Analisa a aplicação do <i>Schema.org</i> no contexto das instituições de patrimônio cultural. Aborda o caso do <i>Europeana Data Model</i> para as coleções da <i>Europeana</i> .
FREIRE et al. (2018)	<i>Cultural heritage metadata aggregation using web technologies: IIIF, Sitemaps and Schema.org.</i>	Analisa tecnologias utilizadas na <i>Web</i> para instituições de patrimônio cultural baseadas no <i>International Image Interoperability Framework (IIIF)</i> e <i>Sitemaps</i> . Aborda o <i>Schema.org</i> para agregação de metadados no contexto dessas instituições.
GARCÍA-MARCO (2013)	<i>Schema.org: la catalogación revisitada.</i>	Discute as possibilidades de inovação na catalogação, a partir da utilização do <i>Schema.org</i> para representação de recursos no contexto das bibliotecas.
GRACY; ZENG; SKIRVIN (2013)	<i>Exploring methods to improve access to music resources by aligning library data with Linked Data: a report of methodologies and preliminary findings.</i>	Apresenta um projeto de coleta, análise e mapeamento das propriedades utilizadas na descrição de gravações musicais em catálogos de bibliotecas e em outras

		instituições culturais. Analisa os registros de gravações musicais no <i>Worldcat</i> estruturados com extensões em <i>Schema.org</i> .
HAN <i>et al.</i> (2015)	<i>Exposing library holdings metadata in RDF using Schema.org semantics.</i>	Analisa os benefícios do <i>Schema.org</i> para exposição dos dados dos acervos. Expõe os problemas identificados nas extensões do vocabulário para os dados de acervos e sugere melhorias.
HYOUNGJOO; KIPP (2019)	<i>Library Linked Data Models: library data in the Semantic Web.</i>	Analisa esquemas <i>Linked Data</i> , ontologias e modelos de dados propostos ou utilizados no contexto das bibliotecas. Discute as implicações da adoção de formatos distintos para implementação de LD, o <i>Schema.org</i> é uma das tecnologias analisadas.
LAMPRON; MIXTER; HAN (2016)	<i>Challenges of mapping digital collections metadata to Schema.org: working with CONTENTdm.</i>	Apresenta os desafios para o mapeamento de metadados de coleções digitais em <i>Schema.org</i> e discute as experiências da Biblioteca da Universidade de Illinois em <i>Champaign-Urbana</i> e OCLC.
MICHEL <i>et al.</i> (2020)	<i>Unleash the potential of your website! 180,000 webpages from the French Natural History Museum marked up with Bioschemas/Schema.org biodiversity types.</i>	Trata dos esforços da comunidade de <i>Bioschemas</i> na criação de extensões em <i>Schema.org</i> para sites da área de Ciências da Vida. Aborda o caso do Museu Nacional de História Natural da França que teve seus dados estruturados com <i>Schema.org</i> .
NAM; KEJRIWAL (2018)	<i>How Do Organizations Publish Semantic Markup? Three Case Studies Using Public Schema.org Crawls.</i>	Analisa a aplicação do <i>Schema.org</i> nos dados dos setores: hospital, museu e escola. Os autores destacam as tendências e diferenças da estruturação dos dados nesses contextos.
PALETTA; MUCHERONI; SILVA (2016)	<i>Novas tecnologias de uso em nuvens aplicáveis às bibliotecas.</i>	Aborda o desenvolvimento de serviços em nuvens no contexto das bibliotecas. Discute a aplicação do <i>Schema.org</i> e outras tecnologias semânticas no <i>Worldcat</i> .
RAEMY (2020)	<i>Enabling better aggregation and discovery of cultural heritage content for Europeana and its partner institutions.</i>	Aborda a aplicação das tecnologias <i>Linked Open Data</i> no contexto da <i>Europeana</i> e apresenta soluções para o enriquecimento semântico e atualização precisa dos dados. Discute a aplicação do <i>Schema.org</i> em projetos realizados no contexto da <i>Europeana</i> .
SCOTT (2015)	<i>White Hat Search Engine Optimization (SEO): structured Web data for libraries.</i>	Apresenta instruções para estruturação dos dados de bibliotecas virtuais relacionados a localização, horário e informações de contato, com o objetivo de otimizar o processamento desses dados pelos mecanismos de busca. Fornece diretrizes para estruturação dos dados a partir da aplicação do <i>Schema.org</i> .
SUOMINEN; HYVÖNEN (2017)	<i>From MARC silos to Linked Data silos?</i>	Apresenta modelos de dados bibliográficos conectados em RDF e as ferramentas para conversão do formato MARC. Discute a abordagem realizada pela <i>National Library of Finland</i> com o

		<i>Schema.org</i> para representação de dados bibliográficos.
WALLIS <i>et al.</i> (2017)	<i>Recommendations for the application of Schema.org to aggregated Cultural Heritage metadata to increase relevance and visibility to search engines: the case of Europeana.</i>	Fornecer recomendações para aplicação do <i>Schema.org</i> e dos princípios <i>Linked Data</i> nas coleções da <i>Europeana</i> .

Fonte: Elaboração da autora.

Os documentos apresentados no Quadro 4 estabelecem a conexão entre as temáticas tratadas nesta pesquisa: “*Schema.org*” e “catálogos digitais” no universo bibliográfico. De modo geral, observou-se que a *Online Computer Library Center* (OCLC) e a *Europeana* são referências no que tange a implementação de tecnologias semânticas para representação no universo bibliográfico, incluindo a aderência ao *Schema.org*.

Discutem sobre as iniciativas da OCLC: Agenjo-Bullón e Hernández-Carrascal (2017), Arakaki (2019), Fons, Penka e Wallis (2012), Gracy, Zeng e Skirvin (2013) e Palleta, Mucheroni e Silva (2016), e sobre os projetos da *Europeana*: Freire, Charles e Isaac (2018); Freire *et al.* (2018), Raemy (2020) e Wallis *et al.* (2017).

Além dos estudos relacionados a OCLC e a *Europeana*, foram identificados estudos sobre a utilização do *Schema.org* no universo bibliográfico associado a técnicas de SEO, abordados por Clark e Rossmann (2017), Fox (2014) e Scott (2015). Assim como nos estudos apresentados no Quadro 3 sobre SEO, o objetivo consiste em melhorar o posicionamento dos resultados de busca e consequentemente aumentar o número de acessos ao conteúdo estruturado.

Foram identificados ainda, projetos de aplicação do vocabulário no contexto dos museus, como em Michel *et al.* (2020) em relação a estruturação dos dados do *French Natural History Museum* e Nam e Kejriwal (2018) que discorrem sobre três casos de aplicação do *Schema.org* nos domínios: hospital, museu e escola.

A partir dos estudos apresentados no Quadro 4, observou-se que o *Schema.org* é analisado no universo bibliográfico como uma tecnologia complementar aos instrumentos que regem a descrição bibliográfica, como: MARC 21, BIBFRAME, RDA e FRBR. Hyoungjoo e Kipp (2019) apresentam uma análise sobre a aplicação dos princípios *Linked Data* e formatos semânticos em um conjunto de bibliotecas que utilizam o formato MARC 21, os autores observaram que as bibliotecas adotam modelos e vocabulários diferentes para estruturação dos dados, como *Dublin Core* (DC), *Schema.org*, *Resource Description Framework* (RDF) e *MarcOnt*,

rdagroup1elements, alguns modelos são baseados em FRBR. Suominen e Hyvönen (2017), identificaram a incompatibilidade na variedade de formatos semânticos adotados no contexto das bibliotecas.

Destaca-se que as discussões realizadas ao longo da presente pesquisa são pautadas principalmente nos projetos da OCLC, devido à compatibilidade com as temáticas “*Schema.org*” e “catálogos digitais”, e a completude dos documentos oficiais disponibilizados pela cooperativa, como: “*Common ground: exploring compatibilities between the Linked Data Models of the Library of Congress and OCLC*” por Godby e Denenberg (2015); “*The relationship between BIBFRAME and OCLC’s Linked-Data Model of bibliographic description: A working paper*” por Godby (2013) e em “*Library Linked Data in the Cloud OCLC’s experiments with new models of resource description*” por Godby, Wang e Mixter (2015). Estes documentos descrevem o processo de desenvolvimento de modelos semânticos baseados em *Schema.org*, trazendo observações relevantes em relação à modernização da arquitetura dos dados nos catálogos.

3 A REPRESENTAÇÃO DE RECURSOS INFORMACIONAIS NO CONTEXTO DAS BIBLIOTECAS

Os registros do conhecimento produzidos ao longo do desenvolvimento da humanidade têm sido armazenados e disponibilizados em diferentes formatos, suportes e fontes de informação. Para que se viabilize a recuperação e o acesso aos registros, o processo de representação é fundamental. No universo bibliográfico, a representação documental, também denominada catalogação propicia a descrição formal e padronizada das principais características de um recurso informacional, permitindo a identificação e a compreensão do recurso, sem necessariamente consultá-lo na íntegra. Catarino e Souza (2012, p. 84) definem a catalogação como

[...] o processo por meio do qual se descreve formalmente um documento ou recurso e se estabelece um variado e variável número de pontos de acesso, objetivando proporcionar, ao usuário final, a condição de encontrar, identificar, selecionar e obter o documento ou o recurso descrito, ou a informação nele contida.

Desse modo, a partir de elementos descritivos e codificados, a catalogação utiliza códigos, modelos e padrões reconhecidos internacionalmente, para estabelecer pontos de acesso que viabilizam a localização, identificação e a recuperação dos recursos informacionais como registros bibliográficos.

O universo bibliográfico é compreendido como o universo ou escopo relacionado às coleções das bibliotecas, incluindo outras comunidades de informação, como arquivos e museus (INTERNATIONAL FEDERATION OF LIBRARY ASSOCIATIONS, 2016). Nesse contexto, os instrumentos utilizados no processo de catalogação visam garantir a qualidade dos dados e a comunicação entre os registros e os sistemas, assim como entre as unidades de informação e seus usuários, seja em meio analógico ou digital.

O processo comunicativo é mediado pelos catálogos, um dos instrumentos mais antigos utilizados pelas bibliotecas. Ao longo da história da catalogação, os catálogos passaram por três fases: catálogos em livros, catálogos em fichas e catálogos digitais (ASSUMPÇÃO; SANTOS; ZAFALON, 2017).

Na primeira fase, os catálogos eram impressos em formato de livro e organizados por autor em ordem alfabética. Esse tipo de catálogo era elaborado em uma única impressão encadernada, para inclusão de novas obras de um autor ou para

registrar novas edições, reservava-se um espaço, caso as descrições ultrapassassem o limite, eram acrescentadas folhas soltas ou o catálogo era reimpresso (JESUS, 2021).

Com o aumento do volume de itens nos acervos, os catálogos em livro foram gradativamente substituídos por catálogos em fichas, permitindo registrar diferentes edições de uma obra em fichas individuais, organizadas alfabeticamente (ASSUMPÇÃO; SANTOS; ZAFALON, 2017). Os itens de um acervo eram descritos em “[...] fichas em papel, medindo 7,5 x 12,5 cm, armazenadas em gavetas e agrupadas basicamente por ordem alfabética, pelo nome de autores, títulos e assuntos.” (SERRA; SANTARÉM SEGUNDO, 2017, p. 170).

Nesse contexto, em meados do século XIX, Charles Ammi Cutter¹ definiu os objetivos do catálogo, descritos por Mey e Silveira (2009, p. 12)

1. Permitir a uma pessoa encontrar um livro do qual o autor, o título ou o assunto seja conhecido. 2. Mostrar o que a biblioteca possui de um autor determinado, de um assunto determinado e de um tipo determinado de literatura. 3. Ajudar na escolha de um livro de acordo com sua edição (bibliograficamente) e de acordo com o seu caráter (literário ou tópico).

Segundo Jesus (2021) esse período deu início a Biblioteconomia moderna, marcada pela criação de dados bibliográficos com base em instrumentos de padronização. O catálogo passou a ser visto sob uma nova perspectiva, assumindo o “[...] *status* de ambiente informacional criado para a busca e a estruturação do conhecimento produzido; e a Catalogação Descritiva passa a ser tratada como uma disciplina que vai além da técnica [...]” (JESUS, 2021, p. 35).

Com a evolução dos catálogos, houve a necessidade de estabelecer princípios internacionais de catalogação, visando a padronização do tratamento descritivo dos recursos. Essas questões foram discutidas na Conferência de Paris, em 1961, resultando na Declaração dos Princípios Internacionais de Catalogação ou Princípios de Paris, a última versão desse documento foi publicada em 2016.

A Declaração dos Princípios Internacionais de Catalogação objetiva normalizar os procedimentos de catalogação internacionalmente, a partir do desenvolvimento de códigos de catalogação, como o *Anglo-American Cataloguing Rules* (AACR), publicado em 1967. O AACR foi projetado para facilitar o intercâmbio de dados

¹ CUTTER, C. A. **Rules for a dictionary catalog**. 4. ed. Washington: Government Printing office, 1904.

bibliográficos e de autoridade, assim como estabelecer pontos de acesso para recuperação.

A primeira versão do AACR apresentava diferenças entre as normas britânicas e norte-americanas, com o intuito de unificar as normas, em 1978 foi lançada a segunda versão do código, o AACR2. Em 1988 o código foi revisado, atualizado e denominado AACR2, em 1998 o código sofreu novas alterações (MEY; SILVEIRA, 2009). A última versão foi publicada em 2002, onde foram incluídas instruções para descrição de recursos eletrônicos (SILVA *et al.*, 2017).

Desse modo, os avanços no âmbito da catalogação, direcionavam-se para informatização dos procedimentos descritivos, visando a adequação dos catálogos ao contexto tecnológico daquele período. Até a década de 1960, as bibliotecas utilizam os catálogos em fichas, o acesso à informação era limitado aos acervos físicos. Em 1970, os procedimentos de tratamento informacional passaram a ser realizados com o uso de computadores, dando origem aos catálogos digitais. Neste período, foi desenvolvido o *Machine-Readable Cataloging* (MARC), um formato de intercâmbio de dados, proposto pela *Library of Congress* (LC) para conversão dos registros impressos para registros digitais.

O formato MARC foi desenvolvido para “[...] acomodar os elementos presentes no AACR e, portanto, ser aderente aos princípios de descrição.” (SERRA *et al.*, 2018, p. 65). A aplicação do MARC para criação de registros deu origem a catalogação cooperativa, permitindo o intercâmbio de metadados entre as instituições. Segundo Assumpção e Santos (2015) a importância do MARC 21 é evidenciada, pois, o compartilhamento de registros padronizados contribuiu significativamente para redução de custos e retrabalhos. Ao final do século XX, o formato MARC foi atualizado, e como uma alusão ao século XXI, passou a ser denominado MARC 21.

Com surgimento dos sistemas *online*, ao final da década de 70, os catálogos digitais ficaram conhecidos como *Online Public Access Catalogs* (OPACs), em português Catálogos de Acesso Público *Online* (ASSUMPÇÃO; SANTOS; ZAFALON, 2017). Cabe destacar que os objetivos definidos por Cutter permanecem constituindo a base para estruturação dos OPACs, entretanto, passaram por reformulações, como na versão mais recente da Declaração dos Princípios Internacionais de Catalogação, publicada pela *International Federation of Library Associations* (IFLA) em 2016.

Conforme a Declaração, os catálogos devem permitir ao usuário: encontrar um único recurso ou um conjunto de recursos; identificar se a entidade descrita

corresponde ao recurso desejado, assim, deve ser possível distinguir entre recursos semelhantes; selecionar recursos que melhor satisfaçam as necessidades informacionais individuais, descartando aqueles que não correspondam as mesmas; adquirir ou obter, promovendo as condições necessárias para garantir o acesso ao recurso, a partir da compra, empréstimo, acesso *online*, entre outras possibilidades. Por fim, deve-se permitir ao usuário navegar e explorar a partir da organização lógica dos dados bibliográficos, dados de autoridade e dos relacionamentos entre entidades, em outros catálogos e fontes de informação (INTERNATIONAL FEDERATION OF LIBRARY ASSOCIATIONS, 2016).

Na Declaração dos Princípios Internacionais de Catalogação, os objetivos dos catálogos foram adaptados para os sistemas *online*. Observa-se que os últimos objetivos, navegar e explorar, destacam-se por apresentar uma nova perspectiva no que tange às possibilidades de navegação, com enfoque na comunicação entre os catálogos com outros catálogos e fontes externas, isto é, fora do domínio das bibliotecas. Nesse sentido, o princípio 2.10 da Declaração traz recomendações acerca da interoperabilidade dados dos catálogos, dentro e fora do contexto das bibliotecas.

2.10 Interoperabilidade. Deve-se fazer todos os esforços possíveis para assegurar o intercâmbio e a reutilização dos dados bibliográficos e de autoridade dentro e fora da comunidade bibliotecária. É extremamente recomendável o uso de vocabulários que facilitem a tradução automática e a desambiguação, para o intercâmbio de dados e ferramentas de descoberta (INTERNATIONAL FEDERATION OF LIBRARY ASSOCIATIONS, 2016, p. 6).

A disponibilização dos catálogos no formato *online*, permitiu a ampliação do alcance dos registros bibliográficos. Com o surgimento da *internet*, os registros migraram para o formato digital, enquanto os novos passaram a ser originados e disponibilizados em meio digital (MARCONDES, 2001). Assim, os catálogos que antes eram limitados a um público específico e a consulta ao acervo físico, passaram a ser acessados via *Web*, possibilitando que os usuários pudessem optar pela versão impressa ou digital dos recursos desejados (SOUZA; FUJITA, 2012).

A conversão dos catálogos para o meio digital e *online* representou um importante avanço para as bibliotecas, entretanto, com a evolução da tecnologia, os instrumentos de representação tornaram-se obsoletos. Segundo Díez (2012) apesar da evolução dos OPACs, as alterações realizadas não foram suficientes para suprir

as necessidades do século XXI, os serviços de busca oferecidos pelos catálogos são limitados se comparados com as ferramentas de recuperação oferecidas na *Web*.

Desde a década de 60, o MARC 21 ainda é o principal formato utilizado para nortear a estruturação dos catálogos. Conforme Serra *et al.* (2018, p. 65) o MARC foi projetado com enfoque nos elementos presentes no AACR2r, embora o formato viabilize a comunicação de elementos descritivos legíveis por máquinas “[...] sua concepção e tipo de funcionamento ainda possuem forte resquício analógico [...]”. Além disso, a lógica de criação de registros, tanto em relação à recuperação, quanto à importação e armazenamento baseiam-se no protocolo Z39.50 (ASSUMPÇÃO; SANTOS, 2015).

O Z39.50 é um protocolo de comunicação entre computadores, desenvolvido em um projeto entre a OCLC e a *Library of Congress* (LC) para recuperação de recursos informacionais. O protocolo define as regras para o intercâmbio de dados e viabiliza que uma “máquina cliente” pesquise em bancos de dados de uma “máquina servidor” para recuperar registros, que retornam como resultados de busca (SANTARÉM SEGUNDO; SILVA; MARTINS, 2018). O Z39.50 foi implementado para auxiliar na realização de buscas em um conjunto de catálogos, restringindo os resultados apenas aos bancos de dados compatíveis (WOODLEY, 2016). Dessa maneira, os dados são interoperáveis entre sistemas compatíveis com o protocolo, entretanto, não podem ser recuperados pelos mecanismos de busca da *Web*.

Em vista disso, apesar da reformulação dos propósitos dos catálogos, a estrutura fornecida para recuperação de informações no século XXI, permanece a mesma definida para o atendimento das necessidades dos usuários do século XX. Nesse sentido, Díez (2012, p. 184, tradução nossa) traz a seguinte reflexão:

[...] de certo modo a perspectiva dos bibliotecários permanece muito apegada aos princípios clássicos de organização da informação, enquanto as expectativas do usuário são claramente influenciadas pelas ferramentas e serviços da web que utilizam.

Com as tecnologias disponíveis, a criação, manipulação e gerenciamento de recursos informacionais na *Web* deixou de ser competência apenas dos profissionais da informação. A *Web* é caracterizada como um ambiente heterogêneo e dinâmico, com ferramentas sofisticadas e intuitivas que viabilizam o compartilhamento de informações em escala global. Diferentemente dos sistemas de informação tradicionais gerenciados por profissionais, na *Web* os próprios usuários podem

participar ativamente da criação, descrição, organização e disponibilização de conteúdos informacionais. Segundo Zeng e Qin (2016), a *Web* passou a ser utilizada como os catálogos de biblioteca, enciclopédias e jornais, tornando-se a principal fonte de informação para recuperação de informações.

Nesse contexto, o aumento expressivo de informações disponibilizadas na *Web* dificulta a identificação, seleção e a recuperação de informações relevantes e fidedignas. Se por um lado, o dinamismo característico da *Web* favoreceu a disseminação do conhecimento, por outro, tornou a recuperação de informações de qualidade cada vez mais complexa.

Gill (2016) discorre sobre a veracidade das informações publicadas na *Web*, destacando que publicações antiéticas podem ser criadas facilmente, como estratégia para atrair acessos. Gilliland (2016) ressalta que uma das desvantagens dos metadados criados pelos usuários está na ausência de controle de qualidade, impactando negativamente na confiabilidade e interoperabilidade entre os recursos e os metadados que os descrevem. Diante dessa realidade, Gill (2016) argumenta sobre a importância das bibliotecas, arquivos e museus como instituições reconhecidas por produzirem e publicarem metadados de qualidade na *Web*. Para Sousa e Fujita (2012, p. 68) “[...] os catálogos on-line tornaram-se instrumentos plurifuncionais, pelo fato das várias funções que desempenham e por ser uma ferramenta confiável de busca e recuperação da informação.”

Sendo assim, para que o conteúdo dos catálogos seja melhor aproveitado, é necessário que as diretrizes e padrões que norteiam os procedimentos de catalogação sejam repensados e atualizados, de modo que a representação oferecida ao usuário, satisfaça as novas necessidades informacionais. Contudo, alinhar os catálogos às exigências de um novo cenário informacional, é uma tarefa complexa que demanda uma série de alterações no que tange ao tratamento descritivo das informações. Como apontam Ramalho e Ouchi (2011, p. 61) “[...] os profissionais da informação se deparam com novos desafios na constante busca para atender as necessidades informacionais de uma nova geração de usuários que já nasceram imersos em ambientes computacionais.”

Para elaboração de registros bibliográficos nos catálogos, leva-se em consideração as particularidades da instituição e do público a que se destina, assim como, são estabelecidos critérios específicos para descrição de recursos distintos. Conforme Zafalon (2017) as atividades das instituições de patrimônio cultural são

pautadas na conveniência do usuário. No processo de catalogação considera-se o formato dos recursos, os padrões e métodos mais adequados ao contexto. Assim, exige-se do catalogador competências que vão além da técnica “[...] desde a modelagem do catálogo, a escolha do código de catalogação, a definição do nível de descrição, a composição da descrição, a definição dos atributos e dos relacionamentos.” (SANTOS, 2013, p. 5).

Desde o desenvolvimento dos primeiros catálogos, nas áreas de Biblioteconomia e Ciência da Informação é recorrente a elaboração de propostas para otimizar o tratamento descritivo de recursos informacionais em consonância com os códigos de catalogação e com as tecnologias de representação emergentes.

Segundo Buckland (1991) a Ciência da Informação (CI) lida com a informação como coisa, isto é, com o conhecimento registrado, portanto, passível de manipulação. A CI pode ser compreendida como um campo científico que se dedica à investigação dos problemas relacionados à comunicação dos registros do conhecimento (SARACEVIC,1995). Nesse sentido, a Ciência da Informação concentra-se no desenvolvimento de estratégias focadas no ciclo informacional, que inclui a produção, organização, armazenamento, representação, disseminação, recuperação, acesso e o uso da informação (NHACUONGUE; FERNEDA, 2015).

Conforme Araújo (2017, p. 24) os estudos da área são direcionados à investigação de diferentes realidades, como “[...] os fluxos de informação no planeta; os contextos organizacionais; a organização de documentos em bibliotecas, arquivos, bases de dados; a apropriação da informação no cotidiano [...]”. Em relação ao tratamento descritivo da informação, Araújo (2009, p. 198) destaca que o objetivo da Ciência da Informação consiste em identificar linguagens que sejam realmente úteis para representação

O conceito de informação que emana das várias pesquisas e aplicações relaciona-se essencialmente com a ideia de representação – da possibilidade de melhorar os processos representacionais, construindo linguagens melhores, notações mais mnemônicas, classes mais consistentes, terminologias menos ambíguas.

Os avanços tecnológicos exercem influência constante na mudança de perspectiva sobre o gerenciamento da informação. Desse modo, a CI busca encontrar nas tecnologias emergentes, alternativas para solucionar a questão da obsolescência dos catálogos.

No universo bibliográfico, destacam-se algumas iniciativas com enfoque no aperfeiçoamento do tratamento descritivo nos catálogos digitais, como o vocabulário de dados *Resource Description and Access* (RDA). O RDA foi desenvolvido para substituir o código de catalogação, AACR2r, propondo um conjunto de regras elaboradas para descrição de recursos em ambientes digitais. Silva *et al.* (2017) destaca ser possível aplicar essas regras em diversos esquemas de metadados, inclusive no formato MARC 21, entretanto, não há garantia de que esse formato atenderá as necessidades de representação no contexto da *Web*. A versão inicial do RDA foi baseada nas entidades conceituais do FRBR, na versão mais recente, foi adotado o modelo *Library Reference Model* (IFLA LRM).

