



Universidade Estadual Paulista Julio de Mesquita Filho (UNESP)

Faculdade de Filosofia e Ciências, Campus de Marília

Programa de Pós-Graduação em Ciência da Informação

Liliana Giusti Serra

# **A Web Semântica na gestão de livros digitais licenciados: uma proposta de modelo**

Marília

2019

Liliana Giusti Serra

# **A Web Semântica na gestão de livros digitais licenciados: uma proposta de modelo**

Tese apresentada ao Programa de Pós-Graduação  
em Ciência da Informação da UNESP, Campus  
de Marília, como requisito parcial para obtenção  
do título de doutor em Ciência da Informação

Bolsista CAPES /PDSE/88881.190379/2018-01

Linha de Pesquisa: Informação e Tecnologia

Orientador: José Eduardo Santarém Segundo

Co-orientadora: Jodi Schneider (UIUC)

Marília

2019

lgiustiserra@gmail.com

---

Serra, Líliliana Giusti.

S487a

A Web Semântica na gestão de livros digitais licenciados: uma proposta de modelo / Líliliana Giusti Serra. – Marília, 2019.  
152 f. : il. ; 30 cm.

Orientador: José Eduardo Santarém Segundo. Co-orientadora: Jodi Schneider.

Tese (Doutorado em Ciência da Informação) – Universidade Estadual Paulista (Unesp), Faculdade de Filosofia e Ciências, 2019.

Bibliografia: f. 143-152.

1. Web Semântica. 2. Linked data. 3. Linked Open Data. 4. Linked Enterprise Data. 5. Livros digitais. 6. Livros digitais – Gestão. I. Santarém Segundo, José Eduardo (orientador). II. Schneider, Jodi (co-orientadora). III. Título.

CDD 005.73

---

Ficha catalográfica elaborada por  
Denise Mancera Salgado  
CRB 8/6601

Liliana Giusti Serra

**A Web Semântica na gestão de livros digitais licenciados:  
uma proposta de modelo**

Tese apresentada ao Programa de Pós-Graduação  
em Ciência da Informação da UNESP, Campus  
de Marília, como requisito parcial para obtenção  
do título de doutor em Ciência da Informação

Banca examinadora:

Prof. Dr. José Eduardo Santarém Segundo  
Universidade Estadual Paulista (UNESP) - Orientador

Profa. Dra. Plácida Leopoldina Ventura Amorim da Costa Santos  
Universidade Estadual Paulista (UNESP)

Prof. Dr. José Fernando Modesto da Silva  
Universidade de São Paulo, Escola de Comunicações e Artes (USP/ECA)

Profa. Dra. Silvana Aparecida Borsetti Gregório Vidotti  
Universidade Estadual Paulista (UNESP)

Prof. Dr. Fabrício Silva Assumpção  
Universidade Federal de Santa Catarina (UFSC)

UNESP, Campus de Marília, 26 de agosto de 2019.

## Agradecimentos

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

Aos meus orientadores, professor Santarém (UNESP) e Jodi Schneider (*University of Illinois at Urbana-Champaign*, UIUC) por terem acreditado em meu potencial e mostrarem-me outras possibilidades de explorar a tecnologia em catálogos de bibliotecas, instigando a curiosidade por novos conhecimentos.

Aos membros da banca, Plácida L. V. A. da C. Santos, José Fernando Modesto da Silva, Silvana Aparecida B. G. Vidotti e Fabrício S. Assumpção, e os suplentes Leonardo C. Botega, Márcia R. da Silva e Ana Carolina Simionato Arakaki, por sua disposição em participar da avaliação do resultado desta pesquisa e por suas contribuições em minha formação acadêmica.

Aos professores do Programa de Pós-Graduação em Ciência da Informação da UNESP e da *School of Information* da UIUC, onde pude ter contato com pessoas de diferentes áreas e oportunidades de amadurecimento acadêmico.

Aos diretores e colegas do SophiA Biblioteca por incentivarem meus estudos e proporcionarem condições para o período de estudos no exterior, principalmente Eduardo Voigt, Walter Saliba e Daniela Barreto. Agradecimento especial a Fabrício Biajoli, ao me ensinar como preparar códigos e grafos para esta pesquisa.

À família acadêmica que encontrei na UNESP, pelo companheirismo, apoio, troca de conhecimentos, amizade que levarei para sempre. Agradecimento especial a Jacquelin Teresa Camperos Reyes por ser meu braço direito em Marília.

Aos familiares e amigos que acompanharam todo o percurso, por vezes turbulento, proporcionando apoio para seguir em frente.

À família Henry que me acolheu em *Grand Forks* (ND) no Natal de 2018, proporcionando um ambiente caloroso frente às temperaturas externas de -45°C.

E principalmente a Deus, por ter permitido que eu chegasse até aqui.

Para

minha avó Edith,  
meu cunhado Fábio Henrique, e  
meu padrinho Celso,

Que assistiram esta conquista  
por outro ângulo.

*Where is the Life we have lost in living?*

*Where is the wisdom we have lost in knowledge?*

*Where is the knowledge we have lost in information?*

T.S. Eliot

The rock, 1934

## Resumo

A inclusão de livros digitais em bibliotecas proporciona outras formas de oferta de serviços e acesso à informação aos usuários. Se por um lado amplia a oferta de acervo, por outro demanda ajustes de metadados bibliográficos e de gestão, acompanhamento da movimentação de livros digitais presentes na(s) plataforma(s) dos fornecedores e atualização constante dos dados e recursos disponíveis aos usuários. A Web Semântica, seus conceitos e tecnologias vêm sendo pesquisadas no ambiente de bibliotecas, analisando as possibilidades para sua utilização em rotinas diversas como, por exemplo, o enriquecimento de catálogos e o reaproveitamento de dados que estão disponíveis na Web. Este cenário despertou o interesse em averiguar se a Web Semântica pode ser utilizada para a criação de modelo de dados (*framework*) para contribuir com o tratamento, enriquecimento, publicação e reuso de registros bibliográficos e de licenciamento entre fornecedores e bibliotecas e condicionou a formulação do problema de pesquisa. O objetivo geral desta tese consiste na elaboração de proposta de modelo de dados com uso da Web Semântica para gestão de livros digitais licenciados. Para construir este modelo foi realizada pesquisa bibliográfica visando identificar as principais dificuldades observadas com a gestão de livros digitais licenciados e os elementos da Web Semântica que permitam o enriquecimento, publicação e reutilização de dados. A partir dos resultados desta revisão de literatura foi realizada pesquisa exploratória para identificar se o *Linked Open Data* pode ser utilizado para tratamento, enriquecimento e padronização de dados bibliográficos e o *Linked Enterprise Data* para dados de licenciamento. Foram identificados os metadados que são necessários para descrição de recursos e formas de contratação e realizado mapeamento em formatos diferentes para conversão de dados de fornecedores (planilhas de dados ou registros no formato ONIX) e de bibliotecas (MARC 21) ao modelo RDF. Com o mapeamento dos metadados foram definidas as etapas para a publicação de *datasets* como elaboração de política, seleção, formalização, formatos, licenças, conversão ferramental, processo de recuperação, marketing e feedback. Após a elaboração e construção do modelo foi realizado exemplo de validação a partir de tratamento e enriquecimento de um registro bibliográfico. Como resultado, esta tese apresentou a viabilidade da construção de modelo de dados para gestão de livros digitais licenciados, proporcionando enriquecimento e troca de dados entre bibliotecas e fornecedores de conteúdo digital. A gestão proporcionada pelo modelo permite que as bibliotecas tenham controle da coleção digital licenciada, acompanhando sua disponibilidade no catálogo, as formas de licenciamento vigentes, o controle sobre as contratações duplicadas ou redundantes, aferindo o investimento realizado em recursos de informação em quantidade e qualidade adequados para atender à demanda dos usuários.

**Palavras-chave:** Livros digitais – Gestão; Web Semântica; *Linked data*; *Linked Open Data*; *Linked Enterprise Data*; Livros digitais licenciados.



## Abstract

Adding digital books to a library's collection allows offers of new services and access to the new information for patrons. On one hand, it is possible to enhance the collection but, on the other hand, it is necessary to update bibliographic and management metadata, in accordance to the moving of resources in providers' platforms. The Semantic Web and its concepts and technologies have been studied in libraries, analyzing the possibilities to use it in several activities such as catalog enrichment and data reuse from sources available on the Web. This scenario sparked interest in whether the Semantic Web can be used to create a framework for the treatment, enrichment, reuse, and publishing bibliographic and licensing records by providers and libraries. This situation results in the formulation of the research problem. The main objective of this dissertation relies on the proposal of a framework using the Semantic Web for digital books management. To construct this framework, bibliographic research was performed focusing on identifying the main difficulties in digital books management and the core elements of the Semantic Web that allow the enrichment, publishing and reusing of data. From the results of this literature review, exploratory research was done to identify if the Linked Open Data can be used for the treatment, enriching, and patterning of bibliographic data, and if the Linked Enterprise Data can be used for licensing data. The needed metadata for resource description and business models were identified, and data mapping in different formats were done for the conversion of providers' data (data sheets or records in ONIX format) and libraries (MARC 21) to the RDF model. With this metadata mapping, the steps for dataset publishing were defined, such as preparing the politics, selection, formalization, formats, licenses, data conversion, retrieval, marketing, and feedback. After planning and structuring the framework, the model's validation was done from the treatment and enrichment of one bibliographic record. As a result, this dissertation shows the possibility of enriching and exchanging data from libraries and digital content providers. The management provided by the model allows libraries to have control of the licensed digital collection, keeping track of its availability in the catalog, the current forms of licensing, the control over duplicate or redundant hiring, and gauging the investment made in information resources in adequate quantity and quality to meet user demand.

**Keywords:** E-books – Management; Semantic Web, Linked data; Linked Open Data; Linked Enterprise Data; Licensed e-books.

## Lista de Figuras

Figura 1: Livro digital em MARC 21 oferecido por fornecedor .....	42
Figura 2: Apresentação de obras concorrentes no catálogo.....	48
Figura 3: Visualização de representações diferentes de obras concorrentes no catálogo .....	48
Figura 4: Dados de licenciamento no MARC 21 .....	52
Figura 5: Diversos significados para a palavra “Kiwi” .....	58
Figura 6: Representação de Recursos, Propriedades e Valores entre biblioteca e usuários .....	74
Figura 7: Fragmento de registro MARC 21 com URIs no subcampo zero (0).....	78
Figura 8: Fragmento do <i>Schema:Book</i> dentro da classe <i>Thing</i> .....	83
Figura 9: Fragmento de informações sobre Monteiro Lobato no VIAF.....	85
Figura 10: Modelo para gestão de livros digitais licenciados.....	89
Figura 11: Modelagem de dados dos <i>datasets</i> dos fornecedores.....	93
Figura 12: Fluxo organizacional para publicação de dados.....	95
Figura 13: Registro convertido ao <i>Schema</i> .....	125
Figura 14: Registro convertido do <i>Schema</i> ao formato JSON-LD .....	127
Figura 15: Desenho do grafo com os vocabulários e propriedades .....	130
Figura 16: Grafo com os dados bibliográficos do <i>Schema:Book</i> .....	131
Figura 17: Grafo com os dados de licenciamento do <i>Schema:Offer</i> .....	132
Figura 18: Grafo com dados de validade do licenciamento do <i>Schema:Intangible</i> ..	132
Figura 19: Grafo com dados da situação do livro no <i>Schema:Action</i> .....	132
Figura 20: Grafo com dados sobre a licença <i>Creative Commons</i> no <i>Schema:CreativeWork</i> .....	132

## Lista de tabelas

Tabela 1: Principais problemas reportados na literatura.....	32
Tabela 2: Algumas fontes de dados RDF e Não-RDF.....	73
Tabela 3: Tipos de dados necessários ao modelo .....	96
Tabela 4: Definições – Política e seleção .....	97
Tabela 5: Boas práticas da W3C na etapa de Política e seleção .....	97
Tabela 6: Especificações de serialização .....	100
Tabela 7: Boas práticas da W3C na etapa de Formalização, enriquecimento de dados, formatos e licenças.....	101
Tabela 8: Mapeamento – Dados bibliográficos - ONIX – SchemaBook – MARC 21 .....	102
Tabela 9: Mapeamento – Dados bibliográficos – Planilha de dados – SchemaBook – MARC 21.....	105
Tabela 10: Mapeamento - Dados de licenciamento – Planilha de dados - Schema - ONIX.....	108
Tabela 11: Definições para conversão ferramental.....	109
Tabela 12: Boas práticas da W3C na etapa de Conversão ferramental .....	109
Tabela 13: Definições para a publicação e recuperação dos dados .....	110
Tabela 14: Boas práticas da W3C na etapa de Processo de recuperação.....	111
Tabela 15: Definições das estratégias para marketing e <i>feedback</i> .....	113
Tabela 16: Boas práticas da W3C na etapa de Marketing e feedback .....	113
Tabela 17: Dimensões e métricas de qualidade para dados conectados .....	115
Tabela 18: Definições para controle de qualidade e preservação .....	117
Tabela 19: Boas práticas da W3C na etapa de Qualidade e preservação.....	117
Tabela 20: Regras de descrição de Autorias .....	120
Tabela 21: Regras de descrição de Edição.....	121
Tabela 22: Regras de descrição de Número de páginas (extensão).....	121
Tabela 23: Regras de descrição de ISBN da versão impressa .....	121
Tabela 24: Regras de descrição de Local de publicação.....	121
Tabela 25: Regras de descrição de Editora .....	122
Tabela 26: Regras de descrição de Assuntos .....	122
Tabela 27: Regras de descrição de URLs para acesso.....	122
Tabela 28: Dados originais recebidos de fornecedor em planilha de dados .....	123
Tabela 29: Dados bibliográficos no modelo <i>Schema</i> mapeados para o MARC 21 ...	133

## Lista de siglas

AACR2	Código de Catalogação Anglo Americano, 2ª edição
AAT	<i>Getty's Art and Architecture Thesaurus</i>
ALA	<i>American Library Association</i>
API	<i>Applicance Program Interface</i>
APP	Aplicações para dispositivos móveis
BIBFRAME	<i>Bibliographic Framework</i>
bibliotek-o	<i>BIBFRAME Extension Ontology</i>
BFI	<i>British Film Institute</i>
BISG	<i>Book Industry Study Group</i>
BnF	<i>Bibliothèque Nationale de France</i>
CAPES	Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
CC	<i>Creative Commons</i>
CC-BY	CC Atribuição de autoria
CC-NC	CC Não comercial
CC-ND	CC Não permite derivação
CC-SA	CC Compartilha da mesma forma
CIC	<i>Committee on Institutional Cooperation</i>
CIP	<i>Statement of International Cataloguing Principles</i>
CLI	<i>Center of Library Initiatives</i>
COUNTER	<i>Counting Online User NeTworked Electronic Resources</i>
DC	<i>Dublin Core</i>
DCAT	<i>Data Catalog Vocabulary</i>
DCMI	<i>Dublin Core Metadata Initiative</i>
DNB	<i>Deutsche Nationalbibliothek</i>
DOI	<i>Digital Object Identifiers</i>
DRM	<i>Digital Rights Management</i>
DTD	<i>Document Type Definition</i>
EBA	<i>Evidence Based Acquisition</i>
EC	<i>European Commision</i>
FaBIO	<i>FRBR aligned Bibliographic Ontology</i>
FAST	<i>Faceted Application of Subject Headings</i>
FOAF	<i>Friend of a Friend</i>
FRBR	<i>Functional Requirements for Bibliographic Records</i>
GNIS	<i>Geographic Names Information System</i>
GW	<i>George Washington University Libraries</i>
HTTP	<i>Hyper Text Transfer Protocol</i>
IBDB	<i>Internet Broadway Database</i>
IEEE	<i>Institute of Electrical and Electronics Engineers</i>
IFLA	<i>International Federation of Library Associations</i>
ILS	<i>Integrated Library System</i>
IMDb	<i>Internet Movie Database</i>

IRI	<i>International Resource Identifier</i>
ISBN	<i>International Standard Book Number</i>
ISNI	<i>International Standard Name Identifier</i>
JSON	<i>JavaScript Object Notation</i>
LCSH	<i>Library of Congress Subject Headings</i>
LD	<i>Linked Data</i>
LD4L	<i>Linked Data for Libraries</i>
LD4L Labs	<i>Linked Data for Libraries Lab</i>
LD4P	<i>Linked Data for Production</i>
LD4P2	<i>Linked Data for Production</i>
LDN	<i>Linked Data Notifications</i>
LED	<i>Linked Enterprise Data</i>
LOC	<i>Library of Congress</i>
LOD	<i>Linked Open Data</i>
LOV	<i>Linked Open Vocabularies</i>
MADS	<i>Metadata Authority Description Schema</i>
MARC	<i>MAchine Readable Cataloging</i>
MeSH	<i>Medical Subject Headings</i>
METS	<i>Metadata Encoding and Transmission Standard</i>
MODS	<i>Metadata Object Description Schema</i>
N3	<i>Notation3</i>
NACO	<i>Name Authority Cooperative Program of the PCC</i>
NISO	<i>National Information Standards Organization</i>
NT	<i>N-Triples</i>
OCLC	<i>Online Computer Library Center</i>
ONIX	<i>Online Information eXchange</i>
ONIX PL	<i>ONIX Publication Licenses</i>
OPAC	<i>Online Public Access Catalog</i>
ORCID	<i>Open Research and Contributor ID</i>
OWL	<i>Ontology Web Language</i>
PCC	<i>Program for Cooperative Cataloging</i>
PDSE	<i>Programa de Doutorado Sanduíche no Exterior</i>
PPGCI	<i>Programa de Pós-Graduação em Ciência da Informação</i>
PREMIS	<i>Preservation Metadata Maintenance Activity</i>
RDA	<i>Resource Description and Access</i>
RDF	<i>Resource Description Framework</i>
RDFS	<i>Resource Description Framework Schema</i>
SHACL	<i>Shapes Constraint Language</i>
SIG	<i>Special Interest Group</i>
SKOS	<i>Simple Knowledge Organization System</i>
SRSIS	<i>Scholarly Resource Semantic Information Store</i>
SUSHI	<i>Standardized Usage Statistics Harvesting Initiative</i>
UNESP	<i>Universidade Estadual Paulista Julio de Mesquita Filho</i>

UIUC	<i>University of Illinois at Urbana-Champaign</i>
URI	<i>Uniform Resource Identifier</i>
URL	<i>Uniform Resource Locator</i>
VIAF	<i>Virtual International Authority File</i>
XML	<i>eXtensible Markup Language</i>
XSD	<i>XML Schema Definition</i>
W3C	<i>World Wide Web Consortium</i>

## Sumário

<b>1</b>	<b>Introdução .....</b>	<b>20</b>
1.1	Justificativa e motivação da pesquisa.....	24
1.2	Metodologia e estrutura da pesquisa .....	26
1.3	Trabalhos relacionados.....	28
<b>2</b>	<b>Livro digital, bibliotecas e fornecedores.....</b>	<b>35</b>
2.1	Qualidade dos metadados.....	39
2.2	Desenvolvimento da coleção .....	44
2.3	Gestão dos livros digitais contratados e geração de indicadores.....	50
<b>3</b>	<b>Web Semântica, <i>Linked Data</i>, <i>Linked Open Data</i> e <i>Linked Enterprise Data</i>..</b>	<b>55</b>
3.1	Linked Data (LD) .....	60
3.2	Linked Open Data.....	62
3.3	Linked Enterprise Data (LED) .....	64
3.4	Publicação de datasets.....	66
<b>4</b>	<b>Web Semântica e bibliotecas .....</b>	<b>68</b>
4.1	Conversão de metadados .....	75
4.2	Algumas iniciativas da Web Semântica em bibliotecas.....	80
4.2.1	Projetos LD4 .....	80
4.2.2	<i>Schema.org</i> .....	82
4.2.3	<i>Virtual International Authority File</i> .....	84
<b>5</b>	<b>Modelo para gestão de coleção digital por meio da Web Semântica .....</b>	<b>86</b>
5.1	Política e seleção.....	95
5.2	Formalização, enriquecimento de dados, formatos e licenças.....	98
5.3	Conversão ferramental .....	101
5.3.1	Inclusão de dados bibliográficos em ONIX.....	102
5.3.2	Inclusão de dados bibliográficos por planilha de dados.....	105
5.3.3	Dados de licenciamento .....	107
5.4	Processo de recuperação .....	109
5.5	Marketing e feedback.....	111
5.6	Qualidade e preservação.....	113
<b>6</b>	<b>Exemplo de validação do modelo .....</b>	<b>119</b>
<b>7</b>	<b>Considerações finais .....</b>	<b>136</b>
	<b>Referências.....</b>	<b>145</b>

## 1 Introdução

---

Esta pesquisa é resultado de estudos realizados no Programa de Pós-Graduação em Ciência da Informação (PPGCI) da Universidade Estadual Paulista “Júlio de Mesquita Filho” (UNESP), Campus de Marília, na linha de pesquisa “Informação e Tecnologia”. Parte da pesquisa foi realizada na *University of Illinois at Urbana-Champaign* (UIUC), em decorrência de financiamento da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), por meio da concessão de bolsa sanduíche do Programa de Doutorado Sanduíche no Exterior (PDSE), que teve duração de seis meses, no período de outubro de 2018 a março de 2019.

De acordo com Saracevic (1995), a Ciência da Informação possui três características gerais: 1) é interdisciplinar; 2) a tecnologia contribui com sua evolução; e 3) é participante do avanço da sociedade da informação. É inegável que as transformações tecnológicas também ocorreram no suporte livro e estas mudanças demandam ajustes e adequações nas bibliotecas. Segundo Frederick (2015), apesar da maturidade das bibliotecas em adquirir, descrever e gerir suas coleções, são inegáveis os desafios impostos na gestão dos metadados dos livros digitais.

As bibliotecas são consideradas organismos vivos, onde serviços e a guarda de informações, fisicamente caracterizadas por documentos impressos textuais, são centralizados visando atender à comunidade usuária. Este cenário é conhecido desde a identificação de coleções de documentos na Antiguidade, passando pela invenção de Gutenberg, até os dias de hoje. Com o avanço de tecnologias, novas formas de propagar informações e conteúdos foram desenvolvidas, representando um desafio aos envolvidos neste processo, que começa com o autor e segue pelo editor/fornecedor, bibliotecário e, finalmente, o usuário final. De acordo com Svenonius (2001, p. ix, tradução nossa), “o acesso instantâneo à informação digital é o mais destacado atributo da era da informação”<sup>1</sup>.

O livro digital afeta a realidade das bibliotecas e sua inclusão nos acervos deve ser pensada na forma de garantir a função original de preservação de publicações e

---

<sup>1</sup> *Instant electronic access to digital information is the single most distinguishing attribute of the information age* (SVENONIUS, 2001, p. ix).



acesso ao público, proporcionando condições para gestão específica do conteúdo digital.

Nesta pesquisa foram identificados três aspectos em relação às bibliotecas e os livros digitais licenciados que necessitam de reflexões: qualidade dos metadados, desenvolvimento da coleção e gestão do licenciamento realizado. A qualidade dos metadados é um ponto destacado por Frederick (2015):

Entre os bibliotecários catalogadores, a preocupação pode estar na qualidade dos registros do catálogo em MARC, ou nas inconsistências e irregularidades dos metadados dos livros digitais nas bases de conhecimento (FREDERICK, 2015, p. 11, tradução nossa)<sup>2</sup>

A disponibilização dos registros contratados no formato MARC 21 (*MAchine-Readable Cataloguing*) é ofertada por muitos fornecedores, porém a qualidade destes metadados nem sempre corresponde às expectativas de bibliotecários catalogadores (ARMSTRONG; LONSDALE, 2011). Muitas vezes estes registros são criados pelos próprios fornecedores, sem a adoção de regras, padronização ou trabalho de um bibliotecário. Ao licenciar um pacote de livros digitais, o fornecedor encaminha ao contratante os registros para que os metadados sejam importados para o catálogo.

Em relação ao desenvolvimento de coleções, o licenciamento de pacotes de livros digitais apresenta como vantagens o custo menor em relação à contratação de livros digitais individuais; e a reunião de obras sobre o mesmo tema, resultando em economia de tempo na seleção do acervo. A seleção e alterações nos pacotes podem ocorrer durante o período de vigência do licenciamento. Independentemente da forma como o livro digital foi licenciado se observa a instabilidade na manutenção dos livros digitais nos acervos. Isto é decorrente da possibilidade de não renovação ou rompimento de contratos entre autores, editores, fornecedores e bibliotecas, resultando em restrições legais para oferta de livros digitais ou não renovação de serviços contratados, aspectos discutidos por Frederick (2015):

As reclamações vão desde a frustração sobre a falta de consistência de funcionalidades e gestão de direitos digitais de uma plataforma de livros

---

<sup>2</sup> Among metadata and cataloguing librarians, the concern may be with the quality of MARC catalogues records or inconsistencies and irregularities of ebook metadata in knowledge bases (FREDERICK, 2015, p. 11).

digitais a outra, a falta de compreensão sobre o que significa “acesso perpétuo”, e confusões sobre questões sobre o licenciamento em geral, até a frustração com as mudanças nas plataformas pouco divulgadas e livros digitais que parecem “desaparecer” do pacote contratado (FREDERICK, 2015, p. 17, tradução nossa)<sup>3</sup>.

Caso ocorram limitações entre editores, autores e fornecedores, livros digitais podem ser removidos das coleções à revelia da biblioteca. Livros esgotados deixam de ser localizados, visto que nem sempre podem ser disponibilizados na plataforma dos fornecedores, mesmo que a biblioteca tenha optado por um licenciamento por aquisição perpétua. Obras podem ficar fora do catálogo das editoras e, conseqüentemente, indisponíveis aos fornecedores. Novos títulos digitais podem ser agregados a um pacote, resultando na necessidade de realização de ajustes no catálogo.

A comunicação de alterações de pacotes de livros digitais licenciados usualmente é realizada por meio de mensagem eletrônica (e-mail) entre os fornecedores e seus clientes, informando sobre os livros digitais que foram incluídos e removidos do conjunto. Esta situação obriga a atualização dos registros no catálogo (inclusão, edição ou remoção), que pode ser realizada de forma manual ou por meio de nova importação de registros.

Ao fazer um licenciamento ou renová-lo são necessários dados sobre a negociação como a identificação do fornecedor, a forma de contratação e o período de vigência. Estes metadados não são bibliográficos, mas de gestão e, portanto, não devem ser representados no formato MARC 21. Entretanto é fundamental aos bibliotecários identificarem quais os livros digitais foram contratados de quais fornecedores.