O FRBR surge com a proposta de auxiliar a construção de regras de catalogação e a projeção de sistemas, baseado nas tarefas do usuário, as *user tasks*. Essas tarefas referem-se às possibilidades que um sistema de informação deve oferecer ao usuário, pautadas nos objetivos do catálogo, definidos por Cutter. Segundo Simionato (2017) os modelos conceituais servem de base para o desenvolvimento de padrões e esquemas de metadados, alterando a estrutura dos padrões utilizados no universo bibliográfico.

O FRBR foi publicado em 1998 e posteriormente expandido em mais dois modelos, o *Functional Requirements for Authority Data* (FRAD) e o *Functional Requirements for Subject Authority Data* (FRSAD). O FRAD define os requisitos funcionais para dados de autoridade que representam pontos de acesso controlados, utilizados pelas instituições para “[...] coletar obras de uma determinada pessoa, família, organização, edições corporativas ou múltiplas de um título.” (SIMIONATO, 2017, p. 217), enquanto o FRSAD trata das relações entre dados de autoridade referentes ao assunto dos recursos bibliográficos.

Esses modelos são estruturados conforme a modelagem entidade-relacionamento, entretanto, adotaram perspectivas diferentes para resolução de problemas relacionados aos registros bibliográficos. Em vista disso, tem sido discutida a implementação do IFLA LRM, cuja proposta consiste em harmonizar e unificar o FRBR, FRAD e FRSAD, apresentando soluções para as deficiências identificadas nesses modelos (ARAKAKI, 2020).

Outra iniciativa frequentemente discutida no universo bibliográfico, é o BIBFRAME, desenvolvido em 2011 pela *Library of Congress* (LC) com o intuito de substituir o formato MARC 21. O BIBFRAME propõe realizar a transição dos dados do

MARC 21 para dados conectados na *Web*, segundo Library of Congress (2012), a iniciativa BIBFRAME visa implementar um novo ambiente para bibliotecas centrado na *Web* e que propicie a interconexão de informações.

Com base na revisão de literatura, é possível afirmar que essas temáticas têm sido frequentemente discutidas nos últimos anos. Observou-se que as soluções encontradas para o aperfeiçoamento do tratamento descritivo em meio digital, são encontradas no escopo da *Web Semântica*. Nessa perspectiva, a próxima seção versa sobre as especificidades da descrição de recursos informacionais nos ambientes digitais, apresentando algumas tecnologias essenciais nesse contexto.

4 REPRESENTAÇÃO E METADADOS EM AMBIENTES DIGITAIS

O constante aperfeiçoamento tecnológico impacta as atividades de diversos setores, tornando a sociedade contemporânea cada vez mais dependente de sistemas computacionais. Os metadados destacam-se como elementos centrais no processo comunicativo entre máquinas e seres humanos. Pomerantz (2015) aponta que os dispositivos eletrônicos são onipresentes nos ambientes informacionais, e seu funcionamento depende de metadados, seja fazendo uso, produzindo ou ambos. Os metadados permitem que os usuários encontrem itens de interesse, criem seu próprio conteúdo e o compartilhem (RILEY, 2017). Além de serem elementos intrínsecos aos sistemas de informação, os metadados estão presentes em nosso cotidiano, constituindo, por exemplo, rótulos de produtos, notas fiscais e documentos pessoais.

O termo “metadado” é geralmente associado ao tratamento descritivo de recursos informacionais. Gill (2016, não paginado, tradução nossa) define os metadados, como “[...] a descrição estruturada dos atributos essenciais de um objeto de informação.” Segundo Santos, Simionato e Arakaki (2014, p. 150) os metadados explicitam os principais aspectos de um recurso informacional, incluindo “[...] sua estrutura, conteúdo, qualidade, contexto, origem, propriedade e condição.”

Desse modo, a descrição é realizada a partir da seleção de elementos descritivos, isto é, de metadados que explicitam as principais características de um recurso informacional, viabilizando a identificação, localização e acesso, seja em ambientes analógicos ou digitais. Santos, Simionato e Arakaki (2014, p. 148) definem um recurso informacional como

[...] a informação objetivada no contexto de um campo do conhecimento podendo ser apresentado em uma estrutura analógica e/ou digital, com valor informacional que caracteriza a sua concepção intelectual expressa na corporificação de manifestações estruturadas na forma de itens. O recurso informacional também é nominado de: ‘item informacional’, ‘informação registrada’, ‘coisa física, ou uma coisa não-física’ e ‘artefato’.

Vale ressaltar que o termo “recurso informacional” não é limitado a documentos textuais, podendo referir-se a qualquer objeto com potencial informativo. Para Zafalon (2017, p. 127) um recurso informacional é o registro de uma representação intelectual ou artística “[...] Assim, tem-se que o recurso informacional seja a ‘ideia corporificada’, a ‘inscrição de uma ideia em um suporte’.”

A representação propicia a descrição padronizada e simplificada das principais características de um recurso informacional permitindo a identificação e compreensão do mesmo sem a necessidade de consultá-lo na íntegra. Desse modo, a representação deve promover as condições necessárias para que o usuário recupere e verifique a relevância de um recurso. Marcondes (2001) destaca que a representação deve ser rica sob o aspecto cognitivo e, ao mesmo tempo, sintética, de modo a economizar a energia do usuário.

No processo de representação são considerados os aspectos intrínsecos e extrínsecos dos recursos. Independentemente da forma física ou intelectual que assumem, os recursos possuem três principais características: conteúdo, contexto e estrutura (GILLILAND, 2016). O conteúdo é uma característica intrínseca relacionada ao assunto tratado, enquanto o contexto consiste em uma característica extrínseca que indica quais aspectos estão associados à criação do recurso. A estrutura refere-se tanto a características intrínsecas, como extrínsecas, tratando do conjunto formal de associações entre os recursos informacionais (GILLILAND, 2016).

Os metadados são categorizados conforme as características e funções que desempenham, Riley (2017) estabelece quatro categorias aos metadados: metadados descritivos, metadados estruturais, metadados administrativos e linguagem de marcação. Com base em Riley (2017), Ninin e Simionato (2018) descrevem os tipos de metadados, exemplos de propriedades e utilidades, apresentadas no Quadro 5.

Quadro 5 - Tipos de metadados, propriedades e funções

Tipos de metadados		Exemplos de propriedades	Usos primários
Descritivos		título, autor, assunto, gênero, data de publicação.	Descoberta, apresentação e interoperabilidade.
Estruturais		Sequência posição na hierarquia.	Navegação
Linguagens de marcação		parágrafo, cabeçalho, lista, nome, data.	Navegação Interoperabilidade
Administrativos	Técnicos	tipo do arquivo, formato do arquivo, tempo/data de criação, esquema de compressão.	Interoperabilidade Manutenção de objetos digitais Preservação
	Preservação	verificação dos dados ação de preservação.	Interoperabilidade Manutenção de objetos digitais Preservação
	Direitos autorais	<i>status</i> de <i>copyright</i> , termos de licença, detentores dos direitos.	Interoperabilidade Manutenção de objetos digitais.

Fonte: Ninin e Simionato (2018, p. 6).

Os metadados descritivos são utilizados para identificar, compreender e localizar um recurso; os metadados estruturais descrevem os relacionamentos entre partes dos recursos; as linguagens de marcação mesclam metadados e conteúdo, a partir da atribuição de *tags* que explicitam elementos estruturais e semânticos; os metadados administrativos constituem uma categoria mais ampla, englobando subcategorias, tais como: metadados técnicos, metadados de preservação e metadados de direito (RILEY, 2017).

Nas instituições de patrimônio cultural, os metadados descritivos são predominantes, elaborados para descrição detalhada dos recursos informacionais nos catálogos. Para descrição de livros, destacam-se metadados como, título, autor, publicação e detalhes do assunto; para musicais, filmes e obras de arte, são normalmente registrados o título, autor, gênero e informações de desempenho; para documentos e registros de arquivo são registrados detalhes de criação (RILEY, 2017).

Os metadados descritivos permitem identificar, autenticar e descrever coleções e recursos informacionais, contemplando metadados gerados por um criador ou sistema, pacotes de informações de envio, registros de catalogação, controle de versões, relações entre recursos, recursos de localização, índices especializados, informações curatoriais e anotações de criadores e outros usuários (GILLILAND, 2016; ZENG; QIN, 2016). Essa categoria de metadado pode ainda ser empregada na descrição de informações referentes a manutenção e origem de um recurso, informando sobre sua história e estado atual (POMERANTZ, 2015). Além de descrever os vários aspectos que envolvem um recurso informacional, os metadados descritivos permitem a contextualização do mesmo em um ambiente informacional, favorecendo a recuperação eficiente de informações pertinentes ao usuário.

Nos sistemas de informação, as diferentes tipologias de metadados devem ser estruturadas conforme padrões e vocabulários reconhecidos internacionalmente. Alves (2010, p. 47) define os padrões de metadados como

[...] estruturas de descrição constituídas por um conjunto predeterminado de metadados (atributos codificados ou identificadores de uma entidade) metodologicamente construídos e padronizados. O objetivo do padrão de metadados é descrever uma entidade gerando uma representação unívoca e padronizada que possa ser utilizada para recuperação da mesma.

Os padrões de metadados podem ainda ser considerados documentos formais que estabelecem “[...] critérios, métodos, processos e práticas uniformes para orientar o projeto, criação e implementação de estruturas de dados, valores de dados, conteúdo de dados e troca de dados de uma maneira eficiente [...]” (ZENG; QIN, 2016, p. 23, tradução nossa).

Desse modo, os padrões de metadados são estruturas que propiciam a descrição uniforme dos recursos informacionais, favorecendo a recuperação de informações e a interoperabilidade entre sistemas. A escolha de um padrão de metadados deve ser baseada nos propósitos e necessidades específicas de um ambiente informacional, visto que existem diferentes níveis de granularidade, portanto, o padrão deve ser suficientemente abrangente e coerente em um dado contexto.

Em relação aos vocabulários, estes são constituídos por termos padronizados para descrição, determinando possíveis conexões e restrições de uso (WORLD WIDE WEB CONSORTIUM, 2019). Os vocabulários evitam o uso de termos redundantes e auxiliam na redução da ambiguidade e conflitos entre elementos semelhantes, são constituídos por classes e propriedades expressas no formato RDF.

O World Wide Web Consortium (2019) recomenda a adoção de vocabulários existentes, utilizados amplamente na *Web*, pois dessa maneira, as comunidades estabelecem um consenso, de modo a favorecer a interoperabilidade e a reutilização dos dados. No ambiente *Web*, os dados são estruturados a partir de protocolos, diretrizes e padrões recomendados pelo W3C. O W3C é um consórcio internacional que regulamenta a publicação de dados na *Web*, visando promover a acessibilidade e interoperabilidade entre diferentes formatos.

Nesse contexto, destaca-se o *Hypertext Markup Language* (HTML), derivado do *Standard Generalized Markup Language* (SGML), um padrão internacional para definir linguagens de marcação. O HTML é utilizado para estruturar a formatação das páginas da *Web*, explicitando características estruturais a partir de *tags* que determinam elementos, tais como: parágrafos, listas, imagens, tabelas, *links*, etc. (WORLD WIDE WEB CONSORTIUM, 2016). Essas marcações são processadas pelos navegadores da *Web* que exibem as páginas conforme a formatação estabelecida pelas *tags*. Além disso, o HTML permite a navegação para além de uma página, a partir de *hiperlinks* que direcionam o usuário para recursos relacionados ao mesmo conteúdo.

À medida que novos tipos de recursos surgiram na *Web*, foram desenvolvidas linguagens mais completas, como o *Extensible Markup Language* (XML), uma linguagem de marcação extensível recomendada pelo W3C. Por ser derivado do SGML, o XML é semelhante ao HTML, entretanto, não possui elementos predefinidos como o HTML, permitindo a utilização de *tags* personalizadas. Além de ser uma linguagem versátil, possui regras rígidas que impedem erros de codificação. Desse modo, garante-se a qualidade na estruturação, permitindo o processamento eficiente dos dados pelos agentes de *software*, além de fornecer a base para especificar outros tipos de codificações (WORLD WIDE WEB CONSORTIUM, 2015a).

Para possibilitar a obtenção de resultados de busca relevantes em meio a quantidade expressiva de informações na *Web*, é necessário que os dados sejam estruturados. Conforme Pomerantz (2015) nem todas as estruturas de metadados são acessíveis aos algoritmos, por este motivo, para que os agentes de *software* sejam realmente eficientes, é crucial que os dados sejam explicitamente estruturados.

Sousa, Martins e Ramalho (2018) apontam que as propriedades do XML podem ser utilizadas para criação de formatos semânticos. Entretanto, somente o XML não é suficiente para construir metadados consistentes, pois para compartilhar dados de modo significativo, é necessário que as práticas de estruturação sejam consensuais entre todos os participantes de uma comunidade (SOUZA; ALVARENGA, 2004).

Nesse contexto, as tecnologias da *Web Semântica*, também denominada *Web de Dados*, surgem como soluções para otimizar a comunicação entre agentes de *software* e seres humanos. Berners-Lee, Hendler e Lassila (2001, p. 2, tradução nossa) definem a *Web Semântica* como uma extensão da *Web* atual “[...] na qual as informações recebem um significado bem definido, permitindo que os computadores e as pessoas trabalhem em cooperação.” Godby, Wang e Mixter (2015) ressaltam que o propósito da *Web Semântica* consiste em atribuir significado aos dados, permitindo que os agentes de *software* executem tarefas sofisticadas para os usuários. Dessa maneira, os computadores podem atuar com maior eficiência, a partir de tecnologias que auxiliam no desenvolvimento de sistemas para interações confiáveis na rede (WORLD WIDE WEB CONSORTIUM, 2015b).

Sendo assim, a *Web Semântica* representa um avanço significativo para representação de informações nos ambientes digitais. A *Web* baseada em HTML foi projetada principalmente para interpretação humana, ficando a critério do usuário definir a relevância de um conteúdo. Em contrapartida, a *Web Semântica* visa o

desenvolvimento de tecnologias que viabilizem tanto a interpretação humana, como de sistemas computacionais.

Conforme Pomerantz (2015, p. 157, tradução nossa) “A *Web* semântica requer mais do que colocar dados *online*; requer a criação de *links* entre os dados estruturados.” A interligação dos dados é possível a partir de um conjunto de práticas denominadas *Linked Data*, em português dados conectados, interligados ou vinculados. O *Linked Data* visa a padronização e disponibilização de dados para o ambiente *Web*. Heath e Bizer (2011) explanam, que assim como os hiperlinks conectam as informações na *Web* clássica, o *Linked Data* permite o estabelecimento de vínculos entre recursos informacionais de fontes de informações distintas.

As tecnologias que compõem a arquitetura da *Web Semântica* e norteiam os princípios *Linked Data* são representadas graficamente por uma estrutura conhecida como, “*layer cake*”, constituído por tecnologias essenciais para atribuição de valor semântico aos dados. Essa estrutura é apresentada detalhadamente em Horrocks e Patel-Schneider (2003); Isotani e Bittencourt (2015) e Idehen (2017). Segundo World Wide Web Consortium (2015c) o *Linked Data* é o cerne da *Web Semântica*, as tecnologias semânticas são projetadas para promover a acessibilidade e integração de dados conectados em níveis de complexidade distintos. Para Riley (2017) o *Linked Data* é a etapa prática para concretização da *Web Semântica*, envolvendo a publicação de dados estruturados e conectados.

A criação de dados conectados, depende de quatro requisitos fundamentais, conhecidos como princípios *Linked Data*. Berners-Lee (2006, não paginado, tradução nossa) definiu as seguintes recomendações para publicação em *Linked Data*:

- 1) Utilize URIs como nomes para as coisas.
- 2) Utilize URIs HTTP para que as pessoas possam procurar por esses nomes.
- 3) Quando alguém procura um URI, forneça informações úteis, utilizando os padrões (RDF, SPARQL).
- 4) Forneça *links* para outros URIs para que mais coisas possam ser encontradas.

O primeiro princípio recomenda a atribuição de um *Uniform Resource Identifier* (URI) aos recursos para serem identificados de forma unívoca na *Web*. Um URI pode ser classificado como um *Uniform Resource Locator* (URL) que define um localizador ou endereço para um determinado recurso, ou como um *Unified Resource Name* (URN) cuja função consiste em representar um nome para um recurso, garantindo a

univocidade e persistência (ISOTANI; BITTENCOURT, 2015). Os URIs podem identificar qualquer tipo de recurso, incluindo conceitos abstratos, segundo Heath e Bizer (2011) esse princípio pode ser considerado como uma extensão da *Web*, onde é possível representar qualquer recurso ou conceito no mundo.

O segundo princípio propõe a utilização do *Hypertext Transfer Protocol* (HTTP) para se referir a URIs, pois, dessa maneira podem ser consultados por uma ampla gama de *softwares* e serviços (POMERANTZ, 2015). A identificação do tipo URIs HTTP permite a combinação entre identificadores únicos, favorecendo a interpretação eficiente dos mecanismos de busca da *Web*.

A codificação dos dados conectados na *Web Semântica* é orientada por um modelo padrão, o terceiro princípio defende o uso do *Resource Description Framework* (RDF), um conjunto de especificações projetadas pelo W3C para estruturação dos dados na *Web*. Enquanto, a *Web* clássica faz uso do HTML, permitindo a navegação por *hiperlinks*, na *Web Semântica*, os recursos são acessados por intermédio de *links* em RDF que propiciam o acesso a fontes de informação relacionadas (ISOTANI; BITTENCOURT, 2015). Os *hiperlinks* são denominados *links* RDF para distingui-los dos *hiperlinks* entre documentos da *Web* clássica (HEATH; BIZER, 2011).

O RDF consiste em uma estrutura lógica e genérica com a finalidade de organizar os dados e fazer declarações sobre entidades e seus relacionamentos (POMERANTZ, 2015). As declarações RDF são descritas por triplas constituídas por sujeito-predicado-objeto. O sujeito refere-se ao URI que identifica um recurso, enquanto o objeto pode ser “[...] um valor literal simples, como uma *string*, número ou data; ou o URI de outro recurso que de alguma forma está relacionado ao sujeito.” (HEATH; BIZER, 2011, p. 15, tradução nossa). Desse modo, além de auxiliar na descrição sintática de maneira mais precisa, o RDF proporciona a descrição semântica, representando os relacionamentos e os significados dos recursos (ISOTANI; BITTENCOURT, 2015).

Ainda em relação ao terceiro princípio, é sugerido o uso do *Simple Protocol and RDF Query Language* (SPARQL), um conjunto de especificações para consulta e manipulação do conteúdo publicado em RDF. Por fim, o quarto princípio defende o uso de *hiperlinks* para relacionar qualquer tipo de recurso informacional. É possível, por exemplo, estabelecer a relação entre uma pessoa e um local ou entre um local e uma empresa (HEATH; BIZER, 2011). Para que os dados modelados em RDF sejam

legíveis para os agentes computacionais, podem ser utilizados formatos, tais como: RDF/XML, RDFa, *Turtle*, *N-triples* e JSON-LD.

O RDF/XML é a serialização de um grafo RDF que utiliza a estrutura em XML, esse tipo de serialização possui uma sintaxe complexa para manipulação e compreensão humana, por essa razão, pode-se optar por outras alternativas (HEATH; BIZER, 2011; RILEY, 2017). A serialização em RDFa adiciona atributos às *tags* HTML, nesse caso, as informações sobre os dados ficam separadas das informações sobre a formatação dos dados (ISOTANI; BITTENCOURT, 2015). *Turtle* é uma representação textual de um grafo RDF que permite a apresentação de vários predicados do mesmo assunto de modo simplificado (HEATH; BIZER, 2011; RILEY, 2017). Enquanto o *N-triples* é um subconjunto de *Turtle*, limitado a uma sintaxe mais estruturada e previsível (RILEY, 2017). Por fim, a serialização em JSON-LD é baseada na sintaxe JSON, um formato intuitivo e conhecido pelos desenvolvedores de *software*, esse tipo de codificação faz uso de pares simples de chave-valor para representar as informações de registro (RILEY, 2017).

Os dados modelados em RDF podem ser publicados conforme a serialização mais conveniente aos propósitos e necessidades de um ambiente informacional, esses formatos permitem que os agentes de *software* processem os dados eficientemente, auxiliando na interpretação e na recuperação de informações.

Atrelado ao *Linked Data*, surge o *Linked Open Data*, um conjunto de práticas para publicação de dados conectados sob uma licença aberta. Segundo Torino *et al.* (2020, p. 4) “[...] dados abertos são aqueles disponibilizados sem barreiras de acesso e em formatos abertos e legíveis por humanos e aplicações computacionais [...]”. Dessa maneira, os dados podem ser melhor aproveitados, a partir da livre reutilização dos mesmos.

O World Wide Web Consortium (2014) estabelece dez princípios básicos, conhecidos como “As melhores práticas para publicação *Linked Data*” para auxiliar na reutilização e acesso aos dados estruturados. Segundo World Wide Web Consortium (2014, não paginado, tradução nossa) os princípios para publicação de dados abertos e conectados são respectivamente:

- 1) **Preparar as partes interessadas:** esclarecer o processo de criação e manutenção do *Linked Open Data*.
- 2) **Selecionar um conjunto de dados:** selecionar um conjunto de dados que ofereça benefícios para reutilização.

- 3) **Modelar os dados:** modelar dados conectados envolve a representação de dados e seus relacionamentos independentemente do contexto.
- 4) **Especificar uma licença apropriada:** a reutilização de dados ocorre de maneira mais eficiente quando há uma declaração clara sobre a origem, propriedade e termos relacionados ao uso dos dados abertos e conectados.
- 5) **Estabelecer bons URIs:** o núcleo do *Linked Data* consiste em uma estratégia para implementação de URIs HTTP, considerando a nomeação de objetos, suporte multilíngue, a realização de alterações nos dados ao longo do tempo [...].
- 6) **Utilizar vocabulários padrões:** descreva objetos com vocabulários previamente definidos sempre que possível. Estenda vocabulários padrão quando necessário e crie vocabulários (somente quando necessário) que sigam as melhores práticas.
- 7) **Realizar a conversão para dados conectados:** a conversão é geralmente realizada por *script* ou por outros processos automatizados.
- 8) **Promover o acesso aos dados conectados:** forneça várias maneiras para que os mecanismos de pesquisa e outros processos automatizados tenham acesso aos dados.
- 9) **Anunciar os novos conjuntos de dados conectados:** lembre-se de anunciar novos conjuntos de dados em um domínio autorizado. É importante lembrar que, como publicador de *Linked Open Data*, um contrato social implícito está em vigor.
- 10) **Estabelecer um contrato social:** reconheça sua responsabilidade em manter os dados após sua publicação. Certifique-se de que o (s) conjunto (s) de dados permaneçam disponíveis onde foi estabelecido e sejam preservados.

Em âmbito governamental esses princípios foram estabelecidos pelo W3C *Government Linked Data Working Group* para promover a reutilização, a persistência dos dados na *Web* e garantir a transparência de informações no domínio público.

No contexto das bibliotecas, são produzidos e gerenciados produtos e serviços de informação à sociedade, portanto, a aderência as tendências mundiais de publicação de dados conforme as iniciativas *Linked Data* e *Linked Open Data* podem ampliar consideravelmente o alcance do conteúdo dos catálogos digitais, promovendo melhor aproveitamento dos recursos disponíveis.

Ressalta-se que os catálogos digitais permitem a navegação por *hiperlinks*, entretanto, os dados não são recuperados pelos mecanismos de busca e não há conexão com fontes externas. Em uma busca na *Web* é possível encontrar a interface de um catálogo de biblioteca, mas não o seu conteúdo, reduzindo consideravelmente o alcance dessas informações, pois, geralmente os usuários não estão buscando especificamente pelas bibliotecas, e sim, pelos recursos oferecidos (WALLIS, 2022).

Peset, Ferrer-Sapena e Subirats-Coll (2011) destacam que a iniciativa *Linked Open Data* pode revolucionar o acesso aos dados, começando pela aderência das Melhores Práticas pelos órgãos governamentais ao redor do mundo. As autoras argumentam que as bibliotecas prezam pela persistência e reutilização dos dados,

portanto, as mesmas devem caminhar em consonância com as tendências mundiais de publicação de dados para o compartilhamento de seu conteúdo.

Em uma pesquisa sobre o novo cenário informacional identificou-se que a busca por recursos informacionais é realizada majoritariamente pelos mecanismos de busca (DEMPSEY *et al.*, 2005 *apud* GODBY; WANG; MIXTER, 2015). Sendo assim, para que os dados dos catálogos sejam valorizados, devem ser estruturados em consonância com as tecnologias emergentes alinhadas ao escopo da *Web Semântica*. Caso os catálogos permaneçam restritos a comunidade de bibliotecas, seus valiosos recursos serão perdidos em meio a inúmeras outras opções na *Web* (GODBY; WANG; MIXTER, 2015).

Diante dessa realidade, identifica-se a necessidade de reavaliar e revitalizar os procedimentos e instrumentos de representação utilizados no âmbito das bibliotecas, considerando as diretrizes e formatos recomendados pelo W3C para publicação de dados estruturados. Godby, Wang e Mixter (2015) argumentam que a reformulação dos procedimentos descritivos com formatos legíveis por máquina permitiria que as bibliotecas se tornassem “[...] mais 'da *Web*' e não apenas 'na *Web*’.” (GODBY; WANG; MIXTER, 2015, não paginado, tradução nossa).

A próxima seção versa sobre o *Schema.org*, os metadados desse vocabulário são identificados e interpretados pelos mecanismos de busca de maneira mais eficiente, permitindo a recuperação de informações contextualizadas. Segundo World Wide Web Consortium (2019, não paginado, tradução nossa) o modo como o *Schema.org* é estruturado o torna “[...] uma boa escolha como vocabulário a ser usado na codificação de dados para o compartilhamento.”

5 SCHEMA.ORG

O *Schema.org* consiste em um vocabulário desenvolvido a partir da parceria entre os mecanismos de busca, *Google*, *Yahoo*, *Microsoft* e *Yandex*. Publicado oficialmente em 2011, o *Schema.org* surgiu com a proposta de padronizar a representação de informações na *Web*, visando otimizar a interpretação e a recuperação dos dados pelos mecanismos de busca e assim promover melhores experiências aos usuários.

Segundo *Schema.org* (2021a, não paginado, tradução nossa) “*Schema.org* é uma atividade colaborativa comunitária, com a missão de criar, manter e promover esquemas para dados estruturados na *internet* em páginas da *Web*, em mensagens de *e-mail* e além.” Wallis (2012) define o *Schema.org* como um vocabulário genérico de alto nível que favorece a identificação e a interpretação dos dados pelos mecanismos de busca e ainda permite simplificar o processo de estruturação das páginas da *Web*.

Nesse sentido, Guha, Brickley e Macbeth (2016) e Mika (2015) afirmam que a multiplicidade de vocabulários utilizados na *Web* complexifica o trabalho dos *webmasters*, pois, estes profissionais precisam lidar com codificações distintas, nomeando recursos semelhantes de modos diferentes. Esse processo influencia na qualidade dos dados, visto que podem ser geradas duplicações e termos incorretos.

Em vista disso, o *Schema.org* foi projetado visando proporcionar maior praticidade aos *webmasters*, a partir de um vocabulário padrão entre os mecanismos de busca. Segundo *Schema.org* (2021a, não paginado, tradução nossa) “Um vocabulário compartilhado torna mais fácil para *webmasters* e desenvolvedores decidirem sobre um esquema e obterem o máximo benefício por seus esforços.” Para Pomerantz (2015) o *Schema.org* se destaca na tarefa de simplificar a criação de dados estruturados na *Web*, nesse viés, Mika (2015, p. 54, tradução nossa) destaca que “[...] o *Schema.org* abordou um problema crítico para *Web* - simplificando a anotação dos dados em páginas da *web*, pelo menos para os tipos mais populares de conteúdo.”

O vocabulário foi projetado para representar qualquer tipo de informação, a partir de classes, subclasses e propriedades. Os elementos do *Schema.org* são organizados de maneira hierárquica para auxiliar a compreensão e a navegação pelo vocabulário (GUHA; BRICKLEY; MACBETH, 2016).

Segundo World Wide Web Consortium (2019) o *Schema.org* foi adotado amplamente em um curto período, principalmente devido à abordagem de *design*, visto que, não há regras rígidas para definição dos tipos de recursos com os quais as propriedades podem ser utilizadas, as relações são apenas sugeridas. Mika (2015) ressalta que outro fator que pode ter influenciado na rápida aderência do *Schema.org* refere-se à possibilidade de incorporar o vocabulário em páginas codificadas em HTML5.

O modelo de dados utilizado pelo *Schema.org* é o RDF *Schema*, uma extensão do RDF compatível com formatos, como, RDFa, Microdados, JSON-LD e CSVW. Esses formatos são incorporados nas páginas HTML para otimizar o processamento dos dados pelos agentes de *software*. Desse modo, os dados estruturados em *Schema.org* podem ser empregados, por exemplo, em mensagens de *e-mail* para confirmação de reservas em restaurantes e hotéis, assim como em recibos de compra para apresentação de detalhes de transação (GUHA; BRICKLEY; MACBETH, 2016). Com base nas anotações em *Schema.org*, os assistentes de *e-mail* fornecem lembretes aos usuários, informando sobre horários, datas e localizações.

Estima-se que um número superior a 10 milhões de *sites* estejam utilizando o *Schema.org*, assim como ferramentas da *Google*, *Microsoft*, *Yandex*, entre outras (SCHEMA.ORG, 2021a). Guha, Brickley e Macbeth (2016) mencionam a aplicação do vocabulário nas tecnologias: *Cortana* (*Windows 10* e *Windows phone*), *Siri* (*Apple*), *Pinterest* e *Yandex*.

Na assistente virtual *Cortana* da *Microsoft*, o *Schema.org* é codificado em Microdados e JSON-LD. A *Cortana* identifica os metadados do vocabulário em mensagens de *e-mail*, extraíndo dados de reserva de voos aéreos, após o recebimento do primeiro *e-mail* de confirmação do voo, as atualizações são rastreadas automaticamente, assim como outras alterações relacionadas (MICROSOFT, 2016). A assistente virtual *Siri* da *Apple* utiliza o *Schema.org* em recursos de pesquisa, incluindo avaliações, ofertas, produtos, preços, organizações, imagens e números de telefone, assim como no *Really Simple Syndication* (RSS) para marcação de notícias (GUHA; BRICKLEY; MACBETH, 2016).

O *Pinterest* é uma rede social que permite aos usuários criar, salvar, organizar e armazenar *pins* que podem ser imagens, vídeos ou produtos. A rede social utiliza o *Schema.org* para fornecer *pins* enriquecidos (*rich pins*), o vocabulário possibilita que uma página da *Web* seja incorporada ao *Pinterest* como um *pin*. Para cada categoria

de *pin* existe um conjunto de propriedades em *Schema.org*, que podem ser obrigatórias ou opcionais (PINTEREST, 2021). O mecanismo de busca *Yandex* faz uso do *Schema.org* em Microdados para representação de receitas, automóveis, revisões, organizações e outros serviços voltados ao consumidor (GUHA; BRICKLEY; MACBETH, 2016).

Identificou-se ainda, a aderência do *Schema.org* pela rede social, *Facebook*, onde o *Schema.org* é uma das codificações recomendadas para descrição de anúncios sobre serviços e produtos. Para que os anúncios de um *site* sejam exibidos aos usuários, é necessária a instalação de um código denominado *pixel*. Segundo Facebook (2021a) para que os anúncios sejam identificados com o código é necessário adicionar Microdados ao *site* nos formatos: *OpenGraph*, *Schema.org* ou JSON-LD para *Schema.org*. A utilização do *pixel* com os formatos recomendados permite que as empresas mensurem o impacto de suas campanhas de *marketing* na rede social, viabilizando que as ações de um indivíduo em um *site* sejam identificadas e analisadas (FACEBOOK, 2021b).

Cabe ressaltar que o *Schema.org* é constituído por metadados genéricos, isto é, não foi projetado para um domínio específico. Sendo assim, o vocabulário não possui uma infraestrutura suficientemente abrangente para os diferentes níveis de detalhamento exigidos em determinados contextos, como no caso dos catálogos digitais. Conforme Github (2021) o *Schema.org* foi projetado para simplificar realidades de representação complexas. Com o propósito de permitir a ampliação e adaptação do vocabulário, o *Schema.org* permite o desenvolvimento de mecanismos de extensão para o atendimento de diferentes necessidades de representação.

O *Schema.org* é um projeto independente, elaborado de maneira colaborativa por grupos de profissionais de diferentes setores. A administração do vocabulário é realizada por dois grupos, o *Steering Group* e o *W3C Schema.org Community Group*. O *Steering Group* é composto por representantes dos buscadores que patrocinam o *Schema.org*, um representante do W3C e outros profissionais responsáveis pela supervisão, avaliação e aprovação de novos lançamentos (SCHEMA.ORG, 2021b). O *W3C Schema.org Community Group*, é uma iniciativa comunitária e aberta que engloba profissionais interessados em contribuir para o aprimoramento do vocabulário. Os membros do grupo comunitário participam das discussões, integração de esquemas e compartilham os resultados obtidos com o *Steering Group* para avaliação e aprovação (SCHEMA.ORG, 2021b).

As discussões do grupo comunitário ocorrem no fórum do W3C *Schema.org Community Group*, na lista de discussão *public-schemaorg@w3.org*, enquanto os problemas do vocabulário são rastreados no repositório do projeto, o *GitHub*. O *GitHub* é uma plataforma direcionada para desenvolvedores de *software*, onde são fornecidas ferramentas para o gerenciamento de projetos. No *GitHub* são armazenados todos os esquemas, exemplos e *softwares* utilizados para publicação do *Schema.org* (GITHUB, 2021). Com base na lista de problemas rastreados no *GitHub*, são elaboradas novas propostas para realização de aprimoramentos (SCHEMA.ORG, 2021c).

O *Schema.org* disponibiliza fontes oficiais, como o *site* e o *blog* do vocabulário, assim como as listas de discussões, publicadas no *GitHub*. Como complemento, pode ser consultado o *blog* “*Data Liberate*” de Richard Wallis, presidente do *Schema Bib Extend Community Group* e membro do *W3C Schema.org Community Group*. No *blog*, são abordadas várias questões relacionadas ao *Schema.org* e a outras tecnologias para representação na *Web*.

A estrutura completa do vocabulário pode ser visualizada no *site* oficial, onde são apresentados todos os termos das hierarquias, organizados em classes (*types*), propriedades (*properties*) e enumerações ou valores enumerados (*enumerated values*), assim como diretrizes e exemplos de uso. A próxima subseção aborda a estrutura, organização e funcionamento do *Schema.org*.

5.1 Estrutura e organização do *Schema.org*

O *Schema.org* consiste em esquemas de metadados constituídos por termos genéricos padronizados e organizados hierarquicamente em classes, propriedades e enumerações. As versões do vocabulário são documentadas no *site* oficial, onde são registradas as principais alterações. No momento de elaboração desta pesquisa, a versão 13.0² é a mais recente do *Schema.org*, entretanto, ressalta-se que as atualizações são realizadas com frequência, o vocabulário está em constante evolução.

² SCHEMA.ORG. **Schema.org version 13.0**: versão publicada em 07 julho de 2021. 2021. Disponível em: <https://schema.org/version/latest>. Acesso em: 29 set. 2021.

Atualmente, o vocabulário possui 792 classes, 1447 propriedades, 15 tipos de dados, 83 Enumerações e 445 membros de enumeração (SCHEMA.ORG, 2021f). No *site* oficial do *Schema.org*, as classes do vocabulário também são denominadas “tipos” (*types*), termo utilizado com maior frequência. Nesta pesquisa, optou-se por utilizar o termo “classe”, portanto, “tipo” é utilizado apenas nas citações diretas.

Em *Schema.org* (2021d, não paginado, tradução nossa) a estrutura do vocabulário é explicada da seguinte maneira:

1. Temos um conjunto de tipos, organizados em uma hierarquia de herança múltipla, onde cada tipo pode ser uma subclasse de vários tipos.
2. Temos um conjunto de propriedades:
 1. cada propriedade pode ter um ou mais tipos como seus domínios. A propriedade pode ser usada para instâncias de qualquer um desses tipos.
 2. cada propriedade pode ter um ou mais tipos como seus intervalos. O valor ou valores da propriedade devem ser instâncias de pelo menos um desses tipos.

Desse modo, a hierarquia principal compreende uma coleção de classes e cada classe compreende uma ou mais subclasses (SCHEMA.ORG, 2021e). A principal classe do *Schema.org* é “*Thing*”, a mais genérica do vocabulário. *Thing* possui as seguintes classes subordinadas: *Action*, *BioChemEntity*, *CreativeWork*, *Event*, *Intangible*, *MedicalEntity*, *Organization*, *Person*, *Place*, *Product* e *Taxon*. O Quadro 6 apresenta a descrição de cada uma dessas classes.

Quadro 6 - Subclasses de *Thing*

Subclasses de <i>Thing</i>	Descrição
<i>Action</i>	Ações executadas por um indivíduo. Podem ser definidas ações concluídas e ações potenciais (<i>potentialAction</i>), isto é, ações que são passíveis de ocorrer.
<i>BioChemEntity</i>	Elementos biológicos, químicos ou bioquímicos. Ex: proteína, gene ou substância química.
<i>CreativeWork</i> ,	Obras, como: livros, filmes, fotografias, <i>softwares</i> , etc.
<i>Event</i>	Eventos com horário e local definidos. Ex: shows, palestras e festivais.
<i>Intangible</i>	Valores intangíveis, como: quantidades e valores estruturados.
<i>MedicalEntity</i> ,	Entidades ligadas a área da saúde e a atividade de medicina.
<i>Place</i>	Entidades que possuem uma extensão fixa. Ex: um endereço físico, ou coordenadas geográficas.
<i>Organization</i>	Entidades como escolas, ONGS e corporações.
<i>Person</i>	Um indivíduo, vivo, morto ou ficcional.
<i>Product</i>	Produtos e serviços.
<i>Taxon</i>	Conjunto de organismos que representam uma unidade biológica.

Fonte: Elaboração da autora com base em *Schema.org* (2021e).

As classes derivadas de *Thing* possuem suas próprias subclasses, por exemplo: *CreativeWork* é subordinada a *Thing* que inclui a subclasse *Book*, que

compreende a subclasse *Audiobook*. As classes e subclasses englobam um conjunto de propriedades que não são necessariamente exclusivas, por exemplo, a propriedade *duration* de *Audiobook* pode pertencer às classes: *Event*, *Episode*, *Movie*, entre outras. A organização hierárquica das classes, *Thing*, *CreativeWork*, *Book* e *Audiobook* podem ser visualizadas na Figura 1.

Figura 1 - Hierarquia principal *Schema.org*



Fonte: Schema.org (2021e, não paginado).

A Figura 1 apresenta parte da hierarquia correspondente as classes e subclasses mencionadas, no *site* oficial do *Schema.org* é possível visualizar e navegar pela hierarquia completa. Ao clicar em cada uma das classes, é possível acessar de modo individualizado as subclasses, propriedades, tipos de dados esperados (*expected types*) e o escopo de cada elemento. O Quadro 7 apresenta as propriedades, tipos esperados e as respectivas descrições para *Book*.

Quadro 7 - Propriedades, tipos esperados e descrição para *Book*

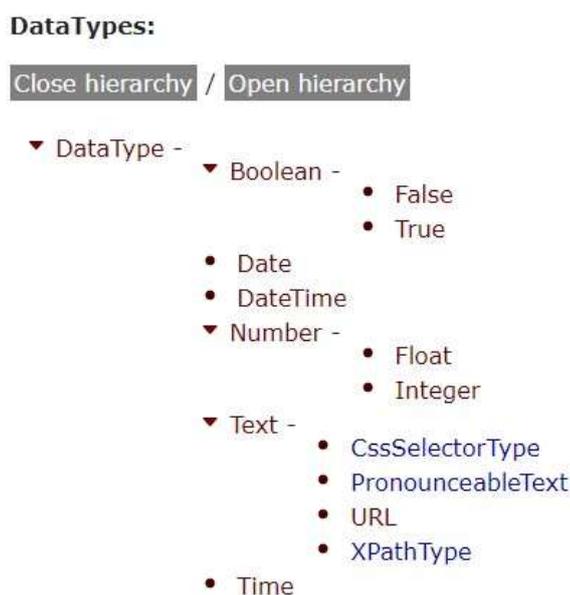
Propriedades de <i>Book</i>		
Propriedade	Tipos esperados	Descrição
<i>Abridged</i>	<i>Boolean</i>	Indica se o livro é uma edição resumida.
<i>BookEdition</i>	<i>Text</i>	Indica a edição.
<i>BookFormat</i>	<i>BookFormatType</i>	Indica o formato.
<i>Illustrator</i>	<i>Person</i>	Indica o indivíduo responsável pelas ilustrações.
<i>Isbn</i>	<i>Text</i>	Indica o ISBN.
<i>NumberOfPages</i>	<i>Integer</i>	Indica o número de páginas.

Fonte: Adaptação e tradução de Schema.org (2021h, não paginado).

Os tipos esperados, isto é, os valores das propriedades, são organizados em uma hierarquia paralela. As propriedades podem incluir: um valor, uma enumeração ou uma entidade e as enumerações referem-se a um conjunto limitado de valores

possíveis para determinadas propriedades, por exemplo, um produto à venda em uma loja *online* pode ser descrito com a classe *Offer*, as propriedades válidas nesse caso, possuem opções limitadas de valores, tais como: *In stock*, *Out of stock* e *Pre-order* (SCHEMA.ORG 2021i). Os tipos esperados e seus desdobramentos são representados na hierarquia *DataTypes*, apresentada na Figura 2.

Figura 2 – Hierarquia *DataTypes*



Schema.org (2021e, não paginado)

As propriedades abrangem os tipos: *boolean*, *data*, *datetime*, *number*, *text* e *time*. O tipo *boolean* representa os valores *false* e *true*; *data* representa datas no formato ISO 8601; *datetime* representa a combinação de data e hora; *number* representa números, incluindo tipos específicos, como *float* e *integer*; *text* representa documentos no formato texto (*CSSSelectorType*, *PronounceableText*, *URL*, *XPathType*) e *time* representa momentos que se repetem ao longo dos dias.

Os nomes das propriedades devem ser configurados do seguinte modo: o primeiro termo deve estar em letra minúscula e os termos subsequentes em letra maiúscula, todos em uma única *string*, ou seja, em uma sequência de caracteres, como, por exemplo: *openingHours*. Como pode ser visualizado no Quadro 7, as propriedades possuem tipos esperados, por exemplo, no caso da propriedade *servesCuisine*, o tipo esperado é *foodestablishment*. Geralmente, as propriedades abrangem mais de um tipo esperado, por exemplo, *acceptsReservations* inclui: *boolean*, *text* ou *URL* (SCHEMA.ORG, 2021g).

O *Schema.org* fornece ainda a classe *Role*, que apresenta como principal característica a flexibilidade para inserção de informações adicionais na descrição de relacionamentos e propriedades. Segundo *Schema.org* (2021c) *Role* pode ser utilizada com qualquer classe ou propriedade do vocabulário, permitindo que anotações sejam adicionadas arbitrariamente em qualquer parte da codificação.

As possibilidades fornecidas por *Role* são especialmente úteis para descrição de relacionamentos entre afiliações organizacionais, como no contexto da música e esporte (HOLLAND; JOHNSON, 2014). A classe possui propriedades próprias como, *startDate*, *endDate* e *roleName*, que permitem descrever funções específicas e marcos temporais.

No vocabulário, a classe *OrganizationRole* é uma subclasse de *Role*, utilizada para descrever relacionamentos entre pessoas e organizações. Na Figura 3 é possível visualizar um exemplo de codificação no formato RDFa, representando a função de um atleta em um time durante determinado período.

Figura 3 - Exemplo de codificação em RDFa para relacionamentos com *Role*

```
<div vocab="https://schema.org/" typeof="SportsTeam">
  <span property="name">San Francisco 49ers</span>
  <div property="member" typeof="OrganizationRole">
    <div property="member" typeof="https://schema.org/Person">
      <span property="name">Joe Montana</span>
    </div>
    <span property="startDate">1979</span>
    <span property="endDate">1992</span>
    <span property="roleName">Quarterback</span>
  </div>
</div>
```

Fonte: *Schema.org* (2021m, não paginado).

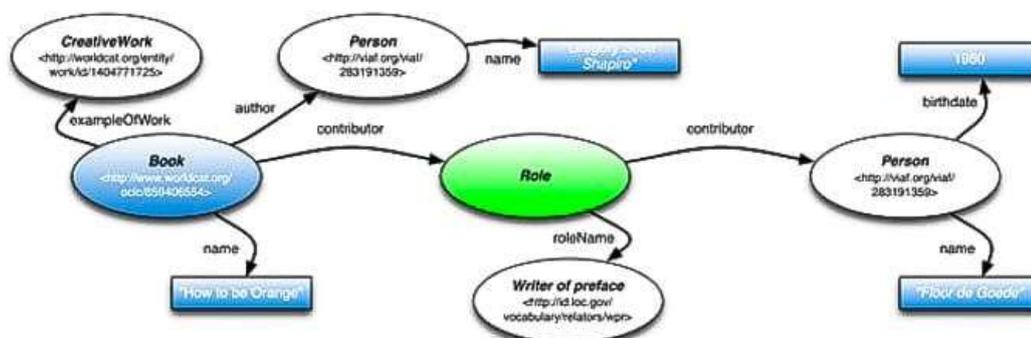
A Figura 3 apresenta um exemplo de possíveis conexões entre o atleta Joe Montana e o time *San Francisco 49ers*, o exemplo indica que o atleta foi integrante do time na função de *quarterback* no período de 1979 a 1992.

A classe *Role* funciona como um nó intermediário que conecta os valores adicionais às entidades principais, nesse caso a propriedade *athlete* é associada ao valor Joe Montana, assim como a propriedade *startDate* indica o ano em que o atleta se tornou membro do time. Para descrever a função do atleta no time, uma opção seria utilizar a propriedade *namesPosition* conectada ao valor “*quarterback*” (HOLLAND; JOHNSON, 2014).

Role ainda inclui a classe *PerformanceRole*, utilizada para representar, por exemplo, o papel interpretado por um ator em determinado filme com a propriedade *characterName* indicando o nome do personagem. Dessa maneira, as propriedades adicionais podem ser conectadas a uma entidade de *Role*, associada às entidades principais, com propriedades, como: membro ou ator (SCHEMA.ORG, 2021m).

No que tange a descrição do conteúdo das bibliotecas, Wallis (2015) afirma que esse mecanismo pode ser útil para descrição de recursos bibliográficos. Na figura 4, Wallis (2015) ilustra como *Role* poderia ser incluída na descrição de relacionamentos acerca do livro intitulado “*How to be Orange: an alternative Dutch assimilation course*”, representando a conexão com o autor do prefácio Floor de Goede.

Figura 4 - Role para descrição da função do autor de um prefácio

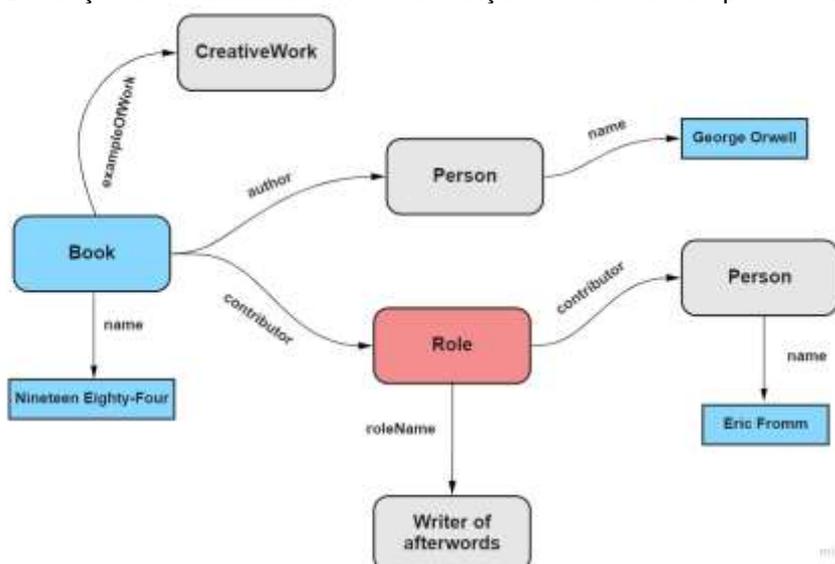


Fonte: Wallis (2015, não paginado).

No *Schema.org* não há uma propriedade para representar especificamente o autor de um prefácio, portanto, no exemplo de Wallis (2015) esse autor é classificado como colaborador da obra, nesse caso, *Role* é utilizada para especificar a função do colaborador e estabelecer a sua conexão com a obra. Observa-se que para cada entidade foi indicado um URI, identificando cada elemento de maneira unívoca.

Com base no exemplo de Wallis (2015), na Figura 5 esquematizou-se alguns dos possíveis relacionamentos entre o livro *Nineteen Eighty-Four* do autor George Orwell e entre um dos autores do posfácio, Eric Fromm, a função desse autor é representada a partir de *Role*.

Figura 5 – Descrição dos relacionamentos e da função do autor de um pós-fácio com *Role*



Fonte: Elaboração da autora com base em Wallis (2015).

Na figura 5, o livro (*Book*) *Nineteen Eight-Four* é representado pela classe *Book*, conectada a classe superior *CreativeWork*, a propriedade *exampleOfWork* indica que existem outras edições da mesma obra disponíveis. A classe *Role* é inserida entre essas descrições, representando a contribuição de Eric Fromm (*Person*) como escritor do pós-fácio (*Writer of afterwords*) do livro.

Para auxiliar a interpretação precisa desses relacionamentos e funções pelos mecanismos de busca, Wallis (2015) recomenda a utilização de URIs para vinculação das entidades, visando evitar ambiguidades e possibilitar a conexão com outras fontes de informação. O autor ressalta que esse mecanismo é flexível e pode ser utilizado em diversos casos, entretanto, é necessário certificar-se que se não há maneiras mais adequadas, indicadas pelo *Schema.org* para representar determinados conceitos, caso não haja, podem ser propostas novas extensões aos administradores do vocabulário.

Conforme *Schema.org* (2021d) a estrutura do vocabulário não foi projetada com o intuito de criar uma “ontologia global”, pois, o foco inicial do projeto consistiu em representar as entidades mais relevantes para os fundadores do *Schema.org*, visando fornecer um tratamento diferenciado aos dados e otimizar a interpretação dos mecanismos de busca. No entanto, a partir da colaboração de equipes heterogêneas, novos mecanismos de extensão foram introduzidos, expandindo o escopo do vocabulário.