A forma de contratação pode sofrer alterações, seja por aplicação de outro modelo de negócio ou por atualização de datas de vigência do licenciamento, entre outras, e é de extrema importância aos bibliotecários a identificação destes dados, justificando investimentos realizados na formação da coleção ou a contratação de acessos simultâneos a alguns livros digitais, por exemplo. A partir destes metadados de gestão será possível a geração de indicadores que orientarão a continuidade ou não do

---

<sup>3</sup> *Complaints range from frustrations about the lack of consistency of functionality and digital rights management from eBook platform to eBook platform, a lack of a common agreement about what “perpetual access” means, and confusion about licensing issues in general, to frustration with poorly publicized platform changes and titles that appear to “disappear” from package purchases (FREDERICK, 2015, p. 17).*

licenciamento, evitando contratações redundantes (livros impressos e no formato digital; mesmo livro digital licenciado com mais de um fornecedor) etc., ou que não apresentam demanda dos usuários.

Tim Berners-Lee, James Hendler e Ora Lassila apresentaram o conceito da Web Semântica em artigo publicado em 2001. Segundo os autores, coleções de informações marcadas com regras de inferência podem prover condições para compartilhar qualquer sistema de representação existente e delegar a computadores a realização de tarefas de forma automática, explorando o conteúdo disponível em sítios e conjuntos de dados dispostos na Web (BERNERS-LEE; HENDLER; LASSILA, 2001).

Ambientes semânticos são identificados como locais onde o dado está disposto de forma a ser legível e processado por máquinas e permitem a ligação entre elementos descritivos, favorecendo a geração, reuso e relacionamento entre outros grupos de dados. Conjuntos de dados (*datasets*) podem ser disponibilizados na Web, permitindo que eles sejam interligados com outros, favorecendo o reuso de dados. Os *datasets* podem ser abertos ou de uso restrito.

Esta pesquisa busca solução para o seguinte **problema**: é possível a criação de um modelo de dados (*framework*) construído por meio da Web Semântica que permita a publicação e reuso de *datasets* favorecendo a troca de registros bibliográficos e de licenciamento entre fornecedores e bibliotecas?

Este problema suscitou os seguintes questionamentos:

- Quais iniciativas da Web Semântica podem ser adotadas na gestão dos livros digitais licenciados e quais padrões de metadados podem ser utilizados?
- É possível publicar os registros bibliográficos de fornecedores em *datasets* para que sejam consumidos por bibliotecas?
- Como ter maior controle sobre os livros digitais licenciados que fazem parte da coleção da biblioteca?

Este problema de pesquisa promoveu a defesa da seguinte **tese**: a Web Semântica permite o enriquecimento e melhoria de metadados para que registros de livros digitais comercializados com fornecedores sejam publicados na forma de *datasets* para

consumo, facilitando a comunicação entre fornecedores e bibliotecas, favorecendo a gestão da coleção.

Assim, a **proposta** de criação de modelo de dados com uso da Web Semântica será construído a partir da (1) compreensão do cenário do livro digital licenciado para bibliotecas e fornecedores; (2) identificação de aplicações da Web Semântica que podem ser utilizadas; (3) orientar a publicação dos dados de fornecedores e de bibliotecas para que sejam consumidos e reutilizados.

Para corroborar esta proposta foi definido um objetivo geral e três específicos, que serão apresentados a seguir.

Embora aspectos da catalogação sejam abordados, o **objetivo geral** desta pesquisa é a proposta de um modelo de dados para gestão de livros digitais licenciados por meio da Web Semântica, que permita a troca de metadados de registros de forma que sejam publicados, consumidos e trocados entre fornecedores e bibliotecas.

Os **objetivos específicos** são:

1. Identificar as principais dificuldades observadas por bibliotecas na contratação de livros digitais licenciados por fornecedores;
2. Identificar aspectos da Web Semântica que permitam a publicação e reutilização de dados entre fornecedores de livros digitais licenciados e bibliotecas;
3. Desenvolver e validar modelo que, por meio da troca e reuso de dados, facilite a gestão de livros digitais licenciados.

### ***1.1 Justificativa e motivação da pesquisa***

A contratação de livros digitais licenciados por bibliotecas tem sido tema de discussões no Brasil e no exterior. A possibilidade de criação de modelo que possa ser adotado por fornecedores, independentemente de suas plataformas de leitura, para que sejam publicados e consumidos por bibliotecas representa uma contribuição relevante para a Ciência da Informação em nível acadêmico e profissional.

Questões conceituais da catalogação não serão discutidas, porém algumas abordagens e exemplos serão apresentados para ilustrar situações localizadas. Tampouco a proposta almeja o desenvolvimento de produto ou serviço que possa ser

oferecido por empresas ou instituições. A proposta visa a construção de um modelo de dados aberto e genérico que permita a publicação de *datasets* para troca de dados entre fornecedores e bibliotecas e que possa ser utilizado para o desenvolvimento de outras aplicações.

A relevância acadêmica baseia-se na consolidação de conhecimentos sobre a Web Semântica em bibliotecas, como a publicação de *datasets* abertos e fechados, para que sejam consumidos entre fornecedores e bibliotecas, permitindo a reutilização e troca de dados. Mapeamentos de dados bibliográficos e de licenciamento para os formatos MARC 21, ONIX e formato genérico (planilha de dados) com o *Schema*. Aprofundamento dos estudos sobre livros digitais em bibliotecas.

A relevância profissional calca-se na possibilidade de proporcionar controle e gestão dos livros digitais licenciados para bibliotecas, dispensando os esforços de importação e atualização de metadados, reunindo dados bibliográficos e de licenciamento, além de permitir identificação de livros digitais incluídos ou removidos dos pacotes contratados, auxiliando no desenvolvimento da coleção. Também permite a construção de ferramentas que podem ser utilizadas por bibliotecas e editores, dispensando conhecimentos específicos de descrição dos metadados relacionados.

Esta pesquisa foi motivada por diversos fatores, a saber:

1. Aprofundamento na temática do livro digital, iniciada no mestrado;
2. Aprofundar os conhecimentos sobre Web Semântica e suas possibilidades de utilização na Ciência da Informação;
3. Contato com fornecedores de livros digitais e a dificuldade deles para a geração de registros bibliográficos no formato MARC 21 (GODBY, 2012);
4. Contato com bibliotecários no Brasil e exterior que contratam livros digitais licenciados e enfrentam dificuldades para importação, manutenção e atualização dos livros digitais no catálogo;
5. Necessidade de controle dos livros digitais licenciados presentes na coleção, identificando os livros digitais novos e os que foram retirados do conjunto, não dependendo somente de comunicações dos fornecedores;
6. Necessidade de geração de indicadores que permitam a criação de relatórios e estatísticas para controle dos livros digitais licenciados e sua utilização pelos usuários;

7. Identificação de poucos estudos sobre o uso da Web Semântica relacionada aos livros digitais licenciados.

## ***1.2 Metodologia e estrutura da pesquisa***

Trata-se de pesquisa aplicada de natureza qualitativa, bibliográfica, descritiva e exploratória.

De acordo com McClure (1989, p. 282, tradução nossa) “pesquisa aplicada usa a teoria e os conceitos da pesquisa básica e, por meio de métodos formais de perguntas, investiga fenômenos do “mundo real””<sup>4</sup>. Segundo Connaway e Powel (2010), pesquisas aplicadas costumam ser mais pragmáticas ao proporcionar informações que podem ter utilização imediata para resolver problemas atuais, mas que não precisam, necessariamente, ter uma aplicação. De acordo com Walliman (2011), a pesquisa qualitativa não envolve números, mas baseia-se em informações expressas em palavras, relatos e opiniões que descrevem o objeto a ser analisado, enfatizando a interpretação dos dados coletados e analisados.

Conforme analisado por Raupp e Bueren (2006) a pesquisa exploratória possui a seguinte característica:

[...] a pesquisa exploratória normalmente ocorre quando há pouco conhecimento sobre a temática a ser abordada. Por meio de estudo exploratório, busca-se conhecer com maior profundidade o assunto, de modo a torná-lo mais claro ou construir questões importantes para a condução da pesquisa (RAUPP; BUEREN, 2006, p. 80).

A escolha por estes métodos é decorrente da pequena presença de literatura sobre o tema pesquisado. Existem estudos sobre a Web Semântica, assim como sobre livros digitais licenciados em bibliotecas. Porém, poucos textos que aventam a utilização da Web Semântica para favorecer a gestão de livros digitais licenciados foram localizados.

A pesquisa bibliográfica foi conduzida com revisão de literatura produzida na Ciência da Informação e na Ciência da Computação. Foram realizadas pesquisas em bases de dados nacionais e internacionais, livros, artigos, teses, dissertações e trabalhos

---

<sup>4</sup> *Applied research takes the theory and concepts from basic research and, by formal methods of inquiry, investigates “real world” phenomena* (MCCLURE, 1989, p. 282).

apresentados em eventos, nos idiomas português, inglês e espanhol, com recorte temporal principal de 2001 a 2019; sobre os temas pesquisados.

Para sustentação do Capítulo 2 foram feitas pesquisas sobre livros digitais, livros digitais licenciados, fornecedores de livros digitais para bibliotecas e implicações observadas com a inclusão de livros digitais nas coleções. Neste capítulo também foram utilizados dados obtidos por meio de entrevistas e conversas com fornecedores e bibliotecários, buscando identificar casos concretos para ilustrar as situações relatadas.

A pesquisa bibliográfica para elaboração do Capítulo 3 foi realizada sobre Web Semântica, *Linked Data* (LD), *Linked Open Data* (LOD) e *Linked Enterprise Data* (LED). Os textos foram selecionados para delimitar o objeto de estudo e identificar os referencias teóricos adequados.

No Capítulo 4 foram utilizados textos que discorriam sobre Web Semântica e bibliotecas, com destaque a projetos em curso como o LD4P, por exemplo.

Com o referencial teórico coletado sobre livros digitais licenciados e as possibilidades de uso da Web Semântica, foi possível a elaboração do modelo para troca de dados entre fornecedores e bibliotecas, proporcionando formas de gestão.

A tese está **construída** em sete capítulos, onde o primeiro apresenta o problema da pesquisa, a proposta, os objetivos geral e específicos, a justificativa, a motivação, a metodologia aplicada para obter os resultados alcançados e os trabalhos relacionados.

No Capítulo 2 são apresentados os livros digitais, os fornecedores e as bibliotecas, proporcionando uma visão ampla do cenário analisado, destacando as questões de fragilidades identificadas e que nortearão a criação do modelo.

O Capítulo 3 analisa a Web Semântica englobando *Linked Data*, *Linked Open Data* e *Linked Enterprise Data*, destacando a diferenciação entre *datasets* abertos e fechados. O Capítulo conclui com as etapas para publicação de *datasets* na Web.

O Capítulo 4 discute a Web Semântica em bibliotecas, buscando reunir aspectos abordados no Capítulo 3, identificando as possibilidades de utilização para solução do problema de pesquisa. Foram enfatizadas somente iniciativas que podem ser utilizadas no modelo.

O Capítulo 5 apresenta o modelo de dados proposto para a gestão de livros digitais licenciados por meio da Web Semântica. São destacadas as etapas para criação do modelo como: política e seleção, formalização, formatos e licenças; conversão ferramental; processo de recuperação; e marketing e *feedback*. São feitos mapeamentos de dados em formatos diferentes, buscando a adoção de um esquema único para fornecedores e bibliotecas.

O Capítulo 6 busca um exemplo de validação do modelo com a realização de tratamento de registro bibliográfico enriquecido e demais funcionalidades que podem ser utilizadas.

No Capítulo 7 são apresentadas as considerações finais sobre a pesquisa, resultados alcançados e possibilidades de novos estudos.

### **1.3 *Trabalhos relacionados***

Algumas pesquisas já foram realizadas sobre os temas da Web Semântica e os livros digitais. Somente dois trabalhos foram localizados que discutem a possibilidade do *Linked Data* ser utilizado na gestão de livros digitais. Outros textos abordam alguns dos aspectos sobre os livros digitais que motivaram esta pesquisa.

Byrne e Goddard (2010) discorrem sobre a aplicação de LD em bibliotecas enfocando exatamente a possibilidade de comunicação entre fornecedores de conteúdo e bibliotecas para compartilhamento de dados bibliográficos e de autoridades. Uma integração desta natureza resultaria em melhorias nas atividades de aquisição, catalogação, controle de autoridades, vocabulário controlado etc. As autoras sugerem que as bibliotecas devem exigir dos fornecedores a disponibilização de dados em formato semântico.

Pode-se facilmente imaginar bibliotecas trabalhando com seus fornecedores para desenvolver colaborativamente uma grande base de conhecimento compartilhada que poderia atuar como um “eixo de vinculação”. Este eixo de vinculação poderia expor uma estreita rede de informações entre editores, agregadores, fornecedores de livros e periódicos, autoridades de assuntos, nome, e outras bibliotecas. Essas grandes lojas de dados vinculados poderiam proporcionar um verdadeiro “sistema integrado de bibliotecas” lidando com todas as funções de seleção, encomenda, catalogação, controle de autoridade,



desenvolvimento de taxonomia e pesquisa (BYRNE; GODDARD, 2010, não paginado, tradução nossa)<sup>5</sup>.

MacEwan (2004) apresenta o projeto piloto *InterParty*, lançado em 2003 e que teria duração de apenas um ano. Este projeto foi fundado pela *European Commission* (EC) em parceria com a EDItEUR, *British Library*, *Royal Swedish Library*, *International Federation of Library Associations* (IFLA), *Library of Congress* (LOC), *Online Computer Library Center* (OCLC), *Digital Object Identifiers* (DOI) *Foundation*, entre outros, e previa a possibilidade de construção de uma rede para distribuição e acesso de metadados comuns à indústria do livro, apresentando ganhos para todos os envolvidos. A interoperabilidade entre os atores do mercado do livro proporcionaria condições para descoberta de recursos, identificação de direitos de propriedade, negociação de contratos, pagamento de *royalties*, desenvolvimento de modelos de negócios e propostas de governança.

O projeto cita a presença de silos de informação<sup>6</sup> existentes em conjuntos de dados que estão isolados, mas que podem ser compartilhados, porém que esbarram em diferenças de formatos e esquemas de representação. Um dos aspectos deste projeto seria a criação de identificadores de autoridades, utilizando dados de bibliotecas, aumentando a confiabilidade dos registros (MACEWAN, 2004).

O *InterParty* seria de utilização restrita para assinantes. Não foram localizadas outras informações sobre o projeto, porém o autor do artigo é atualmente integrante do *Governing Board of Directors* do *International Standard Name Identifier* (ISNI).

Muitos textos discorrem sobre os problemas enfrentados pelas bibliotecas com os metadados enviados por fornecedores, como baixa qualidade da descrição de livros ausência de dados essenciais, links quebrados etc. Os problemas no MARC 21

---

<sup>5</sup> *One can easily imagine libraries working with their vendors to collaboratively develop a large shared knowledgebase that could act as a library "linking hub". The linking hub would expose a network of tightly linked information from publishers, aggregators, book and journal vendors, subject authorities, name authorities, and other libraries. These large stores of linked data could provide a true "integrated library system" handling all of the functions of selection, ordering, cataloguing, authority control, taxonomy development, and search* (BYRNE; GODDARD, 2010, não paginado).

<sup>6</sup> De acordo com Bermès (2013), silos de informação são conjuntos de dados fechados e que não se comunicam com outros dados dispostos na Web ou outros sistemas de dados. As descrições dos elementos foram feitas, porém de forma fechada, restringindo ou impedindo que estes dados sejam reaproveitados ou vinculados a dados externos.

frequentemente apontados são a presença de campos obsoletos, campos e subcampos inexistentes no formato, falta de indicadores etc.

Outros aspectos relatados são a presença de duplicação de livros digitais, atualização dos metadados no catálogo, obras concorrentes e representação dos metadados de licenciamento.

A duplicação de livros digitais pode ocorrer no momento da renovação da assinatura. A cada novo ciclo o fornecedor encaminha arquivo com todos os livros digitais licenciados, onde constam os que já faziam parte da coleção e os novos que foram acrescentados. Como a maioria dos bibliotecários realiza ajustes nas descrições, nem sempre é possível identificar que são obras iguais e mais um cadastro é incluído na base da biblioteca. Também foram relatados questionamentos sobre a duplicação de livros digitais quando a biblioteca possui a versão impressa e licencia a versão digital. Algumas bibliotecas relataram que o fato da coleção possuir a versão impressa de uma obra dispensaria o cadastro da versão digital, podendo incluir a URL (*Uniform Resource Locator*) no cadastro da versão impressa. Esta conduta não possui aderência com o *Statement of International Cataloguing Principles* (CIP, Princípios Internacionais de Catalogação) da IFLA que orienta no item 5.1: “Em geral, uma descrição bibliográfica separada deve ser criada para cada manifestação” (IFLA, 2009, p. 4, tradução nossa)<sup>7</sup>.

A contratação de pacotes de livros digitais é usual. Porém estes pacotes são dinâmicos, ocorrendo a inclusão e exclusão de livros digitais, mesmo durante a vigência do licenciamento. Isto demanda atualizações no catálogo. Como os pacotes podem ser bem grandes, com mais de 300.000 livros digitais (SPURGIN, 2014), muitas bibliotecas questionam se realmente todas as obras licenciadas foram disponibilizadas pelo fornecedor. Devido ao volume, é complexo às bibliotecas verificarem os livros digitais individualmente. Outro fator reportado consiste em falhas de comunicação dos fornecedores ao não notificar alterações (inclusões e exclusões) nos pacotes.

Um aspecto identificado na literatura diz respeito às obras concorrentes e livros digitais duplicados, que serão detalhados no Capítulo 2. Outro aspecto está relacionado com os metadados de licenciamento. Alguns fornecedores se utilizam de campos

---

<sup>7</sup> 5.1. *In general, a separate bibliographic description should be created for each manifestation.*

MARC 21 para armazenar estes dados, ou então a criação de *holding* para contabilizar a coleção.

A **Tabela 1** reúne os principais problemas com livros digitais identificados na literatura consultada.

**Tabela 1:** Principais problemas reportados na literatura

<b>Problemas relatados</b>	<b>Fonte</b>
Ausência de assunto	THOMPSON; TRAILL (2017); SAPON-WHITE (2014); MUGRIDGE; EDMUNDS (2012); ROSSMANN, FOSTER; BABBITT (2009)
Ausência de descrição física	SAPON-WHITE (2014); TRAILL (2013); WALTERS (2013); MARTIN (2007)
Ausência de número de classificação	THOMPSON; TRAILL (2017); VAN KLEECK et al. (2017); SAPON-WHITE (2014); ROSSMANN, FOSTER; BABBITT (2009)
Ausência de notas (conteúdo)	THOMPSON; TRAILL (2017); VAN KLEECK et al. (2017); TRAILL (2013); ROSSMANN, FOSTER; BABBITT (2009)
Atualização de livros digitais no catálogo	CHEN; KIM; MONTGOMERY (2016); SAPON-WHITE (2014); HODGE; MANOFF; WATSON (2013); TRAILL (2013); WALTERS (2013); MUGRIDGE; EDMUNDS (2012); ROSSMANN, FOSTER; BABBITT (2009); MARTIN (2007)
Dados de licenciamento	THOMPSON; TRAILL (2017); CHEN; KIM; MONTGOMERY (2016); WALTERS (2013); WU; MITCHELL (2010); ROSSMANN, FOSTER; BABBITT (2009)
Demora na atualização dos livros digitais	MARTIN; MUNDLE (2010)
Dificuldade na gestão da coleção	CHEN; KIM; MONTGOMERY (2016); MUGRIDGE; EDMUNDS (2012)
Disponibilidade dos livros digitais na plataforma	SPURGIN (2014)
Falta de notificações por parte dos fornecedores	CHEN; KIM; MONTGOMERY (2016); GEORGAS (2015); WALTERS (2013); MUGRIDGE; EDMUNDS (2012); MARTIN; MUNDLE (2010); ROSSMANN, FOSTER; BABBITT (2009); MARTIN (2007)
Falhas no controle de autoridades	VAN KLEECK et al. (2017); SAPON-WHITE (2014); TRAILL (2013); MUGRIDGE; EDMUNDS (2012); MARTIN; MUNDLE (2010); ROSSMANN, FOSTER; BABBITT (2009)
<i>International Standard Book Number (ISBN)</i> [ausência, dois ISBNs (impresso e digital), ISBN do impresso no lugar do digital etc.]	VAN KLEECK et al. (2017); SAPON-WHITE (2014); WALTERS (2013); MARTIN (2007)

Links quebrados	SAPON-WHITE (2014); SPURGIN (2014); TRAILL (2013); MUGRIDGE; EDMUNDS (2012); MARTIN; MUNDLE (2010)
Não aderência ao MARC 21 e/ou normas catalogação	THOMPSON; TRAILL (2017); CHEN; KIM; MONTGOMERY (2016); SAPON-WHITE (2014); TRAILL (2013); WALTERS (2013); MUGRIDGE; EDMUNDS (2012); MINCIC-OBRADOVIC (2011); MARTIN; MUNDLE (2010); ROSSMANN, FOSTER; BABBITT (2009); MARTIN (2007)
Necessidade de ajustes nos registros	CHEN; KIM; MONTGOMERY (2016); GEORGAS (2015); SPURGIN (2014); WALTERS (2013); MUGRIDGE; EDMUNDS (2012); MARTIN; MUNDLE (2010); ROSSMANN, FOSTER; BABBITT (2009); MARTIN (2007)
Obras concorrentes	CHEN; KIM; MONTGOMERY (2016); SAPON-WHITE (2014); MUGRIDGE; EDMUNDS (2012); MARTIN; MUNDLE (2010); MARTIN (2007)
Livros digitais duplicados	CHEN; KIM; MONTGOMERY (2016); WALTERS (2013); MARTIN; MUNDLE (2010); ROSSMANN, FOSTER; BABBITT (2009)
Livros digitais removidos	CHEN; KIM; MONTGOMERY (2016); GEORGAS (2015); SPURGIN (2014); WALTERS (2013); MUGRIDGE; EDMUNDS (2012); ROSSMANN, FOSTER; BABBITT (2009); MARTIN (2007)

Fonte: desenvolvido pela autora

Em 2011 a *National Information Standards Organization* (NISO) formou um grupo de interesse para discutir livros digitais, o *NISO e-Book SIG* (*Special Interest Group*, Grupo Especial de Interesse, tradução nossa). O grupo foi criado para discutir as melhores práticas e padrões relacionados à criação, distribuição, descoberta e preservação de livros digitais, buscando reunir bibliotecas, editores, agregadores e demais organizações que atuam na área de ensino e pesquisa, gestão e curadoria do conhecimento. Também estavam previstas análises sobre temas relacionados ao livro digital, como acessibilidade e metadados, abrangendo os formatos ONIX, MARC 21, *Preservation Metadata Maintenance Activity* (PREMIS), *Metadata Encoding and Transmission Standard* (METS), Dublin Core (DC) etc. (NISO, 2011).

A ideia de criação do grupo de interesse foi motivada por reunião na Conferência em 2011 da *American Library Association* [ALA] *Midwinter* que identificou que a NISO poderia apoiar as iniciativas relacionadas aos livros digitais. As postagens na página do grupo datam de 2011. Em contato realizado em novembro de 2018 com a NISO, foi informado que o grupo estava inativo.

O próximo capítulo apresenta um panorama sobre livros digitais licenciados, contextualizando fornecedores, formas de contratação e de acesso. Os problemas identificados entre fornecedores e bibliotecas que nortearam esta pesquisa são discutidos a seguir.

## 2 Livro digital, bibliotecas e fornecedores

---

### Preâmbulo

Este capítulo apresenta um panorama geral dos livros digitais descrevendo os tipos de fornecedores para bibliotecas, as formas de acesso à plataforma e ao conteúdo, a possibilidade ou não de uso simultâneo dos recursos, a contratação por seleção de livros digitais individuais ou pacotes, e a necessidade de inserir e recuperar os dados de licenciamento nos registros bibliográficos.

Ao analisar este cenário, três principais problemas foram identificados: a qualidade dos metadados encaminhados pelos fornecedores, a fragilidade do desenvolvimento da coleção, e as dificuldades para a gestão dos livros digitais contratados.

A geração de registros em MARC 21 é complexa aos fornecedores, visto que não utilizam este formato, mas o ONIX, ou ainda controlam os metadados por meio de planilhas de dados ou sistemas proprietários. Esta situação contribui para que metadados de livros licenciados sejam encaminhados às bibliotecas sem aderência ao MARC 21, demandando ações de ajustes e correções. Outro ponto observado é a utilização do MARC 21 para registro de dados de licenciamento.

Em relação ao desenvolvimento da coleção é apresentada a insegurança da permanência do livro digital, uma vez pode ser removido da plataforma do fornecedor. Alterações na coleção contratada ou na forma de licenciamento realizado demandam ajustes de metadados, onerando a equipe da biblioteca. A presença de obras concorrentes e/ou duplicadas sugerem ações de unificação de *links* de acessos para livros digitais iguais oferecidos por fornecedores diferentes, e inibição ou exclusão destes no catálogo *online*. A comunicação entre fornecedores e bibliotecas é, usualmente, irregular e falha, o que dificulta a gestão.

Como os dados de fornecedores estão disponíveis nas plataformas, a biblioteca precisa coletar informações que estão distribuídas para consolidar e analisar os dados de uso, estatísticas e indicadores de todos os fornecedores contratados para, a partir destes dados, gerir a coleção dos livros digitais licenciados.

O capítulo encerra abordando o *Standardized Usage Statistics Harvesting Initiative* (SUSHI) e o *Counting OnlineUser NeTworked Electronic Resources* (COUNTER), ferramentas que permitem aos fornecedores a coleta de dados de uso de recursos digitais.

Segundo Frederick (2015), o modelo teórico de Christensen discute a existência de tecnologias sustentáveis – que contribuem com o trabalho desenvolvido, agilizando suas atividades e diminuindo custos -, e existem as tecnologias disruptivas, que acarretam transformações em práticas já estabelecidas, e que podem proporcionar impactos negativos e instabilidade nos serviços prestados. Considerando que nenhuma tecnologia é totalmente disruptiva e que o grau de disrupção depende da área afetada, observa-se que os livros digitais proporcionam benefícios às bibliotecas, mas, por outro lado, exigem revisão de atividades já estabelecidas. Situações de mudanças devem ser encaradas por fornecedores e bibliotecas, ajustando suas rotinas e estabelecendo novas práticas, visto que as existentes e estabelecidas foram afetadas pela tecnologia.

Os livros digitais são recursos cujo acesso é realizado de forma mediada por plataforma de leitura. As plataformas dos fornecedores são desenvolvidas de acordo com o interesse destes, sem adoção de padrão, muitas vezes em formatos proprietários, o que dificulta a troca de informações de e sobre os livros digitais. O *Digital Rights Management* (DRM) controla o uso que será feito dos livros digitais, permitindo o *download* do arquivo ou não, impressão de algumas páginas, a leitura *off-line* e demais restrições (ZHANG, 2018).

Embora possam contratar livros digitais em livrarias virtuais, as bibliotecas usualmente realizam licenciamentos por meio de fornecedores específicos. Isto é decorrente das livrarias virtuais atrelarem o cadastro do contratante a um dispositivo de leitura, o que limitaria as possibilidades de utilização dos usuários por não permitir o acesso por *links* no catálogo em linha (SERRA; SILVA, 2017). Os fornecedores para bibliotecas são os editores, os distribuidores e os agregadores de conteúdo.