Devido à base extensível para dados estruturados, o modelo de dados do *Schema.org* é caracterizado como pragmático e flexível. A flexibilidade do vocabulário pode ser observada na configuração das classes e propriedades, descritas em *Schema.org* (2021d, não paginado, tradução nossa)

[...] algo pode ser simultaneamente um Livro e um Produto e ser descrito de forma útil com propriedades de ambos os tipos. É útil, mas não obrigatório, que os tipos relevantes sejam incluídos em tal descrição. Essa flexibilidade permite que os tipos do *Schema.org* sejam desenvolvidos com alguma descentralização e que o vocabulário seja reutilizado e combinado de maneiras úteis. Quando listamos os tipos esperados associados a uma propriedade (ou vice-versa), pretendemos indicar as principais formas como esses termos serão combinados na prática.

Sendo assim, os relacionamentos entre as classes, subclasses e propriedades são considerados como diretrizes para descrição, e não como regras rígidas, permitindo que o *Schema.org* seja reutilizado e adaptado conforme as necessidades de cada contexto informacional.

Um dos principais benefícios proporcionados pelo *Schema.org* está no enriquecimento semântico dos dados. As páginas estruturadas com metadados do vocabulário nos formatos JSON-LD, RDFa ou Microdados viabilizam a interpretação eficiente dos mecanismos de busca. Dessa maneira, conteúdo das páginas da *Web* pode ser recuperado como um resultado visualmente enriquecido, integrado a tecnologias, como, *Rich Snippets* e *Knowledge Graph*, abordadas na seção seguinte.

5.2 Resultados enriquecidos

A *Google* é uma empresa multinacional que além de propiciar a busca por informações na *Web*, oferece uma gama de produtos e serviços *online*. A missão da *Google* consiste em “[...] organizar as informações do mundo para serem universalmente acessíveis e úteis para todos.” (GOOGLE, 2021a, não paginado, tradução nossa).

Os dados estruturados são fundamentais para que a *Google* colete, interprete e classifique o conteúdo da *Web*. A partir de padrões, esquemas e vocabulários compartilhados, são fornecidos resultados mais precisos e interativos, logo, mais atraentes do que uma lista tradicional de resultados de busca.

O mecanismo de busca recomenda a utilização do *Schema.org* para aperfeiçoar os resultados de busca, nem todos os elementos do vocabulário são exigidos pela pesquisa *Google*, entretanto, podem ser úteis para outros serviços, ferramentas e plataformas (GOOGLE SEARCH CENTRAL, 2021a). Para que os dados sejam incluídos na busca aprimorada da *Google* devem ser estruturados nos formatos: JSON-LD, RDFa ou Microdados. Riley (2017, p. 21, tradução nossa) ressalta que o *Schema.org* promove:

[...] a codificação de pequenos, mas vitais *bits* de conhecimento em páginas da *Web* [...]. São utilizadas tecnologias existentes que levam a *Web* a codificar os blocos de construção do conhecimento humano de uma forma estruturada e legível por máquina.

No mesmo ano de lançamento do *Schema.org*, em 2011, a *Rich Snippets* da *Google* foi a primeira aplicação a identificar os metadados do vocabulário (GUHA; BRICKLEY; MACBETH, 2016). Essa tecnologia permite que o conteúdo codificado com anotações em *Schema.org* se destaque visualmente apresentando as principais informações sobre um recurso informacional, diretamente na lista de resultados de busca. Segundo Google Search Central (2021d) existem sete categorias gerais de resultados, apresentadas no Quadro 8.

Quadro 8 - Categorias gerais de resultados de busca

Categorias	Descrição
<i>Plain blue link</i>	Apresenta três elementos: título da página, URL e um <i>snippet</i> que resume o conteúdo.
<i>Enhancement</i>	Permite a inserção de elementos complementares, como: <i>breadcrumbs</i> , <i>sitelinks</i> , <i>searchbox</i> e logotipos corporativos.
<i>Rich result</i>	Exibe elementos gráficos, como: avaliações, imagens em miniatura, etc. Pode aparecer individualmente ou em um carrossel de resultados.
<i>Knowledge panel entry</i>	Apresenta um conjunto de informações de várias páginas, aprimoradas com elementos gráficos e reconhece qualquer metadado do <i>Schema.org</i> .
<i>Featured snippet</i>	Exibe respostas objetivas para as perguntas do usuário, destacando o trecho de um <i>site</i> com a resposta.
<i>OneBox result</i>	Exibe informações em uma linha ou em uma ferramenta. Ex: <i>local time OneBox</i> e <i>translation OneBox</i> .
<i>Discover</i>	Não integra a lista de resultados de busca, mas aparece em uma lista de rolagem personalizada conforme o perfil do usuário. Essa funcionalidade é acessível em sistemas <i>Android</i> .

Fonte: Google Search Central (2021d).

Vale ressaltar que a ordenação e a forma de apresentação dos resultados enriquecidos podem variar para cada usuário, pois, o design de apresentação é

dinâmico, os resultados são apresentados em designs e posições diferentes conforme o dispositivo, país, horário de consulta, entre outros fatores (GOOGLE SEARCH CENTRAL, 2021c).

Para que as informações sejam identificadas e apresentadas nos resultados enriquecidos, a *Google* fornece diretrizes técnicas, diretrizes para *webmasters* e diretrizes gerais para dados estruturados. Essas diretrizes especificam as classes e propriedades obrigatórias do *Schema.org* para que o mecanismo de busca inclua os dados de uma página nos resultados enriquecidos. Além dos elementos obrigatórios, podem ser adicionados elementos complementares que favorecem a contextualização de um recurso, proporcionando experiências de busca mais satisfatórias ao usuário.

Na Figura 6, é possível visualizar uma das formas de apresentação de um *Rich Result*, esse resultado foi obtido a partir da busca pelo livro *Nineteen Eighty-Four* do autor George Orwell, na interface exibida em dispositivos móveis.

Figura 6 - Resultado enriquecido para *Nineteen Eighty-Four*



Fonte: Google (2021).

Além da apresentação de informações como título, autor, sinopse, data de publicação e gênero, é possível acessar parte do recurso, visualizar as avaliações dos usuários e mais detalhes sobre a obra. É possível ainda iniciar o processo de aquisição de um item diretamente nos resultados de busca. Para que isso seja possível, as lojas e bibliotecas devem utilizar as seguintes classes do *Schema.org*: *ReadAction* para compra e *BorrowAction* para empréstimo (GOOGLE SEARCH CENTRAL, 2021c).

A opção “compra” disponibiliza um campo para inserção do CEP do usuário, possibilitando a localização da biblioteca mais próxima para o empréstimo do item, abaixo desse campo, é exibida uma lista de lojas que possuem o item em estoque, permitindo a compra. Após selecionada essa opção, habilita-se um campo para localização da biblioteca mais próxima ao usuário para o empréstimo do item, como apresenta a Figura 7.

Figura 7 - Resultado enriquecido com a opção de empréstimo



Fonte: Google (2021).

Uma das funcionalidades habilitadas pelos metadados do *Schema.org* é o *snippet* de avaliação, um pequeno trecho de classificação construído a partir de uma média de classificações de vários revisores (GOOGLE SEARCH CENTRAL, 2021b).

Essa categoria de *snippet* pode ser empregada na descrição dos seguintes categorias de recursos: livros, cursos, eventos, instruções, empresas, filmes, produtos, receitas e aplicativos de *software* (GOOGLE SEARCH CENTRAL, 2021b). Na Figura 8 é apresentado um exemplo de resultado de busca aprimorado com o *snippet* de avaliação, o resultado corresponde a busca pelo livro Histórias extraordinárias de Edgar Allan Poe.

Figura 8 - *Snippet* de avaliação para Histórias extraordinárias



Fonte: Google (2021).

O resultado apresentado na Figura 8, contém o *link* para loja virtual, onde o recurso está disponível para compra, acompanhado de uma breve descrição, por último, a avaliação é representada por cinco estrelas. Os *snippets* de avaliação também podem ser habilitados para as seguintes classes do *Schema.org*: *CreativeWorkSeason*, *CreativeWorkSeries*, *Episode*, *Game*, *MediaObject*, *MusicPlaylist*, *MusicRecording* e *Organization* (GOOGLE SEARCH CENTRAL, 2021b).

Em relação as diretrizes da *Google* para habilitação dessas funcionalidades para as bibliotecas, é recomendado que sejam criados *feeds*. No sistema hierárquico do *Schema.org* a entidade *Library* é uma subclasse de *Organization* que inclui a subclasse *LibrarySystem*, utilizada para representar uma rede colaborativa de bibliotecas. Segundo Google Search Central (2021b) para representar as bibliotecas, a classe *LibrarySystem* é obrigatória, portanto, deve ser utilizada mesmo se a biblioteca não fizer parte de uma rede.

O conjunto de dados codificados com os metadados do *Schema.org*, constitui um *Datafeed*, onde são fornecidas informações sobre uma ou mais entidades em uma

única estrutura de visualização, denominada *feed*. Para que as bibliotecas ofereçam a opção de empréstimo nessa estrutura, devem codificar os dados em dois arquivos independentes para as entidades *Book* e *LibrarySystem*. Cada arquivo deve conter uma única entidade no nível raiz (GOOGLE SEARCH CENTRAL, 2021c). A Figura 9 apresenta um exemplo de descrição para *Book* em um *feed*.

Figura 9 - Exemplo de estrutura para entidade *Book* em um *feed*

```

{
  "@context": "https://schema.org",
  "@type": "DataFeed",
  "dataFeedElement": [
    {
      "@context": "https://schema.org",
      "@type": "Book",
      "@id": "http://example.com/work/the_catcher_in_the_rye",
      "url": "http://example.com/work/the_catcher_in_the_rye",
      "name": "The Catcher in the Rye",
      "author": {
        "@type": "Person",
        "name": "J.D. Salinger"
      },
      "sameAs": "https://en.wikipedia.org/wiki/The_Catcher_in_the_Rye",
      "workExample": [
        {
          "@type": "Book",
          "@id": "http://example.com/edition/the_catcher_in_the_rye_paperback",
          "isbn": "9787543321724",
          "bookEdition": "Mass Market Paperback",
          "bookFormat": "https://schema.org/Paperback",
          "inLanguage": "en",
          ...
        },
        ...
      ]
    },
    ...
  ]
},
"dateModified": "2018-09-10T13:58:26.892Z"
}

```

Fonte: Google Search Central (2021c).

O exemplo na Figura 9 apresenta as propriedades obrigatórias para dados estruturados em um *feed*, como: *@context*, *@type*, *@dataFeedElement* e *@dateModified*, todas as entidades de *Book* devem ser listadas no campo *dataFeedElement*. Para vincular um *feed* aos dados das páginas da *Web*, é fundamental atribuir um URI para cada entidade, como ISBN ou o identificador da organização (GOOGLE SEARCH CENTRAL, 2021c). No exemplo apresentado na Figura 10 é possível visualizar um exemplo de descrição para entidade *LibrarySystem* em formato de *feed*.

Figura 10 – Descrição de *LibrarySystem* em um *feed*

```

{
  "@context": "https://schema.org",
  "@type": "DataFeed",
  "dataFeedElement": [
    {
      "@context": "https://schema.org",
      "@type": "LibrarySystem",
      "@id": "http://example.com/library-systems/100",
      "name": "Santa Clara County Library District",
      "additionalProperty": [
        {
          "@type": "PropertyValue",
          "name": "librarytype",
          "value": "public"
        }
      ],
      ...
    },
    ...
  ],
  "dateModified": "2018-09-10T13:58:26.892Z"
}

```

Fonte: Google Search Central (2021c).

Existem ferramentas que permitem a verificação das páginas estruturadas conforme os critérios estabelecidos pelo mecanismo de busca, tais como, *Data Feed validation tool*, *Rich Results Test* e *Schema Markup Validator* (SMV). Antes de serem enviados para a *Google*, os arquivos de *feed* podem ser revisados com a ferramenta *Data Feed validation tool*, posteriormente, a equipe de suporte da *Google* avalia a qualidade dos arquivos para publicação.

Existem ferramentas para revisão e validação das páginas estruturadas, como a *Rich Results Test* que indica se a codificação está em conformidade com os critérios para resultados enriquecidos, apontando erros de codificação. Para validar as anotações em *Schema.org*, sem necessariamente considerar os requisitos específicos da pesquisa *Google*, pode ser utilizada a ferramenta *Schema Markup Validator*. Essa ferramenta indica se os formatos JSON-LD, RDFa ou Microdados foram estruturados adequadamente, e se as classes e propriedades do *Schema.org* foram combinadas de modo a expressar o significado desejado (LEVERING, 2021).

Além dos resultados em *Rich Snippets*, os metadados do *Schema.org* alimentam o *Knowledge Graph*, um banco de dados que reúne e apresenta as principais informações sobre entidades, como: logotipos, contatos e informações sociais (GUHA; BRICKLEY; MACBETH, 2016). O *Knowledge Graph* coleta, estabelece relacionamentos e organiza sistematicamente fatos sobre entidades,

segundo Google (2021b) as informações são coletadas de fontes públicas ou de proprietários confiáveis, são geralmente exibidas conforme a estrutura visual, denominada *Knowledge panel*. Essa estrutura é apresentada conforme o exemplo na Figura 11 que apresenta informações sobre escritor Edgar Allan Poe.

Figura 11 – Informações sobre Edgar Allan Poe em um *Knowledge panel*

The image shows a Google search interface for 'Edgar Allan Poe'. On the right side, a Knowledge Panel is displayed, featuring a portrait of Poe, a grid of smaller images, and a list of key facts: birth (1809, Boston), death (1849, Baltimore), and notable works like 'The Raven' and 'The Hound in the Moon'. Below the facts are sections for 'Livros' (books) and 'Citações' (quotes). On the left side, search results are shown, including a blog post titled '7 curiosidades sobre Edgar Allan Poe' and several YouTube videos with thumbnails and titles related to Poe's life and works.

Fonte: Google (2021).

Como pode ser visualizado na Figura 11, além dos resultados enriquecidos, no canto direito são apresentadas as principais informações sobre Edgar Allan Poe, tais como: nome, data e local de nascimento, data de falecimento, poemas, peças teatrais e miniaturas de livros do escritor.

Para que essas informações sejam compiladas no *Knowledge Graph*, é necessário que as páginas contenham dados estruturados a partir de vocabulários compartilhados entre os mecanismos de busca, como o *Schema.org*. Wallis (2016)

afirma que os buscadores, *Google*, *Bing*, *Yahoo* e *Yandex*, apoiam a iniciativa *Schema.org* e participam ativamente do desenvolvimento do vocabulário, visando favorecer a integração de seus dados ao *Knowledge Graph*. Godby, Wang e Mixer (2015, não paginado, tradução nossa) ressaltam que

[...] uma entidade do mundo real, ao contrário de uma lista de documentos, é um objeto que pode ser descrito em várias fontes de dados ou repositórios de conhecimento. Se a entidade for única e publicamente identificada, essas fontes de dados podem ser mescladas com mais facilidade. Esta é a chave para o bom funcionamento das aplicações da *Web Semântica*.

Os esforços investidos na construção da *Web Semântica* beneficiam consideravelmente a sociedade, no que tange ao acesso a informações de qualidade e na obtenção de respostas instantâneas, objetivas e facilmente compreensíveis para seres humanos e agentes de *software*. Sendo assim, as bibliotecas podem se beneficiar de tecnologias semânticas para representação, tendo como base os princípios *Linked Data* e as diretrizes do *Schema.org*.

Desse modo, viabiliza-se a exposição e o enriquecimento semântico dos dados dos catálogos digitais na *Web*. Os registros podem ser recuperados diretamente na lista de resultados de busca, e ainda há a possibilidade de serem exibidos como resultados enriquecidos. Gilliland (2016, não paginado, tradução nossa) destaca que “As comunidades de informação estão cientes que quanto mais altamente estruturado é um objeto de informação, mais essa estrutura pode ser explorada para pesquisar, manipular e se relacionar com outros objetos.”

Nesse viés, a seção seguinte discorre sobre os metadados da classe *CreativeWork* do *Schema.org* em relação à representação do conteúdo das bibliotecas.

6 A REPRESENTAÇÃO A PARTIR DA CLASSE *CREATIVEWORK*

Tendo em vista a catalogação como a atividade que viabiliza a identificação, localização e a recuperação de recursos bibliográficos em um catálogo, o acesso e a descoberta desse conteúdo depende das tecnologias empregadas na estruturação dos dados. De modo geral, o ponto de partida para busca por informações é a *Web*, são os mecanismos de busca que direcionam o usuário as informações desejadas e permitem a descoberta de recursos relacionados.

Conforme mencionado anteriormente, as bibliotecas utilizam estruturas de *software* e protocolos que não permitem a interoperabilidade entre os catálogos e os mecanismos de busca da *Web*. Em vista disso, os catálogos são ambientes restritos, essa característica reduz consideravelmente a possibilidade de descoberta do conteúdo por usuários potenciais.

Wallis (2013) explana que a função de um mecanismo de busca é identificar se a informação solicitada existe e exibir uma descrição suficientemente informativa nos resultados, direcionando o usuário para o recurso mais adequado, enquanto nas bibliotecas “[...] as interfaces de pesquisa são fontes inerentemente pobres de tais informações, com os rastreadores da *web* tendo que diferenciar o texto do assunto da página.” (WALLIS, 2013, não paginado, tradução nossa).

Observou-se que na literatura sobre a adequação dos catálogos ao contexto atual, a iniciativa *Linked Data* é discutida com frequência, no sentido de enriquecer os dados e ampliar o alcance do conteúdo das bibliotecas na *Web*. Segundo Neish (2015) um dos principais interesses das bibliotecas em relação ao *Linked Data* está em tornar os dados dos catálogos mais detectáveis. Nesse aspecto, o *Schema.org* é uma tecnologia relevante, pois permite a inserção de informações estruturadas nas páginas da *Web*, viabilizando o enriquecimento semântico dos dados e a interpretação dos mecanismos de busca.

O *Schema.org* é frequentemente associado ao *Linked Data*, segundo Neish (2015) uma das maneiras mais simples para publicação de dados conectados, consiste na utilização de ontologias amplamente adotadas na *Web*, como, *Schema.org* ou *Open Graph Protocol*. Conforme Godby, Wang e Mixer (2015) embora a principal proposta do *Schema.org* seja favorecer a recuperação de informações pelos mecanismos de busca, o vocabulário também é amplo, extensível e compatível com os objetivos do *Linked Data* (GODBY; WANG; MIXTER, 2015).

Um dos pontos fortes do *Schema.org* consiste no alinhamento do vocabulário com as regras lógicas para descrição na *Web Semântica* e a compatibilidade com os formatos, RDFa e *JavaScript Object Notation for Linked Data* (GUHA; BRICKLEY; MACBETH, 2016). Segundo Nogales *et al.* (2016) o *Schema.org* compartilha do objetivo de estruturar os dados e viabilizar o estabelecimento de conexões a partir de *links*. Segundo os autores, “Como o *Schema.org* é um tipo de ontologia, tem o potencial de se tornar uma ponte para a *Web* de dados conectados.” (NOGALES *et al.*, 2016, p. 90, tradução nossa). Wallis (2022, p. 40, tradução nossa) reitera esse argumento, afirmando que o *Schema.org* “É um vocabulário aberto para a *Web*, um vocabulário *Linked Data* (embora eles não falem muito sobre isso), baseado em RDF.”

Sendo assim, além de tornar os dados mais acessíveis aos mecanismos de busca e favorecer a obtenção de resultados enriquecidos, as classes, propriedades e relacionamentos do vocabulário permitem a conexão entre dados estruturados de fontes de informação distintas.

Cabe ressaltar que o *Schema.org* é um vocabulário genérico, portanto, não abrange suficientemente os diferentes níveis de detalhamento exigidos para descrição em domínios específicos, como no caso das bibliotecas. Segundo Godby (2016), pesquisadora sênior da OCLC, as comunidades de bibliotecas não consideravam o *Schema.org* viável para descrição bibliográfica, devido à superficialidade do vocabulário e o enfoque no setor comercial. Ainda segundo Godby (2016) para os pesquisadores da OCLC, o *Schema.org* poderia suprir apenas parcialmente às necessidades curatoriais das bibliotecas, pois, o vocabulário parecia muito focado em produtos comerciais.

O *Schema.org* é utilizado no setor de *e-commerce* em técnicas de *Search Engine Optimization* (SEO), cujo objetivo consiste em otimizar o posicionamento de um *site* nas páginas de resultados de busca orgânicos, isto é, naqueles não patrocinados, ou pagos. As técnicas de SEO em conjunto com estratégias de *marketing* auxiliam na classificação e posicionamento dos *links* nos resultados de busca, aumentando as chances de acesso, com o aumento do tráfego nos *sites*, as vendas podem ser impulsionadas. Segundo Price (2019) não há evidências de que o *Schema.org* tenha de fato influência no *ranking* de resultados orgânicos, entretanto, os *rich snippets* permitem que os resultados se destaquem, conseqüentemente as taxas de cliques tendem a aumentar.

No contexto das bibliotecas, não é comum a aderência de padrões e vocabulários genéricos, como o *Schema.org*. Entretanto, com a constante evolução da *Web*, as bibliotecas e outros centros de informação são direcionados para uma mudança de perspectiva em relação à associação dos instrumentos tradicionais de representação com tecnologias desenvolvidas fora do universo bibliográfico. Nesse sentido, Fox (2014) argumenta que apesar das motivações e necessidades distintas entre o setor comercial e as bibliotecas, ambos podem se beneficiar das tecnologias sofisticadas desenvolvidas no âmbito das companhias de mecanismos de busca.

No *Schema.org*, a descrição de obras pode ser realizada a partir dos metadados da classe *CreativeWork*, os tipos de recursos contemplados nessa classe são compatíveis com o conteúdo dos catálogos digitais. A classe compreende metadados para descrição de recursos, como: livros, artigos, trabalhos acadêmicos, pinturas, esculturas, manuais, fotografias, gravações musicais, filmes, séries, jogos, entre outros.

No sistema hierárquico do vocabulário, as classes principais herdam um conjunto de propriedades da classe superior, *Thing*. Na versão atual do *Schema.org*, *CreativeWork* herda 12 propriedades e tipos esperados de *Thing*, apresentadas no Quadro 9.

Quadro 9 - Propriedades de *CreativeWork* herdadas de *Thing*

Propriedades de <i>Thing</i>	Tipos esperados
<i>additionalType</i>	<i>URL</i>
<i>alternateName</i>	<i>Text</i>
<i>Description</i>	<i>Text</i>
<i>disambiguatingDescription</i>	<i>Text</i>
<i>Identifier</i>	<i>PropertyValue or Text or URL</i>
<i>Image</i>	<i>ImageObject or URL</i>
<i>mainEntityOfPage</i>	<i>CreativeWork or URL</i>
<i>Name</i>	<i>Text</i>
<i>potentialAction</i>	<i>Action</i>
<i>SameAs</i>	<i>URL</i>
<i>SubjectOf</i>	<i>CreativeWork or Event</i>
<i>url</i>	<i>URL</i>

Fonte: Schema.org (2021j).

Em relação as propriedades apresentadas no Quadro 9, *additionalType* indica uma classe adicional para um recurso, geralmente utilizada para inclusão de classes mais específicas provenientes de vocabulários externos; *alternateName* corresponde a um nome alternativo atribuído a um recurso; *description* corresponde a descrição de um recurso; *disambiguatingDescription* é uma subpropriedade de *description*, utilizada na desambiguação de descrições de recursos semelhantes; *identifier* refere-se a um

identificador único, como um URL, ISBN, etc.; *image* corresponde a imagem que representa um recurso; *mainEntityOfPage* indica uma página ou obra, onde o recurso descrito atua como entidade principal; *name* corresponde ao nome/título de um recurso; *potentialAction* refere-se a uma ação potencial, isto é, existe a possibilidade de ser realizada; *sameAs* indica o URL de uma página da *Web* que identifica um recurso (ex: página da Wikipédia do recurso); *subjectOf* indica uma obra ou evento relacionado ao recurso descrito; *url* corresponde ao endereço eletrônico para localização do recurso. Além das propriedades herdadas de *Thing*, *CreativeWork* abrange mais 106 propriedades, apresentadas no Quadro 10.

Quadro 10 - Propriedades e tipos esperados de *CreativeWork*

Propriedades de <i>CreativeWork</i>	Tipos esperados
<i>About</i>	<i>Thing</i>
<i>Abstract</i>	<i>Text</i>
<i>AccessMode</i>	<i>Text</i>
<i>AccessModeSufficient</i>	<i>ItemList</i>
<i>AccessibilityAPI</i>	<i>Text</i>
<i>accessibilityControl</i>	<i>Text</i>
<i>AccessibilityFeature</i>	<i>Text</i>
<i>AccessibilityHazard</i>	<i>Text</i>
<i>AccessibilitySummary</i>	<i>Text</i>
<i>accountablePerson</i>	<i>Person</i>
<i>acquireLicensePage</i>	<i>CreativeWork</i> or <i>URL</i>
<i>aggregateRating</i>	<i>AggregateRating</i>
<i>alternativeHeadline</i>	<i>Text</i>
<i>archivedAt</i>	<i>URL</i> or <i>WebPage</i>
<i>assesses</i>	<i>DefinedTerm</i> or <i>Text</i>
<i>associatedMedia</i>	<i>MediaObject</i>
<i>audience</i>	<i>Audience</i>
<i>audio</i>	<i>audio Object</i> or <i>Clip</i> or <i>MusicRecording</i>
<i>author</i>	<i>Organization</i> or <i>Person</i>
<i>award</i>	<i>Text</i>
<i>character</i>	<i>Person</i>
<i>citation</i>	<i>CreativeWork</i> or <i>Text</i>
<i>comment</i>	<i>Comment</i>
<i>commentCount</i>	<i>Integer</i>
<i>conditionsOfAccess</i>	<i>Text</i>
<i>contentLocation</i>	<i>Place</i>
<i>contentRating</i>	<i>Rating</i> or <i>Text</i>
<i>contentReferenceTime</i>	<i>DateTime</i>
<i>contributor</i>	<i>Organization</i> or <i>Person</i>
<i>copyrightHolder</i>	<i>Organization</i> or <i>Person</i>
<i>copyrightNotice</i>	<i>Text</i>
<i>copyrightYear</i>	<i>Number</i>
<i>correction</i>	<i>CorrectionComment</i> or <i>Text</i> or <i>URL</i>
<i>countryOfOrigin</i>	<i>Country</i>
<i>creativeWorkStatus</i>	<i>DefinedTerm</i> or <i>Text</i>
<i>creator</i>	<i>Organization</i> or <i>Person</i>
<i>creditText</i>	<i>Text</i>
<i>dateCreated</i>	<i>Date</i> or <i>DateTime</i>
<i>dateModified</i>	<i>Date</i> or <i>DateTime</i>

<i>datePublished</i>	<i>Date or DateTime</i>
<i>discussionUrl</i>	<i>URL</i>
<i>editEIDR</i>	<i>Text or URL</i>
<i>editor</i>	<i>Person</i>
<i>educationalAlignment</i>	<i>AlignmentObject</i>
<i>educationalLevel</i>	<i>DefinedTerm or Text or URL</i>
<i>educationalUse</i>	<i>DefinedTerm or Text</i>
<i>encoding</i>	<i>MediaObject</i>
<i>encodingFormat</i>	<i>Text or URL</i>
<i>exampleOfWork</i>	<i>CreativeWork</i>
<i>expires</i>	<i>Date</i>
<i>funder</i>	<i>Organization or Person</i>
<i>genre</i>	<i>Text or URL</i>
<i>hasPart</i>	<i>CreativeWork</i>
<i>headline</i>	<i>Text</i>
<i>inLanguage</i>	<i>Language or Text</i>
<i>interactionStatistic</i>	<i>InteractionCounter</i>
<i>interactivityType</i>	<i>Text</i>
<i>interpretedAsClaim</i>	<i>Claim</i>
<i>isAccessibleForFree</i>	<i>Boolean</i>
<i>isBasedOn</i>	<i>CreativeWork or Product or URL</i>
<i>isFamilyFriendly</i>	<i>Boolean</i>
<i>isPartOf</i>	<i>CreativeWork or URL</i>
<i>keywords</i>	<i>DefinedTerm or Text or URL</i>
<i>learningResourceType</i>	<i>DefinedTerm or Text</i>
<i>license</i>	<i>CreativeWork or URL</i>
<i>locationCreated</i>	<i>Place</i>
<i>mainEntity</i>	<i>Thing</i>
<i>maintainer</i>	<i>Organization or Person</i>
<i>material</i>	<i>Product or Text or URL</i>
<i>materialExtent</i>	<i>QuantitativeValue or Text</i>
<i>mentions</i>	<i>Thing</i>
<i>offers</i>	<i>Demand or Offer</i>
<i>pattern</i>	<i>DefinedTerm or Text</i>
<i>position</i>	<i>Integer or Text</i>
<i>producer</i>	<i>Organization or Person</i>
<i>Provider</i>	<i>Organization or Person</i>
<i>Publication</i>	<i>PublicationEvent</i>
<i>Publisher</i>	<i>Organization or Person</i>
<i>PublisherImprint</i>	<i>Organization</i>
<i>publishingPrinciples</i>	<i>CreativeWork or URL</i>
<i>recordedAt</i>	<i>Event</i>
<i>releasedEvent</i>	<i>PublicationEvent</i>
<i>review</i>	<i>Review</i>
<i>schemaVersion</i>	<i>Text or URL</i>
<i>sdDatePublished</i>	<i>Date</i>
<i>sdLicense</i>	<i>CreativeWork or URL</i>
<i>sdPublisher</i>	<i>Organization or Person</i>
<i>size</i>	<i>DefinedTerm or QuantitativeValue or SizeSpecification or Text</i>
<i>sourceOrganization</i>	<i>Organization</i>
<i>spatial</i>	<i>Place</i>
<i>spatialCoverage</i>	<i>Place</i>
<i>sponsor</i>	<i>Organization or Place</i>
<i>teaches</i>	<i>DefinedTerm or Text</i>
<i>temporal</i>	<i>DateTime or Text</i>
<i>temporalCoverage</i>	<i>DateTime or Text or URL</i>
<i>text</i>	<i>Text</i>

<i>thumbnailUrl</i>	<i>URL</i>
<i>timeRequired</i>	<i>Duration</i>
<i>translationOfWork</i>	<i>CreativeWork</i>
<i>translator</i>	<i>Organization or Person</i>
<i>typicalAgeRange</i>	<i>Text</i>
<i>usageInfo</i>	<i>CreativeWork or URL</i>
<i>version</i>	<i>Number or Text</i>
<i>video</i>	<i>Clip or VideoObject</i>
<i>workExample</i>	<i>CreativeWork</i>
<i>workTranslation</i>	<i>CreativeWork</i>

Fonte: Schema.org (2021j)

A multiplicidade de propriedades de *CreativeWork* conferem detalhes para descrição de diversos tipos de recurso, são contemplados desde elementos essenciais para identificação, como: *author* (autor), *publisher* (editora) e *datePublished* (data de publicação), como metadados mais específicos referentes a direitos autorais, representados pelas propriedades: *license* (licença) que inclui propriedades mais específicas, como: *copyrightHolder* (detentor de direitos autorais) e *copyrightYear* (ano oficial dos direitos autorais).

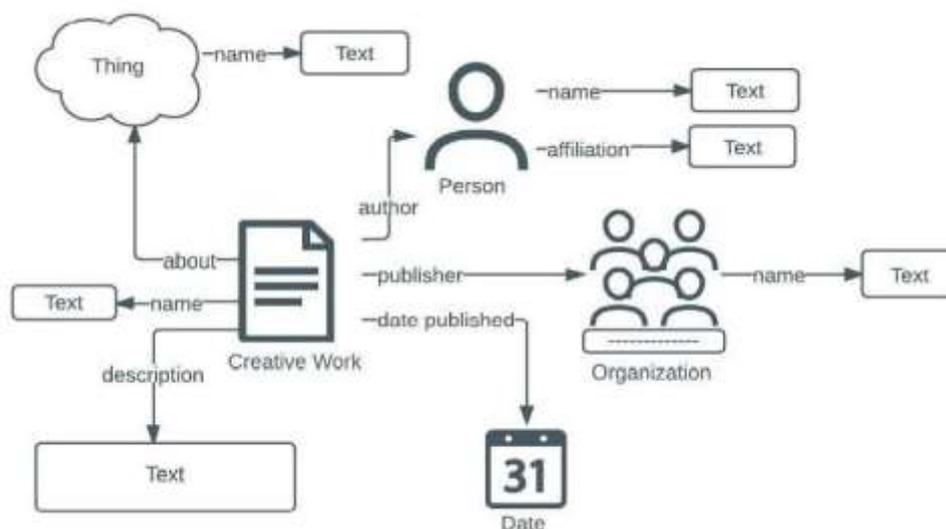
A classe ainda abrange metadados de acessibilidade em um conjunto de modos de acesso relacionados à percepção auditiva, tátil, textual e visual, representados pelas propriedades: *accessModeSufficient* (modos de acesso únicos ou combinados para compreender o conteúdo intelectual de um recurso); *accessibilityAPI* (recurso compatível com a API de acessibilidade referenciada) e *accessibilitySummary* (resumo legível por humanos com deficiências específicas, podendo expressar sutilezas de um conteúdo).

Além disso, são contemplados metadados específicos para representação de recursos educacionais, com propriedades, como: *educationalAlignment* (alinhamento com uma estrutura educacional estabelecida), *educationalLevel* (níveis educacionais), *educationalUse* (finalidade de um trabalho no contexto educacional), entre outras.

No *site* oficial do *Schema.org*, todas as classes e propriedades são descritas individualmente com a indicação dos tipos esperados e sub-propriedades, em alguns casos há exemplos práticos para codificação nos formatos: Microdados, RDFa e JSON-LD. Por ser um vocabulário genérico, os termos são de fácil interpretação, isto é, são compreensíveis para além do universo bibliográfico, favorecendo a reutilização dos metadados. Desse modo, diferentes tipos de recursos podem ser descritos e relacionados com as propriedades de *CreativeWork*.

Barker e Campbell (2014) apresentam o estabelecimento de possíveis relacionamentos entre as classes *Thing*, *CreativeWork*, *Person* e *Organization* e suas respectivas propriedades.

Figura 12 - Relacionamentos a partir de *CreativeWork*



Fonte: Barker e Campbell (2014, não paginado).

Na Figura 12 é possível visualizar as propriedades e respectivos tipos esperados vinculados às classes *Thing*, *CreativeWork*, *Person* e *Organization*. Essas propriedades fornecem as principais informações sobre um recurso e estabelecem relacionamentos entre entidades e recursos informacionais. Conforme Barker e Campbell (2014) se os metadados do *Schema.org* estiverem disponíveis, qualquer mecanismo de pesquisa será capaz de processá-los para otimizar sua interface de pesquisa, distinguindo entre recursos diferentes com o mesmo nome; permitindo que as informações mais relevantes sejam exibidas com destaque na página de resultados; possibilitando que os resultados sejam filtrados por propriedades como, preço, fornecedor ou data de publicação, além de propiciar *links* para localização de resultados relacionados ao mesmo assunto ou editor.

Segundo Godby (2016) apesar da resistência e ceticismo em relação a adoção do *Schema.org* pelas bibliotecas, no contexto da OCLC, o vocabulário foi considerado rico o suficiente para ser incluído nas discussões sobre a modernização da arquitetura dos dados. Conforme a autora, identificou-se que a classe *CreativeWork* poderia ser estruturada em triplas RDF, constituindo um modelo semântico para representação do conteúdo das bibliotecas.

Desse modo, foi estabelecida uma parceria entre a OCLC e o W3C, dando origem ao *Schema Bib Extend Community Group*, cuja função consiste em “[...] discutir e preparar propostas para estender os esquemas do *Schema.org* para representação aprimorada da marcação e compartilhamento de informações bibliográficas.” (WORLD WIDE WEB CONSORTIUM, 2021, não paginado, tradução nossa). O *Schema Bib Extend* é constituído por membros de bibliotecas, desenvolvedores de sistemas de bibliotecas, editores, universidades, cooperativas e consórcios (WALLIS, 2013). As extensões propostas pelo *Schema Bib Extend* são baseadas na ontologia *Bibliographic Ontology* (BIBO), desenvolvida para modelagem de dados bibliográficos em RDF conforme a lógica de descrição na *Web Semântica*, a BIBO fornece os principais conceitos e propriedades para descrever citações e referências bibliográficas (STRUCTURED DYNAMICS, 2013).

Visando expandir o vocabulário e fornecer maior detalhamento para descrição de recursos informacionais, o *Schema Bib Extend* criou novas propriedades para a classe *CreativeWork*. Dentre as possibilidades de aperfeiçoamento e ampliação da classe, identificou-se a necessidade de criar propriedades para descrição de obras publicadas em volumes ou publicadas por várias editoras, pois, segundo Wallis (2014) o *Schema.org* fornecia a base para descrição de livros, entretanto, apresentava lacunas para edições e volumes de periódicos, trabalhos em vários volumes, citações e relacionamentos entre uma obra e suas edições.

Com o propósito de preencher essas lacunas, foram adicionadas novas propriedades para descrição e estabelecimento de relacionamentos entre periódicos e artigos. Geralmente, artigos são publicados em números e volumes, um número pode abranger vários artigos relacionados, assim, foram incluídas as propriedades *PublicationVolume*, *PublicationIssue*, *volumeNumber* e *issueNumber*. As funções dessas propriedades são semelhantes a *seasonNumber* e *EpisodeNumber*, utilizadas para representar recursos de TV ou rádio (WALLIS, 2014). Essas propriedades identificam a posição de uma temporada ou episódio em um conjunto ordenado.

Para representação de um recurso como parte de uma coleção, como no caso de um conjunto de artigos em um periódico, foram adicionadas as propriedades *hasPart* e a propriedade inversa *isPartOf* (GODBY, 2016). Em relação a representação de obras publicadas em várias edições, foram incluídas as propriedades *workExample* e *exampleOfWork*, baseadas em FRBR e RDA (GODBY, 2016).

No que tange às contribuições do *Schema Bib Extend*, Wallis (2014) ressalta que esses esforços permitiram o estabelecimento de uma base ampla para a descrição de recursos bibliográficos em detalhes suficientes para serem compreendidos pelos mecanismos de busca e se tornarem parte de seus *Knowledge Graphs* (WALLIS, 2014).

Muitas das extensões propostas pelo *Schema Bib Extend* foram incluídas oficialmente no *Schema.org*. Wallis e Scott (2014) afirmam que o *Schema.org* apoia o desenvolvimento de extensões para promover maior nível de granularidade aos metadados bibliográficos, complementando iniciativas como, RDA e BIBFRAME.

Nesse contexto, a OCLC é referência no desenvolvimento de projetos inovadores relacionados à representação no contexto das bibliotecas. Com base nos princípios *Linked Data* e nas diretrizes do *Schema.org*, a OCLC em parceria com o *W3C Schema Bib Extend* desenvolveu um modelo experimental para estruturação dos dados do *Worldcat*. Paralelamente, o modelo foi analisado em relação ao BIBFRAME, criado pela *Library of Congress* para substituir o formato MARC 21.

Para elucidar como *Schema.org* pode ser utilizado para descrição do conteúdo das bibliotecas, a próxima seção discorre sobre os projetos de modelagem de dados da OCLC, destacando os principais aspectos e resultados obtidos.

7 O SCHEMA.ORG NO CONTEXTO DAS BIBLIOTECAS

A OCLC gerencia o *Worldcat*, um catálogo que promove o acesso aos registros bibliográficos de uma rede de bibliotecas de vários países. Em conjunto com comitês, padrões e projetos comunitários, a OCLC dedica-se em explorar os benefícios do *Linked Data* para descoberta, controle de autoridade e desambiguação (ONLINE COMPUTER LIBRARY CENTER, 2022).

Segundo Fons, Penka e Wallis (2012) enquanto os projetos da OCLC relacionados ao *Linked Data* estavam em desenvolvimento, o *Schema.org* foi lançado, oportunizando o aperfeiçoamento desses projetos com enfoque na exposição dos dados estruturados na *Web*.

Em 2012 a OCLC desenvolveu um modelo experimental de dados vinculados derivados do *Schema.org* e codificados em RDFa, resultando na publicação de aproximadamente 300.000.000 registros MARC em HTML no *Worldcat* (GODBY; WANG; MIXTER, 2015). Conforme os princípios *Linked Data*, foram atribuídos URIs vinculados a conjuntos de dados RDF para representação da *Dewey Decimal Classification* (DDC), *Virtual International Authority File* (VIAF) e *Faceted Application of Subject Terminology* (FAST) (GODBY; WANG; MIXTER, 2015).

O DDC consiste em um sistema numérico de classificação e organização do conteúdo de um acervo por áreas do conhecimento, permitindo a identificação e localização dos itens. O VIAF combina arquivos de autoridade de nomes em um único serviço hospedado pela OCLC para vinculação de arquivos de autoridade amplamente utilizados (ONLINE COMPUTER LIBRARY CENTER, 2021a). O FAST é um esquema de cabeçalho de assunto enumerativo e facetado, derivado do *Library of Congress Subject Headings* (LCSH) (ONLINE COMPUTER LIBRARY CENTER, 2021b).

Godby, Wang e Mixter (2015) explanam que os registros MARC foram convertidos em RDF, a primeira etapa consistiu no estabelecimento de correspondências entre os campos e subcampos MARC com os metadados do *Schema.org*. Segundo os autores, os experimentos demonstraram que a conversão exige mais do que apenas renomear os campos MARC com metadados *Schema.org*, as classes, propriedades e relações devem ser explicitamente reconhecidas em um modelo formal. Sendo assim, a OCLC desenvolveu um modelo experimental baseado

em *Schema.org* com enfoque em seis entidades principais: *People, Places, Organizations, Concepts, Works* e *Objects* (GODBY; WANG; MIXTER, 2015).

Visando ampliar a cobertura do *Schema.org* para descrição bibliográfica, a OCLC em parceria com o *Schema Bib Extend* criou o *BiblioGraph.net* (BGN), um vocabulário de extensão que permite a compreensão dos conceitos definidos pelas bibliotecas para além da comunidade restrita de catalogadores (GODBY, 2016). O BGN é constituído por uma coleção de esquemas de metadados que descrevem conceitos, entidades e relacionamentos para coleções de patrimônio bibliográfico e cultural, incluindo bibliotecas e outros centros de informação (BIBLIOGRAPH.NET, 2019). O vocabulário é baseado nas entidades principais definidas pela OCLC e os relacionamentos são pautados no modelo conceitual FRBR.

Segundo Simionato (2017, p. 212) os modelos conceituais servem de base para o desenvolvimento de padrões e esquemas de metadados, permitindo a desconstrução dos registros bibliográficos e a representação dos dados a partir de relações “[...] isto significa que a estrutura fixa de preenchimento de campos e atributos modifica-se para relacionamentos entre entidades e atributos.”

O modelo FRBR abrange entidades, atributos e relações, as entidades são segmentadas em três grupos, descritos no Quadro 11.

Quadro 11 - Grupos FRBR

Grupo	Entidade
Grupo 1	Produtos de uma obra intelectual ou artística. Entidades: obra, expressão, manifestação e item.
Grupo 2	Responsabilidade intelectual da obra, a produção física e disseminação, ou a custódia das entidades do Grupo 1. Entidades: pessoas físicas e jurídicas (organizações).
Grupo 3	Assuntos de uma obra. Entidades: conceito (ideia ou noção abstrata), objeto (coisas materiais), evento (ação ou ocorrência) e lugar (localização).

Fonte: International Federation of Library Associations (2009).

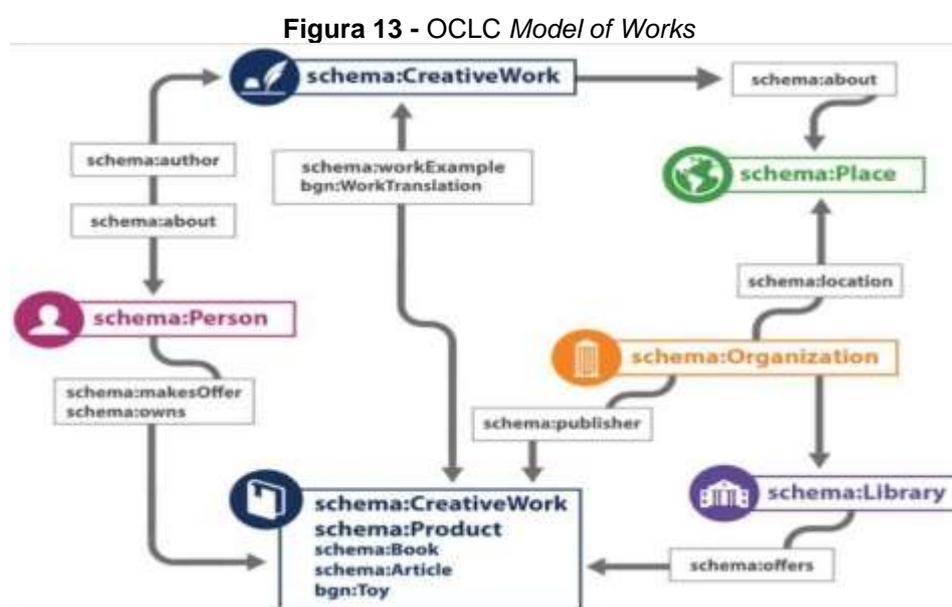
O modelo de dados da OCLC concentrou-se nas entidades do Grupo 1: obra, expressão, manifestação e item. A obra, refere-se a uma criação intelectual ou artística; a expressão trata da realização intelectual ou artística de uma obra; a manifestação consiste na personificação física da expressão de uma obra; e por fim, o item é considerado como um único exemplar de uma manifestação que reflete a sua forma física (INTERNATIONAL FEDERATION OF LIBRARY ASSOCIATIONS, 2009).

As entidades do Grupo 1 representam diversos aspectos dos objetos de interesse do usuário e relacionam-se com todos os elementos do universo bibliográfico. Os atributos representam as características das entidades, os usuários formulam suas buscas a partir dessas características, as entidades são conectadas a partir de relações (INTERNATIONAL FEDERATION OF LIBRARY ASSOCIATIONS, 2009).

Os conceitos definidos no modelo FRBR são considerados como abstratos, compreendidos apenas no domínio bibliográfico, em vista disso, representar as entidades do Grupo 1 a partir das classes, propriedades e relacionamentos do *Schema.org* é uma tarefa complexa. Do ponto de vista do *Schema.org*

[...] o conceito de 'Obra' é algo concreto que os usuários da *Web* procuram quando buscam no *Google* por livros, DVDs, etc. Mas para os especialistas em modelagem na comunidade bibliotecária, uma Obra é muito mais abstrata – é um esforço intelectual único com autoria atribuída, imperfeitamente representada em sistemas de biblioteca como um conjunto de descrições ou um conjunto de propriedades comuns entre edições, formatos ou traduções (GODBY, 2013, p. 11, tradução nossa).

No modelo de dados da OCLC, os conceitos do FRBR foram adaptados e harmonizados com o *Schema.org* e com vocabulário adicional, o BGN. Na Figura 13 é possível visualizar alguns dos possíveis relacionamentos entre as principais entidades definidas no modelo.



Fonte: Godby, Wang e Mixter (2015, não paginado).

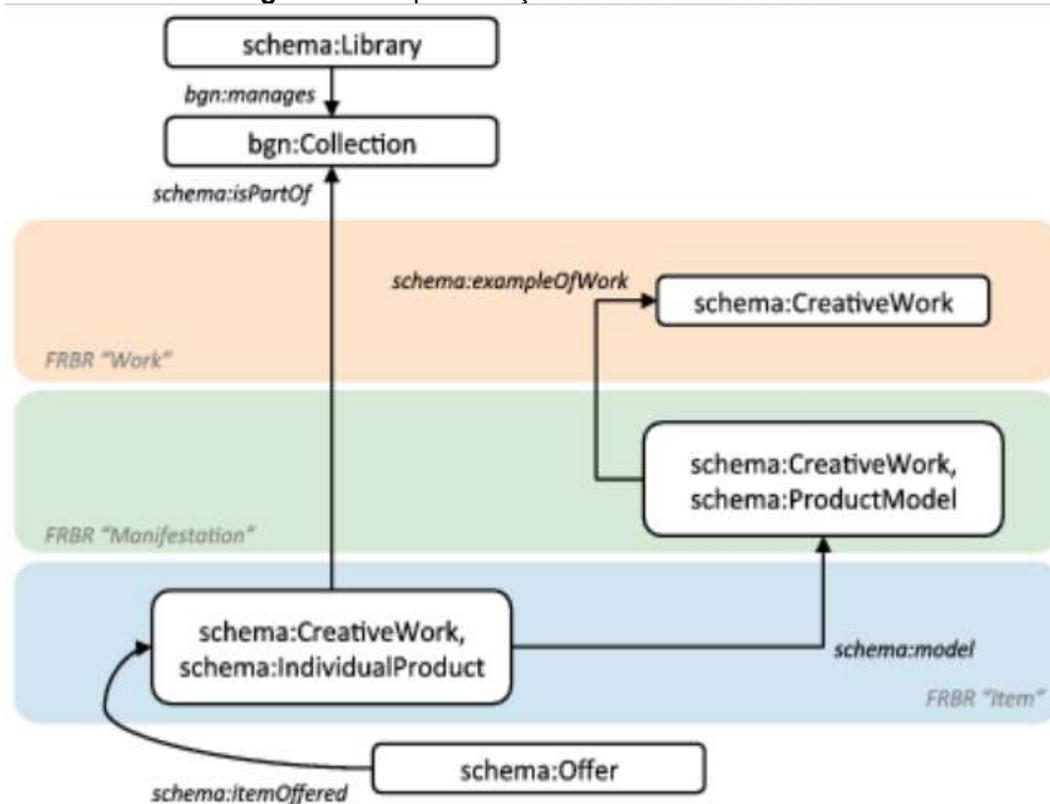
O esquema na Figura 13 apresenta a distinção entre uma obra (*Work*) como conteúdo e como um Produto ou Item. Uma obra pode ser a tradução de outra, pode derivar-se de uma obra, constituindo-se como uma nova obra e pode ainda ser representada como um Produto/Objeto disponível para empréstimo em uma biblioteca; o conteúdo pode ser sobre uma Pessoa, Lugar ou Organização (GODBY; WANG; MIXTER, 2015).

Os relacionamentos são estabelecidos, tanto entre as obras, como entre obras e entidades. A entidade Pessoa (*Person*) representa o autor responsável pela criação do conteúdo; a entidade Organização (*Organization*) representa a editora, responsável por transformar o conteúdo em um Produto (*Product*); uma biblioteca também é caracterizada como Organização, responsável por adquirir e disponibilizar o Produto via empréstimo; as entidades podem estabelecer relacionamentos entre si, como entre uma Organização e um Lugar (*Place*) (GODBY; WANG; MIXTER, 2015).