Os editores são os profissionais do livro. São empresas que representam os autores e publicam suas obras, respondendo pela criação editorial, divulgação, distribuição e venda. Distribuidores são empresas que não publicam obras, mas as comercializam, com permissão dos editores ou agregadores que representam. [...] Os agregadores de conteúdo são empresas que comercializam títulos de editores, porém em plataformas próprias. (SERRA, 2017, p. 232).

Editores que possuem plataformas de leitura podem realizar licenciamentos direto com as bibliotecas. Editoras que não desenvolveram plataformas podem disponibilizar seus livros digitais por meio de agregadores. Caso a editora tenha



plataforma própria, além de comercializar seus próprios livros digitais, também pode ser representada por agregadores. Os distribuidores não possuem plataformas e comercializam os livros de editores e/ou agregadores. A distinção entre agregadores e distribuidores é que somente o agregador possui plataforma. O distribuidor permite o acesso aos livros por meio da plataforma do editor ou do agregador.

O conteúdo licenciado é contratado de forma individual ou em pacotes de publicações. A contratação é definida por meio de aplicação de modelo de negócio, que pode ser perpétuo - sem data para expirar -, transitório - com a realização de licenciamentos periódicos -, ou orientado ao usuário - com seleção e acesso sendo definidos pelos usuários (SERRA; SANTARÉM SEGUNDO, 2018).

Nos modelos transitórios, a definição dos livros digitais que serão incluídos no acervo é realizada, na maioria das vezes, pelo fornecedor, por meio de pacotes ou conjunto de livros. Cada fornecedor tem sua plataforma e o processo de licenciamento varia, sendo único para cada um.

As plataformas podem ser acessadas por computadores, dispositivos de leitura (*e-readers, tablets*) ou *smartphones*. A descoberta da coleção digital pode ocorrer no catálogo em linha (*Online Public Access Catalog - OPAC*), nas plataformas dos fornecedores ou ainda por meio de ferramentas de serviços de descoberta (WIKOFF, 2012). Como uma biblioteca não conseguirá licenciar todos os livros digitais de seu interesse com um único fornecedor, observa-se um cenário de contratação de diversos fornecedores de conteúdo digital, de acordo com as editoras representadas ou pela temática ofertada. Entretanto, diferentemente dos livros impressos, a relação com o fornecedor não se encerra quando os livros digitais contratados são entregues, visto que a leitura ocorrerá pela plataforma do fornecedor, sem a qual não existe acesso ao conteúdo.

Permitir a busca dos livros por meio de consulta na plataforma de cada fornecedor não é interessante às bibliotecas, uma vez que os usuários devem primeiramente saber quais fornecedores possuem determinado livro ou obras de uma editora para, a partir desta informação, acessar a plataforma adequada, fazer a pesquisa, localizar a obra desejada e realizar o acesso. Esta situação é confusa aos usuários e demanda esforços na localização dos livros digitais desejados. Também não é prático aos bibliotecários, pois não permite a visualização da coleção como um todo,

contemplando as obras impressas e digitais, além de demandar a coleta e o tratamento dos dados de acessos realizados na plataforma de cada fornecedor, quando os dados estão disponíveis ou são fornecidos, dificultando ou comprometendo sua compilação e, conseqüentemente, a análise de indicadores e geração de relatórios.

Uma alternativa é a inclusão dos livros digitais que foram licenciados no catálogo da biblioteca, com identificação das obras ocorrendo em um único local, independente do fornecedor contratado (NISO, 2014; BEALL, 2009). De acordo com Armstrong e Lonsdale (2011), o OPAC é a forma preferida dos usuários para o acesso aos livros digitais. O mesmo entendimento foi apresentado por Miao Jin na *Mississippi State University Libraries' eResources & Emerging Technologies Summit* em 2013, em palestra que lembrou a função do catálogo da biblioteca como o local adequado para localização dos livros digitais: 1) é um inventário da coleção; 2) permite aos usuários descobrirem os recursos presentes no acervo; e 3) provê informação de como acessar o recurso (ZHANG; JIN, 2014).

Os livros digitais demandam dois tipos distintos de acesso: à plataforma e ao conteúdo. Sem acessar a plataforma não é possível realizar a leitura do livro. Somente usuários identificados e validados conseguem passar pela barreira da plataforma para então poder consultar o conteúdo. A identificação dos usuários ocorre por *login*, integração entre sistemas ou permissão de acesso a redes privadas.

As formas de acesso ao conteúdo são atributos dos modelos de licenciamento e são elementos essenciais para determinar o valor da contratação e se os livros podem ser acessados de forma simultânea (concorrente) ou não. Os atributos de acesso são:

- 1) Monousuário: sem acesso simultâneo (SERRA; SANTARÉM SEGUNDO, 2018; WALTERS, 2014);
- 2) Multiusuário: possibilidade de acesso simultâneo limitado a uma quantidade estipulada (SERRA; SANTARÉM SEGUNDO, 2018). Emula a aquisição de exemplares impressos que podem ser consultados. Caso todos os “exemplares digitais” estejam sendo utilizados, o acesso é restrito até que o empréstimo digital em curso expire. Alguns fornecedores permitem a formação de lista de reservas caso não existam acessos disponíveis, ou então uma devolução antecipada;

- 3) Número de usuários: fornecedor e biblioteca estipulam a quantidade de usuários que utilizarão a plataforma, com o valor da contratação norteados pela quantidade de potenciais acessos que podem ocorrer. Este atributo pode permitir o uso concorrente ou não (SERRA; SANTARÉM SEGUNDO, 2018);
- 4) Ilimitado: acesso livre aos livros digitais com uso concorrente (SERRA; SANTARÉM SEGUNDO, 2018).

O fornecedor pode ampliar a oferta de livros digitais que estão presentes no pacote disponível permitindo, inclusive, que todo seu catálogo esteja acessível. Dependendo do modelo contratado, a biblioteca fará reajustes periódicos (modelos transitórios) ou de acordo com a quantidade de acessos realizados (modelos orientados aos usuários), efetuando pagamentos totais ou parciais pelos conteúdos que foram utilizados.

Os problemas observados entre fornecedores e bibliotecas que são analisados nesta pesquisa são:

- 1) Qualidade dos metadados;
- 2) Desenvolvimento da coleção;
- 3) Gestão dos livros digitais contratados e geração de indicadores.

Estes itens serão discutidos a seguir.

## **2.1 *Qualidade dos metadados***

De acordo com Van Kleeck et al. (2017) a gestão da qualidade engloba controles como garantia de qualidade e controle de qualidade. Enquanto a garantia de qualidade preocupa-se com a entrada de dados e a correção de erros pré-existentes, o controle de qualidade ocupa-se com ajustes após a inclusão dos dados. Para os autores, enquanto a catalogação manual atua no controle de qualidade durante o processo de inclusão do registro, a importação de metadados de livros digitais sofre da ausência de atenção na descrição.

A qualidade dos metadados oferecidos pelos fornecedores de livros digitais, em geral, é baixa (THOMPSON; TRAILL, 2017; VAN KLEECK et al., 2017; SAPON-

WHITE, 2014; WALTERS, 2013; MUGRIDGE; EDMUNDS, 2012; MINCIC-OBRADOVIC, 2011; MARTIN; MUNDLE, 2010; BEALL, 2009; ROSSMANN; FOSTER; BABBITT, 2009), visto que os padrões descritivos adotados pelas bibliotecas não são os mesmos adotados por editores, sem o rigor definido pelas normas internacionais de representação desenvolvidos pela Ciência da Informação. A baixa qualidade dos metadados dos livros digitais compromete a descoberta de recursos no catálogo. Esta situação faz com que metadados oferecidos por fornecedores resultem na necessidade de ajustes dos dados, tornando a catalogação mais custosa.

Como um membro do CLI [*Center of Library Initiatives*<sup>8</sup>] comentou, os registros “gratuitos” acabam sendo muito caros, dado o trabalho inicial para melhorar o arquivo em MARC e o trabalho feito em cada biblioteca para limpar os pontos de acesso e ajustar os registros nos sistemas locais (MARTIN; MUNDLE, 2010, p. 234-235, tradução nossa).<sup>9</sup>

Orientações de catalogação para editores foram publicadas (BOOK INDUSTRY STUDY GROUP (BISG), 2015; REGISTER; MCLROY, 2015; BRAND; DALY; MEYERS, 2003), como, por exemplo, o MARC *Record Guide for Monograph Aggregator Vendors* (Guia de descrição de monografias no MARC 21 para agregadores e vendedores, tradução nossa), desenvolvido pelo *Program for Cooperative Cataloging - PCC* (2011). Entretanto, este e outros esforços não parecem ter contribuído com a qualidade dos metadados disponibilizados pelos fornecedores.

Existe, entretanto, demanda para que os fornecedores ofereçam dados de livros digitais no padrão utilizado pelas bibliotecas (WALTERS, 2013; LUTHER, 2009), afinal é inviável fazer catalogação manual ou mesmo importação dos metadados disponibilizados em outras bases de dados de forma individual, visto que os pacotes assinados podem conter milhares de livros digitais.

Os fornecedores usualmente utilizam o formato ONIX que é aplicado para livros impressos e digitais, reunindo dados descritivos de registros bibliográficos acrescido de informações sobre disponibilidade em estoque, valores de frete, material

---

<sup>8</sup> O CLI é uma iniciativa de membros do *Committee on Institutional Cooperation* (CIC) para facilitar o uso de recursos presentes em bibliotecas universitárias, economizando orçamentos, tempo e espaço.

<sup>9</sup> *As one CLI member commented, the “free” records ended up being quite expensive, given the initial work to improve the MARC file and the work done by each library to clean up access points and set records in the local systems* (MARTIN; MUNDLE, 2010, p. 234-235).

para divulgação, dados complementares etc. O *ONIX for Books Product Information Format Specification* (ONIX para especificação de formato de livros e produtos de informação, tradução nossa) foi compilado para a EDItEUR, grupo internacional que busca a padronização de infraestrutura para a comercialização de livros impressos, digitais e periódicos. A EDItEUR também publicou um guia de boas práticas que, entre outras abordagens, estimula a adoção de padrões descritivos e a interoperabilidade de dados de comercialização para atuação em nível mundial (EDITEUR, 2012).

O ONIX é desenvolvido em linguagem *eXtensible Markup Language* (XML), possui o formato *Document Type Definition* (DTD) e o esquema *XML Schema Definition* (XSD), que são recomendações da *World Wide Web Consortium* (W3C).

Assim como o formato MARC 21, existe mais de um formato ONIX, tratando informações distintas. O *ONIX for Books* é empregado para livros em quaisquer formatos, o *ONIX for Subscription Products*, para assinaturas de periódicos, bases de dados, *softwares* etc., o *ONIX for Licensing Terms and Rights Information* descreve características de comercialização, e o *ONIX Identifier Registration Format* que engloba a descrição de ISBN, DOI etc., vinculados às respectivas agências de registro.

Contudo, nem todos os fornecedores trabalham com o ONIX, utilizando-se de planilha de dados para cadastrar e controlar seus livros digitais, ou ainda de sistemas proprietários ou listas impressas (LUTHER, 2009). Isto acarreta na distribuição de dados não estruturados, com fornecedores criando planilhas e colunas com quantidades, tipos de informação e conteúdos próprios, sem nenhuma padronização. Para que seja possível uma importação de dados de fornecedores por meio de planilha de dados é necessária a adoção de um modelo básico, que será detalhado no Capítulo 5.

Os fornecedores não utilizam o formato MARC 21, visto que é utilizado no ambiente de bibliotecas, assim como as bibliotecas não utilizam o ONIX, que por sua vez atende somente aos editores e livrarias. Apesar de existir mapeamento entre o ONIX e o MARC 21 (GODBY, 2012), este não será aplicado nesta pesquisa, uma vez que a proposta é utilizar-se da Web Semântica, que se serve dos modelos de dados *Schema* e *Resource Description Framework* (RDF), que serão discutidos nos Capítulos 3 e 4.

Para oferecer metadados em MARC 21 os fornecedores precisam recorrer a ferramentas terceiras que propiciem a conversão dos dados, desenvolver o MARC 21 em seus sistemas proprietários ou contratar bibliotecários para catalogar os livros digitais, resultando em aumento de custos e esforços redundantes para representação de uma informação que já existe em outro formato.

A **Figura 1** exemplifica a não aderência aos padrões de representação descritiva livro digital em MARC 21 encaminhado por editor internacional que atua no Brasil.

**Figura 1:** Livro digital em MARC 21 oferecido por fornecedor

---

```

000 04890nam a2200145 4500
001 000055675
005 20170915104435.9
008 170915|2016|||||||||||||||||||d
020 __ |a 9788535284188
040 __ |c NyNyIMP
100 1_ |a NEWMAN, Michael G.; TAKEI, Henry; KLOKKEVOLD, Perry R.;
CARRANZA, Fermin A., Jr., |e Author
245 10 |a Carranza Periodontia Clínica 12ED |c NEWMAN, Michael G.; TAKEI,
Henry; KLOKKEVOLD, Perry R.;CARRANZA, Fermin A., Jr.,.
260 __ |c 2016
520 2_ |a Totalmente revisado e atualizado para atender agrave;s demandas do
seacute;culo XXI, Carranza Periodontologia Cliacute;nica oferece o que haacute;
de mais recente em ciecirc;ncia baacute;sica, assim como procedimentos
periodontais baacute;sicos e avancedil;ados de faacute;cil aprendizagem. Esta
12a ediccedil;atilde;o incorpora alteraccedil;otilde;es compatiacute;veis com
nossa era digital, ao mesmo tempo em que manteacute;m seu excelente
conteuacute;do, traccedit;o caracteriacute;stico do livro ao longo dos anos.

Padratilde;o ouro desde 1947, Carranza Periodontologia Cliacute;nica eacute;
mais do que apenas um livro-texto. Ele conta com a orientaccedil;atilde;o de
especialistas, uma melhor organizaccedil;atilde;o, novo conteuacute;do on-line,
descreve teacute;cnicas em terapias reconstrutiva, esteacute;tica e de implante,
aleacute;m de explicar a evidecirc;ncia de apoio de cada tratamento.

Profissionais de renome ajudam o leitor a compreender os fundamentos, tomar as
melhores decisotilde;es cliacute;nicas, obter os melhores resultados de cada
procedimento, evitar complicaccedil;otilde;es e exceder as expectativas dos
pacientes.

PRINCIPAIS CARACTERIacute;STICAS:
Mais de 5.000 imagens ilustram os detalhes das condiccedil;otilde;es e
tratamentos especiacute;ficos.

```

- Conteacute;m um Atlas de Patologia Periodontal com a mais abrangente coleccedit;atilde;o de casos encontrados em toda parte.

- Inclui a etiologia e o tratamento de doenças periodontais, a relação entre doença periodontal e saúde sistêmica, assim como a implantodontia oral.
- Casos clínicos baseados em cenários reais mostram como solucionar problemas e melhorar os resultados do paciente.
- Ciência atualizada e conteúdo clínico com base em evidências científicas ajudam na revisão de tópicos importantes e na preparação para provas.

**520** |a Totalmente revisado e atualizado para atender às demandas do século XXI, Carranza Periodontologia Clínica oferece o que há de mais recente em ciência atualizada, assim como procedimentos periodontais básicos e avançados de fácil aprendizagem. Esta 12ª edição incorpora alterações compatíveis com nossa era digital, ao mesmo tempo em que mantém seu excelente conteúdo, o caráterístico do livro ao longo dos anos.

Padrão ouro desde 1947, Carranza Periodontologia Clínica é mais do que apenas um livro-texto. Ele conta com a orientação de especialistas, uma melhor organização, novo conteúdo on-line, descreve técnicas em terapias reconstrutiva, estética e de implante, além de explicar a evidência de apoio de cada tratamento.

Profissionais de renome ajudam o leitor a compreender os fundamentos, tomar as melhores decisões clínicas, obter os melhores resultados de cada procedimento, evitar complicações e exceder as expectativas dos pacientes.

#### **PRINCIPAIS CARACTERÍSTICAS:**

- Mais de 5.000 imagens ilustram os detalhes das condições e tratamentos específicos.
- Contém um Atlas de Patologia Periodontal com a mais abrangente coleção de casos encontrados em toda parte.
- Inclui a etiologia e o tratamento de doenças periodontais, a relação entre doença periodontal e saúde sistêmica, assim como a implantodontia oral.
- Casos clínicos baseados em cenários reais mostram como solucionar problemas e melhorar os resultados do paciente.
- Ciência atualizada e conteúdo clínico com base em evidências científicas ajudam na revisão de tópicos importantes e na preparação para provas.

**856 4** |u <http://fornecedor.ipublish.com/ebook/9788535284188>

Fonte: Metadados originais de editor com atuação no Brasil

Zhang e Jin (2014) relatam que não é incomum a ocorrência de uso de campos obsoletos do MARC 21 ou codificações não autorizadas. Rossmann, Foster e Babbitt (2009) apontam outras falhas ou omissões, conforme descrito a seguir:

Alguns vendedores não criam números de chamadas e detalhados cabeçalhos de assuntos em seus registros gratuitos, portanto a navegabilidade é reduzida (a não ser que a biblioteca invista tempo

ou dinheiro acrescentando números de chamada e pontos de acessos de assuntos adequados (ROSSMANN; FOSTER; BABBITT, 2009, p. 547, tradução nossa)<sup>10</sup>.

Dentre os problemas apontados, o mais grave está relacionado com as *tags* 100|a e 700|a pois, além de comprometer o registro bibliográfico, inclui dados não padronizados no cadastro de autoridade do tipo Pessoa e pontos de acesso.

Critérios para estabelecer o nível de qualidade são relativos à política descritiva. Segundo Rossmann, Foster e Babbitt (2009), os bibliotecários devem definir qual o nível de qualidade que é aceitável nos registros em MARC 21 oferecidos pelos fornecedores, porém é necessário computar as atividades que serão feitas pelos bibliotecários nos metadados recebidos.

Esta seção fez uma breve análise da qualidade dos metadados encaminhados às bibliotecas por fornecedores. A seguir serão discutidos aspectos relacionados ao desenvolvimento da coleção e livros digitais.

## **2.2 *Desenvolvimento da coleção***

O desenvolvimento da coleção é bastante afetado com a inclusão de livros digitais licenciados em decorrência da constante entrada e saída de livros digitais nos pacotes contratados (FREDERICK, 2015), aliado a falhas de comunicação entre fornecedores e bibliotecas, e a presença de obras concorrentes e livros digitais duplicados. Estes aspectos serão analisados a seguir.

Com os livros impressos, após a aquisição e recebimento dos mesmos é feita a representação no catálogo e sua disponibilização para consultas. O fornecedor não vai até a estantes da biblioteca para remover os exemplares que foram adquiridos caso um autor transfira seus direitos autorais para outra editora. Também não removem edições antigas quando novas são lançadas. Muito menos substituem livros digitais por obras similares e não definem eventuais restrições de acesso aos leitores. Entretanto, estas situações ocorrem com os livros digitais (FREDERICK, 2015).

---

<sup>10</sup> *Some vendors do not create call numbers and detailed subject headings in their free records, so browsability is reduced (unless libraries spend time or money adding proper call numbers and additional subject access points* (ROSSMANN; FOSTER; BABBITT, 2009, p. 547).



Ao contratar um pacote de livros digitais normalmente o fornecedor encaminha um arquivo com todos os livros digitais em MARC 21 para inclusão no catálogo. Como a quantidade de livros digitais costuma ser volumosa, nem sempre a biblioteca consegue conferir se todos os contratados estão no conjunto e pode ocorrer da quantidade do pacote não coincidir com o número de registros bibliográficos no arquivo importado (ZHANG; JIN, 2014). A mudança de livros digitais licenciados nos pacotes é constante, com alteração das obras e a sua disponibilidade ocorrendo de acordo com as negociações entre fornecedores e autores ou seus representantes.

Durante a vigência do licenciamento os livros digitais presentes em um catálogo podem ser alterados, independentemente do modelo de negócio que foi aplicado. Esta situação exige acompanhamento de livros digitais que são incluídos ou removidos dos pacotes e que, conseqüentemente, devem ter sua situação atualizada no catálogo. Nem sempre o fornecedor informa a biblioteca que ocorreram alterações nos pacotes (CHEN; KIM; MONTGOMERY, 2016; GEORGAS, 2015; SPURGIN, 2014; HODGE; MANOFF; WATSON, 2013; WALTERS, 2013; MUGRIDGE; EDMUNDS, 2012; WIKOFF, 2012; MINCIC-OBRAĐOVIC, 2011; ROSSMANN; FOSTER; BABBITT, 2009; WU; MITCHELL, 2010).

A interrupção do acesso aos livros digitais licenciados pode ocorrer se editores romperem com os agregadores, ou autores trocarem de editoras, ou ainda se bibliotecas cancelarem contratos de licenciamento ou optarem pela não renovação. Também pode ocorrer de editoras encerrarem suas atividades, causando instabilidade no acesso ao livro digital, uma vez que ele é condicionado ao uso da plataforma do fornecedor. Se a editora encerrar sua plataforma, o acesso ao livro licenciado é finalizado (SERRA; SILVA, 2016).

Foram localizados relatos sobre a remoção ou indisponibilidade de metadados de livros digitais nas plataformas dos fornecedores. O primeiro caso para ilustrar esta situação não afetou diretamente bibliotecas, mas os leitores. Em 2009 as obras “A Revolução dos bichos” (*Animal farm*) e “1984” de George Orwell foram removidas do leitor digital dedicado (*e-reader*) *Kindle*, da *Amazon*. Os leitores que haviam licenciado estes livros digitais foram surpreendidos com a remoção das obras de seus equipamentos e ressarcidos com a devolução dos valores pagos. A remoção foi motivada porque a *Amazon* foi informada que a editora que disponibilizou os livros

digitais não possuía direitos das obras (WALTERS, 2014). Evitando futuros questionamentos legais, a *Amazon* optou por simplesmente remover as obras dos dispositivos dos leitores.

Em bibliotecas foram identificados alguns casos de remoção de livros digitais de pacotes contratados por meio de entrevistas com bibliotecários e fornecedores. Uma biblioteca de uma universidade privada em Washington, DC, por exemplo, reportou em novembro de 2018 a remoção, entre outros, dos títulos “*Learning to save the future*”<sup>11</sup>, de Alexander J. Means e “*Family firms*”<sup>12</sup>, de Brännback e Carsrud. Não foram comunicados os motivos da remoção dos livros digitais da plataforma. A bibliotecária de aquisições desta instituição reportou que, em média, 1 ou 2 livros digitais são removidos por mês, por fornecedor.

Em conversa com uma das bibliotecárias responsável pela aquisição de recursos eletrônicos da UIUC em dezembro de 2018, foi informado que cerca de 15 livros digitais são removidos ou alterados nas plataformas de agregadores mensalmente. Este fato tem motivado o interesse da biblioteca em priorizar o licenciamento diretamente com os editores, evitando a contratação de agregadores de conteúdo, exatamente pela volatilidade dos pacotes.

Um caso relatado por biblioteca universitária privada brasileira e confirmado pelo agregador que detinha os direitos de representação foi a remoção dos livros digitais do autor Guilherme Nucci, como “Curso de direito penal” v.1, 2, 3, “Processo penal e execução penal”, “Crime organizado”, entre outros, publicados pelo Grupo *Gen*. Esta situação foi confirmada por meio de entrevista com o agregador em outubro de 2018. Como os livros digitais deste autor constam na bibliografia de cursos de direito, diversas bibliotecas universitárias foram afetadas.

Em outro caso, duas bibliotecas de instituições privadas de ensino superior de São Paulo relataram em entrevistas em 2017 que fornecedores removeram alguns livros digitais na área de saúde que constavam nas bibliografias de diversos cursos. Quando identificaram a ausência dos livros digitais, os bibliotecários questionaram os

---

<sup>11</sup> MEANS, A. J. **Learning to save the future**: rethinking education and work in an era of digital capitalism. New York: Routledge, 2018. ISBN 9781315450194.

<sup>12</sup> CARSRUD, A. L.; BRÄNNBACK, M. **Family firms in transition**: case studies on the management of growth, decline, and transition. New York: Springer, 2012. ISBN 9781461460466.

fornecedores, que não haviam comunicado a remoção. Um dos fornecedores alegou que precisou remover alguns livros digitais, mas que tinha inserido outros como forma de compensação. Mesmo mantendo a quantidade de livros digitais igual a contratada, a alteração de obras não confere tranquilidade às bibliotecas, sem poder assegurar que os livros digitais estarão disponíveis aos usuários em longo prazo.

A comunicação das alterações ocorridas nos catálogos é feita de forma manual, com os fornecedores relatando por mensagem os itens que foram incluídos, alterados ou removidos (CHEN; KIM; MONTGOMERY, 2016), não proporcionando atualização dinâmica entre contratado e contratante, demandando ajustes frequentes nos metadados de descrição bibliográfica e de licenciamento. Esta rotina é trabalhosa para fornecedores e bibliotecas.

Segundo Mugridge e Edmunds (2012) e Wu e Mitchell (2010) as notificações dos fornecedores são problemáticas devido a ausência de sistema para comunicar as bibliotecas contratantes sobre as alterações realizadas. Como a notificação de alterações do fornecedor pode não ser atualizada com celeridade, corre-se o risco de atrasar a disponibilização de novos livros digitais licenciados no catálogo, deixando-os disponíveis somente na plataforma do fornecedor. Esta situação, entretanto, compromete as possibilidades de descoberta e, conseqüentemente, de utilização dos livros por parte dos usuários.

De acordo com Sapon-White (2014) a OCLC e a editora *Springer* oferecem um serviço de notificação que avisa quando novos livros digitais são disponibilizados. Chen, Kim e Montgomery (2016) relatam o caso da *University of Texas at Dallas* onde, quando as informações não são recebidas, a equipe verifica no sítio de cada fornecedor sobre alterações dos livros digitais e atualiza seu catálogo.

Outros problemas relatados por bibliotecas são a presença de obras concorrentes e a duplicação de livros digitais. Ao contratar diversos agregadores as bibliotecas ficam vulneráveis a que livros digitais iguais sejam incluídos no catálogo porque uma mesma editora disponibilizou suas obras para mais de um agregador. Estes casos são identificados como obras concorrentes (VAN KLEECK et al., 2017; CHEN; KIM; MONTGOMERY, 2016).


A presença de obras concorrentes pode ser analisada por duas perspectivas. Se existe alta demanda para utilização do livro digital, o mesmo constar em duas

plataformas amplia as possibilidades de atendimento dos usuários. Por outro lado, se a obra não apresenta demanda, um investimento redundante foi realizado. Caso a obra exista em versão impressa, contratar a versão digital passa por estas mesmas perspectivas.

A **Figura 2** apresenta a duplicação de livros digitais oferecidos por fornecedores diferentes (Wiley e Safari) no catálogo *online* da rede de bibliotecas da UIUC.