No que tange a representação das manifestações de uma obra, estas podem ser conectadas com as propriedades *exampleOfWork* e *workExample*, baseadas em RDA e FRBR. A manifestação de uma obra corresponde a personificação física da mesma, conforme International Federation of Library Associations (2016, p. 16) “Uma manifestação pode conter uma coleção de obras, uma obra individual ou uma parte componente de uma obra.” As manifestações são individualizadas a partir de URIs, como: ISBN, ISSN, DOI, entre outros.

No *Schema.org*, uma obra enquanto conteúdo é diferenciada ontologicamente de uma obra como Item ou Produto, eliminando a necessidade de adicionar ao modelo uma classe como “*frbr:Manifestation*” (GODBY; WANG; MIXTER, 2015). A representação da manifestação de uma Obra como Produto/Item pode ser expressa com as propriedades: *ProductModel* e *IndividualProduct*. Na Figura 14 é possível visualizar a representação da entidade Obra como um Produto.

Figura 14 - Representação de Obra como Produto



Fonte: Godby, Wang e Mixter (2015, não paginado).

Conforme a Figura 14, a entidade Biblioteca, representada como `schema:Library` é responsável por gerenciar uma Coleção (*Collection*), como indica a propriedade `bgn:manages`, derivada do vocabulário de extensão BGN; a propriedade `isPartOf` vincula uma obra (como Produto/Item) a uma coleção; se o Item for único (ex: diário), pode ser descrito com a propriedade `schema:IndividualProduct`; caso integre um conjunto de produtos, a conexão é realizada com a propriedade `schema:ProductModel`. A propriedade *Offer*, proveniente da ontologia *GoodRelations*, é utilizada para “oferecer” o produto ao usuário, essa propriedade é vinculada a hierarquia `schema:Product` pela propriedade `schema:itemOffered`.

A ontologia *GoodRelations* foi projetada para *e-commerce* e integrada oficialmente ao *Schema.org*, embora seja direcionada para produtos e serviços comerciais, proporcionou novas possibilidades de representação para as bibliotecas. Segundo Godby (2013) os metadados da *GoodRelations* são úteis para descrição de recursos bibliográficos, pois, é possível descrever um recurso como um objeto tangível que pode ser emprestado, comprado, vendido, negociado e administrado.

Na medida que novas extensões foram adicionadas ao *Schema.org* e que outros vocabulários alinhados aos princípios *Linked Data* se estabeleceram, a necessidade de utilização do BGN foi reduzida (BIBLIOGRAPH.NET, 2019). O BGN foi descontinuado em 2019 e substituído pela extensão “*bib.schema.org*” (BIBLIOGRAPH.NET, 2019).

Com o intuito de visualizar um exemplo concreto sobre os resultados do trabalho da OCLC, buscou-se no *Worldcat* pelo livro *Pride and Prejudice* da autora Jane Austen. A Figura 15 apresenta um dos resultados obtidos para essa busca.

Figura 15 – Resultado de busca no *Worldcat* para *Pride and Prejudice*



Fonte: Worldcat (2022).

Observa-se que o registro apresenta a capa do livro e fornece informações essenciais como, autor, editora, gênero, resumo e formato do item. São fornecidas ainda informações adicionais como a classificação representada por estrelas, os assuntos tratados na obra, itens similares e a opção de visualizar todas as edições e formatos disponíveis. Em relação à classificação por estrelas, o *Worldcat* está vinculado ao *GoodReads*, na parte inferior dos registros, há comentários provenientes dessa plataforma, onde os usuários podem contribuir com suas opiniões para avaliação dos recursos.

No que tange as edições e formatos derivados de uma obra, ao selecionar a opção “ver todas as edições e formatos” é apresentada uma lista com as opções disponíveis, indicando a editora, ano de publicação, idioma, gênero e formato do item (impresso ou *e-book*). Na figura 16 é possível visualizar a lista de edições disponíveis no *Worldcat* para *Pride and Prejudice*.

Figura 16 - Edições e formatos de *Pride and Prejudice* no *Worldcat*

	Título / Autor	Tipo	Língua	Data / Edição	Publicação
<input type="checkbox"/> 1.	 Pride and Prejudice by Jane Austen Autor: Jane Austen	imprimir livro	inglês	2010	Broughton- in-Furness : Coordination Group Publishers Ltd : [distributor] ; Coordination Group Publishers
<input type="checkbox"/> 2.	 Pride and Prejudice by Jane Austen Autor: Jane Austen	e-book : Documento	inglês	2007	Global Media
<input type="checkbox"/> 3.	 Pride and Prejudice by Jane Austen Autor: Jane Austen	e-book : Documento	inglês	2006	[Place of publication not identified] Global Media
<input type="checkbox"/> 4.	 Pride and Prejudice by Jane Austen Autor: Jane Austen	imprimir livro : Ficção	inglês	1998	New York : Penguin Books
<input type="checkbox"/> 5.	 Pride and Prejudice by Jane Austen Autor: Jane Austen	imprimir livro	inglês	1978	[Place of publication not identified] : Heron Books

Fonte: Worldcat (2022).

Com base no modelo experimental da OCLC, explanado por Godby, Wang e Mixter (2015), compreende-se que os formatos e edições apresentados na Figura 16 correspondem às manifestações de uma obra. A figura apresenta uma lista com as edições da obra de Jane Austen, publicadas em anos e formatos distintos, permitindo ao usuário selecionar a opção desejada. Godby, Wang e Mixter (2015) ressaltam que muitas das características dos recursos publicados por editoras comerciais podem ser descritas com propriedades do *Schema.org*, entretanto, é necessário utilizar extensões para descrição de obras específicas, como manuscritos exclusivos, arquivos, teses ou coleções digitalizadas.

No que se refere a utilização do *Schema.org* nos registros do *Worldcat*, foram identificadas duas marcações no código fonte do registro apresentado na Figura 15, sobre uma das edições de *Pride and Prejudice*. Na figura 17 é apresentado o recorte desse código fonte, indicando em destaque as marcações identificadas.

Figura 17 - Código fonte de *Pride and Prejudice* no Worldcat

```

</head>

<body id="worldcat" typeof="http://schema.org/WebPage">
<div property="http://schema.org/about" resource="http://www.worldcat.org/oclc/664391841"/>
<!-- BEG div-skip -->
<div id="div-skip">
<a href="#content"></a>
</div>
<!-- END div-skip --><!-- BEG div-javascript -->
<noscript>
<div id="div-javascript">
Seu browser não está habilitado para JavaScript. Algumas funcionalidades do WorldCat não estarão disponíveis.
</div>
</noscript>
<!-- END div-javascript --><div id="doc3" class="yui-t7" style="padding: 8px 4px">

  <div id="hd">
    <!-- stuff from navigation -->
<script type="text/javascript" src="/wcpa/re120220204/html/menuBar.js">&nbsp;</script>

```

Fonte: Worldcat (2022).

A Figura 17 apresenta parte do código fonte do registro de *Pride and Prejudice*, a primeira marcação identifica o catálogo como uma *webpage*, com a classe (*type*) “*schema.org/WebPage*”, a segunda marcação apresenta a propriedade “*schema.org/about*” com a numeração OCLC (664391841), como valor, isto é, o URI que identifica o recurso no catálogo.

Em Schema.org (2021h) são fornecidos exemplos para codificação de livros em catálogos de biblioteca, a Figura 18 apresenta os principais metadados recomendados pelo *Schema.org* para esse tipo de registro na estrutura JSON-LD, o exemplo refere-se ao livro *Le concerto* de Guy Ferchault.

Figura 18 - Estrutura para *Le concerto* em JSON-LD

@id	http://example.org/#record
@type	Book
name	Le concerto
offers	
@type	Offer
sku	780 R2
serialNumber	CONC91000937
offeredBy	
@id	http://library.anytown.gov.uk
@type	Library
name	Anytown City Library
itemOffered	#record
businessFunction	http://purl.org/goodrelations/v1#LeaseOut
availability	https://schema.org/InStock
author	Ferchault, Guy
additionalType	http://example.org/Product

Fonte: Schema.org (2021h).

Na estrutura apresentada na Figura 18, além dos metadados descritivos que identificam o recurso, como, *name* e *author*, há a identificação da biblioteca, *Anytown City Library*, a qual o recurso pertence, assim como metadados que classificam o recurso como um produto (*Product*) oferecido aos usuários pela biblioteca, a partir da classe *Offer* e da propriedade correspondente, *itemOffered*. Ainda sobre a Figura 18, para o registro de *Le concerto*, o *Schema.org* fornece exemplos de codificação nos formatos, Microdados, RDFa e JSON-LD. Na Figura 19 é possível visualizar a codificação em JSON-LD.

Figura 19 - Codificação em JSON-LD para *Le concerto*

```
<script type="application/ld+json">
{
  "@context": "https://schema.org/",
  "@id": "#record",
  "@type": "Book",
  "additionalType": "Product",
  "name": "Le concerto",
  "author": "Ferchault, Guy",
  "offers": {
    "@type": "Offer",
    "availability": "https://schema.org/InStock",
    "serialNumber": "DCN091000937",
    "sku": "780 R2",
    "offeredBy": {
      "@type": "Library",
      "id": "http://library.anytown.gov.uk",
      "name": "Anytown City Library"
    },
    "businessFunction": "http://purl.org/goodrelations/v1#LeaseOut",
    "itemOffered": "#record"
  }
}
</script>
```

Fonte: Schema.org (2021h).

Os registros codificados conforme os exemplos fornecidos pelo *Schema.org* podem ser identificados e localizados pelos mecanismos de busca, possibilitando a descoberta do conteúdo das bibliotecas pelos usuários, consequentemente, os acessos aos catálogos tendem a aumentar.

Ao realizar uma busca no mecanismo *Google* para o título "*Rules for a dictionary catalog*", livro escrito por Charles Ammi Cutter, identificou-se que um dos resultados correspondia a um item proveniente do *Worldcat*. A Figura 20 apresenta esse resultado, destacado em vermelho.

Figura 20 - Registro *Worldcat* recuperado como resultado de busca

https://quod.lib.umich.edu > moa ▾ Traduzir esta página
 Rules for a dictionary catalogue, by Charles A. Cutter ...
 Rules for a dictionary catalogue, by Charles A. Cutter 4. Form-entry f (for F). 5. Giving edition and imprint, with notes when necessary (for G). 6.

https://www.worldcat.org > oclc ▾ Traduzir esta página
Rules for a dictionary catalog, (Book, 1904) [WorldCat.org]
 [Charles A Cutter; W P Cutter; Worthington Chauncey Ford; ... Print book : National government publication : English : 4th ed. ... Dictionary catalogs.

As pessoas também pesquisaram

cutter rules for a dictionary catalogue	dictionary catalogue in library science
cutters rules	what is catalogue
charles ammi cutter	cutter library system

https://digital.library.unt.edu > ark: ▾ Traduzir esta página
 Rules for a Dictionary Catalog - UNT Digital Library
 1 de jun. de 2022 — Thumbnail image of item number 4 in: 'Rules for a Dictionary Catalog' ... Cutter, Charles A. (Charles Ammi), 1837-1903 1904. ... Edition: 4 ...

https://openlibrary.org > books > R... ▾ Traduzir esta página
 Rules for a dictionary catalog (1904 edition) - Open Library
 15 de jul. de 2010 — Rules for a dictionary catalog by Charles Ammi Cutter, 1904. Govt. Print. Off. edition, in English - 4th ed., rewritten.




Fonte: Google (2022)

Ao acessar o resultado de busca em destaque na Figura 20, observou-se no código fonte da página, as mesmas marcações em *Schema.org* apresentadas na Figura 17 para o registro de *Pride and Prejudice*, identificando a *webpage* e o recurso com um URI.

Esse exemplo demonstra que os registros das bibliotecas codificados com metadados do *Schema.org* realmente podem ser recuperados nos resultados orgânicos, isto é, diretamente na lista de resultados de busca por ordem de relevância. Nesse caso, o registro foi recuperado como um resultado comum e não como um *rich snippets*, entretanto, a classificação do mesmo na primeira página de resultados representa o potencial do *Schema.org* para descoberta desses recursos. Compreende-se que é de interesse das bibliotecas explorar as possibilidades de descoberta na *Web*, pois, se os recursos forem facilmente localizados, podem ser melhor aproveitados pelos usuários.

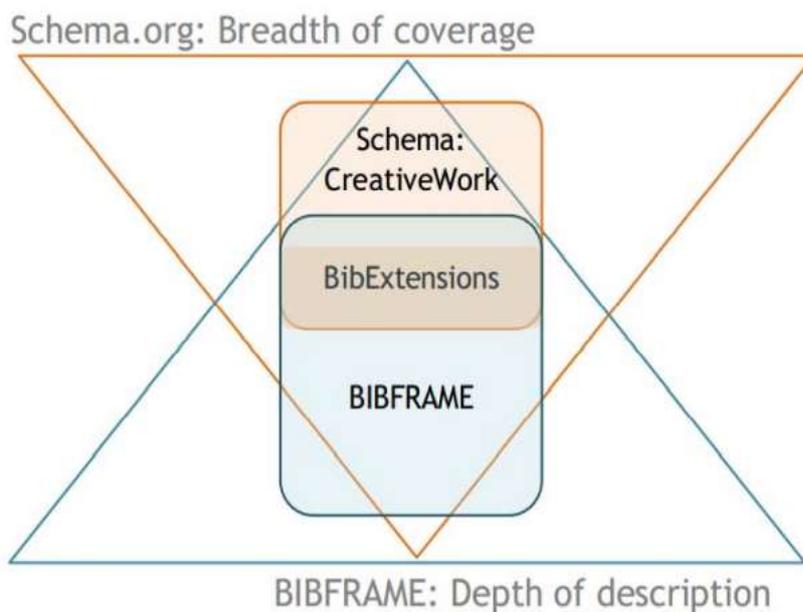
Com o lançamento do BIBFRAME pela *Library of Congress*, a OCLC desenvolveu projetos para analisar a compatibilidade do padrão com o modelo baseado em *Schema.org* (*The OCLC Schema Model*), segundo Godby (2013), considerou-se que os modelos deveriam ser alinhados, pois foram desenvolvidos para

descrição dos mesmos tipos recursos.

O BIBFRAME consiste em um modelo de dados vinculados para dados bibliográficos, projetado para substituir o formato MARC 21. A iniciativa propõe evoluir os padrões de biblioteca para “[...] tornar a informação bibliográfica mais útil dentro e fora da comunidade de bibliotecas.” (LIBRARY OF CONGRESS, 2016, não paginado, tradução nossa). O BIBFRAME é baseado em RDF e nos modelos conceituais FRBR, FRAD e FRSAD, nas diretrizes do RDA e na linguagem de marcação XML, que permitem a adequação do modelo as tendências atuais de tratamento da informação (ARAKAKI *et al.*, 2017).

Os resultados da análise sobre o modelo baseado em *Schema.org* e BIBFRAME foram descritos por Godby (2013), segundo a autora ambos trabalham com conceitos semelhantes, o *Schema.org* fornece uma cobertura ampla para as principais entidades definidas no modelo, porém de maneira superficial, enquanto o BIBFRAME é mais completo nesse sentido. Sendo assim, os modelos foram considerados complementares, como ilustra a Figura 21.

Figura 21 – Modelo baseado em *Schema.org* e modelo BIBFRAME



Fonte: Godby (2013, p. 30).

A Figura 21 apresenta os modelos como complementares e ilustra a amplitude de cobertura do modelo baseado em *Schema.org* e extensões (*BibExtensions*), assim como apresenta a profundidade da descrição fornecida pelo BIBFRAME. Segundo Godby (2013) a partir da análise dos modelos, concluiu-se que nenhum dos modelos

permite a representação adequada dos conceitos do Grupo 1 do FRBR. A autora ressalta que a ausência de um consenso sobre os conceitos de obra e expressão é problemática, enquanto para as bibliotecas esses conceitos são abstratos, para o *Schema.org* é algo tangível que pode ser pesquisado pelos usuários a partir dos mecanismos de busca da *Web*.

Taniguchi (2017) explica que a classe *CreativeWork* do *Schema.org* é semelhante a *Work* do BIBFRAME, entretanto, inclui propriedades que não representam a essência conceitual de Obra definida no modelo, como “*publication*”, “*Publisher*” e “*material*”. O autor salienta que *CreativeWork* trabalha com um conceito definitivamente diferente do BIBFRAME. Certamente, os modelos apresentam objetivos distintos, enquanto o modelo baseado em *Schema.org* visa, principalmente otimizar a interpretação dos dados pelos mecanismos de busca *Web*, o BIBFRAME possui enfoque na curadoria dos dados bibliográficos, com uma estrutura mais completa para descrição.

Segundo Godby (2013) vislumbrou-se o alinhamento entre os modelos, enquanto os principais conceitos seriam representados conforme o *Schema.org* para descoberta na *Web*, os detalhes necessários para curadoria e gerenciamento dos dados poderiam ser expressos em BIBFRAME e padrões associados, definidos no contexto das bibliotecas.

É importante considerar que a permanência do *Schema.org* como vocabulário padrão entre os mecanismos de busca não é garantida, pois, o mesmo pode ser alterado ou substituído. Nesse sentido, no que tange ao alinhamento entre os modelos, Godby (2013) aponta que o BIBFRAME deve ser projetado de modo a minimizar a dependência do *Schema.org*, pois caso, o vocabulário mude radicalmente ou caia em desuso, as bibliotecas não seriam prejudicadas.

Sobre esse aspecto, Godby, Wang e Mixter (2015) argumentam que embora não seja possível garantir a sobrevivência do *Schema.org*, a estruturação dos dados com o vocabulário serviria como base para modelagem conforme os princípios da *Web Semântica*.

“[...] o *Schema.org* já serviu como um veículo para consolidar as melhores ideias sobre engenharia do conhecimento em larga escala, começando com suas raízes em *Cyc* no início dos anos 1980 e continuando com as inovações introduzidas por sua crescente comunidade de usuários, incluindo o grupo *Schema Bib Extend* [...]” (GODBY; WANG; MIXTER, 2015, não paginado, tradução nossa).

Ainda segundo Godby, Wang e Mixter (2015), dessa maneira, a comunidade de bibliotecas estaria mais familiarizada com a modelagem de dados conforme a lógica da *Web Semântica*, de modo a favorecer o mapeamento do *Schema.org* com qualquer tipo de ontologia que o substituísse. Sendo assim, o vocabulário seria apenas um ponto de partida para modernização dos catálogos e para a exposição dos dados na *Web*.

O modelo FRBR, no qual as análises da OCLC foram baseadas, foi repensado em conjunto com o FRAD e FRSAD a partir da iniciativa IFLA *Library Reference Model* (LRM), cuja proposta consiste na harmonização dos três modelos. Cabe destacar que na revisão de literatura não foram encontrados estudos que associam o LRM ao *Schema.org*, portanto, as discussões realizadas nesta pesquisa são baseadas nas análises da OCLC com o modelo FRBR.

No que tange a modelagem dos dados com o BIBFRAME, considerou-se a primeira versão, BIBFRAME 1.0, lançada em 2012. Em 2016 o BIBFRAME foi reformulado e publicado como BIBFRAME 2.0, o progresso da conversão dos registros MARC 21 a partir dessa versão é abordado por Ávila-Barrientos (2020), Taniguchi (2018) e Xu, Hess e Akerman (2017). Certamente, a adoção do BIBFRAME será realizada de maneira gradual, a implementação depende da flexibilidade do modelo para adaptação de uma quantidade significativa de dados codificados no formato MARC 21 (ÁVILA-BARRIENTOS, 2020).

A *Library of Congress* fornece ferramentas de apoio para conversão dos dados em MARC para BIBFRAME, há bibliotecas que já converteram e disponibilizaram conjuntos de dados baseados no modelo, como *British Library* (Reino Unido), *Deutsche National Bibliothek* (Alemanha), *George Washington University Library* (EUA), *National Library of Medicine* (EUA) e *Princeton Library* (EUA) (ARAKAKI, 2016).

Segundo Wallis (2022), os esforços investidos na publicação do BIBFRAME 2.0 podem ser aproveitados, o autor argumenta que, da mesma forma que existem *softwares* para conversão de registros MARC para BIBFRAME, é possível obter saídas BIBFRAME para metadados *Schema.org*, incorporados nas interfaces das bibliotecas. Essas questões são tratadas no "*Bibframe2Schema.org*", cujo propósito consiste em mapear a versão mais recente do BIBFRAME para *Schema.org*, o grupo propõe desenvolver um protótipo para analisar os resultados da conversão, e assim,

identificar os possíveis benefícios da estruturação dos dados com esse formato (WALLIS, 2022).

Observa-se que as tendências para representação no universo bibliográfico possuem enfoque na exposição dos dados dos catálogos, visando possibilitar a interoperabilidade com a *Web* e outros sistemas de informação. As soluções para revitalização desses instrumentos são encontradas nas tecnologias semânticas, em padrões, vocabulários e ontologias que permitem a contextualização e a conexão dos dados, de um modo legível pelos agentes de *software*.

Smith-Yoshimura (2018) apresenta um *survey* sobre projetos de implementação e publicação do *Linked Data* no universo bibliográfico, onde os resultados obtidos em 2015 e 2014 foram comparados com os resultados de 2018. Nesse estudo, observou-se que as instituições fazem uso de múltiplos vocabulários e ontologias baseadas em RDF para publicação de dados conectados, segundo Smith-Yoshimura (2018) identificou-se um aumento na publicação de dados vinculados em *Schema.org* e BIBFRAME, conforme o relato das instituições entrevistadas, a porcentagem de adoção do *Simple Knowledge Organization System (SKOS)* “[...] diminuiu de 60% em 2015 para 44% em 2018, refletindo no aumento do uso do *Schema.org* (30% em 2015 vs. 46% em 2018) [...], a utilização do BIBFRAME aumentou, de 15% em 2015 para 27% em 2018.” (SMITH-YOSHIMURA, 2018, p.14, tradução nossa).

Ainda em relação ao *survey*, as instituições entrevistadas destacaram algumas das motivações para adoção do *Linked Data*, como a exposição dos dados na *Web* para atingir um público maior e a otimização da disseminação dos dados a partir de técnicas de SEO (SMITH-YOSHIMURA, 2018). Nessa perspectiva, um dos benefícios almejados em relação à aderência do *Schema.org* para publicação de dados conectados, refere-se à possibilidade de descoberta de informações como resultados enriquecidos.

Segundo Wallis (2013) os benefícios da publicação dos dados codificados em *Schema.org* só serão percebidos, se bibliotecas e outras instituições publicarem seus metadados conforme os princípios *Linked Data*, dessa maneira, os mecanismos de busca poderão identificar e relacionar as diversas descrições de um mesmo recurso publicadas por páginas distintas na *Web*. O autor argumenta que “Muitas instâncias (como dados) da mesma coisa sendo publicada na *Web*, naturalmente aumentarão

sua visibilidade nas listagens de resultados.” (WALLIS, 2013, não paginado, tradução nossa).

Segundo Godby, Wang e Mixter (2015) há uma pluralidade de interpretações sobre o conceito de ‘visibilidade’, conforme os autores, o conceito pode ser associado aos resultados de busca com informações enriquecidas, pode ainda ser relacionado a quantidade de acessos em uma página, assim como pode ser atrelado a aparência dos resultados com as ferramentas *Google*. Independente das interpretações e objetivos em relação ao modo de apresentação das informações, a necessidade de adequação dos catálogos ao contexto atual e os benefícios da exposição dos dados são evidentes, entretanto, existem desafios significativos para concretização das propostas de adequação.

Conforme Smith-Yoshimura (2018) as instituições que objetivam publicar dados conectados relatam dificuldades para concretização desse propósito, dentre os fatores, está a escassez de recursos financeiros para atualizar e ampliar os serviços fornecidos, a ausência de ferramentas adequadas para medir e avaliar o impacto dos dados estruturados, assim como a satisfação dos usuários.

Certamente, a transição dos dados em MARC 21 para modelos baseados no escopo da *Web Semântica* e *Linked Data* envolve diversos fatores que influenciam no progresso de atualização dos catálogos, apresentando desafios consideráveis para as bibliotecas, entretanto, é crucial a adequação dos instrumentos e procedimentos de catalogação em consonância com o cenário tecnológico atual.

Sendo assim, as tecnologias semânticas surgem como soluções para o desenvolvimento de novas estratégias e adaptações na estrutura dos dados dos catálogos. Nesse aspecto, o *Schema.org* apresenta-se como uma alternativa para otimizar a interpretação dos dados, tanto para agentes de *software*, como para pessoas, conforme propõe a *Web Semântica*. Desse modo, a aderência do vocabulário em conjunto com os instrumentos de representação das bibliotecas poderia viabilizar a descoberta do conteúdo dos catálogos na *Web*.

8 CONSIDERAÇÕES FINAIS

Tendo em vista a constante evolução da *Web* e a gama de possibilidades para representação de informações nesse ambiente, o repensar dos instrumentos e procedimentos de catalogação tornou-se crucial. Em virtude do constante avanço tecnológico, a *Web* tornou-se o principal serviço utilizado para o compartilhamento, busca e recuperação de informações, beneficiando diversos setores da sociedade.

Os usuários contemporâneos habituados a praticidade e a gama de possibilidades oferecidas na era digital, anseiam por respostas rápidas, precisas e dinâmicas. Nesse contexto, as tecnologias semânticas viabilizam a estruturação dos dados, visando favorecer a interpretação eficiente dos agentes de *software*, e assim propiciar resultados de busca mais satisfatórios para os usuários. Dessa maneira, as tecnologias semânticas oportunizam que as bibliotecas explorem seus potenciais como provedoras de metadados relevantes e confiáveis.

Sendo assim, o objetivo da presente pesquisa consistiu em apresentar o *Schema.org*, como uma alternativa para atualização dos catálogos digitais em consonância com o cenário tecnológico atual. Foram discutidas as principais características do vocabulário, benefícios, limitações e possibilidades de extensão, com o intuito de responder à questão de pesquisa: “Como o *Schema.org* pode contribuir para adequação dos catálogos digitais?”

Em relação aos objetivos específicos, na Seção 3 buscou-se atingir o primeiro “Identificar os principais aspectos da representação de recursos informacionais no contexto das bibliotecas nos catálogos digitais”, em complemento, na Seção 4 abordou-se a importância e as funções dos metadados no contexto digital. O segundo objetivo “Descrever o *Schema.org* a partir da estrutura e metadados do vocabulário para representação” foi abordado na Seção 5 e nas subseções 5.1 e 5.2, onde a proposta e as principais características e benefícios do vocabulário foram explicitados.

O terceiro objetivo “Investigar a aplicação do *Schema.org* para estruturação dos dados dos catálogos digitais” e o quarto objetivo “Apresentar os possíveis benefícios e desafios da aderência do *Schema.org* no contexto das bibliotecas” foram tratados nas seções 6 e 7. Na Seção 6 discutiu-se sobre a representação de recursos a partir da classe *CreativeWork*. Na Seção 7 abordou-se a representação dos dados dos catálogos com base nos projetos da OCLC acerca do desenvolvimento de modelos semânticos baseados em *Schema.org*.

A partir da literatura analisada, pode-se afirmar que um dos pontos fortes do *Schema.org* consiste na compatibilidade do vocabulário com formatos semânticos, que favorecem a conexão e a interpretação do conteúdo estruturado. Dessa maneira, viabiliza-se a integração dos dados ao *Knowledge Graph* e a descoberta de informações em *Rich snippets*, essas tecnologias permitem a exibição de informações de modo mais atraente em comparação a uma lista tradicional de resultados de busca. Outra característica positiva do *Schema.org* está na estrutura e diretrizes flexíveis que viabilizam a adaptação de classes, propriedades e relacionamentos conforme necessidades específicas.

Em contraponto, devido ao caráter genérico, o *Schema.org* apresenta limitações para descrição em domínios específicos, como no caso do universo bibliográfico. Nesse aspecto, cabe ressaltar que o vocabulário não foi projetado para substituir padrões complexos, como o formato MARC 21, mas pode ser empregado como uma tecnologia complementar com enfoque na descoberta do conteúdo dos catálogos na *Web*.

Embora o *Schema.org* seja frequentemente caracterizado como um vocabulário superficial e atrelado a objetivos comerciais, apresenta flexibilidade para descrição em qualquer domínio. Além disso, é importante considerar a relevância e credibilidade do *Schema.org*, visto que, o vocabulário foi projetado pelos principais mecanismos de busca em consonância com as diretrizes do W3C para publicação de dados estruturados.

Em vista disso, pondera-se que a aderência do *Schema.org* pelas bibliotecas, poderia viabilizar a contextualização, interpretação e o compartilhamento dos dados dos catálogos com os mecanismos de busca da *Web*. Consequentemente, os recursos informacionais seriam melhor aproveitados, pois, seriam facilmente localizados e descobertos por usuários potenciais. Com a estruturação dos catálogos, as bibliotecas se beneficiariam com o enriquecimento semântico dos dados, além de contribuir com a construção e evolução da *Web Semântica*, a partir da disponibilização de informações estruturadas, fidedignas e compreensíveis para além do universo bibliográfico.

Além dos benefícios da adequação dos catálogos, é importante considerar os desafios envolvidos nesse processo. Os projetos desenvolvidos pela OCLC demonstram que a modernização da arquitetura dos catálogos é uma realidade, entretanto, as mudanças são graduais. A atualização dos catálogos envolve diversos

fatores, como a necessidade de profissionais capacitados para lidar com as novas tecnologias de representação e com questões delicadas sobre quais dados devem ser expostos na *Web*, considerando direitos autorais e termos de licença, além dos recursos financeiros envolvidos em todo processo de adequação.

Considera-se que os tópicos abordados nesta pesquisa contribuem com novos conhecimentos para área de Ciência da Informação, especialmente para o eixo de Representação da Informação. A pesquisa permite ampliar a visão da área em relação as novas formas de representação, a partir do uso estratégico das tecnologias semânticas disponíveis.

Para realização de pesquisas futuras, sugere-se o aprofundamento sobre a utilização do *Schema.org* em técnicas de SEO em relação à visibilidade de recursos e serviços no universo bibliográfico.

REFERÊNCIAS

- AGENJO-BULLÓN, X.; HERNÁNDEZ-CARRASCAL, F. Avances de Bibframe en 2016: perspectivas del nuevo modelo bibliográfico. **Anuario ThinkEPI**, v. 11, p. 310-318, abr. 2017. DOI 10.3145/thinkepi.2017.58. Disponível em: <https://recyt.fecyt.es/index.php/ThinkEPI/article/view/thinkepi.2017.58>. Acesso em: 22 set. 2021.
- ALRASHED, T. *et al.* Dataset or not? A study on the veracity of semantic markup for dataset pages. *In: INTERNATIONAL SEMANTIC WEB CONFERENCE, 20., 2021, [s.l.]. Proceedings [...]. [S.l.]: IBM, 2021. p. 2-17.* Disponível em: <http://people.csail.mit.edu/tarfah/papers/dataset.pdf>. Acesso em: 20 set. 2021.
- ALVES, R. C. V. **Metadados como elementos do processo de catalogação.** 2010. 132 p. Tese (Doutorado em Ciência da Informação) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista Júlio Mesquita Filho, Marília, 2010. Disponível em: <http://repositorio.unesp.br/handle/11449/103361>. Acesso em: 05 ago. 2021.
- ARAKAKI, A. C. S. O modelo IFLA Library Reference Model e o Linked Data. **Informação & Informação**, v. 25, n. 3, p. 163-186, jul./set. 2020. DOI 10.5433/1981-8920.2020v25n3p163. Disponível em: <http://www.uel.br/revistas/uel/index.php/informacao/article/view/42011/28232>. Acesso em: 01 ago. 2021.
- ARAKAKI, F. A. **Metadados administrativos e a proveniência dos dados:** modelo baseado na família PROV. 2019. 139 f. Tese (Doutorado em Ciência da Informação) - Faculdade de Filosofia e Ciências, Universidade Estadual Paulista Júlio de Mesquita Filho, Marília, 2019. Disponível em: <https://repositorio.unesp.br/handle/11449/180490>. Acesso em: 23 set. 2021.
- ARAKAKI, F. A. *et al.* BIBFRAME: tendência para a representação bibliográfica na web. **Revista Brasileira de Biblioteconomia e Documentação**, v. 13, n. esp., 2017. Disponível em: <https://portal.febab.org.br/anais/article/view/1921>. Acesso em: 20 mar. 2022.
- ARAÚJO, C. A. A. Correntes teóricas da Ciência da Informação. **Ciência da Informação**, v. 38, n. 3, p.192-204, 2009. Disponível em: https://www.scielo.br/scielo.php?pid=S0100-19652009000300013&script=sci_abstract&tlng=pt. Acesso em: 2 jun. 2021.
- ARAÚJO, C. A. A. Teorias e tendências contemporâneas da Ciência da Informação. **Inf. Pauta**, v. 2, n. 2, jul./dez. 2017. Disponível em: <http://periodicos.ufc.br/informacaoempauta/article/view/20162>. Acesso em: 02 jun. 2021.
- ASSUMPÇÃO, F. S.; SANTOS, P. L. V. A. da C. Representação no domínio bibliográfico: um olhar sobre os formatos MARC 21. **Perspectivas em Ciência da Informação**, v. 20, n. 1, p. 54-74, jan./mar. 2015. DOI 10.1590/1981-5344/2054.

Disponível em:

<http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/2054/1582>. Acesso em: 04 jul. 2021.

ASSUMPÇÃO, F. S.; SANTOS, P. L. V. A. da C.; ZAFALON, Z. R. O controle de autoridade no domínio bibliográfico: os catálogos digitais. **Biblios**, n. 68, jul./set. 2017. DOI 10.5195/biblios.2017.342. Disponível em:

<https://brapci.inf.br/index.php/res/download/111669>. Acesso em: 05 jun. 2021.

ÁVILLA-BARRIENTOS, E. BIBFRAME y el futuro de la descripción, vinculación y recuperación de los recursos de información. **TransInformação**, Campinas, v. 32, n. e 190069, jul. 2020. DOI 10.1590/1678-9865202032e190069. Disponível em:

<https://www.scielo.br/j/tinf/a/BCcpB5wkY5YQQGSktjbtDq/?lang=es>. Acesso em: 11 mar. 2022.

BARKER, P.; CAMPBELL, L. Whats is *Schema.org*? **Cetis**, n. B01, jun., 2014.

Disponível em: <http://publications.cetis.org.uk/2014/960>. Acesso em: 06 jan. 2022. 2021.

BARRETO, A. de. A. A condição da informação. **São Paulo em Perspectiva**, v. 16, n. 3, p. 67-74, 2002. Disponível em:

https://www.scielo.br/scielo.php?script=sci_abstract&pid=S0102-88392002000300010&lng=en&nrm=iso&tlng=pt. Acesso em: 05 jun. 2021.

BERNERS-LEE, T. **Linked Data**. 2006. Disponível em:

<https://www.w3.org/DesignIssues/LinkedData.html>. Acesso em: 31 jul. 2021.

BIBLIOGRAPH.NET. **BiblioGraph.net Vocabulary**: deprecated. 2019. Disponível em:

<https://bibliograph.github.io/BibloGraph-Frozen/bibliograph.net/vocabularydeprecated.html>. Acesso em: 02 fev. 2022.

BIZER, C., HEATH, T.; BERNERS-LEE, T. Linked Data: the story so far.

International Journal on Semantic Web and Information Systems, v. 5, n. 3, p. 1-22, jul. 2009. DOI 10.4018/jswis.2009081901. Disponível em:

<https://www.semanticscholar.org/paper/Linked-Data-The-Story-So-Far-Bizer-Heath/9f54a0057d0694bc7d1dcf69d186e313ca92775c>. Acesso em: 11 jan. 2022.

BUCKLAND, M. Information as thing. **Journal of the American Society for Information Science**, v. 45, n. 5, p. 351-360, jun. 1991. DOI 10.1002/(SICI)1097-4571(199106)42:5<351::AID-ASI5>3.0.CO;2-3. Disponível em:

[https://asistdl.onlinelibrary.wiley.com/doi/10.1002/\(SICI\)1097-4571\(199106\)42:5%3C351::AID-ASI5%3E3.0.CO;2-3](https://asistdl.onlinelibrary.wiley.com/doi/10.1002/(SICI)1097-4571(199106)42:5%3C351::AID-ASI5%3E3.0.CO;2-3). Acesso em: 06 jul. 2021.

CATARINO; M. E.; SOUZA, T. B. A representação descritiva no contexto da Web Semântica. **TransInformação**, Campinas, v. 2, n. 24, p. 77-90, maio/ago. 2012.

Disponível em:

<https://www.scielo.br/j/tinf/a/LNXBFHmzhdhTKkswBqry58R/abstract/?lang=pt>.

Acesso em: 06 mar. 2022.

CLARK, J. A.; ROSSMANN, D. The open SESMO (Search Engine & Social Media Optimization) Project: linked and structured data for library subscription databases to enable webscale discovery in search engines. **Journal of Web Librarianship**, v. 11, n. 3-4, p. 172-193, nov. 2017. DOI 10.1080/19322909.2017.1378148. Disponível em: <https://www.tandfonline.com/doi/full/10.1080/19322909.2017.1378148>. Acesso em: 22 set. 2021.

DÍEZ, L. A. Redefiniendo el catálogo: expectativas de las interfaces de descubrimiento centradas en el usuario. **Investigación Bibliotecológica**, v. 26, n. 56, p. 181-204, jan/abr. 2012. Disponível em: http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S0187-358X2012000100009. Acesso em: 03 out. 2021.

FACEBOOK. **Sobre tags de microdados para catálogos**. 2021a. Disponível em: <https://www.facebook.com/business/help/1175004275966513?id=725943027795860>. Acesso em: 10 out. 2021.

FACEBOOK. **Sobre o pixel do Facebook**. 2021b. Disponível em: <https://www.facebook.com/business/help/742478679120153id=1205376682832142>. Acesso em: 10 out. 2021.

FONS, T.; PENKA, J.; WALLIS, R. OCLC's Liked Data initiative: using *Schema.org* to make library relevant on the Web. **Information Standards Quarterly**, v. 24, n. 2-3, p. 29-33, jun. 2012. DOI 10.3789/ISQV24N2-3.2012.05. Disponível em: <https://www.niso.org/niso-io/2012/06/oclc-linked-data-initiative>. Acesso em: 22 set. 2021.

FOX, R. Digital libraries: the systems analysis perspective. **OCLC Systems & Services: International digital library perspectives**, v. 30, n. 4, p. 206-211, nov. 2014. DOI 10.1108/OCLC-07-2014-0029. Disponível em: <https://www.emerald.com/insight/content/doi/10.1108/OCLC-07-2014-0029/full/html>. Acesso em: 22 set. 2021.

FREIRE, N.; CHARLES, V.; ISAAC, A. Evaluation of *Schema.org* for aggregation of cultural heritage metadata. *In: EUROPEAN SEMANTIC WEB CONFERENCE, 15., 2018, Crete. Proceedings [...]*. Crete: Springer, 2018. p. 225-239. Disponível em: https://link.springer.com/chapter/10.1007%2F978-3-319-93417-4_15. Acesso em: 22 set. 2020.

FREIRE, N. *et al.* Cultural heritage metadata aggregation using web technologies: IIIF, Sitemaps and *Schema.org*. **International Journal on Digital Libraries**, v. 21. p. 19-30, out. 2018. DOI 10.1007%2Fs00799-018-0259-5. Disponível em: <https://link.springer.com/article/10.1007%2Fs00799-018-0259-5>. Acesso em: 22 set. 2021.

GARCÍA, R.; GIL, R. Social media copyright management using Semantic Web and Blockchain. *In: INTERNATIONAL CONFERENCE ON INFORMATION INTEGRATION AND WEB-BASED APPLICATIONS & SERVICES, 21., 2019, Munich. Proceedings [...]*. Munich: Association for Computing Machinery, 2019. p.

339-343. Disponível em: <https://dl.acm.org/doi/10.1145/3366030.3366128>. DOI 10.1145/3366030.3366128. Acesso em: 22 set. 2021.

GARCÍA-MARCO, F. J. *Schema.org*: la catalogación revisitada. In: BAIGET, T. (ed.); GUALLAR, J. (coord.). **Anuario ThinkEPI**. Barcelona: EPI SCP, 2013, v. 7, p. 169-172, 2013. Disponível em: <https://recyt.fecyt.es/index.php/ThinkEPI/article/view/30355>. Acesso em: 21 set. 2021.

GIL, A. C. **Métodos e técnicas de pesquisa**. 6 ed. São Paulo: Atlas, 2008. 216 p.

GILL, T. Metadata and the web. In: BACA, M. (ed.). **Introduction to metadata**. 3rd ed. Los Angeles: Getty Publications, 2016. Disponível em: <http://www.getty.edu/publications/intrometadata/metadata-and-the-web>. Acesso em: 05 ago. 2021.

GILLILAND, A. J. Setting the stage. In: BACA, M. (ed.). **Introduction to metadata**. 3rd ed. Los Angeles: Getty Publications, 2016. Disponível em: <http://www.getty.edu/publications/intrometadata/setting-the-stage>. Acesso em: 05 ago. 2021.

GITHUB. **Welcome to Schema.org**. 2021. Disponível em: <https://github.com/schemaorg/schemaorg>. Acesso em: 15 ago. 2021.

GODBY, C. J. **The relationship between BIBFRAME and OCLC's Linked-Data Model of bibliographic description**: a working paper. Dublin, Ohio: OCLC Research. 2013. Disponível em: <http://www.oclc.org/content/dam/research/publications/library/2013/2013-05.pdf>. Acesso em: 01 out. 2021.

GODBY, C. J. **A Division of Labor**: The role of *Schema.org* in a Semantic Web model of library resources. In: *Linked Data for Cultural Heritage*. Chicago: Amer Library Assn Editions, 2016. 152 p. Disponível em: <https://www.oclc.org/research/publications/2016/schema-role-semantic-weblibrary.html>. Acesso em: 21 jan. 2020.

GODBY, C. J.; DENENBERG, R. **Common ground**: exploring compatibilities between the Linked Data models of the Library of Congress and OCLC. Dublin, Ohio: Library of Congress and OCLC Research. 2015. Disponível em: <https://www.oclc.org/content/dam/research/publications/2015/oclcresearch-loc-linked-data-2015.pdf>. Acesso em: 01 out. 2021.

GODBY, C. J.; WANG, S.; MIXTER, J. K. **Library linked data in the cloud**: OCLC's experiments with new models of resource description. *Synthesis Lectures on the Semantic Web: Theory and Technology*, 2015, 240 p. Disponível em: <https://www.oclc.org/research/publications/books/library-linked-data-in-the-cloud/chapter1.html#x1-40001>. Acesso em: 20 jul. 2021.

GOOGLE. **About**. 2021a. Disponível em: <https://about.google/intl/ALL>. Acesso em: 20 ago. 2021.

GOOGLE. **How Google's Knowledge Graph works**. 2021b. Disponível em: <https://support.google.com/knowledgepanel/answer/9787176?hl=en>. Acesso em: 23 ago. 2021.

GOOGLE SEARCH CENTRAL. **Understand how structured data works**. 2021a. Disponível em: https://developers.google.com/search/docs/advanced/structured-data/intro-structured-data?visit_id=637644593458278607-2636045658&hl=en&rd=1. Acesso em: 19 ago. 2021.

GOOGLE SEARCH CENTRAL. **Review Snnipet**. 2021b. Disponível em: <https://developers.google.com/search/docs/advanced/structured-data/review-snippet>. Acesso em: 19 ago. 2021.

GOOGLE SEARCH CENTRAL. **Book**. 2021c. Disponível em: <https://developers.google.com/search/docs/advanced/structured-data/book#book-work>. Acesso em: 20 ago. 2021.

GOOGLE SEARCH CENTRAL. **Enable search result features for your site**. 2021d. Disponível em: <https://developers.google.com/search/docs/advanced/appearance/search-result-features#why-specify-search-features-for-a-page>. Acesso em: 20 ago. 2021.

GRACY, K. F.; ZENG, M. L.; SKIRVIN, L. Exploring methods to improve access to music resources by aligning library data with Linked Data: A report of methodologies and preliminary findings. **Journal of the American Society for Information Science and Technology**, v. 64, n. 10, p. 2078–2099, jun. 2013. DOI 10.1002/asi.22914. Disponível em: <https://onlinelibrary.wiley.com/doi/abs/10.1002/asi.22914>. Acesso em: 22 set. 2021.

GRANT, M. J.; BOOTH, A. A typology of reviews: an analysis of 14 review types and associated methodologies. **Health Information and Libraries Journal**, v. 26, n. 2, jun. p. 91-108, 2009. DOI 10.1111/j.1471-1842.2009.00848.x. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/19490148>. Acesso em: 10 jun. 2022.

GUHA, R.; BRICKLEY, D.; MACBETH, S. *Schema.org*: evolution of structured data on the Web. **Communications of the ACM**, v. 59, n. 2, p.1-28, fev. 2016. DOI 10.1145/2844544. Disponível em: <https://dl.acm.org/doi/fullHtml/10.1145/2844544>. Acesso em: 13 jul. 2021.

HAN, M. J. K. *et al.* Exposing library holdings metadata in RDF using *Schema.org* semantics. *In*: INTERNATIONAL CONFERENCE ON DUBLIN CORE AND METADATA APPLICATIONS, 2015, São Paulo. **Proceedings [...]**. São Paulo: DCMI, 2015, p. 41-49. Disponível em: <https://dcpapers.dublincore.org/pubs/article/view/3772>. Acesso em: 22 set. 2021.

HEATH, T.; BIZER, C. **Linked Data**: evolving the Web into a global space. [S.l.]: Morgan & Claypool, 2011. (Synthesis Lectures on the Semantic Web: Theory and Technology) v. 1, n.1. DOI 10.2200/S00334ED1V01Y201102WBE001. Disponível em:

<https://www.morganclaypool.com/doi/abs/10.2200/S00334ED1V01Y201102WBE001>. Acesso em: 29 jul. 2021.

HEPP, M. The Web of data for e-commerce: *Schema.org* and GoodRelations for researchers and practitioners. *In: INTERNATIONAL CONFERENCE ON WEB ENGINEERING*, 15., 2015, Rotterdam. **Proceedings [...]**. Rotterdam: Springer, 2015. p. 723-727. Disponível em: https://link.springer.com/chapter/10.1007/978-3-319-19890-3_66. Acesso em: 20 set. 2021.

HOLLAND, V. T.; JOHNSON, J. **Introducing 'Role'**. 2014. Disponível em: <http://blog.Schema.org/2014/06/introducing-role.html>. Acesso em: 27 nov. 2021.

HORROCKS, I.; PATEL-SCHNEIDER, P. F. **Three theses of representation in the Semantic Web**. 2003. Disponível em: <http://www2003.org/cdrom/papers/refereed/p050/p50-horrocks.html>. Acesso em: 06 ago. 2021.

HYOUNGJOO, P.; KIPP, M. Library Linked Data Models: library data in the Semantic Web. **Cataloging & Classification Quarterly**, v. 57, n. 5, jun. p. 261-277, 2019. DOI 10.1080/01639374.2019.1641171. Disponível em: <https://www.tandfonline.com/doi/full/10.1080/01639374.2019.1641171>. Acesso em: 21 set. 2021.

IDEHEN, K. U. **Semantic Web layer cake tweak, explained**. 2017. Disponível em: <https://medium.com/openlink-software-blog/semantic-web-layer-cake-tweak-explained-6ba5c6ac3fab>. Acesso em: 04 ago. 2021.

INTERNATIONAL FEDERATION OF LIBRARY ASSOCIATIONS. **Functional Requirements for Bibliographic Records**. 2009. 137 p. Final report. Disponível em: <https://www.ifla.org/publications/functional-requirements-for-bibliographic-records>. Acesso em: 15 jan. 2021.

INTERNATIONAL FEDERATION OF LIBRARY ASSOCIATIONS. **Declaração dos Princípios Internacionais de Catalogação (PIC)**. Tradução: Marcelo Votto Texeira. 2016. 21 p. Disponível em: https://www.ifla.org/files/assets/cataloguing/icp/icp_2016-pt.pdf. Acesso em: 08 set. 2020.

ISOTANI, S.; BITTENCOURT, I. I. **Dados abertos conectados**. São Paulo: Novatec, 2015. Disponível em: <https://ceweb.br/livros/dados-abertos-conectados>. Acesso em: 02 ago 2021.

JESUS, A. F. **Recomendações teórico-metodológicas para a publicação de dados bibliográficos abertos e conectados**. 2021. 166 p. Dissertação (Mestrado em Ciência da Informação) – Universidade Federal de São Carlos, São Carlos, 2021. Disponível em: <https://repositorio.ufscar.br/handle/ufscar/14228>. Acesso em: 20 ago. 2021.

JONES, S, M. *et al.* It's all about the cards: sharing on social media probably encouraged HTML metadata growth. **arXiv**, 2021. *In press*. DOI

<https://doi.org/10.48550/arXiv.2104.04116>. Disponível em:
<https://arxiv.org/abs/2104.04116>. Acesso em: 15 mar. 2022.

KARLE, E. *et al.* Extending the *Schema.org* vocabulary for more expressive accommodation annotations. *In: INFORMATION AND COMMUNICATION TECHNOLOGIES IN TOURISM*, 2017, Rome. **Proceedings [...]**. Rome: Springer, 2017. p. 31-41. Disponível em: https://link.springer.com/chapter/10.1007%2F978-3-319-51168-9_3. Acesso em: 21 set. 2021.

KEJRIWAL, M. *et al.* Empirical best practices on using product-specific *Schema.org*. *In: AAAI CONFERENCE ON ARTIFICIAL INTELLIGENCE*, 35., 2021, Palo Alto. **Proceedings [...]**. Palo Alto: Association for the Advancement of Artificial Intelligence, 2021. p. 15452-15457. Disponível em: <https://ojs.aaai.org/index.php/AAAI/article/view/17816>. Acesso em: 21 set. 2021.

LAMPRON, P.; MIXTER, J.; HAN, M. J. K. Challenges of mapping digital collections metadata to *Schema.org*: working with CONTENTdm. *In: METADATA AND SEMANTICS RESEARCH*, 10., 2016, Göttingen. **Proceedings [...]**. Göttingen: Springer, 2016, p. 181-186. Disponível em: https://link.springer.com/chapter/10.1007/978-3-319-49157-8_15. Acesso em: 22 set. 2021.

LEVERING, R. **Announcing Schema Markup Validator**: validator.*Schema.org* (beta). 2021. Disponível em: <http://blog.Schema.org/2021/05/announcing-schema-markup-validator.html>. Acesso em: 26 nov. 2021.

LIBRARY OF CONGRESS. **Bibliographic framework as a Web of data**: Linked Data Model and supporting services. Washington, 2012. Disponível em: <https://www.loc.gov/bibframe/pdf/marclid-report-11-21-2012.pdf>. Acesso em: 10 ago. 2021.

LIBRARY OF CONGRESS. **Overview of the BIBFRAME 2.0 Model**. Disponível em: <https://www.loc.gov/bibframe/docs/bibframe2-model.html>. Acesso em: 27 fev. 2022.