**Figura 2:** Apresentação de obras concorrentes no catálogo

---

<p><b>3: Supply chain performance and evaluation models (E-book)</b>        ● Status: (E-book) <a href="#">Wiley Online Library – Full text</a>        Estampe, Dominique        Wiley 2014</p>	
<p><b>4: Supply chain performance and evaluation models (E-book)</b>        ● Status: (E-book) <a href="#">Safari – Online access</a>        Estampe, Dominique        2014</p>	

---

Fonte: OPAC da UIUC<sup>13</sup>

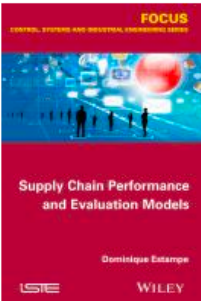
A **Figura 3** apresenta a diferença descritiva dos metadados dos livros digitais da **Figura 2**, o que também pode influenciar na descoberta e utilização dos recursos.

**Figura 3:** Visualização de representações diferentes de obras concorrentes no catálogo

---

« Back to Search Results 
[Cite this](#) [Email this](#) [Add to favorites](#) [Staff view](#)

**Supply chain performance and evaluation models [electronic resource] /**  
 Dominique Estampe.

	<p>Main Author: Estampe, Dominique.</p> <p>Published: Hoboken : Wiley, 2014.</p> <p>Topics: Business logistics – Management.</p> <p>Genres: Electronic books.</p> <p>Online Access: <a href="#">Wiley Online Library – Full text</a></p> <p>Tags: No Tags, Be the first to tag this record! <a href="#">+ Add</a></p>
---	---

---

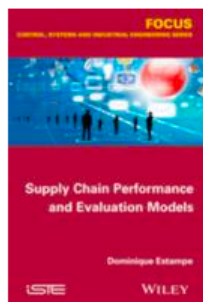
<sup>13</sup> <https://www.library.illinois.edu/>

« Back to Search Results

 Cite this  Email this  Add to favorites  Staff view

## Supply chain performance and evaluation models /

Dominique Estampe.



Author:	Estampe, Dominique
Published:	London : ISTE ; Hoboken, NJ : John Wiley & Sons, 2014.
Series:	Focus series in automation & control.
Topics:	Business logistics – Evaluation.   Delivery of goods – Management.   Industrial procurement.   Marketing channels – Management.   BUSINESS & ECONOMICS / Industrial Management.   BUSINESS & ECONOMICS / Management.   BUSINESS & ECONOMICS / Management Science.   BUSINESS & ECONOMICS / Organizational Behavior.
Genres:	Electronic books.
Online Access:	Safari – Online access
Tags:	No Tags, Be the first to tag this record!  Add

Fonte: OPAC da UIUC<sup>14</sup>

Obras duplicadas representam uma situação diferente de obras concorrentes. De acordo com Chen, Kim e Montgomery (2016), a *University of Texas at Dallas* apresentou duplicidade de livros digitais, pois a cada renovação de licenciamento uma nova carga de metadados era realizada no catálogo. Como a biblioteca tinha realizado ajustes em alguns livros digitais, a localização dos duplicados foi complexa, demandando bastante trabalho da equipe.

Van Kleeck et al. (2017) relatam a situação observada na *University of Florida* que identificou a presença de obras concorrentes e duplicadas em diversas coleções de fornecedores. As duplicidades foram constatadas devido a participação da instituição em consórcio, onde as instituições que fizeram o licenciamento incluíram os metadados em seus catálogos, resultando na redundância. Os fornecedores *Institute of Electrical and Electronics Engineers (IEEE)* e PB possuíam 90% de livros digitais duplicados: *Referex* com 70% de ocorrências, *Knovel* com 60%, *SerSol* e *Springer* com 40%, *Digitalia* com 30%, *B24L7* com 20% e os fornecedores *Ebrary* e *HEB* com 10%. Somente três fornecedores (*EIV*, *HST* e *LCL*) não possuíam duplicidades em seus registros. A alta presença de duplicidades indicadas pelos autores proporciona um inchaço desnecessário no catálogo, dificultando ainda mais a gestão da coleção e o uso que é feito da mesma, além de demandar controle para remoção das duplicidades identificadas.

<sup>14</sup> <https://www.library.illinois.edu/>

A atualização dos dados bibliográficos e de licenciamento, além de serem questões relativas ao desenvolvimento da coleção, também são fundamentais para a gestão dos livros digitais contratados, tema que será discutido adiante.

### ***2.3 Gestão dos livros digitais contratados e geração de indicadores***

O licenciamento de livros digitais pode ser realizado, renovado, cancelado ou expirado, exigindo ações de gestão da coleção disponível para consulta dos usuários. A necessidade de ajustes constantes nos livros digitais presentes no catálogo demanda da equipe da biblioteca o acompanhamento e controle frequentes para atualização. Segundo Thompson e Traill (2017, não paginado, tradução nossa), “À medida em que as coleções de livros digitais crescem (e a equipe das bibliotecas encolhem), os processos de gerenciamento também devem ser escalonáveis”<sup>15</sup>.

A questão da gestão dos livros digitais contratados suscita discussões sobre onde e como registrar estes dados de forma a ter condições de realizar buscas e produzir relatórios e estatísticas.

Os livros digitais podem ser removidos, inibidos ou alterados no catálogo se ocorrerem as seguintes situações:

- a) Não renovação do licenciamento;
- b) Remoção do livro digital da plataforma pelo fornecedor;
- c) Rompimento ou término da contratação entre biblioteca e fornecedor;
- d) Renovação do licenciamento, alterando a data de vigência.

Os livros digitais que não estão mais disponíveis devem ser inibidos no OPAC, evitando aos usuários a tentativa de acesso que resultará em erro e, conseqüentemente, frustração. Ao manter recursos no catálogo que não estão disponíveis, a biblioteca pode comprometer sua credibilidade perante a comunidade atendida, além de ter dificuldades para mensurar o real tamanho de sua coleção. Se optar por inibir os livros digitais removidos, seu sistema de automação guardará a informação dos livros digitais que foram contratados e a quantidade de acessos realizados pelo OPAC, com os dados passíveis de utilização em estatísticas e relatórios. Se por outro lado a biblioteca optar por excluir os livros digitais, os dados

---

<sup>15</sup> *As ebooks collections grow (and libraries' staff shrink), management processes must also be scalable.* (THOMPSON; TRAILL, 2017, não paginado).

de acesso são apagados do banco de dados, comprometendo a geração de indicadores e o histórico de utilização da coleção.

As formas de contratação de livros digitais não são estanques e podem mudar de acordo com a necessidade da biblioteca e/ou opções oferecidas pelos fornecedores. No caso de utilização de modelos de negócios orientados aos usuários é usual a troca de forma de licenciamento, de acordo com a quantidade de acessos que foram realizados em um período. Assim, um livro digital que foi licenciado pelo modelo transitório *Evidence Based Acquisition* (EBA), depois de um ano tem seu modelo de contratação alterado para Aquisição Perpétua, por exemplo. Outra possibilidade é a biblioteca contratar mais acessos simultâneos de livros que possuem alta demanda. Estas alterações nas formas de contratação também precisam ser atualizadas.

A centralização dos livros digitais em um único ponto possibilita melhor gestão dos recursos licenciados, identificando livros digitais de fornecedores, modelos de negócios adotados, datas de inclusão ou validade dos livros digitais etc., além de poder validar os usuários nas plataformas diretamente por meio do sistema integrado de biblioteca (ILS, *Integrated Library System*) utilizado pela instituição. Esta ação também contribui para a mensuração do uso que é feito do conteúdo de um único ponto, a partir da quantidade de livros digitais que foram acessados por meio do catálogo e apresentar mais opções de recursos onde, em alguns casos, a biblioteca pode dispor de um mesmo livro nas formas impressa e digital.

O controle da coleção digital licenciada de livros é complexo devido a instabilidade de manutenção dos dados dos livros digitais contratados. Além das atualizações dos metadados bibliográficos, os metadados de licenciamento também precisam ser registrados e acompanhados, visando suprir as demandas dos usuários por meio de investimentos de acordo com o modelo de negócios adequado à demanda apresentada.

A gestão de livros digitais é diferente da dos livros impressos. Enquanto os livros com volume físico ficam em estantes, os livros digitais estão em nuvem, em plataformas de terceiras partes. A comunicação entre fornecedores e biblioteca termina quando os livros impressos são entregues. Com os livros digitais a dependência da plataforma é constante (ROSSMANN; FOSTER; BABBITT, 2009).

O acompanhamento do número de acessos realizados demanda controle, uma vez que o acesso é o correspondente a um empréstimo. Os dados de acessos podem ser solicitados aos fornecedores, porém cada fornecedor coleta e distribui os dados estatísticos de uma forma, sem padrões. A biblioteca pode solicitar ao fornecedor que encaminhe relatórios de acessos de forma regular, porém cria-se uma dependência junto ao fornecedor, o que não é interessante. Ao receber dados estatísticos de diversos fornecedores, a biblioteca precisa consolidá-los, o que é moroso.

Outro fator de destaque é que os livros são contratados junto a fornecedores diferentes. Isto exige distinguir quais os livros digitais licenciados com cada fornecedor, por qual modelo e com qual vigência, tendo com isso dados para avaliar se os livros estão sendo utilizados e se justifica a continuidade da contratação.

Chen, Kim e Montgomery (2016) relatam a experiência da *University of Texas at Dallas* que definiu pela inclusão de metadados de licenciamento em etiquetas do formato MARC 21. Inicialmente incluíram o nome do fornecedor como outra editora. Depois optaram por utilizar a *tag* 024 para identificar os livros digitais. A instituição também adotou o uso de secundárias de instituições (710 |a) e entrada secundária de série (830 |a) para registrar o nome do fornecedor, com a informação sobre a forma de acesso incluída junto a URL, na *tag* 856 |y, e identificação da interface na *tag* 856 |x, conforme ilustrado na **Figura 4**.

**Figura 4:**Dados de licenciamento no MARC 21

### 1. Add 710 or 830 field, which identifies a specific vendor

• ebrary	710	2_\$a ebrary, Inc.
• EBL	710	2_\$a Ebooks Corporation.
	830	_0\$a Ebook library.
• EbscoHost		\$a EBSCOhost.
• Wiley	710	2_\$a Wiley InterScience (Online Service)
• Cambridge	710	2_\$a Cambridge University Press.
	830	_0\$a Cambridge books online.
• Sage	710	2_\$a Sage eReference (online service)



## 2. Add an Interface Name in 856 \$x:

---

• ebrary	• Ebrary E-Books
• EBL	• Ebook Library
• EbscoHost	• Ebsco Ebooks
• Wiley	• Wiley Online Library
• Cambridge	• Cambridge E-books
• Sage	• Sage Ebooks
• Gale	• Gale Ebooks

---

## 3. Add Simultaneous User information in 856 \$y:

---

• ebrary	• SUPO or MUPO
• EBL	• Nonlinear or Unlimited
• Wiley	• EA or UA
• Cambridge	• EA, 1U, 3U
• Sage	• EA
• EbscoHost	• 1U

---

Fonte: CHEN; KIM; MONTGOMERY, 2016, p. 265.

A experiência da *University of Texas at Dallas* não questiona o uso do MARC 21 para inclusão de dados de licenciamento. Foi a solução adotada pela instituição para identificar os fornecedores. A rede de bibliotecas da *Univeristy of Houston* adotou o uso da *tag* local 910 para fins administrativos e esta forma de gestão está atendendo de forma satisfatória, porém não é simples à equipe (WU; MITCHELL, 2010).

Na opinião desta pesquisadora não é adequada a utilização de *tags* do MARC 21 para fazer o gerenciamento da coleção digital, pois ponto de acesso e nome de fornecedores são informações que demandam tratamentos diferentes e possuem usos distintos, além da confusão que estes dados podem causar no catálogo. Além disso, este recurso não parece ser apropriado para controlar renovações. O MARC Bibliográfico é utilizado para descrever recursos de informação e não dados de proveniência dos recursos. Utilizar *tags* locais para descrever dados de aquisição não proporcionará recuperação confiável, visto que os dados dependem de estratégia de busca aplicada a dados digitados em campos textuais e não em campos com formatos ou configurações apropriadas para representar valores como datas, moedas, preços etc. Por outro lado, o uso da 856 |y com informações sobre as possibilidades de acesso do livros digitais e da 856 |x com a identificação do fornecedor podem ser úteis aos usuários, desde que apresentados no OPAC com significados dos códigos ou siglas de fácil compreensão.

Os dados de licenciamento devem ser registrados em banco de dados, mas não no formato MARC 21. Dados como fornecedor, modelo de negócio aplicado, valor,

período de validade, quantidade de acessos simultâneos contratados etc., podem ser controlados e passíveis de atualizações, preservando dados de transações anteriores e permitindo o manejo de conjuntos de dados de forma global.

As plataformas dos fornecedores podem controlar o uso que é feito dos livros digitais a partir de coleta de dados. O protocolo SUSHI permite que dados de uso de recursos digitais sejam coletados por meio da COUNTER, organização não governamental fundada em 2002 com o objetivo de gravar, trocar e interpretar o uso de recursos *online* (NISO, 2010).

A COUNTER desenvolveu o Código de práticas COUNTER, um *framework* para as principais categorias de recursos digitais como banco de dados e de multimídia, seriados, obras de referência e livros digitais. A aplicação do Código COUNTER é voltada para bibliotecas, vendedores, intermediários e demais interessados que precisam de estatísticas confiáveis de uso de recursos *online* (COUNTER, 2019). O Código de práticas COUNTER e o SUSHI possuem *Schema* em XML específico para livros, com diversos formatos de relatórios como a quantidade de requisições recebidas, número de seções realizadas, quantidade de acessos negados, total de pesquisas feitas etc. (COUNTER, 2019).

Por meio do COUNTER é possível identificar qual, como, quem, quando e por quanto tempo um capítulo ou um livro digital foram acessados. Estes dados contribuem com a identificação de solicitações de usuários e ajudam com as decisões de licenciamento, divulgação e marketing (ZHANG, 2018).

Neste capítulo foram discutidas as principais dificuldades observadas pelas bibliotecas em relação aos livros digitais licenciados. A experiência de outras instituições não aponta para uma única solução, com cada biblioteca buscando formas de realizar a gestão de seus livros digitais. Não foram localizados artigos discutindo a possibilidade de centralização dos livros digitais de todos os editores em um único ponto, o que torna o controle de licenciamento e quantidade de acessos dispersos ou até mesmo desconhecidos. Espera-se resolver estes problemas a partir do modelo de dados para gestão de livros digitais licenciados, que será apresentado no Capítulo 5.

O próximo capítulo discute a Web Semântica e o *Linked Data*, com destaque ao *Linked Open Data* e *Linked Enterprise Data*.

### 3 Web Semântica, *Linked Data*, *Linked Open Data* e *Linked Enterprise Data*

---

#### Preâmbulo

Este capítulo apresenta a Web Semântica e o *Linked Data*, destacando as aplicações *Linked Open Data* e *Linked Enterprise Data*. Embora tenha sido identificado o *Knowledge Graphs*, esta aplicação não será abordada nesta pesquisa por não ter sido selecionada para a elaboração do modelo.

Primeiramente é feito uma breve descrição da evolução histórica da Web Semântica, apresentando alguns de seus elementos como a presença de *Uniform Resource Identifier* (URI)s, o modelo RDF, a construção de triplas e o uso de ontologias.

O modelo RDF oferece condições para a formalização de triplas explicitando as relações existentes entre sujeito (recurso), predicado (propriedade) e objeto (valor). Para que ocorra a comunicação entre máquinas é necessária a adoção de padrões, modelos e boas práticas que permitam a interlocução de dados distribuídos na Web, e a representação destes dados de forma legível para consumo de pessoas.

Enquanto o LOD prevê a interligação entre dados que podem ser utilizados de forma livre, inclusive com *ranking* de estrelas que sinaliza o quão aberto o dado é, o LED é empregado para interligar dados em ambientes empresariais, com acesso restrito. Estes dados podem estar presentes em outras aplicações utilizadas pela empresa, ou ainda podem ser criados a partir do consumo de dados abertos dispostos na Web acrescidos a dados restritos à empresa.

*Datasets* são conjuntos de dados e para sua publicação é necessário passar por uma série de etapas como seleção, curadoria, conversão de formatos, inclusão de entidades e relacionamentos, controle de qualidade e demais ações para que os dados possam ser interligados com outros dados na Web.

O capítulo encerra com a apresentação das etapas que serão utilizadas para a publicação do modelo para gestão de livros digitais licenciados.

De acordo com Gilliland (2008), toda informação, independentemente se é física ou digital, possui três características: conteúdo, contexto e estrutura, e estes aspectos devem ser representados nos metadados. Para a autora:

Conteúdo é relacionado com o que o objeto contém ou sobre do que trata e é intrínseco a um recurso de informação; Contexto indica quem, o que, porque, onde e como os aspectos associados à criação do objeto são extrínsecos a um recurso de informação; Estrutura é relacionada com o conjunto formal de associações dentro e entre recursos de informação individuais e pode ser intrínseca ou extrínseca a ambos (GILLILAND, 2008, p. 2, tradução nossa)<sup>16</sup>.

A Web nasceu a partir da demanda de troca de informações entre cientistas em universidades e instituições, por iniciativas de Berners-Lee. Segundo Marksheffel (2017), o objetivo de Berners-Lee ao desenvolver a Web era tornar real o sonho de *Vannevar Bush*, que em 1945 publicou um artigo na revista *The Atlantic Magazine*, discorrendo sobre o conceito do *Memex*, uma máquina onde poderiam ser armazenadas e consultadas todas as informações desejadas, em qualquer formato, de qualquer lugar e a qualquer momento, relacionando dados de forma que uma informação permitiria a navegação para outra, que por sua vez encaminharia para outra, de forma infinita. Porém a Web não poderia ser somente um local para publicar informações utilizando *hiperlinks* para criar um elo entre elas. Após quase dez anos da publicação do artigo que anunciou o desenvolvimento da Web, Berners-Lee, Hendler e Lassila publicaram em 2001 o conceito da Web Semântica.

De acordo com Breitman, Casanova e Truszkowski (2007), a Web tradicional pode ser compreendida como sintática, uma vez que as informações são apresentadas por computadores, porém sua identificação e interpretação são delegadas aos humanos. Com o crescente aumento da quantidade de informações distribuídas na Web é impossível a seleção e apropriação destas informações por pessoas, demandando o uso de aplicações que permitam a identificação e consumo dos dados de forma contextualizada. Mas, conforme Berners-Lee, Hendler e Lassila (2001), a Web é livre

---

<sup>16</sup> *Content relates to what the object contains or is about and is intrinsic to an information object; Context indicates the who, what, why, where, and how aspects associated with the object's creation and is extrinsic to an information object; Structure relates to the formal set of associations within or among individual information objects and can be intrinsic or extrinsic or both (GILLILAND, 2008, p.2).*

e descentralizada e a construção de uma Web Semântica deve manter esta ideia. Assim, não são criadas regras de como os dados devem ser publicados, mas são apresentadas boas práticas que permitam a criação de modelos de dados ou ontologias que se comuniquem e sejam interoperáveis. Assim, a proposta da Web Semântica foi apresentada não para substituir a Web, mas para ser uma extensão dela. De acordo com Berners-Lee, Hendler e Lassila:

A Web Semântica trará estrutura ao significado do conteúdo de sítios da Web, criando um ambiente onde agentes de softwares podem perambular de um sítio a outro, prontos para executarem tarefas para os usuários (BERNERS-LEE; HENDLER; LASSILA, 2001, p. 1, tradução nossa)<sup>17</sup>.

A noção essencial da Web é que se trata de uma comunidade aberta, que pode ser “burra” ou “esperta”. É “burra” quando não permite a atualização de dados de temas relacionados de forma dinâmica, exigindo esforços para identificação de informações consistentes. Ela pode ser “esperta” se possuir aplicações inteligentes que permitem, por exemplo, a um sítio de comércio eletrônico recomendar produtos que podem ter chances de vendas identificadas a partir do padrão de compras do consumidor, proporcionar detalhadas informações geográficas sobre uma localidade com atrações ao redor, entre outras possibilidades. Para o desenvolvimento desta Web “esperta” são necessárias aplicações construídas de forma a permitir que formas de conexão de informações, identificando o dado correto na fonte adequada (ALLEMANG; HENDLER, 2011).

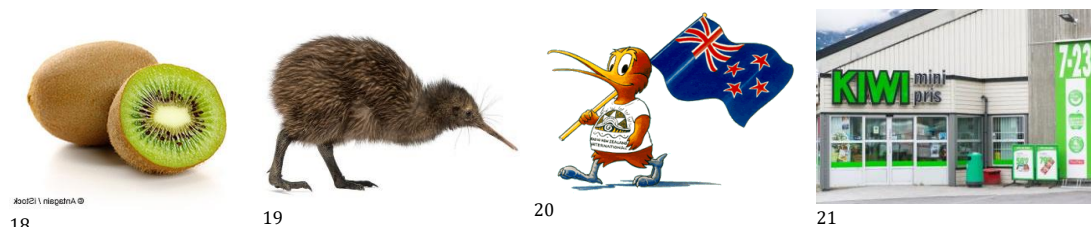
Algumas das complexidades na organização da informação presente na Web são a ambiguidade e a falta de consistência dos dados. É necessário que coisas, pessoas, conceitos etc. tenham uma identificação única para que os dados possam prover um resultado assertivo. A utilização de URIs permite a identificação singular de recursos de informação, possibilitando que sejam utilizados dentro do contexto esperado e não apresentando dados desconstruídos ou sem sentido. Um exemplo que pode ser utilizado para ilustrar a ambiguidade é a palavra “Kiwi”. Este termo pode designar uma fruta,

---

<sup>17</sup> *The Semantic Web will bring structure to the meaningful content of Web pages, creating an environment where software agents roaming from page to page can readily carry out sophisticated tasks for users* (BERNERS-LEE; HENDLER; LASSILA, 2001, p. 1).

um pássaro, uma rede de supermercados da Noruega ou ainda pode ser a identificação de pessoas que nasceram na Nova Zelândia. Estes “kiwis” constam na **Figura 5**.

**Figura 5:** Diversos significados para a palavra “Kiwi”



Fonte: elaborado pela autora. Fontes das imagens nas notas de rodapé 17 a 20.

Ao fazer uma busca na Web por “Kiwi” sem que exista singularização dos dados, os resultados recuperados podem ser completamente variados, dificultando a identificação e seleção das informações desejadas. Agora, ao ser possível criar um identificador único para cada uma dessas ocorrências, é possível prover informações confiáveis que favorecem a recuperação, assegurando que o “Kiwi” desejado, seja ele o pássaro, a fruta, os nativos da Nova Zelândia ou ainda a rede de supermercados na Noruega, seja recuperado.

Entretanto, mesmo com a Web Semântica não é possível às máquinas resolverem completamente as ambiguidades, distinguindo um “Kiwi” de outro. Mesmo com a formação de triplas, a distinção entre palavras iguais com significados diferentes ainda pode ser complexa. O significado é feito a partir da URI que está explícita, identificando que se trata de uma ave e não de uma fruta, por exemplo. Mas uma busca somente por “Kiwi” ainda poderá apresentar como resultado uma variedade de respostas, se mais elementos não forem incluídos na pesquisa de forma a distinguir e eliminar a ambiguidade. Por outro lado, ao vincular os dados a um *dataset* dentro de um contexto, a problemática é minimizada ou até eliminada. Se a pesquisa é relacionada ao reino animal por exemplo, seria consultado somente o “Kiwi” pássaro, não buscando outros resultados porque o significado foi definido no conjunto de dados ao qual está vinculado.

<sup>18</sup> Fruta: <https://foodfacts.mercola.com/kiwifruit.html>

<sup>19</sup> Pássaro: <https://www.sciencedaily.com/releases/2015/07/150723083731.htm>

<sup>20</sup> Pessoas nascidas na Nova Zelândia: <http://thevoiceofthezamorin.blogspot.com/2014/11/why-do-new-zealanders-call-themselves.html>

<sup>21</sup> Rede de supermercados da Noruega: [https://www.123rf.com/photo\\_59322580\\_andalsnes-norway-circa-june-2016-kiwi-mini-pris-supermarket-in-andalsnes-norway-kiwi-is-part-of-norg.html](https://www.123rf.com/photo_59322580_andalsnes-norway-circa-june-2016-kiwi-mini-pris-supermarket-in-andalsnes-norway-kiwi-is-part-of-norg.html)

O RDF é um padrão formal desenvolvido pela W3C para codificar a descrição de recursos da Web de forma a serem compreendidos por computadores, podendo ser compartilhados e reutilizados de diversas formas (MARKSCHEFFEL, 2017; W3C, 2014). No RDF, os *International Resource Identifiers* (IRIs, Identificadores Internacionais de Recursos, tradução nossa) são utilizados para identificar as triplas e as relações existentes entre sujeito (recurso), predicado (propriedade) e objeto (valor), que formam as declarações. A singularização de elementos descritivos é explicada por Berners-Lee, Hendler e Lassila (2001):

Porque o RDF utiliza URIs para codificar a informação em um documento, as URIs garantem que os conceitos não são somente palavras em um documento, mas estão interligados a uma única definição que qualquer pessoa pode localizar na Web (BERNERS-LEE; HENDLER; LASSILA, 2001, p. 3, tradução nossa)<sup>22</sup>.

O funcionamento da Web Semântica depende do uso de *datasets* estruturados e de conjunto de regras de inferência capazes de compreender uma requisição. O desafio consiste em utilizar uma linguagem que expresse os dados e as regras para que a informação seja compreendida. Porém não basta formalizar os dados para que eles sejam legíveis por máquinas. É necessário que os dados também sejam compreendidos por pessoas, apresentando as relações estabelecidas. Segundo Allemang e Hendler (2011), a Web Semântica necessita de condições para distribuição de dados e não somente a criação de vínculos entre sítios.

A ideia principal da Web Semântica é suportar a distribuição da Web em nível de dados ao invés do nível de apresentação. Ao invés de ter uma página da Web apontando para outra, um dado pode apontar para outro, usando uma referência global chamada *Uniform Resource Identifiers* (URIs). A infraestrutura da Web promove um modelo de dados onde a informação de uma única entidade pode ser distribuída na Web (ALLEMANG; HENDLER, 2011, p. 6, tradução nossa)<sup>23</sup>.

A seguir será apresentado o conceito do *Linked Data*.

---

<sup>22</sup> *Because RDF uses URIs to encode this information in a document, the URIs ensure that concepts are not just words in a document but are tied to a unique definition that everyone can find on the Web* (BERNERS-LEE; HENDLER; LASSILA, 2001, p. 3).

<sup>23</sup> *The main idea of the Semantic Web is to support a distributed Web at the level of the data rather than at the level of the presentation. Instead of having one web page point to another, one data item can point to another, using global references called Uniform Resource Identifiers (URIs). The Web infrastructure provides a data model whereby information about a single entity can be distributed over the Web* (ALLEMANG; HENDLER, 2011, p. 6).

### 3.1 *Linked Data (LD)*

De acordo com Bizer, Heath e Berners-Lee (2011), o *Linked Data* consiste em criar *links* entre dados provenientes de diferentes fontes na Web, porém é diferente de bancos de dados de organizações ou demais sistemas. Para que ocorra a interligação entre os dados, é necessário que eles sejam publicados de forma a serem legíveis por máquinas, com significados e as relações entre eles definidas explicitamente. Byrne e Goddard (2010) concordam com a definição de LD oferecida pela *Freie Universitat*, da Alemanha e que é composta de duas ideias simples: 1) o uso do modelo RDF para publicar dados na Web; e 2) o uso de URIs para explicitar a relação entre entidades.

Ao conectar uma informação de um sítio a um conjunto de dados, uma atualização feita neste conjunto acarretará automática atualização do sítio, abrindo possibilidades de geração de novos dados a partir da integração com fontes diferentes. Assim, uma aplicação pode coletar dados de um único ou de diversos *datasets*, de acordo com os propósitos que se deseja alcançar. Para que seja possível esta integração entre *datasets* se faz necessária a utilização de linguagens que sejam compreensíveis por máquinas. Isto permite alterar a configuração de silos de informação que estão distribuídos na Web (DENAUX et al., 2017; WU, 2017; BERNÈS, 2013; ALLEMANG, 2010; HYLAND, 2010), permitindo a interligação de dados que, como consequência, apresentarão mais e novos resultados sobre um termo.

A comunicação entre máquinas ocorre com a adoção de padrões, modelos e de boas práticas que permitam a interconexão entre informações distribuídas na Web, proporcionando que, além da integração entre fontes, ocorra o reuso dos dados. O LD representa o uso de boas práticas para publicação de dados na Web, de forma organizada, utilizando padrões internacionais recomendados pela W3C (LÓSCIO; BURLE; CALEGARI, 2017).

De acordo com Berners-Lee (2010), para que ocorra a publicação de dados que permitam sua interligação com outros, é necessário seguir algumas regras:

1. Utilização de URIs como nomes de coisas;
2. Uso de URIs *Hyper Text Transfer Protocol* (HTTP) para a consulta de nomes por pessoas;



3. A consulta a uma URI deve fornecer informação utilizando um padrão (RDF, SPARQL) para que possa ser utilizada;
4. Inclusão de links em outras URIs, para que outras coisas possam ser descobertas.

Segundo Bermès (2013) a não utilização destas regras não impede a interligação de dados, mas, por outro lado, restringe a expansão da conexão entre eles. Quanto maior o conjunto de dados disponíveis e interligados, maior será a construção de uma rede de informação contextualizada. Os dados são interligados formando grafos.

O RDF é um modelo de dados que armazena a sintaxe formal do recurso, mas não seu componente semântico. A parte semântica é proporcionada por um esquema de domínio, também identificado como vocabulário ou ontologia, que define as propriedades (relacionamentos) que conecta os recursos, as classes dos recursos ou as restrições existentes para combinar classes com propriedades. Caso não seja identificado um vocabulário que atenda ao domínio desejado, um novo pode ser desenvolvido, entretanto o reuso de vocabulários publicados geram *links* para outros vocabulários ou *datasets*, com a remoção de ambiguidades sendo feita com a indicação de vínculo entre os termos, como, por exemplo, a expressão *owl:sameAs* que indica que um dado é o mesmo representado em outro *dataset* (RIETVELD, 2016).

O termo ontologia proporciona interpretações distintas de acordo com a perspectiva em foco. De acordo com Breitman, Casanova e Truskowski (2007), a origem da palavra é grega com a unificação dos termos *ontos* (ser) e *logos* (palavra). A utilização deste termo foi introduzida por filósofos alemães no século 19 para diferenciar o estudo do ser em relação ao estudo de vários seres nas ciências naturais. O termo é utilizado em outras áreas como na ciência da computação, comércio eletrônico e em sistemas de informação em geral.

Na seara da Filosofia, ontologia constitui um particular sistema de categorias de acordo com uma certa visão de mundo. No âmbito da Ciência da Computação caracteriza-se como um artefato constituído por um vocabulário específico que é utilizado para descrever uma realidade, acrescido de um conjunto de suposições explícitas que proporcionam significado a este vocabulário (GUARINO, 1998).

No caso mais simples, uma ontologia descreve a hierarquia de conceitos relacionados por relacionamentos de subsunção [integração]; em casos mais

sofisticados, são acrescentados axiomas adequados para expressar outras relações entre conceitos e para restringir a interpretação pretendida (GUARINO, 1998, p. 3, tradução nossa)<sup>24</sup>.

Na perspectiva da Ciência da Informação, ontologia é:

Um artefato tecnológico que possibilita representar formalmente as propriedades e relacionamentos de um determinado modelo conceitual, favorecendo a utilização de inferências automáticas nos processos de organização e recuperação de recursos informacionais (RAMALHO, 2010, p. 107).

O uso de ontologias contribui com os serviços de referências em bibliotecas ao ampliar a ocorrência de respostas contextualizadas aos usuários. Ao aplicar ontologias aumentam-se as possibilidades de inferências, uma vez que outros elementos podem ser recuperados e não somente os que foram empregados na indexação (RAMALHO, 2010). O vocabulário é expresso em esquemas (RDF *Schemas* ou *Resource Description Framework Schema* (RDFS), ou ainda em linguagens de ontologia como a *Ontology Web Language* (OWL). O RDF pode ser descrito em diversas notações como o *Notation3* (N3), *N-Triples* (NT), *Turtle*, *RDF/XML*, *JavaScript Object Notation* (JSON), entre outras. Também é possível adotar ontologias como *Friend of a Friend* (FOAF) ou o *Simple Knowledge Organization System* (SKOS) ao *Schema* RDF utilizado.

A seguir é analisado o LOD, uma aplicação da Web Semântica.

### 3.2 *Linked Open Data*

O *Linked Open Data* é resultado de um grupo de trabalho do W3C que tem como objetivo a criação de rede de dados interligados que podem ser utilizados livremente. Segundo Berners-Lee (2010), o LOD foi desenvolvido sob uma licença aberta exatamente para a reutilização livre dos dados.

---

<sup>24</sup> *In the simplest case, an ontology describes a hierarchy of concepts related by subsumption relationships; in more sophisticated cases, suitable axioms are added in order to express other relationships between concepts and to constrain their intended interpretation* (GUARINO, 1998, p. 3).

Nem todos os *datasets* precisam ser abertos, visto que existem informações que circulam somente entre um grupo, porém, para ser considerado LOD os dados devem ser publicados para uso aberto, seguindo as boas práticas orientadas pelo W3C.

São considerados dados abertos aqueles que correspondem com a definição de dados bibliográficos abertos (*open bibliographic data*, tradução nossa), que estabelece que são dados que podem ser utilizados, reutilizados e redistribuídos livremente, com as mesmas atribuições de compartilhamento com as quais foram publicados. O W3C orienta que as licenças de uso de dados publicados na Web sejam incluídas nos *datasets*, informando sobre a presença de restrições ou não para uso e reuso dos dados. Recomenda-se que as informações sobre a proveniência e licença dos dados sejam legíveis por máquinas e humanos (LÓSCIO; BURLE; CALEGARI, 2017).

Para Wessels et al. (2017) os dados podem ser coletados de diversas fontes como a Web, mídias sociais, aplicações para dispositivos móveis (APP), dados geoespaciais ou bancos de dados comerciais. Os dados podem ser coletados separadamente e permitem a criação de inferências ditando tendências, comportamentos, movimentos a partir do cruzamento de informações advindas de fontes distintas e que sozinhas poderiam não proporcionar o entendimento de um contexto.

Uma avaliação por estrelas foi proposta para indicar quanto um dado é aberto ou não. Ao obter uma estrela, significa que as etapas anteriores já foram alcançadas, sendo este critério uma demonstração de evolução do dado. Assim, a obtenção de mais estrelas somente é possível após atender completamente ao requisito anterior. Para Berners-Lee (2010), o *ranking* de estrelas consiste de:

- 1 estrela: Disponível na Web (em qualquer formato), mas com uma licença aberta, para ser *Open Data* (dados abertos, tradução nossa);
- 2 estrelas: Disponível de forma que os dados são legíveis por máquina;
- 3 estrelas: Os formatos utilizados de dados não são proprietários;
- 4 estrelas: Utilização de padrões abertos da W3C (RDF e SPARQL) para identificar coisas;
- 5 estrelas: Vincular o seu dado ao de outra pessoa, proporcionando contexto.

Para Berners-Lee (2010) o LOD é o *Linked Data* que permite uso dos dados de forma livre, gratuita e sem restrições para reutilização. Segundo o autor, o *Creative Commons CC-BY*<sup>25</sup> e o *Open Government Licence*, do Reino Unido, são exemplos de licenças abertas.

Dados podem ser abertos ou fechados. Embora exista o movimento para tornar os dados abertos, passíveis de consumo e interoperáveis na Web de forma livre e gratuita, instituições podem trabalhar em ambientes onde os dados estão fechados, com acesso restrito. Isto, porém, não impede que estas instituições também se sirvam de dados abertos.

A seguir será apresentado o *Linked Enterprise Data*, usualmente composto por dados fechados.

### 3.3 *Linked Enterprise Data (LED)*

Para Allemand (2010), a adoção de aplicações da Web Semântica em ambientes empresariais é identificada como *Linked Enterprise Data*. Segundo este entendimento, para Hyland (2010), o LED oferece técnicas eficientes para integrar dados de uma empresa. Segundo a autora, empresas investem altas somas de dinheiro para aquisição ou desenvolvimento de soluções para a gestão e análise de dados. Esse esforço é empregado com o objetivo de obter dados com alta qualidade, confiabilidade e para embasamento para tomadas de decisões.

As razões para manter dados fechados (*proprietary data*) podem ser diversas e normalmente são utilizadas por empresas privadas. As informações consumidas por este público usualmente são relacionadas com segredos de negócios e estratégias da empresa. Portanto, é compreensível que os dados sejam de uso restrito aos participantes do ecossistema da empresa, com aplicação de níveis de acesso de acordo com a hierarquia e cultura institucional.

---

<sup>25</sup> *Creative Commons* (CC): organização não governamental norte-americana que desenvolveu variedades de licenças de uso gratuito que são atribuídos a recursos de informação. As licenças podem ser dos tipos atribuição de autoria (BY), compartilhar da mesma forma (SA), não comercial (NC) e sem permitir derivação (ND). Estes tipos de licenças podem ser combinados e incluídos nos metadados (CREATIVE COMMONS, 2018).

Segundo Hillerbrand (2016), as aplicações da Web Semântica podem ser utilizadas em empresas para integrar informações de fontes diversas diminuindo redundâncias, resolver ambiguidades em terminologias, aumentar a recuperação de informações, identificar informações relevantes dentro de um domínio etc., além de prover instrumentos para tomadas de decisão.

Tanto como ocorre com a Web, em empresas também existem silos de informações. Estas situações são identificadas quando dados são armazenados de forma individual por pessoas ou grupos, utilizando uma forma de organização que atenda às suas demandas. Ou ainda quando uma mesma informação é trabalhada de formas diferentes entre departamentos. Esta característica ocorre mesmo sem a interferência de recursos tecnológicos e pode ser ilustrado com casos em que departamentos possuem sua própria biblioteca, dispensando a consulta ao acervo de uma biblioteca central construída e mantida pela empresa. Informação é poder e tê-la de forma acessível, rápida, ágil e eficiente faz a diferença na concretização de tarefas. Por outro lado, para as empresas, possuir informações dispersas e duplicadas é problemático, pois representa redundância em investimentos e dificuldades para localização de dados (ALLEMAND, 2010).

Segundo Villazon-Terrazas et al. (2017) as empresas podem construir ontologias para suportar as demandas de seus negócios, utilizando vocabulários existentes ou expandindo-os de acordo com necessidades e evoluções da instituição. Ao adotar um mesmo esquema para representar seus dados, é favorecida a integração entre outros sistemas, além da adoção de um único vocabulário, contribuindo com a comunicação entre os envolvidos. Esta característica faz com que os esquemas utilizados na Web sejam mais genéricos, enquanto para as empresas são mais centrados no domínio específico.

Usualmente as empresas manterão seus *datasets* fechados ou com acesso restrito, enquanto a Web luta para cada vez mais disponibilizar dados abertos. Existe um entendimento que os *datasets* fechados possuem maior qualidade e confiabilidade nos dados, pois foram construídos com foco em um domínio específico, contudo não existe impedimento para que empresas também utilizem *datasets* abertos.

A seguir serão abordadas as etapas para publicação de *datasets* na Web.

### 3.4 *Publicação de datasets*

Segundo Hyland (2010), a publicação de dados consiste em disponibiliza-los na Web ou conecta-los a redes públicas ou privadas. Para Villazon-Terrazas et al. (2017a), trata-se da disponibilização dos grafos e metadados armazenados em *triple stores* (armazenamento de triplas, tradução nossa).

A publicação de *datasets* envolve uma série de etapas, que variam de acordo com o escopo do projeto. Assumpção (2018), Santarém Segundo (2018), Denaux et al. (2017), Villazon-Terrazas et al. (2017a), Rietveld (2016), dentre outros, apresentam passos a serem seguidos até a disponibilização do *dataset*. Embora existam variações na literatura, os passos seguidos são bastante semelhantes, com alterações observadas mais nas nomenclaturas atribuídas do que nas ações necessárias.

Os custos para a publicação de dados dependem de fatores que contemplam a seleção, obtenção e curadoria de dados abertos ou fechados, conversão de formatos, inclusão de vínculos entre entidades e relacionamentos, limpeza e manutenção dos livros digitais para ajustes na qualidade (*data cleaning*), armazenamento das triplas (*triple store*), entre outros (WESSELS et al., 2017; HYLAND, 2010). Abertos ou fechados, existem custos para a publicação de *datasets*. Estes custos usualmente estão relacionados a serviços, suporte, capacitação, educação dos envolvidos e infraestrutura necessária.

Como o objetivo desta pesquisa é propor um modelo aberto para publicação de dados, as etapas seguidas serão as propostas por Santarém Segundo (2018), uma vez que apresentam um contexto neutro, não detalhando uma área ou projeto específico, além de seguirem a recomendação da W3C sobre as Melhores Práticas para Dados na Web (LÓSCIO; BURLE; CALEGARI, 2017). Esta recomendação propõe trinta e cinco práticas e oito benefícios que podem ser obtidos com sua aplicação.

As etapas apresentadas por Santarém Segundo (2018) são:

- 1) Política e seleção;
- 2) Formalização, formatos e licenças;
- 3) Conversão ferramental;
- 4) Processo de recuperação;

- 5) Marketing e *feedback*;
- 6) Qualidade e preservação.

O próximo capítulo aborda a Web Semântica no universo de bibliotecas. São apresentadas possíveis aplicações no catálogo e questões relacionadas com a conversão de registros para modelos semânticos. Projetos e iniciativas como o LD4P, VIAF e *Schema Book* serão apresentados.

## 4 Web Semântica e bibliotecas

---

### Preâmbulo

Este capítulo discorre sobre algumas aplicações da Web Semântica em bibliotecas. Para que dados de bibliotecas sejam compatíveis com a Web dos dados são necessárias adaptações e substituição do modelo adotado, de forma a que sejam descritos em RDF e aceitem URIs como identificadores. Embora o MARC 21 não possua a estrutura adequada para descrever dados para o LD, algumas alterações já foram realizadas no formato para permitir que URIs sejam vinculadas a dados, favorecendo a interligação de dados. Alguns *datasets* já estão disponíveis, oferecendo dados em RDF e em não-RDF, contribuindo com a inclusão de conceitos aos registros existentes nos catálogos.

Devido as restrições do MARC 21 em relação ao LD, um novo modelo de dados está em construção, o BIBFRAME. Para esta pesquisa, entretanto, os exemplos adotados serão em MARC 21, formato utilizado por diversas bibliotecas.

Para aplicação do LD é necessária a conversão de registros para o modelo RDF e inclusão de URIs como identificadores de recursos e relações.

Embora a OCLC esteja liderando iniciativas com o LOD, estes esforços não serão empregados nesta pesquisa, ocorrendo a seleção de iniciativas do LD4 e *Schema.org*. O capítulo encerra com a descrição do *dataset* VIAF, que será utilizado no exemplo de validação do modelo a ser proposto, no Capítulo 6.



A Web Semântica permite diversas aplicações no ambiente de bibliotecas. Seu uso pode proporcionar o enriquecimento do catálogo de forma dinâmica, vinculando dados com informações complementares dispostas em *datasets* na Web. Também pode contribuir com ações para agilizar as atividades de catalogação e indexação a partir do momento que fontes externas podem ser consumidas para estas atividades. De acordo com Santos e Alves (2009), a Ciência da Informação e a Web Semântica são correlatas no âmbito da representação dos recursos de informação.

[...] as tecnologias da Web Semântica convergem para a área de Ciência da Informação, estabelecendo uma estreita relação na questão da representação do conhecimento, principalmente no que se refere ao uso de metadados considerados essenciais no estabelecimento dos requisitos para uma boa representação dos recursos informacionais na rede (SANTOS; ALVES, 2009, p.1).

Para Byrne e Goddard (2010) as dificuldades para implantação de LD pelas bibliotecas não são tecnológicas, mas de falta de consciência. Para as autoras o perigo consiste em bibliotecários enxergarem o LD como mais um padrão de metadados, sem vislumbrarem as possibilidades de expansão e visibilidade aos catálogos. Para Shieh e Reese (2015), a inclusão de elementos da Web Semântica como o uso de URIs amplia o alcance das coleções de bibliotecas.

A Web, destacando aqui os buscadores, é uma fonte para pesquisa bastante utilizada e que atende às necessidades imediatas dos usuários. Colocar o catálogo da biblioteca na Web permite inseri-la neste espaço, sendo uma fonte de qualidade na obtenção de dados bibliográficos, fazendo conexões com dados distribuídos na rede, que podem ser de outras bibliotecas, editores ou fontes (SHIEH; REESE, 2015).

A Ciência da Informação tem muito a contribuir com a expansão das aplicações da Web Semântica, uma vez que bibliotecas possuem coleções de recursos que podem ser compartilhadas e reutilizadas. “Catalogadores e indexadores produzem e padronizam metadados desde as primeiras tentativas de organização da informação que se conhece na história da Biblioteconomia” (SANTOS; ALVES, 2009, p.2). Bibliotecários possuem habilidades para trabalhar e organizar a informação a partir de padrões e protocolos internacionais.

Segundo Byrne e Goddard (2010, não paginado, tradução nossa), “com a experiência em pesquisa, geração de metadados e desenvolvimento de ontologias, os

bibliotecários devem estar realmente bem posicionados para entender e implementar o *linked data*<sup>26</sup>. Bibliotecas precisam reconhecer-se como potenciais provedoras de dados confiáveis. Para tanto precisam adequar seus metadados ao ambiente da Web, permitindo interoperabilidade com outros sistemas e conjuntos de dados.

O *W3C Library Linked Data Incubator Group* trabalhou no período de maio de 2010 a agosto de 2011 com o intuito de aumentar a interoperabilidade de dados entre bibliotecas na Web, aproximando especialistas em LD e a comunidade bibliotecária, identificando as possibilidades de cooperação. O foco dos estudos foi calcado na exploração de ferramentas já utilizadas por bibliotecas como modelos de metadados, esquemas, padrões e protocolos para a interoperabilidade. Este movimento ambicionou tornar as bibliotecas provedoras de fontes confiáveis para a Web de dados, principalmente em virtude da possibilidade de disponibilização de conjuntos padronizados que podem ser consumidos, compartilhando instrumentos como catálogos de autoridades, vocabulários controlados etc. (W3C INCUBATOR GROUP REPORT, 2011; BYRNE; GODDARD, 2010; MACEWAN, 2004).

De acordo com Miller (2004) e complementado por Byrne e Goddard (2010), o papel das bibliotecas no LD é:

- 1) Expor coleções: por meio da Web Semântica disponibilizar conteúdo convertendo seus dados ao modelo RDF para que os recursos possam ser compartilhados na Web;
- 2) Disponibilizar tesouros, mapeamentos e serviços legíveis na Web: identificar e converter tesouros, dicionários e vocabulários controlados para modelos semânticos permitindo ampla distribuição e reuso;
- 3) Compartilhar lições aprendidas: projetos e experiências devem ser compartilhados, expandindo as comunicações para a comunidade do LD como um todo;
- 4) Persistência: bibliotecários devem posicionar-se no cenário do LD, contribuindo com a disponibilização de dados, preferencialmente de forma

---

<sup>26</sup> *With their expertise in search, metadata generation, and ontology development, librarians should actually be well positioned to understand and implement linked data* (BYRNE; GODDARD, 2010, não paginado).

aberta. Também devem demandar de seus fornecedores a disponibilização de registros construídos semanticamente (BYRNE; GODDARD, 2010).

Por outro lado, segundo Singer (2009), um fator assustador para bibliotecas é a questão da confiança. Será que as bibliotecas serão capazes de confiar em metadados criados por não-bibliotecários? Será que vinculariam os dados de seus catálogos com dados que não podem controlar? São questões que devem ser discutidas. De acordo com Miller (2004) a confiança é um aspecto de grande relevância para a Web Semântica, afinal na Web é possível que qualquer pessoa fale algo sobre alguma coisa. Por isso é que bibliotecas representam confiança ao disponibilizar os seus dados para que sejam consumidos na Web.

O relatório final do *W3C Library Linked Data Incubator Group* alcançou alguns entendimentos como:

- 1) O termo Biblioteca refere-se ao conjunto de instituições do âmbito cultural e histórico, incluindo bibliotecas, museus e arquivos, independentemente se suas coleções são pequenas ou grandes, se são formadas por recursos analógicos ou digitais, ou sob custódia de instituições públicas ou privadas;
- 2) Dados de biblioteca (*library data*) se referem a qualquer tipo de informação digital produzida e sob curadoria de bibliotecas. Foram consideradas três categorias de dados de biblioteca: conjunto de dados (*datasets*), conjuntos de valores (*element sets*) e vocabulário de classes e propriedades (*value vocabularies*), que serão detalhados adiante;
- 3) São considerados dados em LD aqueles publicados de acordo com os princípios definidos para facilitar o vínculo entre *datasets*, ou seja, utilizando URIs como identificadores únicos de recursos, expressos sob o modelo RDF, especificando as relações existentes entre as coisas;
- 4) São considerados dados abertos aqueles que não possuem restrição legal para uso, reuso ou distribuição. O LD não precisa necessariamente ser aberto, porém ao usar dados desta natureza, potencializam-se as possibilidades de uso.

O relatório final constatou que os dados de bibliotecas não se apresentam integrados com os recursos da Web, conservando a construção de silos de informações

observada até então. Muitos dos padrões adotados por bibliotecas (MARC 21, ISO 2709, protocolo z39.50 etc.) são utilizados somente por esta comunidade, além do fato de muitos dados de recursos bibliográficos serem representados em linguagem natural ou por meio de códigos, armazenados em subcampos ou posições no MARC 21. Bibliotecários e especialistas em Web Semântica utilizam terminologias diferentes ao se referirem a conceitos semelhantes, o que também pode comprometer o entendimento entre as áreas.

Outro aspecto observado no relatório final é que na época em que foi publicado eram poucas as iniciativas de publicação de *datasets* de bibliotecas, destacando que a qualidade dos metadados é variada. Além disso, os esforços para conversão de registros legados ao modelo RDF e a inclusão de vínculos entre coisas e relacionamentos é uma atividade que exige bastante esforço das instituições (W3C INCUBATOR GROUP REPORT, 2011).

Conforme o W3C *Library Linked Data Incubator Group* (2011), e destacado por Assumpção (2018) e DeWeese e Segal (2014), foram definidas três categorias de dados de bibliotecas, a saber:

- 1) Conjunto de dados (*datasets*): descrevem os recursos presentes em uma biblioteca;
- 2) Conjunto de valores (*element sets*): contribuem com a padronização de dados. Catálogos de autoridade, vocabulários controlados etc., fazem parte deste conjunto;
- 3) Vocabulário de classes e propriedades (*value vocabularies*): vocabulários que podem ser empregados na descrição para propriedades e classes no RDF.

Segundo DeWeese e Segal (2014), o LD provê benefícios às bibliotecas ao aumentar as possibilidades de interoperabilidade com o reuso de conjuntos de dados e vocabulários, a inclusão de URIs aos metadados etc.

O PCC (2018a) publicou guia apresentando algumas fontes que podem ser utilizadas pelas bibliotecas para a inclusão de URIs. Na **Tabela 2** as fontes estão separadas entre aquelas que possuem dados em RDF e as que possuem dados não-RDF. Embora as fontes com dados não-RDF não permitem expansão com demais *datasets*, as URIs destes conjuntos podem ser utilizadas como identificadores.

**Tabela 2:** Algumas fontes de dados RDF e Não-RDF

<b>Dados em RDF</b>	<b>Dados não-RDF</b>
DBpedia	<i>AllMovie</i>
<i>Faceted Applications of Subject Terminology (FAST)</i>	<i>AllMusic</i>
<i>GeoNames</i>	<i>British Film Institute (BFI)</i>
<i>Getty's Art and Architecture (AAT)</i>	<i>Canadian Geographical Names</i>
ISNI	<i>Discogs</i>
<i>Library of Congress Subject Heading (LCSH)</i>	<i>Geographic Names Information System (GNIS)</i>
<i>Medical Subject Headings (MeSH)</i>	<i>Internet Broadway Database (IBDB)</i>
<i>Name Authority Cooperative Program of the PCC - NACO</i>	<i>Internet Movie Database (IMDb)</i>
<i>Open Researcher and Contributor (ORCID)</i>	<i>Legal Entity Identifier</i>
<i>RDA Vocabularies</i>	<i>Library of Congress Authorities</i>
VIAF	<i>ResearcherID</i>
<i>Wikidata</i>	<i>Scholar Universe</i>

Fonte: PCC, 2018a, p. 1-2.

Dados de bibliotecas devem ser de uso aberto e compartilhado. Como a geração de metadados é custosa às instituições, muitas vezes existe um sentimento em não os disponibilizar de forma aberta, uma vez que a instituição catalogadora investiu na produção dos mesmos. Outra justificativa para a não abertura dos dados está relacionada com questionamentos sobre a qualidade dos metadados, fazendo com que a biblioteca prefira mantê-los fechados ao invés de disponibiliza-los para uso coletivo, muitas vezes com receio de prejudicar sua imagem ao dispor de metadados cuja qualidade pode não atender a todos os interessados.