LOHVYENENKO, C.; NEDBAL, D. Usage of Semantic Web in austrian regional tourism organizations. *In: INTERNATIONAL CONFERENCE, SEMANTICS*, 15., 2019, Karlsruhe. **Proceedings [...]**. Karlsruhe: Springer, 2019. p. 3-18. Disponível em: https://link.springer.com/chapter/10.1007/978-3-030-33220-4_1#aboutcontent. Acesso em: 22 set. 2021.

MARCONDES, C. H. Representação e economia da informação. **Ciência da Informação**, v. 30, n. 1, p. 61-70, jan./abr., 2001. Disponível em: <http://revista.ibict.br/ciinf/article/view/939/976>. Acesso em: 08 set. 2020.

MENDÉZ, F. J. M.; PASTOR-SANCHÉZ, J. A.; CARREÑO, J. L. Linked open data en bibliotecas: estado del arte. **Information Research**, v. 25, n. 2, jun. 2020. Disponível em: <http://www.informationr.net/ir/25-2/paper862.html>. Acesso em: 27 dez. 2021.

MEY, E. S. A.; SILVEIRA, N. C. **Catálogo no plural**. Brasília, DF: Briquet de Lemos, 2009. 217 p.

MICHEL, F. *et al.* Unleash the potential of your website. **Biodiversity information science and standards**, [s.l.], v. 4, n. e59046, set. 2020. DOI 10.3897/biss.4.59046. Disponível em: <https://biss.pensoft.net/article/59046>. Acesso em: 22 set. 2021.

MICROSOFT. **Sending flight information to Microsoft Cortana with contextual awareness**. 2016. Disponível em: [https://docs.microsoft.com/en-us/previous-versions/bing/contextual-awareness/dn632191\(v=msdn.10\)?redirectedfrom=MSDN](https://docs.microsoft.com/en-us/previous-versions/bing/contextual-awareness/dn632191(v=msdn.10)?redirectedfrom=MSDN). Acesso em: 18 ago. 2021.

MIKA, P. On *Schema.org* and why it matters for the web. **IEEE Computer Society**, n. 15250595, p. 52-55, jun. 2015. DOI 10.1109/MIC.2015.81. Disponível em: <https://ieeexplore.ieee.org/document/7131410>. Acesso em: 13 jul. 2021.

MIXTER, J. K.; O'BRIEN, P.; ARLITSCH, K. Describing theses and dissertations using *Schema.org*. In: DCMI INTERNATIONAL CONFERENCE ON DUBLIN CORE AND METADATA APPLICATIONS, 2014, Austin. **Proceedings [...]**. Austin: DCMI, 2014. p. 136-148. Disponível em: <https://dcpapers.dublincore.org/pubs/article/view/3715>. Acesso em: 21 set. 2020.

NAM, D.; KEJRIWAL, M. How do organizations publish semantic markup? three case studies using public *Schema.org* crawls. **IEE Computer Society**, v. 51, n. 6, p. 42-51, jun. 2018. DOI 10.1109/MC.2018.2701635. Disponível em: <https://ieeexplore.ieee.org/abstract/document/8395081>. Acesso em: 22 set. 2021.

NEISH, P. Linked data: what is it and why should you care? **The Australian Library Journal**, v. 64, n. 1, p. 3-10, 2015. DOI 10.1080/00049670.2014.974004. Disponível em: <https://www.tandfonline.com/doi/full/10.1080/00049670.2014.974004>. Acesso em: 27 de. 2021.

NEVADO-CHINÉ, N.; ALCARAZ-MARTÍNEZ, R.; NAVALÓN, J. A. Análisis de la implementación *Schema.org* en el repositorio RODERIC e impacto en el posicionamiento en Google y Google Scholar. **Revista Española de Documentación Científica**, v. 44, n.3, p. 1-15, jul/sep. 2021. DOI 10.3989/redc.2021.3.1797. Disponível em: <https://redc.revistas.csic.es/index.php/redc/article/view/1329>. Acesso em: 15 mar. 2022.

NHACUONGUE, J. A.; FERNEDA, E. O campo da Ciência da Informação: contribuições, desafios e perspectivas. **Perspectivas em Ciência da Informação**, v. 20, n. 2, p. 3-18, abr./jun. 2015. Disponível em: <http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/1932>. Acesso em: 04 ago. 2021.

NININ, D. M.; SIMIONATO, A. C. Publicación de datos abiertos en instituciones de patrimonios culturales. **Palabra Clave**, La Plata, v. 8, n. 1, 2018. DOI 10.24215/18539912e056. Disponível em: <https://www.palabraclave.fahce.unlp.edu.ar/article/view/PCe056/9974>. Acesso em: 25 ago. 2021.

NOGALES, A. *et al.* Linking from *Schema.org* microdata to the Web of Linked Data: an empirical assessment. **Computer Standards & Interfaces**, v. 45, p. 90-99, mar. 2016. DOI 10.1016/j.csi.2015.12.003. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S0920548915001440>. Acesso em: 19 ago. 2021.

NOURA, M. *et al.* Automatic knowledge extraction to build semantic Web of Things applications. **IEEE Internet of Things Journal**, v. 6, n. 5, p. 8447-8454, out. 2019. DOI 10.1109/JIOT.2019.2918327. Disponível em: <https://ieeexplore.ieee.org/document/8720028>. Acesso em: 21 set. 2021.

ONLINE COMPUTER LIBRARY CENTER. **VIAF**: The Virtual International Authority File. 2021a. Disponível em: <http://viaf.org>. Acesso em: 11 fev. 2022.

ONLINE COMPUTER LIBRARY CENTER. **FAST Linked Data**. 2021b. Disponível em: <http://experimental.worldcat.org/fast>. Acesso em: 11 fev. 2021.

OLIVEIRA, A. R. de. **Metadados como atributos da informação em bases de dados jornalísticas na Web**. 163 f. 2016. Tese (Doutorado em Comunicação Social) - Universidade Metodista de São Paulo, São Bernardo do Campo. 2016. Disponível em: <http://tede.metodista.br/jspui/handle/tede/1585>. Acesso em: 21 set. 2021.

OUCHI, M. T.; SIMIONATO, A. C. Descrição de conjuntos de dados na Web com *Schema.org*. **Informação & Tecnologia**, v. 5, n.1, p. 128-140, jan./jun. 2018. DOI 10.22478/ufpb.2358-3908.2018v5n1.38451. Disponível em: <https://periodicos.ufpb.br/ojs/index.php/itec/article/view/38451>. Acesso em: 09 jul. 2021.

PALETTA, F. C.; MUCHERONI, M. L.; SILVA, J. F. M. Novas tecnologias de uso em nuvens aplicáveis às bibliotecas. *In*: SEMINÁRIO NACIONAL DE BIBLIOTECAS UNIVERSITÁRIAS, 19., 2016, Manaus. **Resumos [...]**. Manaus: UFAM, 2016, p. 2-12. Disponível em: <http://repositorio.febab.org.br/items/show/4464>. Acesso em: 21 set. 2021.

PANASIUK, O. *et al.* Representing GeoData for tourism with *Schema.org*. *In*: INTERNATIONAL CONFERENCE ON GEOGRAPHICAL INFORMATION SYSTEMS THEORY, APPLICATIONS AND MANAGEMENT, 4., 2018, Madeira. **Proceedings [...]**. Madeira: GISTAM, 2018. p. 239-243. DOI 10.5220/0006755102390246. Disponível em: <https://www.scitepress.org/Papers/2018/67551/67551.pdf>. Acesso em: 22 set. 2021.

PEETERS, R. *et al.* Using *Schema.org* annotations for training and maintaining product matchers. *In*: INTERNATIONAL CONFERENCE ON WEB INTELLIGENCE, 10., 2020, Biarritz. **Proceedings [...]**. Biarritz: Association for Computing Machinery, 2020. p. 195-204. Disponível em: <https://dl.acm.org/doi/10.1145/3405962.3405964>. Acesso em: 20 set. 2020.

PESET, F.; FERRER-SAPENA, A.; SUBIRATS-COLL, I. Open data y linked open data: su impacto en el área de bibliotecas y documentación. **El profesional de la**

información, v. 20, n. 2, p. 165-173, mar/abr. 2011. Disponível em: <http://hdl.handle.net/10251/29837>. Acesso em: 29 set. 2021.

PINTEREST. **Getting started**. 2021. Disponível em: <https://developers.pinterest.com/docs/rich-pins/overview>. Acesso em: 15 ago. 2021.

POMERANTZ, J. **Metadata**. Cambridge: The MIT Press, 2015.

PRICE, C. **What is schema markup and why it's important for SEO**. 2019. Disponível em: <https://www.searchenginejournal.com/technical-seo/schema>. Acesso em: 19 jun. 2022.

PRODANOV, C. C.; FREITAS, E. C. de. **Metodologia do trabalho científico: métodos e técnicas da pesquisa e do trabalho acadêmico**. 2. ed. Novo Hamburgo: Feevale, 2013. 276 p. Disponível em: <https://www.doccity.com/pt/livro-metodologia-do-trabalho-cientifico-metodos-e-tecnicas-de-pesquisa-e-do-trabalho-academico-2-ed-prodanov/4851085>. Acesso em: 13 jul. 2021.

PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO. **Linhas de Pesquisa**. 2021. Disponível em: <http://www.ppgci.ufscar.br/sobre/linhas-de-pesquisa>. Acesso em: 27 ago. 2021.

RAEMY, J. A. **Enabling better aggregation and discovery of cultural heritage content for Europeana and its partner institutions**. 2020, 96 p. Dissertation (Master of Science HES-SO in Information Science) - Haute École de Gestion de Genève, Geneva, 2020. Disponível em: <https://doc.rero.ch/record/329698>. Acesso em: 22 set. 2021.

RAMALHO, R. A. S.; OUCHI, M. T. Tecnologias semânticas: novas perspectivas para a representação de recursos informacionais. **Informação & Informação**, Londrina, v. 16, n. 3, p. 60-75, jan./ jun. 2011. DOI 10.5433/1981-8920.2011v16n3p60. Disponível em: <https://brapci.inf.br/index.php/res/v/32023>. Acesso em: 20 jun. 2021.

RILEY, J. **Understanding metadata: what is metadata, and what is it for?** Baltimore: National Information Standards Organization (NISO), 2017. Disponível em: https://groups.niso.org/apps/group_public/download.php/17446/Understanding%20Metadata.pdf. Acesso em: 20 jul. 2021.

ROA-MARTÍNEZ, S. M.; VIDOTTI, S. A. B. G.; PASTOR-SÁNCHEZ, R. A. Mercado semântico enriquecido para programas de posgrado en Latinoamérica. **Perspectivas em Ciência da Informação**, v. 23, n. 3, p. 67-88, jul./out. 2018. Disponível em: <https://www.scielo.br/j/pci/a/xfnjpFD9hVydxsrr6HTMmn/?lang=es>. Acesso em: 21 set. 2020.

SANTARÉM SEGUNDO, J. E.; SILVA, M. F.; MARTINS, D. L. Revisitando a interoperabilidade no contexto dos acervos digitais. **Informação & Sociedade**, João Pessoa, v. 29, n. 2, p. 61-84, abr./jun. 2018. Disponível em: <https://pesquisa.tainacan.org/repositorio-de-pesquisa/revisitando-a-interoperabilidade-no-contexto-dos-acervos-digitais>. Acesso em: 25 abr. 2021.

SANTOS, P. L. V. A. da C. Catalogação, formas de representação e construções mentais. **Tendências da Pesquisa Brasileira em Ciência da Informação**, v. 6, n. 1, 2013. Disponível em: <https://brapci.inf.br/index.php/res/v/119476>. Acesso em: 25 abr. 2021.

SANTOS, P. L. V. A. C.; SIMIONATO, A. C.; ARAKAKI, F. A. Definição de metadados para recursos informacionais: apresentação da metodologia BEAM. **Informação & Informação**, Londrina, v. 19, n. 1, p. 146-163, jan./abr. 2014. DOI <http://dx.doi.org/10.5433/1981-8920.2014v19n1p146>. Disponível em: <http://www.uel.br/revistas/uel/index.php/informacao/article/view/15251/14211>. Acesso em: 04 ago. 2021.

SARACEVIC, T. Interdisciplinary nature of Information Science. **Ciência da Informação**, [s.l.], v. 24, n. 1, 1995. Disponível em: <https://brapci.inf.br/index.php/res/v/22344>. Acesso em: 04 ago. 2021.

SCHEMA.ORG. **Welcome to Schema.org**. 2021a. Disponível em: <https://Schema.org>. Acesso em: 09 jul. 2021.

SCHEMA.ORG. **About Schema.org**. 2021b. Disponível em: <http://Schema.org/docs/about.html>. Acesso em: 09 jul. 2021.

SCHEMA.ORG. **How we work**. 2021c. Disponível em: <https://Schema.org/docs/howwework.html>. Acesso em: 16 ago. 2021.

SCHEMA.ORG. **Data model**. 2021d. Disponível em: <https://Schema.org/docs/datamodel.html>. Acesso em: 16 ago. 2021.

SCHEMA.ORG. **Full hierarchy**. 2021e. Disponível em: <https://Schema.org/docs/full.html>. Acesso em: 17 ago. 2021.

SCHEMA.ORG. **Organization of Schemas**. 2021f. Disponível em: <https://Schema.org/docs/schemas.html>. Acesso em: 07 ago. 2021.

SCHEMA.ORG. **Style guide**. 2021g. Disponível em: <https://Schema.org/docs/styleguide.html>. Acesso em: 17 ago. 2021.

SCHEMA.ORG. **Book: A Schema.org Type**. 2021h. Disponível em: <https://Schema.org/Book>. Acesso em: 17 ago. 2021.

SCHEMA.ORG. **Getting started with Schema.org using Microdata**. 2021i. Disponível em: <https://Schema.org/docs/guide.html>. Acesso em: 18 ago. 2021.

SCHEMA.ORG. **CreativeWork**. 2021j. Disponível em: <https://Schema.org/CreativeWork>. Acesso em: 18 ago. 2021.

SCHEMA.ORG. **workExample**. 2021k. Disponível em: <https://Schema.org/workExample>. Acesso em: 06 jan. 2021.

SCHEMA.ORG. **Schema.org extensions**. 2021l. Disponível em: <https://Schema.org/docs/extension.html>. Acesso em: 11 fev. 2021.

SCHEMA.ORG. **Role**. 2021m. Disponível em: <https://Schema.org/Role>. Acesso em: 15 mar. 2022.

SCOTT, D. White hat search engine optimization (SEO): structured Web Data for libraries. **Partnership: the Canadian Journal of Library and Information Practice and Research**, v. 10, n. 1, p. 2-21, jun. 2015. DOI 10.21083/partnership.v10i1.3328. Disponível em: <https://journal.lib.uoguelph.ca/index.php/perj/article/view/3328>. Acesso em: 23 set. 2021.

SERRA, L. G. *et al.* Os princípios da descrição e os formatos MARC 21 e ONIX. **Ciência da Informação**, Brasília, v. 2, n. 46, p. 51-66, jun. 2018. Disponível em: <http://revista.ibict.br/ciinf/article/view/2327>. Acesso em: 08 set. 2021.

SILVA, L. C. *et al.* O código RDA e a iniciativa BIBFRAME: tendências da representação da informação no domínio bibliográfico. **Em Questão**, Porto Alegre, v. 23, n. 3, p. 130-156, set./dez. 2017. Disponível em: <https://brapci.inf.br/index.php/res/v/11927>. Acesso em: 06 set. 2021.

SILVA, L. C. da. **Publicação de dados de pesquisa científica**: proposta de estruturação semântica de cadernos abertos de pesquisa frente às dimensões da e-Science. 2020. 243 f. Tese (Doutorado em Ciência da Informação) - Faculdade de Filosofia e Ciências, Universidade Estadual Paulista Júlio de Mesquita Filho, Marília, 2020. Disponível em: <https://repositorio.unesp.br/handle/11449/194341>. Acesso em: 20 set. 2020.

SIMIONATO, A. C. Família FR, BIBFRAME e Linked Data: integração de dados bibliográficos. *In*: ZAFALON, Z. R; DAL'EVEDOVE, P. R. (org.). **Perspectivas da representação documental**. São Carlos, SP: CPOI, 2017, v. 1, p. 211-226. Disponível em: <http://eprints.rclis.org/31906>. Acesso em: 13 abr. 2021.

SIMSEK, U.; KARLE, E.; FENSEL, D. Machine readable Web APIs with *Schema.org* action annotations. **Procedia Computer Science**, v. 137, 2018. p. 255-261. DOI 10.1016/j.procs.2018.09.025. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1877050918316302?via%3Dihub>. Acesso em: 21 set. 2021.

SMITH-YOSHIMURA, K. Analysis of 2018 International Linked Data Survey for Implementers. **Code4lib Journal**, n. 42, nov. 2018. Disponível em: <https://journal.code4lib.org/articles/13867>. Acesso em: 10 mar. 2022.

SOUSA, B. P. de; FUJITA, M. S. L. Do catálogo impresso ao on-line: algumas considerações e desafios para o bibliotecário. **Revista ACB**, Florianópolis, v.17, n.1, p. 59-75, jan./jun. 2012. Disponível em: https://revistaacb.emnuvens.com.br/racb/article/viewFile/822/pdf_71. Acesso em: 14 jun. 2021.

SOUSA, J. L.; MARTINS, P. G. M.; RAMALHO, R. A. S. Análise dos padrões XML e RDF para a representação na web sob a perspectiva da Ciência da Informação: um estudo preliminar. **Informação & Tecnologia** (ITEC), Marília/João Pessoa, v.5, n.1, jan./jun. 2018. Disponível em: <https://brapci.inf.br/index.php/res/v/110402>. Acesso em: 28 jul. 2021.

SOUZA, R. R.; ALVARENGA, L. A Web Semântica e suas contribuições para a Ciência da Informação. **Ciência da Informação**, Brasília, v. 33, n. 1, p. 132-141, jan./abr. 2004. Disponível em: <https://brapci.inf.br/index.php/res/v/17559>. Acesso em: 28 jul. 2021.

STRUCTURED DYNAMICS. **Bibliographic Ontology Specification**. 2013. Disponível em: <https://bibliontology.com>. Acesso em: 05 jan. 2022.

SUOMINEN, O.; HYVÖNEN, N. From MARC silos to Linked Data silos? **o bib Das offene Bibliotheksjournal**, v. 4, n. 2, p. 1-13, jul. 2017. Disponível em: <https://www.o-bib.de/article/view/2017H2S1-13>. Acesso em: 23 set. 2021.

TANIGUCHI, S. Is BIBFRAME 2.0 a suitable schema for exchanging and sharing diverse descriptive metadata about bibliographic resources? **Cataloging & Classification Quarterly**, v. 56, n. 1, p. 40-61, 2018. DOI 10.1080/01639374.2017.1382643. Disponível em: <https://doi.org/10.1080/01639374.2017.1382643>. Acesso em: 11 mar. 2022.

THULUVA, A.; ANICIC, D.; RUDOLPH, S. IoT Semantic interoperability with device description shapes. *In*: EUROPEAN SEMANTIC WEB CONFERENCE, 15., 2018, Crete. **Proceedings [...]**. Crete: Springer, 2018. p. 409–422. Disponível em: <https://www.springerprofessional.de/iot-semantic-interoperability-with-device-description-shapes/15995326>. Acesso em: 21 set. 2021.

THULUVA, A. *et al.* Semantic Node-RED for rapid development of interoperable industrial IoT applications. **Semantic Web**, v. 11, n. 6, p. 949-975, 2020. Disponível em: <http://semantic-web-journal.net/content/semantic-node-red-rapid-development-interoperable-industrial-iot-applications>. Acesso em: 15 mar. 2022.

TOMOYOSE, K.; SANTOS, A. A.; ARAKAKI, A. C. S. *Schema.org* para recuperação da informação em redes sociais. *In*: BRAGATO, T. H. B.; TOGNOLI, N. B. (org.). **Organização do conhecimento responsável: promovendo sociedades democráticas e inclusivas**. Belém: ISKO-Brasil; UFPA, 2019, v. 5, p. 176-182. Disponível em: <https://brapci.inf.br/index.php/res/v/123290>. Acesso em: 22 set. 2021.

TORINO, E. *et al.* Enriquecimento semântico para a disponibilização de dados abertos: teoria e prática. **Encontros Bibli**, Florianópolis, v. 25, p. 01-19, 2020. Disponível em: <https://brapci.inf.br/index.php/res/v/139090>. Acesso em: 15 fev. 2022.

WALLIS, R. **A fundamental component of a new Web?** 2016. Disponível em: <https://www.dataliberate.com/2016/03/06/a-fundamental-component-of-a-new-web>. Acesso em: 20 ago. 2021.

WALLIS, R. **OCLC WorldCat Linked Data release**: significant in many ways. 2012. Disponível em: <https://www.dataliberate.com/2012/06/21/oclc-worldcat-linked-data-release-significant-in-many-ways>. Acesso em: 23 ago. 2021.

WALLIS, R. *et al.* Recommendations for the application of *Schema.org* to aggregated cultural heritage metadata to increase relevance and visibility to search engines: the case of Europeana. **The Code4Lib Journal**, v. 36, p. 1-19, abr. 2017. Disponível em: <https://journal.code4lib.org/articles/12330>. Acesso em 22 ago. 2021.

WALLIS, R. **From records to a web of library data**: pt3 beacons of availability. 2013. Disponível em: <https://www.dataliberate.com/2013/03/15/from-records-to-a-web-of-library-data-pt3-beacons-of-availability>. Acesso em: 29 nov. 2021.

WALLIS, R. **A step for Schema.org**: a leap for bib data on the Web. 2014. Disponível em: <https://www.dataliberate.com/2014/09/02/a-step-for-schema-org-a-leap-for-bib-data-on-the-web>. Acesso em: 03 jan. 2022.

WALLIS, R. **The role of Role in Schema.org**. 2015. Disponível em: <https://www.dataliberate.com/2015/04/15/the-role-of-role-in-schema-org>. Acesso em: 16 mar. 2022.

WALLIS, R. Follow me to the library! bibliographic data in a discovery driven world. **JLIS**, v. 13, n. 1, p. 38-44, jan. 2022. Disponível em: <https://www.jlis.it/index.php/jlis/article/view/417>. Acesso em: 31 mar. 2022.

WALLIS, R.; SCOTT, D. **Schema.org support for bibliographic relationships and periodicals**. 2014. Disponível em: <http://blog.Schema.org/2014/09/schemaorg-support-for-bibliographic-2.html>. Acesso em: 03 jan. 2022.

WORLD WIDE WEB CONSORTIUM. **Best practices for publishing Linked Data**. 2014. Disponível em: <https://www.w3.org/TR/ld-bp>. Acesso em: 29 set. 2021.

WORLD WIDE WEB CONSORTIUM. **What is HTML?** 2016. Disponível em: <https://www.w3.org/standards/webdesign/htmlcss>.

WORLD WIDE WEB CONSORTIUM. **XML essentials**. 2015a. Disponível em: <https://www.w3.org/standards/xml/core>. Acesso em 27 jul. 2021.

WORLD WIDE WEB CONSORTIUM. **Semantic Web**. 2015b. Disponível em: <https://www.w3.org/standards/semanticweb>. Acesso em: 27 jul.2021.

WORLD WIDE WEB CONSORTIUM. **Linked Data**. 2015c. Disponível em: <https://www.w3.org/standards/semanticweb/data>. Acesso em: 29 jul. 2021.

WORLD WIDE WEB CONSORTIUM. **Vocabularies**. 2015d. Disponível em: <https://www.w3.org/standards/semanticweb/ontology>. Acesso em: 17 set. 2021.

WORLD WIDE WEB CONSORTIUM. **Data on the web best practices**. 2019. Disponível em: <https://www.w3.org/Translations/DWBP-pt-BR/#dataVocabularies>. 2019. Acesso em: 10 ago. 2021.

WORLD WIDE WEB CONSORTIUM. **Schema Bib Extend Community Group**. 2021. Disponível em: <https://www.w3.org/community/schemabibex>. Acesso em: 24 ago. 2021.

WORLDCAT. **Pride and Prejudice**. 2022. Disponível em: <https://www.worldcat.org/title/pride-and-prejudice-by-jane-austen/oclc/664391841?referer=di&ht=edition>. Acesso em: 22 mar. 2022.

WOODLEY, M. S. Metadata matters: connecting people and information. *In*: BACA, M. (ed.). **Introduction to metadata**. 3rd ed. Los Angeles: Getty Publications, 2016. Disponível em: <http://www.getty.edu/publications/intrometadata/metadata-matters>. Acesso em: 05 ago. 2021.

XU, A.; HESS, H. AKERMAN, L. From MARC to BIBFRAME 2.0: crosswalks. **Cataloging & Classification Quarterly**, v. 56, n. 2-3, 2017. DOI 10.1080/01639374.2017.1388326. Disponível em: <https://www.tandfonline.com/doi/full/10.1080/01639374.2017.1388326>. Acesso em: 11 mar. 2022.

ZAFALON, Z. R. Recurso informacional e representação documental. *In*: ZAFALON, Z. R.; DAL'EVEDOVE, P. R. (org.). **Perspectivas da representação documental: discussão e experiências**. São Carlos, SP: CPOI, 2017. 471p. Disponível em: <http://eprints.rclis.org/31906>. Acesso em: 20 maio 2021.

ZENG, M. L.; QIN, J. **Metadata**. 2nd ed. Chicago: ALA Neal-Schuman, 2016.