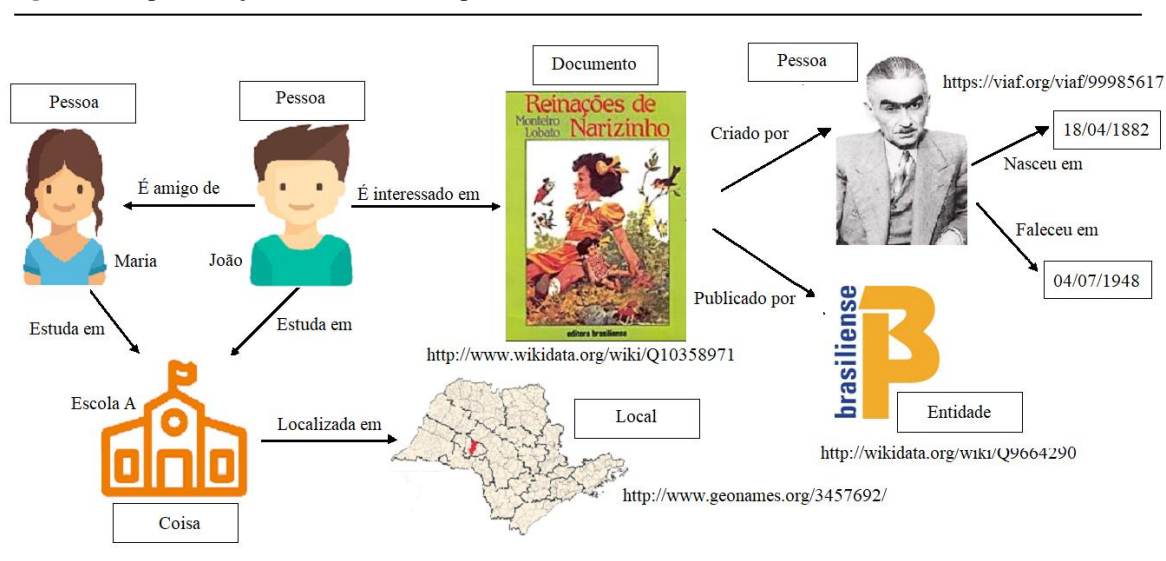
Restringir o acesso a metadados de bibliotecas é um comportamento identificado por Bermès (2013) como “*database hugging*” (abraçar banco de dados, tradução nossa), que consiste em inibir a exposição pública de dados, com consumo restrito somente à instituição catalogadora. Ao optar por este caminho, as bibliotecas mantêm a postura de silos de informações, com dados encerrados em uma única fonte, sem vínculo com *datasets* externos que podem contribuir para o entendimento de contextos e o enriquecimento de dados, restringindo as possibilidades de uso e descobertas.

Para Malmsten (2009) é preferível publicar dados do que não os publicar caso não estejam perfeitos. Segundo Byrne e Goddard (2010, não paginado, tradução nossa), “a Web Semântica é para derrubar silos, não para construir novos<sup>27</sup>”.

Segundo Shieh e Reese (2015) as bibliotecas precisam enxergar que suas coleções agora são descritas com dados e não mais com registros e para realizar a conversão dos dados de MARC 21 para RDF é necessário quebrar os registros em pequenos pedaços que representam elementos ou entidades, e remontá-los em grafos usando predicados (propriedades) para expressar a relação entre o sujeito e o objeto.

A **Figura 6** ilustra algumas relações que podem ser criadas nas triplas dentro do contexto de uma biblioteca. Como exemplo foi utilizado o escritor Monteiro Lobato, que entrou em domínio público em janeiro de 2019. Aqui é analisada a obra “Reinações de Narizinho”, publicada pela editora Brasiliense. A obra é de interesse de um leitor (João), que é amigo da leitora (Maria) e ambos estudam na “Escola A”, situada na cidade de Marília, SP.

**Figura 6:** Representação de Recursos, Propriedades e Valores entre biblioteca e usuários



Fonte: elaborado pela autora a partir de W3C (2014).

No exemplo acima a identificação única (URI) do autor foi retirada do VIAF. As URIs da editora e da obra foram retiradas da *Wikidata*. A relação entre o escritor e sua obra é de criador (*creator*), enquanto a Brasiliense desempenha função de

<sup>27</sup> *The semantic web is about breaking down silos, not building better ones* (BYRNE; GODDARD, 2010, não paginado).

responsável pela edição (*publisher*). A URI do local da escola foi retirado do *GeoNames*, e a relação entre João e Maria é representada pelo FOAF.

Ao fazer relações de itens e autoridades, a biblioteca continua autocentrada, porém não é mais um silo de informações, pois seus dados estão descritos para a Web Semântica. Além de vincular elementos do acervo, esta aplicação pode dar visibilidade à biblioteca, uma vez que os dados da coleção podem ser recuperados por outras fontes.

Outra abordagem que pode ser feita com a Web Semântica é o enriquecimento de dados a partir de fontes externas, proporcionando mais informações aos usuários sobre autoridades e itens presentes no catálogo. Bojars, Zogla e Egíte (2015) possuem a seguinte compreensão em relação ao enriquecimento de dados:

O enriquecimento de dados é uma tarefa complexa e detalhes sobre como isso pode ser feito depende dos *datasets* envolvidos. Por exemplo, dados selecionados de campos de registros podem ser copiados de um *dataset* para outro, convertendo e mesclando dados se necessário. [...] Registros de autoridades podem conter *links*, tanto internos quanto externos, que podem ser recursos valiosos para o enriquecimento de dados (BOJARS; ZOGLA; EGÏTE, 2015, p. 187, tradução nossa)<sup>28</sup>.

De acordo com Singer (2009) é fundamental que as bibliotecas façam parte da Web e o LD é a forma lógica para conseguir isso. Revela-se também uma forma de gestão de dados bibliográficos que podem ser formados por diversas fontes, compartilhados e reutilizados.

A seguir é discutida a conversão de dados para que registros bibliográficos fiquem compatíveis com o LD e que possam ser reutilizados.

#### **4.1 Conversão de metadados**

As regras para representação estabelecidas pelo Código de Catalogação Anglo-Americano, 2ª edição (AACR2), não se mostram eficazes para atender a descrição de recursos digitais. O *Functional Requirements for Bibliographic Records* (FRBR) é um

---

<sup>28</sup> *Data enrichment is a complex task and details of how it can be performed depend on the datasets involved. For example, selected data record fields may be copied from one dataset to the other, converting and merging data as necessary. [...] Authority data records may contain links, both internal and external, that can be a valuable resource for data enrichment* (BOJARS; ZOGLA; EGÏTE, 2015, p. 187).

modelo conceitual que contribuiu com o desenvolvimento do Recurso Descrição e Acesso (RDA, *Resource Description and Access*) que propõe novas regras e modelos para descrição.

Para Byrne e Godard (2010), a implantação do RDA pelas bibliotecas tornará mais simples a aplicação do LD, uma vez que os dados já estarão no modelo RDF, facilitando sua publicação e vínculos com outros *datasets*. O RDA foi desenvolvido com base no FRBR, modelo de entidade – relacionamento.

Nem o FRBR ou o RDA foram desenvolvidos em alinhamento com o padrão do *Linked Data*, mas o modelo FRBR usa entidades (coisas) e relacionamentos, que é conceitualmente similar aos conceitos básicos da Web Semântica (COYLE, 2010, p.7, tradução nossa)<sup>29</sup>.

O formato MARC 21 é utilizado há muito tempo para representar registros bibliográficos e de autoridades, porém ele não possui compatibilidade com a Web dos dados, mantendo os dados das bibliotecas isolados (CHEN, 2017; SCHREURS; LORIMER, 2017; SHIEH; REESE, 2015). Para o W3C *Linked Data Incubator Group* (2011) o MARC 21 executou um papel essencial na criação e compartilhamento de dados entre bibliotecas. Cole et al. (2013) endossam a constatação que o formato MARC 21, criado na década de 1960, não foi concebido para suportar completamente os princípios do LOD, mesmo após atualizações e sua serialização para o formato MARCXML.

Como o formato MARC 21 não é utilizado na Web dos dados é necessária a conversão de registros para o modelo RDF. “A transição [da biblioteca] de centrada em registro para centrada em recurso não será simples ou barata” (SINGER, 2009, p. 119, tradução e grifos nossos)<sup>30</sup>.

Para Coyle (2010), a participação das bibliotecas como provedoras de dados na Web pode vir a alterar a compreensão da representação.

---

<sup>29</sup> Neither FRBR nor RDA was developed to meet the linked data standard, but the FRBR model uses entities (things) and relationships, which is conceptually similar to the basic concepts of the semantic Web (COYLE, 2010, p. 7).

<sup>30</sup> The shift from record-centricity to resource-centricity would not be easy or cheap (SINGER, 2009, p. 119).



A ideia que os metadados de biblioteca serão usados amplamente na Web aberta altera o significado de catalogar: catalogar não será mais limitado à criação de registros para o catálogo da biblioteca, mas servirá também para outras funções e para os usuários que talvez não façam uso do catálogo da biblioteca. Esta é a verdadeira expansão do papel dos dados da biblioteca até o ponto onde podem ser utilizados para qualquer função bibliográfica (COYLE, 2010, p. 8, tradução nossa)<sup>31</sup>.

Para Chen (2017) é necessária a conversão de dados do MARC 21 em banco de dados construídos com URIs. As complexidades encaradas pelas bibliotecas, entretanto, vão desde a dificuldade em escolher os termos apropriados para realizar a conversão para o RDF, até a não consistência dos dados, muitas vezes exigindo intervenção manual nos registros, o que torna a atividade lenta e custosa.

A característica do MARC 21 de nomear subcampos com o mesmo nome, exigindo a identificação do campo para compreender o dado que está sendo descrito, agrega complexidade no momento do mapeamento dos dados para o RDF. Outro aspecto é que, mesmo transformando os dados de MARC 21 para MARCXML, o modelo de dados do vocabulário controlado é apoiado em expressões (*strings*) e não em URIs. Estes são alguns dos aspectos que reforçam a compreensão que o MARC 21 não é o formato adequado para publicação de *datasets* como LOD (COLE et al., 2013). Os identificadores facilitam a exposição e troca de dados, proporcionando interoperabilidade entre sistemas e a modelagem dos dados (SHIEH; REESE, 2015). Identificadores formados por URIs contribuem para a singularização de recursos e relações, base da Web Semântica.

Os dados legados em MARC 21 precisam ser convertidos para o RDF, tornando os metadados passíveis de reuso e interligação com outros dados. De acordo com Singer (2009):

A meta inicial é ser capaz de criar o RDF nos esquemas RDA e FRBR a partir de dados existentes em MARC que estão em uso. Isso é uma necessidade crítica, e é irreal esperar que catalogadores possam retrospectivamente voltar e remodelar cada registro existente para um novo

---

<sup>31</sup> *The idea that library metadata will be used widely on the open Web changes the meaning of cataloging: cataloging will no longer be limited to the creation of records for the library catalog, but will serve other functions as well, and users who may never directly make use of the library catalog. This is a true expansion of the role of library data, to the point where it can be used for any bibliographic function* (COYLE, 2010, p. 8).

formato. Se isso for um requerimento do RDA, isso estaria morto na chegada (SINGER, 2009, p. 124, tradução nossa)<sup>32</sup>.

Além da conversão dos dados legados, também é necessária a inclusão de URIs para que sejam feitas integrações dos dados do catálogo com *datasets* externos. Shieh e Reese (2015) relatam experiência da *George Washington University Libraries* (GW) que incluiu URIs em lotes de registros bibliográficos. Um dos registros convertidos é apresentado na **Figura 7** (destaques nossos).

**Figura 7:** Fragmento de registro MARC 21 com URIs no subcampo zero (|0)

---

```

100 1# |a Neustadt, Richard E. |0 http://id.loc.gov/authorities/names/n79007283 |0
    http://viaf.org/viaf/17283877
650 #1 |a Swine influenza |x Vaccination |z United States.
650 #0 |a Medical policy |z United States. |0
    http://id.loc.gov/authorities/subjects/sh2008107275
650 #2 |a Influenza A vírus |z United States. |0 http://id.nlm.nih.gov/mesh/D009980
650 $2 |a Influenza Vaccines |z United States. |0 http://id.nlm.nih.gov/mesh/D007252
650 #2 |a Influenza, Human |x prevention & control |z United States. |0
    http://id.nlm.nih.gov/mesh/D007251
650 #2 |a Public Policy |z United States. |0 http://id.nlm.nih.gov/mesh/D11640
700 1# |a Fineberg, Harvey V. |e author. |0
    http://id.loc.gov/authorities/names/n790029521 |0 http://viaf.org/viaf/75118372

```

---

Fonte: SHIEH; REESE, 2015, p. 218 (grifos nossos).

Na **Figura 7** observa-se que para a identificação de autores principal (*tag* 100) e secundário (*tag* 700) foram utilizadas duas fontes: VIAF e as autoridades da *Library of Congress* (LOC), cada uma armazenada em um subcampo |0. Na primeira ocorrência de assunto (*tag* 650) a ferramenta não localizou uma URI, porém as demais *tags* 650 tiveram a inclusão das URIs, com dados do LCSH e MESH. Após a disponibilização do subcampo |1 em 2017, a LOC recomenda que os identificadores que representam objetos do mundo real como ResearchID, ORCID etc., sejam armazenados neste subcampo e não mais no |0. Os dados armazenados no |0 são aqueles criados por bibliotecas e com aplicação essencial para elas, como o ISNI, VIAF, autoridades da *Library of Congress*, entre outras. Esta atualização no MARC exigirá que a GW altere algumas URIs do subcampo |0 para o |1.

---

<sup>32</sup> *The initial goal is be able to create the RDF in the RDA and FRBR schemas from the existing MARC records currently in use. This is a critical need; it is unrealistic to expect catalogers to be able to retrospectively go back and remodel every record in existence into a new format. If this were a requirement of the RDA, it would be dead on arrival* (SINGER, 2009, p. 124).

Analisando dezesseis instituições que relataram experiências com LOD, Chen (2017) identificou que a modelagem dos dados foi realizada de três formas:

- 1) Seleção do modelo de dados ou a ontologia básica que será utilizada. O uso do FRBR foi adotado por algumas instituições de acordo com a qualidade dos dados do legado;
- 2) Em alguns casos mais de um vocabulário foi utilizado na modelagem de dados, como o FOAF, SKOS etc.;
- 3) Conversão dos dados ao formato MODS.

Chen (2017) relata que deste conjunto de dezesseis bibliotecas somente três instituições não partiram do formato MARC 21 para realizar a criação de triplas. Foi identificado que quanto mais consistentes forem os dados do legado, menor a necessidade de ajustes manuais na conversão.

A conversão dos dados na UIUC foi feita do MARC 21 para o MODS, cujo padrão estava mais preparado para receber URIs e que pode ser usado como um caminho para a conversão do legado para modelos semânticas como o *Schema.org*. Depois foram incluídas as URIs conectando criadores e contribuintes utilizando o VIAF como vocabulário enquanto para assuntos foi utilizado o LSCH. Após esta etapa foi feita a conversão do MODS para o RDF. A UIUC não converteu seu legado ao BIBFRAME, mas ao *Schema.org*, permitindo futura interoperabilidade entre os modelos (RAYLE, 2016). A seleção do MODS foi calcada na sua característica de permitir a inclusão de URIs aos dados, situação que não era estável no MARC 21 na época em que foi realizada a conversão dos registros.

Segundo Assumpção (2018), o MODS foi desenvolvido pela LOC em 2001 como uma alternativa ao formato MARC 21 para a descrição de registros bibliográficos. O modelo para descrição de autoridades MADS foi publicado em 2005, como um acompanhamento do MODS.

Conforme observado nas experiências relatadas, diversas estratégias para conversão do legado podem ser adotadas. Deve ser levado em conta a qualidade dos dados, o volume, os recursos e esforços necessários para esta atividade.

Após a modelagem é necessário incluir as relações existentes. A partir do momento que as relações de autoria e assunto estão identificadas, Chen (2017) sugere

a utilização de *dataset* externo para inclusão de significado aos recursos. Evidentemente podem acontecer ambiguidades na identificação e nessas situações a escolha do termo apropriado provavelmente terá que ser feita manualmente.

O mapeamento do MARC 21 para um modelo semântico não precisa ser feito com o emprego de um único vocabulário, mas o uso de vários ou extensão dos existentes (COLE et al., 2013). Podem ser utilizadas ontologias já publicadas, ampliar existentes, ou partir para a construção de uma própria.

A seguir serão apresentados os projetos de uso da Web Semântica no escopo de bibliotecas, liderando iniciativas para a participação no cenário do LD.

## **4.2 Algumas iniciativas da Web Semântica em bibliotecas**

Nesta seção serão apresentadas brevemente algumas iniciativas da Web Semântica aplicadas às bibliotecas. Serão relatados os projetos LD4, o *Schema.org* e o *dataset* VIAF. Estas iniciativas estão elencadas nesta pesquisa porque foram selecionadas para utilização no modelo de dados que será proposto no Capítulo 5.

### **4.2.1 Projetos LD4**

Quatro projetos foram patrocinados pela *Andrew W. Mellon Foundation: Linked Data for Libraries (LD4L)*, *Linked Data for Production (LD4P)*, *Linked Data for Libraries Labs (LD4L Labs)* e *Linked Data for Production Pathway to Implementation (LD4P2, Caminho para Implementação, tradução nossa)*. Desde 2016 os relatórios, atualizações e documentações em geral destes projetos estão armazenados em ferramenta colaborativa (*Duraspace Wiki*<sup>33</sup>). Esta página também reúne informações e acessos a diversas ontologias, fontes e demais dados de interesse, sendo um catalisador de ferramentas e possibilidades de aplicação de Web Semântica.

O LD4L buscou contribuir para a utilização do LD por bibliotecas por meio da criação da *Scholarly Resource Semantic Information Store (SRSIS)*, Armazenamento de Informações Semânticas de Recursos Científicos, tradução nossa), que é um modelo e ontologia que visa estabelecer comunidades de bibliotecas como LOD.

---

<sup>33</sup> <https://wiki.duraspace.org/pages/viewpage.action?pageId=74515029>

O LD4P iniciou com a proposta de fazer a transição dos fluxos de trabalho e serviços técnicos para o ambiente do LOD. Seus objetivos eram:

(1) Reinventar a produção de metadados para qualquer recurso que a biblioteca faz curadoria fazendo uso do *Linked Data* em um ambiente cooperativo e distribuído; (2) Criar uma ferramenta essencial e uma configuração de sistema para pilotar a produção destes metadados; (3) Desenvolver uma estrutura comunitária / colaborativa para coordenar a ontologia e o trabalho técnico (WONG, 2017, p. 332, tradução nossa)<sup>34</sup>.

O grupo está trabalhando para desenvolver padrões, guias e infraestrutura para produção de metadados que sejam disponibilizados como LOD, e no engajamento da comunidade bibliotecária para reforçar a expansão da Web Semântica.

O LD4L *Labs* trabalha alinhado com o projeto LD4P e anseia criar ferramentas relacionadas ao LD que contribuam com os serviços desenvolvidos em bibliotecas, além de servir de suporte às instituições que estão iniciando projetos na área. Também está empenhado na modelagem de ontologias entre domínios e produção de metadados, como o desenvolvimento do BIBFRAME e o BIBFRAME *Extension Ontology* (*bibliotek-o*, BIBFRAME Ontologia de extensão, tradução nossa), criando melhores práticas para extensão de modelos de ontologias para a produção de LD (BRANAN; FUTORNICK, 2018).

O LD4L *Labs* possui quatro grandes áreas de atuação: 1) desenvolver e manter ferramentas para criar LD; 2) explorar estratégias para compreensão dos relacionamentos utilizados em grafos para contribuir com a descoberta de recursos; 3) contribuir com o desenvolvimento da ontologia BIBFRAME, com esforços para manutenção de persistência das URIs; e 4) apoiar a conversão de dados criando ferramentas que podem ser utilizadas pela comunidade bibliotecária. Além do apoio dado à LOC em relação ao BIBFRAME, o projeto também apoia iniciativas do *Schema.org*, entre outras.

---

<sup>34</sup> (1) Reinvent metadata production for any resource a library curates making use of linked data in a cooperative, distributed environment; (2) Create an essential tool and a system configuration to pilot this metadata production; (3) Develop a community/collaborative framework for coordinating the ontology and technical work (WONG, 2017, p. 332).

No sítio do LD4L *Labs* estão disponibilizadas ferramentas para conversão de metadados de livros digitais do MARC 21 ao BIBFRAME e do BIBFRAME ao MARC 21.

Os objetivos do LD4P2 são avançar na utilização e aplicação do LD em bibliotecas. Este grupo está trabalhando para criar ferramentas, ontologias e serviços com o intuito de melhorar a descoberta, uso e entendimento de informações acadêmicas (WOODS; WARNER, 2018). O estágio LD4P2 pode ser compreendido como uma segunda fase do LD4P e também está em curso, com período de desenvolvimento definido entre 2018 e 2020.

Uma das ações do projeto LD4P2 é o desenvolvimento do protocolo *Linked Data Notifications* (LDN) que, como recomendação da W3C, permite a troca de mensagens entre servidores e aplicações consumidoras sobre qualquer tipo de informação (MCGEE; WARNER, 2018). Até o momento foram identificadas as seguintes possibilidades de aplicação deste protocolo (USE CASES..., 2018):

- 1) Notificar uma terceira parte vinculada a um provedor de *dataset* quando alguma atualização for feita;
- 2) Permitir que catalogadores reportem erros em recursos de uma terceira parte, recomendando a fonte para que seja feita alteração;
- 3) Notificar o provedor do *dataset* sobre o tipo de *link* estabelecido por uma terceira parte, definindo se deseja que seja criado um *link* inverso;
- 4) Notificar o catalogador se uma instituição catalogadora externa fizer ajustes no recurso que foi descrito pela sua instituição;
- 5) Permitir que provedores de dados notifiquem repositórios ou arquivos sobre novos documentos, *datasets* ou anotações.

Na elaboração do modelo outras possibilidades de uso de notificações serão discutidas.

#### 4.2.2 *Schema.org*

O *Schema.org* é uma comunidade colaborativa cuja missão principal é criar, manter e promover modelo dados para a Web. Esta comunidade foi fundada em 2011 pelo *Google*, *Microsoft*, *Yahoo* e *Yandex* e provê esquemas abertos que podem ser

utilizados livremente. O *Schema* não almeja ser uma ontologia para uso mundial, mas as propriedades desenvolvidas são aquelas que os fundadores acreditam que possuem maior aplicação na Web. Foram criados diversos esquemas, de acordo com o tipo de informação que será descrita, buscando atender uma ampla gama de áreas, permitindo que coisas sejam descritas em ambiente semântico e disponibilizados na Web dos dados (RAYLE, 2016). Assim, existem esquemas para representação de pessoas, eventos, lugares, produtos, saúde etc.

Os esquemas são um conjunto de classes associadas a um conjunto de propriedades. Quando foi lançado, o *Schema.org* contava com 297 classes e 187 relacionamentos (GUHA; BRINCKLEY; MACNETH, 2016). No período em que o sítio do *Schema.org* foi visitado para esta pesquisa estavam disponibilizados 598 classes, 862 propriedades e 114 valores de enumeração (SCHEMA.ORG, 2018).

Em decorrência de terem amplas possibilidades de emprego, os *Schemas* são de interesse de bibliotecas e demais organizações relacionadas com publicações. Dentro das classes de esquema para coisa (*thing*) – a mais genérica das classes - são encontrados diversos vocabulários que podem ter sub-vocabulários, apresentando a hierarquia entre as classes e subclasses. Na **Figura 8** é ilustrado o *Schema Book*, que está dentro da propriedade *CreativeWork*. O *Book*, por sua vez, possui a subpropriedade *Audiobook*.

**Figura 8:** Fragmento do *Schema:Book* dentro da classe *Thing*



Dentro dos *frameworks* são visualizadas outras entidades de acordo com o que será descrito. Por exemplo, o *Schema:Book* é utilizado por bibliotecas, editoras, livrarias etc. Para o uso em bibliotecas existem entidades que são úteis somente a este grupo. Para uso de editores o *Schema* apresenta entidades vinculadas ao comércio eletrônico e demais dados pertinentes aos editores. Isso pode ser feito utilizando a extensão *publisherImprint* do *Schema:Book*. Tanto a biblioteca quanto o editor descrevem um livro, porém, como a necessidade de cada um é distinta, o *Schema* possui flexibilidade para atender aos dois públicos, permitindo o uso de dois ou mais esquemas, visto que seu modelo de dados é entre domínios (*cross-domain*) (GUHA; BRINCKLEY; MACNETH, 2016).

O *Schema:Book* possui três grupos de propriedades: *Book*, *CreativeWork* e *Thing*, cada uma com seu conjunto de subpropriedades. Para cada propriedade é informado o tipo de dado esperado, se booleano, texto, tipo de formato, pessoa e número inteiro, além de descrição da informação que deve ser incluída.

#### 4.2.3 *Virtual International Authority File*

O VIAF é um consórcio formado por instituições de todo o mundo que tem por objetivo prover dados confiáveis de autoridades. O VIAF foi iniciado em 2003 por meio de parceria entre a LOC, OCLC e a *Deutsche Nationalbibliothek* (DNB, Biblioteca Nacional da Alemanha, tradução nossa) visando reduzir os gastos e compartilhar os arquivos de autoridades de bibliotecas (ASSUMPCÃO, 2018). Em 2007 a *Bibliothèque Nationale de France* (BnF, Biblioteca Nacional da França, tradução nossa) juntou-se ao grupo, completando os quatro membros originais do consórcio (OCLC, 2018).

Atualmente 62 instituições disponibilizam seus dados no VIAF. Este grupo é formado por 58 instituições contribuintes presentes em todos os continentes e pelos colaboradores *Perseus Digital Library*, *Wikipedia*, *Syriac Reference Portal* e *ISNI*.

A proposta do VIAF não é consolidar a descrição de autores impondo uma forma única de representação, mas agrupar as formas como as agências catalogadoras estabelecem seus pontos de acesso, respeitando diferenças culturais, linguísticas etc.

A **Figura 9** apresenta as formas utilizadas pelas instituições presentes no VIAF para descrever o autor “Monteiro Lobato”. Além das formas de descrição, também são



visualizadas as agências catalogadoras dentro das formas adotadas. A apresentação das variações dos nomes é feita levando em consideração a maior quantidade de instituições que utilizam a mesma forma de descrição para o autor, contribuindo na seleção da representação adequada. Além dos dados do autor (*personal*), o VIAF também apresenta dados de obras e expressões.

**Figura 9:** Fragmento de informações sobre Monteiro Lobato no VIAF

<a href="#">Lobato, José Bento Monteiro, 1882-1948</a> 	Personal
<a href="#">Lobato, Monteiro, 1882-1948</a> 	
<a href="#">Монтейру Лобату, Ж. Б. 1882-1948</a> 	
<a href="#">Monteiro Lobato escritor brasileiro</a> 	
<a href="#">Lobato, José Bento Monteiro</a> 	
<a href="#">Lobato, Monteiro</a> 	
<a href="#">Монтейру Лобату, Ж. Б. Жозе Бенту 1882-1948</a> 	
<hr/>	
<a href="#">Lobato, José Bento Monteiro (1882-1948).   Reinações de Narizinho</a>  xR	Work
<a href="#">Lobato, José Bento Monteiro, 1882-1948.   Barca de Gleyre</a> 	Work
<a href="#">Lobato, José Bento Monteiro, 1882-1948.   Works. 1959. 1.a série. Literatura g...</a> 	Work
<a href="#">Lobato, José Bento Monteiro, 1882-1948.   Picapau amarelo.   Russian   (Тыни́а̃н ...</a> xR	Expression
<a href="#">Lobato, José Bento Monteiro, 1882-1948.   Correspondence. Selections</a> 	Work
<a href="#">Lobato, José Bento Monteiro, 1882-1948.   Works. 2a. Série. Literatura infanti ...</a> 	Work

Fonte: VIAF<sup>35</sup>

No próximo capítulo será apresentada o modelo para publicação de dados de fornecedores de livros digitais licenciados para consumo de fornecedores e bibliotecas por meio da Web Semântica.

<sup>35</sup> [https://viaf.org/viaf/99985617/#Lobato,\\_Jos%C3%A9\\_Bento\\_Monteiro,\\_1882-1948](https://viaf.org/viaf/99985617/#Lobato,_Jos%C3%A9_Bento_Monteiro,_1882-1948)

## 5 Modelo para gestão de coleção digital por meio da Web Semântica

---

### Preâmbulo

Neste capítulo é apresentada a proposta de modelo para gestão de livros digitais licenciados com utilização de recursos e técnicas da Web Semântica, detalhando sua arquitetura, características, modelagem com relação dos campos necessários para descrição de dados bibliográficos, de fornecedores e de licenciamento, as etapas para publicação do *dataset*, e o envio e recebimento de notificações. As etapas para publicação definidas são: política e seleção; formalização, enriquecimento de dados, formatos e licenças; conversão ferramental; dados de licenciamento; processo de recuperação; marketing e *feedback*; e qualidade e preservação, conforme apresentado por Santarem Segundo (2018).

Na parte de conversão ferramental foram feitos mapeamentos contemplando os formatos ONIX e MARC 21 e o modelo de dados *Schema* para dados bibliográficos e de licenciamento. Como foi identificado que nem todos os fornecedores utilizam o formato ONIX, mas outras ferramentas, foi proposto modelo de planilha de dados com os campos e respectivas sequencias. Estes campos também foram mapeados para o ONIX e *Schema*. Os campos de autoria (criadores e contribuidores), edição, número de páginas (extensão), ISBN da versão impressa, local de publicação, editora, assuntos e a descrição de URLs para acesso foram detalhados na representação na planilha de dados, orientando o uso de regras específicas.

Para cada etapa de publicação foram relacionadas as boas práticas da W3C que foram atendidas, com sua numeração e descrição.

Até o momento esta pesquisa abordou sua motivação, proposta e tese (Capítulo 1) sobre o uso da Web Semântica para a gestão de livros digitais licenciados. Para contextualizar a relevância da pesquisa foram apresentadas situações entre bibliotecas e fornecedores de livros digitais, identificando algumas dificuldades observadas (Capítulo 2). O Capítulo 3 discorreu sobre os conceitos da Web Semântica, abordando o *Linked Data*, *Linked Open Data* e *Linked Enterprise Data*, fornecendo base teórica sobre as aplicações que podem ser empregadas em bibliotecas. No Capítulo 4 foram discutidas as possibilidades de aplicação do *Linked Data* em bibliotecas, as iniciativas LD4P, *Schema.org* e VIAF, e discussões sobre a conversão de metadados bibliográficos. Dentre os projetos e discussões foram considerados somente os que possuem aplicabilidade no modelo para publicação de dados de livros digitais licenciados para troca de informações entre fornecedores e bibliotecas.

Este capítulo apresenta o detalhamento das especificações definidas para elaboração de modelo com intuito de favorecer a gestão de coleções de livros digitais licenciados das bibliotecas e aumentar a qualidade dos dados de fornecedores.

A apresentação do modelo está organizada em dois momentos: primeiramente a descrição de sua base e, posteriormente, seu exemplo de validação (Capítulo 6), com apresentação de registro processado.

Os passos para a elaboração do modelo foram construídos com base no referencial teórico identificado sobre a Web Semântica, visando contribuir para a solução das dificuldades identificadas no Capítulo 2.

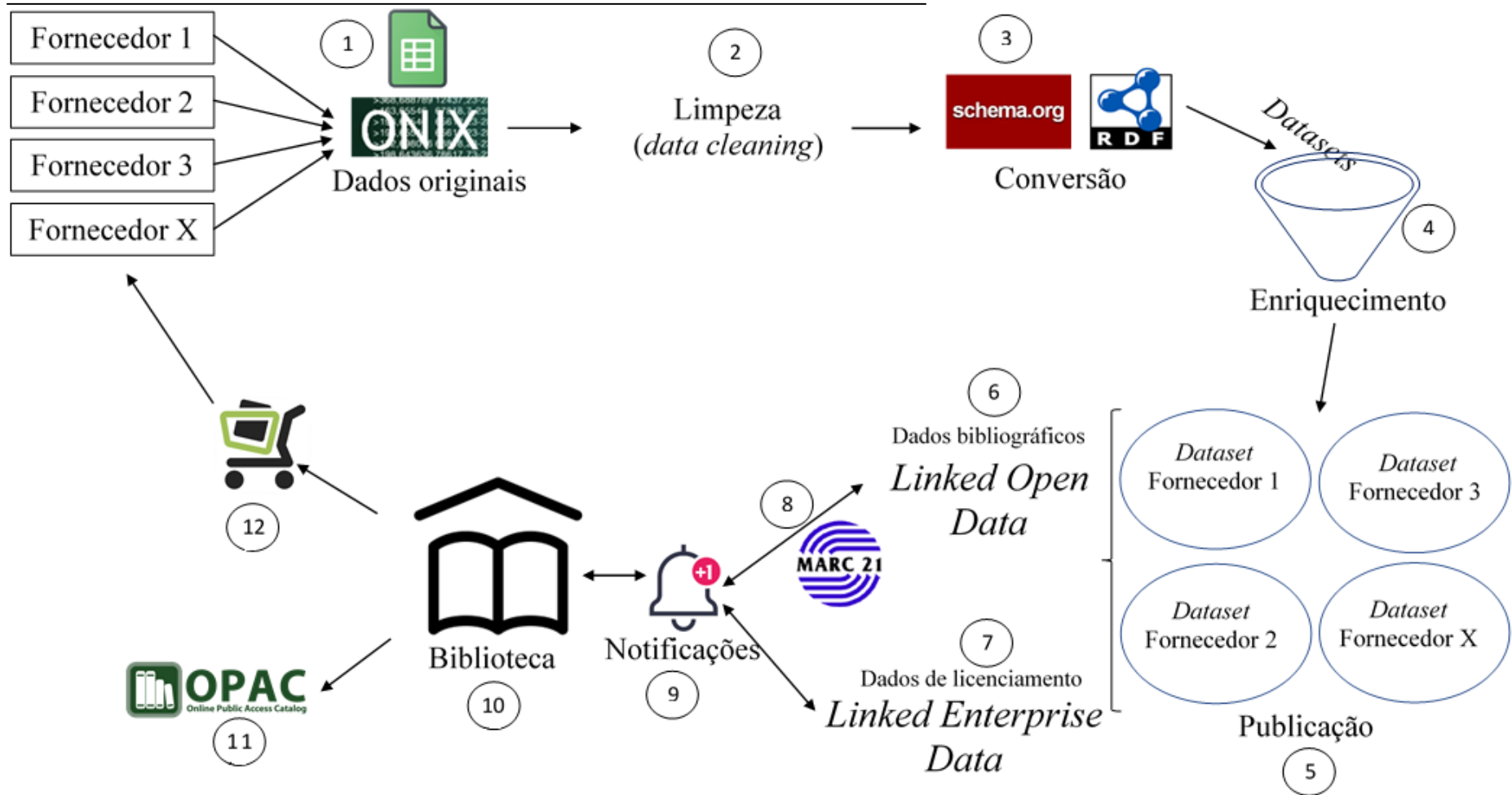
Cada fornecedor terá o seu *dataset*. Inicialmente pensou-se em publicação de *dataset* único reunindo os dados de todos os fornecedores, entretanto esta concepção foi alterada para que cada fornecedor tenha o seu próprio *dataset*. Esta definição foi alcançada devido a:

- a. Manutenção de sigilo sobre os dados de licenciamento trocados entre fornecedores e bibliotecas. Editores usualmente possuem postura conservadora e reunir todos os dados em um único *dataset* poderia proporcionar insegurança aos fornecedores;
- b. Ao optar por um *dataset* único reunindo dados bibliográficos e de licenciamento dos fornecedores corria-se o risco de perda de credibilidade;

- c. As bibliotecas podem selecionar os *datasets* dos fornecedores desejados, evitando a visualização de livros digitais que não são de seu interesse;
- d. Cada fornecedor terá em seu *dataset* os dados bibliográficos e de licenciamento reunidos, podendo realizar transações comerciais com seus clientes;
- e. Fornecedores e bibliotecas podem realizar configurações sobre notificações, informando as alterações realizadas nos registros bibliográficos e nas formas de licenciamento. O fornecedor pode aproveitar ou não as sugestões de correções bibliográficas realizadas pelas bibliotecas. Se aproveitar a sugestão, o *dataset* é atualizado e as demais bibliotecas terão os dados atualizados;
- f. Caso existam obras concorrentes, cada fornecedor terá o seu registro. Isto não seria possível com um *dataset* único. A relevância de manter os dados separados por fornecedores consiste em diferenças descritivas que podem ocorrer de um mesmo livro digital representado por fornecedores diferentes.
- g. Ao encaminhar uma notificação de alteração de registro o fornecedor pode acatar ou não as sugestões recebidas da biblioteca. O modelo deve permitir a visualização dos livros digitais disponíveis por fornecedor, com os dados do licenciamento realizado;
- h. Possibilidade de coleta e geração de estatísticas diretamente no *dataset* de cada fornecedor.

A **Figura 10** apresenta as etapas definidas para o modelo, destacando cada passo do processo, iniciando pelos dados originais dos fornecedores, passando pela limpeza, conversão, enriquecimento, publicação dos *datasets* com dados em LOD e LED, unificação dos dados em ponto único vinculado com o catálogo da biblioteca e para envio de solicitações de livros aos fornecedores.

Figura 10: Modelo para gestão de livros digitais licenciados



Fonte: Elaborado pela autora

As etapas sinalizadas no modelo são:

- 1) Os dados de origem dos fornecedores podem ser obtidos por duas formas: formato ONIX ou planilhas de dados. Os fornecedores podem continuar a utilizar seus sistemas, exportando os dados para que sejam incluídos no *dataset*.
- 2) Os dados selecionados dos fornecedores serão tratados e normalizados (*data cleaning*);
- 3) Os dados tratados serão convertidos ao modelo RDF *Schema*;
  - a. O modelo foi idealizado para receber dados no formato ONIX e também por meio de planilhas de dados. Isto permite que cada fornecedor exporte os dados de seus sistemas para publicação nos *datasets*. O uso de planilhas de dados também foi selecionado para o modelo em virtude de muitos fornecedores não utilizarem o ONIX;
  - b. Para as bibliotecas a definição do conjunto de dados foi construída com a atual manutenção do formato MARC 21, pensando em futura conversão ao BIBFRAME ou outro formato que for recomendado pela *Library of Congress*;
- 4) Os dados serão enriquecidos por *dataset(s)*. Para exemplo de validação do modelo será utilizada somente uma fonte para enriquecimento de dados de autoridades do tipo Pessoa, porém o modelo visa permitir que outros *datasets* sejam utilizados;
- 5) Após o tratamento e enriquecimento dos dados é possível sua publicação como *dataset*;
- 6) A princípio a motivação era a publicação de *datasets* de acesso fechado com os dados dos fornecedores para comunicação e reuso somente para seus clientes. Com a evolução da pesquisa definiu-se pela publicação aberta dos dados bibliográficos. Isto decorre de os livros digitais representados pelos fornecedores não serem sigilosos. O LOD é empregado para os dados bibliográficos;
- 7) Os dados de licenciamento entre fornecedores e bibliotecas é de acesso restrito aos envolvidos. O LED é utilizado para os dados de licenciamento;
- 8) Como bibliotecas utilizam o formato MARC 21 para descrição de registros bibliográficos e os fornecedores também precisam deste formato para

atenderem às demandas de seus clientes bibliotecas, a disponibilização dos metadados é oferecida também no MARC 21;

- 9) Fornecedores e bibliotecas podem trocar notificações, sinalizando alterações na disponibilidade de livros digitais nas plataformas, mudanças nas formas de licenciamento, ou ainda alterações realizadas nas descrições;
- 10) Os dados que transitam entre fornecedores e bibliotecas são relacionados a três conjuntos: bibliográficos, de fornecedores e de licenciamentos. Fornecedores e bibliotecas trocam dados bibliográficos, de licenciamentos e notificações. A biblioteca centraliza os metadados de livros digitais dos fornecedores que foram selecionados
- 11) A biblioteca disponibiliza em seu catálogo os livros digitais que foram licenciados, com as atualizações que ocorrerem durante a vigência do licenciamento;
- 12) A biblioteca pode enviar pedidos de novos licenciamentos ou troca de existentes diretamente ao(s) fornecedor(es), contribuindo com a agilidade em negociações.

Optou-se por não desenvolver uma ontologia específica para o projeto, mas enriquecer e adequar ontologias já publicadas. Esta decisão visa o reaproveitamento de ontologias existentes e em decorrência da construção de modelo de domínio genérico.

O modelo não é uma ferramenta para catalogação de registros bibliográficos, mas para identificação, licenciamento, disponibilização no catálogo e gestão da coleção digital. Para o modelo é indiferente se a catalogação foi feita com aderência ao AACR2 ou RDA. O modelo não define qual regra a biblioteca deve adotar.

Se forem consideradas as regras de catalogação, destaca-se a manutenção do padrão do AACR2 porque nem todas as bibliotecas aderiram ao RDA, principalmente no Brasil. O BIBFRAME é um formato em desenvolvimento, funcionando em versões beta e ainda não está pronto para entrar em produção, o que justifica a escolha pelo MARC 21 neste momento.

Como a *Library of Congress* e demais instituições estão trabalhando para o desenvolvimento de ferramentas para converter legados em MARC 21 para BIBFRAME, não será feito o esforço de converter os metadados a outros formatos agora, como o MODS, por exemplo. Um aspecto considerado é que “o novo código [RDA] é independente de qualquer esquema particular de metadados” (JIN;

SANDBERG, 2014, p. 217, tradução nossa)<sup>36</sup> e “um dos maiores objetivos do RDA é manter compatibilidade com os dados legados e, aparentemente, os dois códigos [AACR2 e RDA] irão coexistir em muitos catálogos de bibliotecas” (JIN; SANDBERG, 2014, p. 219, tradução nossa)<sup>37</sup>. Entretanto, é interessante disponibilizar às bibliotecas os registros em outros formatos como MARC XML ou RDF.

#### O fluxo do processo para publicação (**Figura 10**)

proporcionou a identificação das etapas, permitindo a modelagem dos *datasets* para troca de dados entre fornecedores e bibliotecas, destacando os conjuntos de dados que serão abertos (*Linked Open Data*) e os que serão fechados (*Linked Enterprise Data*). Esta modelagem (**Figura 11**) também contribuiu com o exemplo de validação da proposta do modelo, que é apresentada no Capítulo 6, verificando as possibilidades de implementação e o atendimento dos *stakeholders* do projeto.

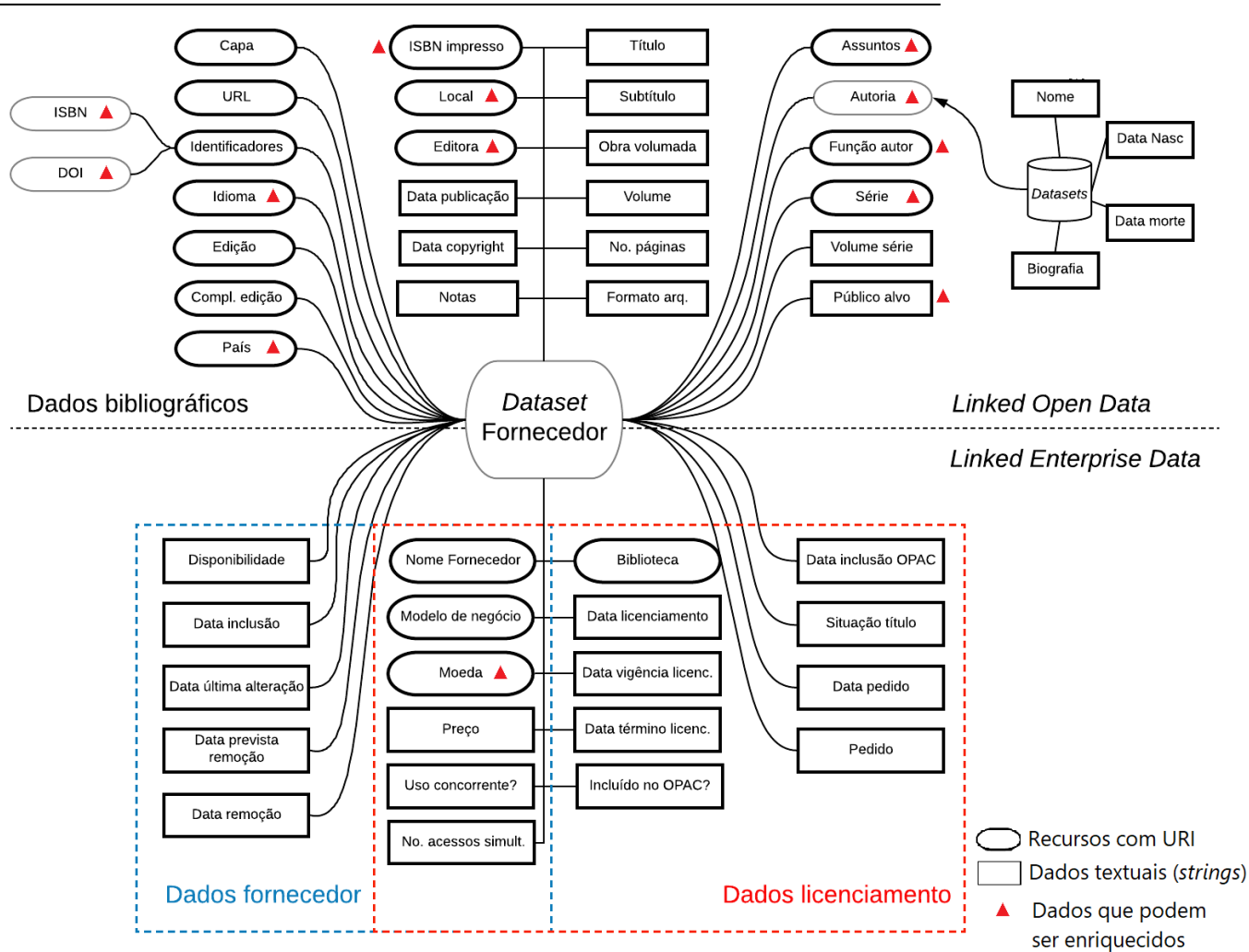
---

<sup>36</sup> *The new code is independent of any particular metadata* (JIN; SANDBERG, 2014, p. 217).

<sup>37</sup> *One of RDA's major goals is to maintain compatibility with legacy AACR2 data, and it is likely that the two codes will co-exist in many library catalogs* (JIN; SANDBERG, 2014, p. 219)



Figura 11: Modelagem de dados dos *datasets* dos fornecedores



Fonte: Elaborado pela autora com a ferramenta *Lucidchart*<sup>38</sup>.

<sup>38</sup> <https://www.lucidchart.com/>

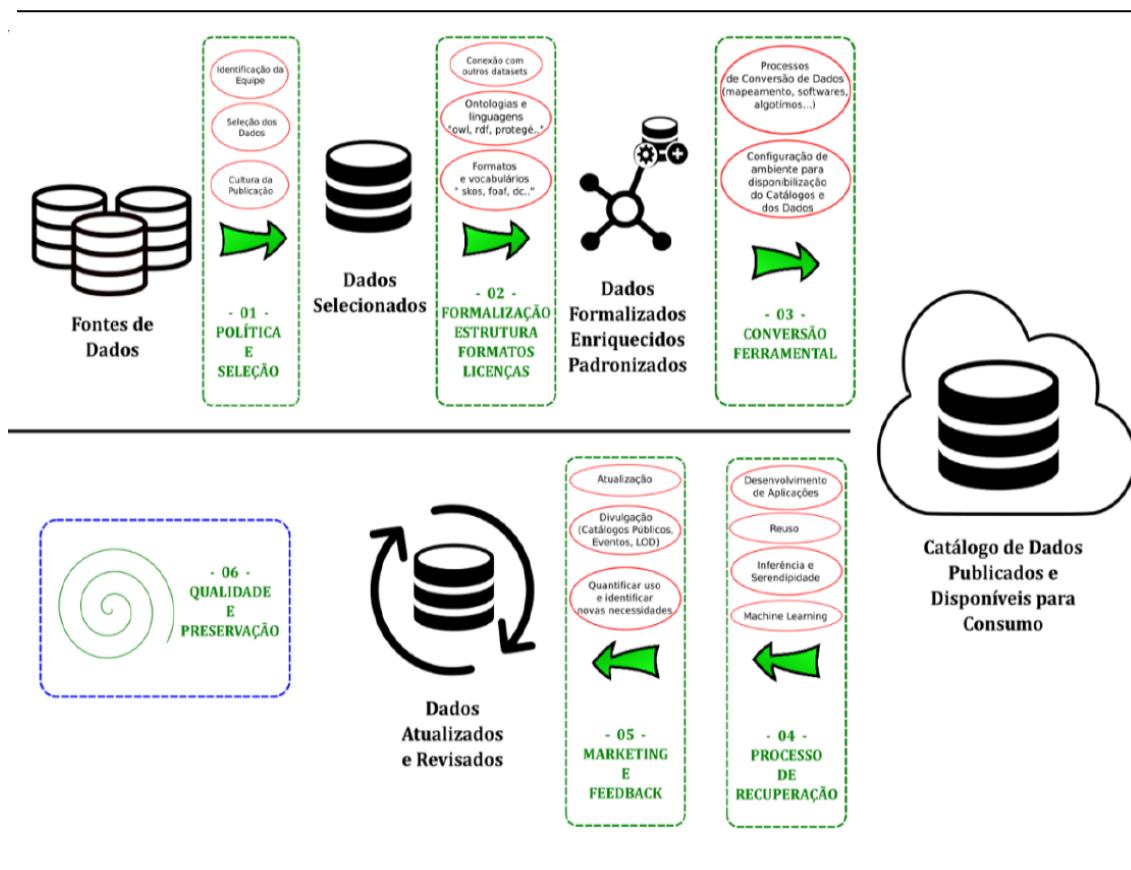
A representação da modelagem abordou as seguintes perspectivas:

- 1) A modelagem elencou os campos necessários para descrição de conjuntos de dados bibliográficos, de fornecedores e de licenciamentos.
- 2) Os dados do fornecedor e de licenciamento estão disponíveis no *dataset* como *Linked Enterprise Data*, enquanto os dados dos registros bibliográficos são disponibilizados como *Linked Open Data*;
- 3) Os *datasets* são retroalimentados com dados de fornecedores e bibliotecas informando sobre alterações nos registros bibliográficos, nas formas de acesso e contratação dos livros digitais e realização de novos pedidos;
- 4) A partir das alterações no *dataset* são enviadas notificações para:
  - a. Fornecedores: retroalimentação com dados sobre alterações realizadas nos registros bibliográficos e de licenciamento;
  - b. Bibliotecas: alterações de livros digitais (novos, alterados ou removidos) e possibilidades de contratação e mudança de dados de licenciamento.
- 5) O recebimento e o tipo de notificação desejada (novos livros digitais, remoção, alteração de modelo de negócios etc.) devem ser passíveis de configuração por bibliotecas e fornecedores, definindo seus interesses;
- 6) Para efeitos de representação somente a autoridade do tipo pessoa teve sua modelagem expandida, mostrando como os dados podem ser enriquecidos por fontes externas, porém este grupo inclui dados de entidade coletiva, evento, assuntos etc.conforme o mapeamento de campos realizado;
- 7) Os campos de notas foram representados por um único elemento na modelagem, porém diversas notas podem ser utilizadas, sendo um retângulo para cada tipo de nota;
- 8) Dados de fornecedores e licenciamentos são compartilhados, evitando redundância de valores;
- 9) Dentre os dados de licenciamento estão presentes as informações sobre pedidos solicitados pelas bibliotecas aos fornecedores.

A seguir serão detalhadas as etapas para publicação e as especificações que foram definidas. A fim de ilustrar a aderência do modelo às boas práticas para publicação de dados abertos da W3C, as 35 boas práticas apresentadas por Lóscio, Burle e Calegari (2017) são relacionadas ao final de cada etapa, indicando sua numeração e descrição.

As etapas para publicação dos dados são as baseadas em Santarém Segundo (2018), conforme definido no Capítulo 3, identificadas como 1) Política e seleção; 2) Formalização, formatos e licenças; 3) Conversão ferramental; 4) Processo de recuperação; 5) Marketing e *feedback*; e 6) Qualidade e preservação. Estas etapas são apresentadas na **Figura 12**. As caixas pontilhadas sinalizam fluxo contínuo.

**Figura 12:** Fluxo organizacional para publicação de dados



Fonte: SANTARÉM SEGUNDO, 2018, p. 125.

A seguir serão detalhadas as etapas para publicação.

### 5.1 Política e seleção

Nesta etapa são definidos os passos iniciais, com o estabelecimento das principais orientações que devem ser seguidas. Esta etapa é essencialmente política e deve estar alinhada aos objetivos da publicação. Não é uma fase que demande definições técnicas (SANTARÉM SEGUNDO, 2018), mas os critérios conceituais que irão nortear o projeto. A equipe responsável pela oferta de dados (fornecedores) deve selecionar as fontes de dados que serão utilizadas e de que forma os dados serão publicados. Além das fontes, foram estabelecidos quais metadados fazem parte do escopo. Tanto os fornecedores de conteúdo quanto as bibliotecas

possuem necessidades descritivas específicas e nem todas são aplicáveis aos dois universos. Foram definidos quais dados são efetivamente interessantes a ambos, evitando o acúmulo e trânsito de informações não pertinentes.

Na **Tabela 3** estão relacionadas as principais informações sobre dados bibliográficos, de fornecedores e de licenciamentos que serão consumidos pelo modelo.

**Tabela 3:** Tipos de dados necessários ao modelo

<b>Bibliográficos</b>	<b>Fornecedores</b>	<b>Licenciamentos</b>
Imagem da capa	Nome do Fornecedor	Nome do Fornecedor
ISBN impresso	Disponibilidade do livro	Biblioteca
ISBN digital	Modelos de licenciamento	Data do licenciamento
DOI	Formas de acesso (uso concorrente?)	Modelo aplicado
Idioma	No. de acessos simultâneos	Forma de acesso (uso concorrente?)
Título: subtítulo	Preço	Situação do livro
Autores	Moeda	Incluído no OPAC
Identificadores de pessoas	Data inclusão na plataforma	Data de inclusão no OPAC
Função de autores	Data remoção na plataforma	Pedido (inclusão de novo livro)
Edição	Data prevista para remoção na plataforma	Data pedido
Complemento Edição	Data da última alteração do livro	Preço
Local de publicação		Moeda
Editora		No. Acessos simultâneos
País de publicação		Data início vigência do licenciamento
Data de publicação		Data término vigência do licenciamento
Data de copyright		
Volume		
Obra volumada		
Assuntos		
Série		
Volume da série		
Notas		
Público alvo		
Número de páginas		
Formato do arquivo		
URL para acesso		

Fonte: elaborado pela autora

Na **Tabela 3** observa-se que alguns campos estão presentes em mais de um conjunto de dados como, por exemplo, nome do fornecedor, modelo de negócio, moeda, preço, uso concorrente e número de acessos simultâneos contratados. O modelo permite que estes dados sejam compartilhados, dispensando a inclusão dos mesmos de forma redundante, reaproveitando um dado que já está presente na base. Assim, não é necessário informar

novamente o nome do fornecedor, por exemplo, uma vez que ele já consta do *dataset* e pode ser reutilizado, sem necessidade de duplicidade.

No caso de dados do fornecedor deve ser permitida a seleção da forma de publicação de acordo com seu interesse, podendo ser aberta ou fechada. As notificações sobre alterações nos dados realizadas pelas bibliotecas e que serão comunicadas aos fornecedores são abertas, uma vez que para dados bibliográficos será utilizado o LOD. Entretanto, a realização de pedidos, alterações nos licenciamentos e demais operações comerciais são de uso restrito, com o LED. A **Tabela 4** reúne as fontes e a forma de publicação que foram estabelecidas para o modelo.

**Tabela 4:** Definições – Política e seleção

Atividades		Definições
Fontes (origem) de dados utilizadas		Legado de editores em ONIX ou planilha de dados
Cultura de publicação		
	Dados bibliográficos	LOD
	Dados de fornecedores	LED
	Dados do licenciamento	LED

Fonte: elaborado pela autora.

Esta é a etapa 1 da **Figura 10**, indicando a seleção e coleta dos originais que serão incluídos no *dataset*. Não existem restrições para que os dados sejam obtidos de outras fontes (planilhas de terceiros ou sistemas próprios). Porém, ao coletar registros de fontes com modelos diferentes, é necessário realizar uma normalização para que todos fiquem com a mesma característica antes de passar à fase seguinte, onde ocorrerá a formalização dos dados. Esta etapa, embora não seja realizada na fase da qualidade, contribui com a consistência dos dados, enxugando termos redundantes ou palavras diferentes que representam a mesma coisa. Esta ação facilita a aplicação de vocabulários que ocorrerá na próxima fase.

Para esta etapa foram identificadas as boas práticas da W3C (LÓSCIO; BURLE; CALEGARI, 2017) que são atendidas (**Tabela 5**):

**Tabela 5:** Boas práticas da W3C na etapa de Política e seleção

No. Boas práticas	Descrição
1	Forneça metadados
2	Forneça metadados descritivos
28	Avalie a cobertura do conjunto de dados.

Fonte: elaborado pela autora a partir de LÓSCIO; BURLE; CALEGARI, 2017.

Uma vez selecionadas as fontes, os dados desejados e a forma como serão publicados, passa-se para a segunda etapa para publicação do *dataset*, onde os dados serão formalizados e enriquecidos. Neste momento também são definidos os formatos que serão utilizados e os tipos de licenças aplicados ao conjunto de registros bibliográficos.

## 5.2 *Formalização, enriquecimento de dados, formatos e licenças*

A formalização inicia-se pela etapa de limpeza, visando produzir dados neutros<sup>39</sup>, diminuindo a incidência de erros. A limpeza busca minimizar inconsistência dos dados, eliminando duplicidades e erros na descrição. Para o *data cleaning* podem ser utilizadas diversas ferramentas, de acordo com o interesse e experiência dos responsáveis por esta etapa. Posterior a limpeza dos dados, inicia-se uma etapa de cunho técnico de transformação dos mesmos, para que fiquem compatíveis com a Web Semântica. Neste momento os dados são trabalhados, ocorrendo mudança de modelo e formato. Após a transformação do legado ocorre o *data lifting*, com o enriquecimento dos dados com outros vocabulários. Diversas ontologias, linguagens e vocabulários podem ser aplicados para a inclusão de URIs e formalização das relações existentes nos grafos.

Como os dados presentes nos *datasets* são de interesse do mercado do livro, a partir do projeto *Linked Open Vocabularies* (LOV) (LINKED, 2018) e das recomendações do W3C (W3C, 2018), foram identificadas algumas linguagens e ontologias que podem ser utilizadas para a formalização do modelo:

- 1) *DC Terms*: modelo genérico de metadados mantido pelo *Dublin Core Metadata Initiative* (DCMI), que abrange propriedades, esquemas de vocabulários e de codificação de sintaxes e classes (DCMI, 2012);
- 2) *FaBiO* (*FRBR-aligned Bibliographic Ontology*): contribui com a descrição de entidades que são publicadas ou que potencialmente serão publicadas (SHOTTON; PERONI, 2018). O uso desta ontologia pode contribuir com uma futura implementação do RDA na biblioteca;
- 3) *FOAF*: utilizado para vincular pessoas e informações. Trata-se de um dicionário de termos, propriedade ou classe, descrevendo pessoas, grupos e documentos (BRICKLEY; MILLER, 2014). Pelo *FOAF* também é possível a identificação

---

<sup>39</sup> De acordo com Culbertson, Mandelstam e Prager (2011), dados neutros possuem ênfase somente em informações aplicáveis a registros com o mesmo conteúdo. Não são contemplados neste conjunto informações específicas de nenhum fornecedor, nem mesmo em notas, qualificadores ou inclusão de títulos uniformes

- (*Account Identifier*) para prestação de serviços, o que pode ser aplicado para reconhecimento de bibliotecas e fornecedores nos licenciamentos;
- 4) *MARC Code list for relators Scheme*: define as URIs com as relações entre autores e registros;
  - 5) *SKOS*: utilizado para sistemas de organização do conhecimento como esquemas de classificação, tesouros, cabeçalhos de assuntos e taxonomias. É uma recomendação da W3C (W3C, 2012);
  - 6) *Schema*: descreve entidades e as relações entre elas. Possui um conjunto de tipos de itens para descrição, como trabalhos criativos (*creative works*), objetos não textuais, eventos, organizações, pessoas, locais, produtos e revisão. Existe um tipo específico para descrição de livros (*SchemaBook*, abordado no Capítulo 4). Como este tipo está presente no ecossistema de fornecedores de conteúdo e bibliotecas, é o principal modelo adotado neste projeto, porém não o único, visto que dados de licenciamento também serão trocados entre fornecedores e bibliotecas e estão presentes em outros *Schemas*.

Na sequência é definida a licença com a qual os dados serão disponibilizados. Os tipos de licenças aplicados aos dados bibliográficos são as *Creative Commons* (CC, 2018a) que são organizadas em três camadas: código legal, legível por máquinas e por humanos. Os tipos de licenças podem ser:

Atribuição (BY): permite a distribuição, adaptação ou construção de uma nova obra (obra derivada), contanto que a autoria da obra original seja creditada;

Compartilhar igual (SA): obras derivadas criadas a partir de uma obra original devem ser compartilhadas com o mesmo tipo de licença atribuída no original;

Não derivadas (ND): não permite a criação de outras obras a partir de um original;

Não comercial (NC): não permite a comercialização da obra.

Os tipos de licenças podem ser combinados de acordo com a definição de uso desejada, com o estabelecimento de restrições ou permissões, formando, por exemplo, as situações:

BY ND: atribuição, porém sem permitir a criação de obras derivadas;

BY ND NC: não permite a criação de obras derivadas, nem o uso comercial;

BY-SA: permite a atribuição da obra com compartilhamento igual à licença da obra original.

Os dados de licenciamento possuem, no modelo, licença de uso restrito, em virtude deste conjunto utilizar aspectos de LED.

A partir do momento que as linguagens e ontologias foram selecionadas, o próximo movimento é preparar os dados para consultas futuras, fazendo a serialização dos mesmos. Para o modelo foram escolhidos os formatos XML, JSON e JSON-LD. O XML e o JSON são formatos bastante conhecidos e com boa utilização. O JSON-LD suporta o *Linked Data*.

A **Tabela 6** reúne as especificações definidas para o modelo nesta etapa. Na **Figura 10** estão representados nos passos 2 a 4.

**Tabela 6:** Especificações de serialização

Atividades	Definições
Limpeza dos dados	Uso de ferramentas para <i>data cleaning</i>
Definição de modelo de dados	RDFS
Esquema adotado	<i>SchemaBook</i>
Fontes para enriquecimento	<i>Dataset(s)</i> selecionado(s) com dados de autoridades
Linguagens e ontologias utilizadas	DC <i>Terms</i> , FaBiO, FOAF, MARC <i>Code List for Relators Scheme</i> , SKOS
Definição de tipos de licenças	
Dados bibliográficos	CC BY AS, identificando a versão utilizada
Dados do fornecedor	Licenças de uso restrito, garantindo privacidade dos dados
Dados do licenciamento	Licenças de uso restrito, garantindo privacidade dos dados
Serialização	XML, JSON, JSON-LD

Fonte: elaborado pela autora.

É interessante aos fornecedores exibirem todos os livros digitais disponíveis às bibliotecas, porém pode ocorrer de alguns quererem repassar a seus clientes somente os livros digitais que foram contratados. Para tanto, os estudos do grupo de interesse sobre o *Content Negotiation by Profile* (SVENSSON; ATKINSON; CAR, 2018), podem ser uma opção para futuras implementações. Este documento orienta a criação de perfis para compartilhamento de dados entre clientes e servidores. Ainda não se trata de uma recomendação da W3C, mas um trabalho em progresso. É necessário acompanhar o projeto para avaliar sua aplicação ao modelo no futuro.



Para esta etapa foram atendidas as seguintes boas práticas da W3C (LÓSCIO; BURLE; CALEGARI, 2017) (**Tabela 7**):

**Tabela 7:** Boas práticas da W3C na etapa de Formalização, enriquecimento de dados, formatos e licenças

No. Boas práticas	Descrição
3	Forneça metadados padronizados
4	Forneça informações sobre a licença dos dados
5	Forneça informações sobre a proveniência dos dados
9	Use URIs persistentes como identificadores dos conjuntos de dados
10	Use URIs persistentes como identificadores dentro dos conjuntos de dados
12	Use formatos de dados legíveis por máquinas padronizados
13	Use representações dos dados neutras no que se refere ao local
14	Forneça os dados em múltiplos formatos
15	Reuse vocabulários, preferencialmente aqueles padronizados
16	Escolha o nível de formalização correto
31	Enriqueça os dados pela geração de novos dados
34	Siga os termos do licenciamento

Fonte: elaborado pela autora a partir de LÓSCIO; BURLE; CALEGARI, 2017.

### 5.3 Conversão ferramental

Com os dados trabalhados, passa-se à etapa de conversão, quando os mesmos são preparados e disponibilizados na Web, passíveis de serem consumidos. Isto demanda conversão de dados, com mapeamento de formatos compatíveis com os padrões utilizados pelos consumidores envolvidos. Os dados em ONIX e os coletados de planilha de dados são mapeados para o *SchemaBook*.

Para que bibliotecas possam consumir estes dados de acordo com seus padrões, é necessária a conversão do *SchemaBook* para o formato MARC 21, com possibilidade de inclusão do(s) livro(s) digital(is) licenciado(s) no catálogo. Esta segunda conversão dispensa o fornecedor de investir na criação de registros no formato MARC 21 para distribuição às bibliotecas sempre que realizar um licenciamento ou renovação.

Destaca-se novamente aqui que os metadados de livros digitais podem ser disponibilizados nos *datasets* em outros formatos, inclusive em RDF no *SchemaBook*, facilitando a atuação de desenvolvedores e evoluções futuras em sistemas de biblioteca, dispensando conversões.

As alterações realizadas pelas bibliotecas nos registros bibliográficos são convertidas do MARC 21 para o padrão da planilha de dados (se o fornecedor utiliza esta ferramenta), ONIX ou ao *SchemaBook* para que possam ser utilizadas pelos fornecedores.

A seguir serão definidos os mapeamentos de acordo com a origem dos dados do fornecedor.

### 5.3.1 Inclusão de dados bibliográficos em ONIX

Os fornecedores que utilizam sistemas com o formato ONIX podem exportar os dados que desejam compartilhar no modelo, demandando uma conversão do ONIX ao *Schema* e deste para o MARC 21, para que os livros digitais sejam incluídos nos catálogos das bibliotecas.

O mapeamento ONIX-*Schema*-MARC 21 está descrito na **Tabela 8**. As propriedades do *SchemaBook* foram majoritariamente selecionadas de *Property/Thing*, entretanto, *Schemas* diferentes do *Book* foram utilizados no mapeamento: *Person* para datas, afiliação de autores e dados biográficos, *Place* para registrar o país de publicação e *CreativeWork/MediaObject* para identificação do arquivo da capa do livro.

**Tabela 8:** Mapeamento – Dados bibliográficos - ONIX – SchemaBook – MARC 21

ONIX	Schema	MARC 21
<i>LanguageCode</i>	<i>inLanguage</i>	008/35-37: código do idioma
<i>ProductForm</i>	<i>typeOfResource</i>	007/01-02 008/23: s (eletrônico) ou o (online) 245 __  h [recurso eletrônico]
ISBN	<i>identifier/ isbn</i> <i>identifier</i>	020 __  a
<i>ProductFormFeatureValue</i>	<i>BookFormatType &gt; identifier</i>	530 __  a 776 __  a  z (outra forma física)
<i>CollectionIDType</i> <i>DOI</i>	<i>identifier</i>	024 7#  a  2 doi
<i>NameIdentifier</i> <i>IDtypeName</i>	<i>identifier</i>	024 7#  2 Fonte (autoridade)
<i>NameIdentifier</i> <i>IDvalue</i>	<i>identifier</i>	024 7#  a
<i>PersonNameInverted</i>	<i>author</i> <i>creator</i>	100 __  a 700 __  a
<i>ContributorStatement</i> <i>PersonName</i>	<i>potentialAction</i>	245 __  c

<i>Contributor</i> <i>SequenceNumber</i> <i>ContributorRole</i>	<i>author</i> <i>creator</i> <i>about/_mainEntity</i> <i>contributor</i> <i>illustrator</i> <i>translator</i>	100 __  a  e  4 700 __  a  e  4
<i>PersonDate</i>	<i>Person/ birthDate</i> <i>Person/ deathDate</i>	100 __  d (autoridade)
<i>Affiliation</i>	<i>Person/ affiliation</i>	100 __  u (autoridade)
<i>CorporateName</i>	<i>sourceOrganization</i>	110 __  a
<i>ConferenceName</i>	<i>publication</i> <i>PublicationEvent</i>	111 __  a
<i>ConferenceNumber</i>	<i>recordedAt</i>	111 __  n
<i>ConferenceDate</i>	<i>releasedEvent</i>	111 __  d
<i>ConferencePlace</i>	<i>releasedEvent</i> <i>PublicationEvent</i>	111 __  c
<i>TitlePrefix</i>	<i>name</i>	245 __  a (Ind 2)
<i>TitleText</i> <i>ProductFormDetail</i> <i>TitleType</i>	<i>name</i>	245 __  a 222 __  a 246 13  a 246 30  a 242 1_  a 201 __  a 247 10  a 246 18  a 246 14  a 246 1_  a
<i>Subtitle</i>	<i>alternativeHeadline</i>	245 __  b
<i>TitleType</i> <i>TitleText</i> <i>01 Distinctive Title (book)</i>	<i>alternativeHeadline</i>	246 _2  a
<i>EditionNumber</i>	<i>bookEdition</i>	250 __  a
<i>Produc SupportingResource</i>	<i>publicationVolume</i>	--
<i>PartNumber</i>	<i>volumeNumber</i>	--
<i>CityOfPublication</i>	<i>pubPlace</i>	260 __  a
<i>PublisherName</i>	<i>publisher</i>	260 __  b
<i>PublicationDate</i>	<i>datePublished</i>	260 __  c
<i>CopyrightYear</i>	<i>copyrightYear</i>	008/07-10

<i>CountryOfPublication</i> <i>RegionCode</i>	<i>Place/ Address/ Country</i>	008/15-17 270 __  d
<i>NumberOfPieces</i>	<i>numberOfPages</i>	300 __  a 1 recurso online ( p.)
<i>ProductFormDetail</i> <i>Daisy, EPUB, OEB, DOC,</i> <i>DOCX, HTML, ODF, PDF,</i> <i>PDF/A, RTF, TXT</i> <i>Amazon Kindle</i>	<i>bookFormat</i>	300 __  b 347 __  a  b
<i>Collection/ Title without prefix</i>	<i>isPartOf</i>	490 __  a
<i>NumberWithinSeries</i> <i>YearOfAnnual</i>	<i>position</i>	490 __  v
<i>Text</i> <i>AncillaryContentDescription</i>	<i>text</i>	500 __  a
<i>Text</i> <i>TExtTypeCode</i>	<i>description</i>	505 0_  a
<i>Main description</i> <i>TextSourceCorporate</i>	<i>description</i>	520 __  a
<i>BiographicalNote</i>	<i>Person/ hasOccupation</i> <i>jobTitle</i> <i>Person/ description</i>	545 0_  a
<i>PrizeName</i> <i>PrizeYear</i>	<i>award</i>	586 __  a
<i>AudienceCode</i>	<i>audience</i> <i>typicalAgeRange</i>	008/22 = 'g': adultos 008/22 = 'j': crianças e jovens 008/22 = 'd': jovens adultos 008/22 = 'e': ensino superior e educação de adultos 008/22 = 'f': ensino profissionalizante
<i>AudienceCode</i> <i>AudienceRangeValue</i> <i>AudienceDescription</i>	<i>contentRating</i>	521 __  a
<i>SubjectHeadingText</i>	<i>about</i> <i>subjectOf</i>	650 _4  a
<i>NameAsSubject</i>	<i>character</i>	600 _4  a
<i>PlaceAsSubject</i>	<i>contentLocation</i> <i>spatialCoverage</i>	651 _4  a
<i>WebsiteLink</i>	<i>resource</i>	856 41  u  y
<i>SupportingResource (Capa)</i>	<i>Thing/ CreativeWork/</i> <i>MediaObject/ ImageObject</i>	--

Fonte: elaborado pela autora a partir de LOC, 2018a, SCHEMA, 2018 e EDITEUR, 2012.

Uma *Appliance Program Interface* (API) para exportação somente destes campos deve ser preparada para que os dados sejam coletados do ONIX para o modelo, facilitando o processo de entrada e atualização dos livros digitais definidos pelos fornecedores.

### 5.3.2 Inclusão de dados bibliográficos por planilha de dados

Abaixo é apresentada sugestão de sequência de planilha de dados para provimento dos metadados dos fornecedores. Esta planilha não pode ser alterada, pois esta ação compromete a carga dos dados. A planilha deve ser disponibilizada aos fornecedores sem a possibilidade de inclusão, edição ou exclusão de colunas. A ordenação das colunas na planilha de dados é fundamental para que os dados sejam coletados, compreendidos e processados. A planilha contempla a descrição de dados de registros bibliográficos e de licenciamento.

Os dados alimentados na planilha de dados devem ser convertidos ao *SchemaBook*. Uma vez no modelo os metadados dos livros digitais podem ser importados para o catálogo *online*, disponibilizando acesso aos usuários. Para tanto é necessário converter os dados do *SchemaBook* ao MARC 21, tornando-os compatíveis com os sistemas de bibliotecas. O mapeamento dos dados bibliográficos das planilhas de dados para o *SchemaBook* e deste para o MARC 21 está especificado na **Tabela 9**. Nem todos os dados do *SchemaBook* possuem correlação com o MARC 21. Os campos e a sequência dos metadados de licenciamento estão descritos na **Tabela 10**.

**Tabela 9:** Mapeamento – Dados bibliográficos – Planilha de dados – SchemaBook – MARC 21

Coluna da planilha	<i>SchemaBook</i>	MARC 21
Idioma	<i>inLanguage</i>	008/35-37 código do idioma
Tipo de recurso (este campo é de preenchimento automático para todos os registros, visto que somente conteúdo digital está no escopo)	<i>typeOfResource</i>	008/23: s (eletrônico) ou o (online) 245 __  h [recurso eletrônico]
ISBN	<i>identifier &gt; isbn identifier</i>	020 __  a
ISBN versão impressa*	<i>BookFormatType &gt; identifier</i>	776 __  z 530 __  a
DOI	<i>identifier</i>	024 7#  a  2 doi
Autorias* (pessoa), função de autor(es) e demais dados de identificação	<i>author creator about/ <u>mainEntity</u> contributor person (editor)</i>	100 __  a,  d 700 __  a,  d,  e

	<i>person (illustrator)</i>	
Entidade coletiva	<i>organization</i>	110 __  a 710 __  a
Dados sobre eventos	<i>publication</i> <i>PublicationEvent</i>	111 __  a 711 __  a
Título	<i>name</i>	245 __  a
Subtítulo	<i>alternativeHeadline</i>	245 __  b
Edição*	<i>bookEdition</i>	250 __  a
Obra volumada (campo lógico Sim / Não)	<i>publicationVolume</i>	--
Volume	<i>volumeNumber</i>	--
Local de publicação*	<i>pubPlace</i>	260 __  a
Editora*	<i>publisher</i>	260 __  b
Data de publicação	<i>datePublished</i> <i>sdDatePublished</i>	260 __  c
Data de copyright	<i>copyrightYear</i>	260 __  c ©
País de publicação	<i>Place/ address/ Country</i>	270 __  d
Número de páginas*	<i>numberOfPages</i>	300 __  a
Formato do arquivo	<i>bookFormat</i>	300 __  b 347 __  b
Série	<i>isPartOf</i>	490 __  a
Volume da série	<i>position</i>	490 __  v
Nota geral	<i>text</i>	500 __  a
Nota de conteúdo	<i>description</i>	505 __  a
Nota de resumo	<i>Description</i>	520 __  a
Dados biográficos	<i>Person/ hasOccupation</i> <i>jobTitle</i> <i>Person/ description</i>	545 0_  a
Nota de premiação	<i>award</i>	586 __  a
Nota de edição impressa	<i>text</i>	530 __  a
Público alvo	<i>audience</i> <i>typicalAgeRange</i> <i>contentRating</i>	008/22 = 'g': adultos 008/22 = 'j': crianças e jovens 008/22 = 'd': jovens adultos 008/22 = 'e': ensino superior e educação de adultos 008/22 = 'f': ensino profissionalizante
Assuntos*	<i>about</i> <i>subjectOf</i>	650 _4  a
Local como assunto	<i>contentLocation</i>	651 _4  a

	<i>spatialCoverage</i>	
URL para acesso*	<i>resource</i>	856 41  u  y
Capa	<i>MediaObject / ImageObject</i>	--

Fonte: elaborado pela autora a partir de LOC, 2018a e SCHEMA, 2018.

O mapeamento apresentado na **Tabela 9** priorizou a presença de campos que permitam a descrição em profundidade no formato MARC 21, minimizando a necessidade de ajustes posteriores pela biblioteca. Este mapeamento também permite a transmissão da capa do livro, contribuindo com sua identificação visual no catálogo. Os formatos ISO2709 e MARC 21 e o protocolo Z39.50 não permitem o trânsito de arquivos de imagens junto aos metadados. Ao incluir a capa aos metadados, o modelo já proporciona ganhos às bibliotecas.

Nem todos os campos são de preenchimento mandatório, dependendo do tipo de recurso e das informações disponíveis. Os campos assinalados com asterisco (\*) são os que foram identificados em planilhas de dados e que necessitam de tratamento adicional de forma a acomodar as informações nas *tags* e subcampos apropriados, sem criar muita complexidade de preenchimento aos fornecedores. As regras definidas para este conjunto estão descritas no Capítulo 6.

Este mapeamento visa ser de aplicação genérica. É recomendado que os campos de datas e demais campos que necessitam de formato especial para identificação dos dados trabalhem com máscaras, minimizando a incidência de erros na inclusão de registros.

### 5.3.3 *Dados de licenciamento*

Além dos dados bibliográficos, o modelo também permite a troca de dados de comercialização dos livros digitais entre fornecedores e bibliotecas. Os dados de licenciamento devem estar presentes na planilha de dados e nos dados oriundos do ONIX. Não é realizado um mapeamento para o MARC 21 pois os dados de contratação de licenciamento não são registrados neste formato. Os dados de licenciamento fazem parte da planilha de dados na sequência em que estão apresentados, porém foram reunidos nesta tabela para permitir uma visualização das conversões necessárias.

Os elementos para descrição de dados de licenciamento foram identificados no *Schema.org* (SCHEMA, 2018) e no formato ONIX, que possuem diversas especificações, de acordo com o tipo de dado que será descrito. Para atender as características do tipo de dado, outros *Schemas* foram utilizados como as especificações de *CreativeWork*, *Action*, *Price*,

*Intangible*, *Organization* e *Product*. Para o mapeamento foram utilizadas as especificações ONIX *for Books* (EDITEUR, 2012), *Book Trade Order* (EDITEUR, 2014), *Sales and Sales Tax Report* (EDITEUR, 2008), *ONIX Best practices* (EDITEUR, 2018a) e *Publications licenses* (ONIX-PL) (EDITEUR, 2008). A **Tabela 10** mapeia os dados de licenciamento com o *Schema*, o formato ONIX e os campos presentes na planilha de dados.

**Tabela 10:** Mapeamento - Dados de licenciamento – Planilha de dados - Schema - ONIX

<b>Coluna da planilha</b>	<b>Schema</b>	<b>ONIX</b>
Moeda	<i>Thing/ Action/ TradeAction/ BuyAction/ priceCurrency</i>	<i>Sales/ B.144 CurrencyCode</i>
Preço	<i>Thing/ Action/ TradeAction/ BuyAction/ price</i>	<i>Sales/ B.126 Price</i>
Modelo de negócio (relacionado com código do cliente)	<i>Thing/ PriceSpecification/ additionalType</i>	<i>ONIX-PL/ 2. ExpressionType</i>
Uso concorrente? (campo lógico Sim / Não)	<i>Thing/ Intangible/ PriceSpecification/ eligibleTransactionVolume</i>	<i>Onix-PL/ 19. UsageQuantity</i>
No. acessos simultâneos (relacionado com código do cliente)	<i>Thing/ Intangible/ PriceSpecification/ eligibleQuantity</i>	<i>ONIX-PL/ UsageQuantityType/ NumberOfSimultaneousUsers</i>
Disponibilidade	<i>Thing/ Action/ TradeAction/ BuyAction/ actionStatus</i>	<i>ONIX for books/ P.26 ProductAvailability</i>
Data inclusão	<i>Thing/ Creative Work/ Book/ dateCreated</i>	<i>Best practices/ h.15 SentDateTime</i>
Data última alteração (plataforma)	<i>Thing/ CreativeWork/ dateModified</i>	<i>ONIX for books/ 17 LastUpdated</i>
Data prevista remoção	<i>Thing/ Intangible/ Offer/ eligibleDuration</i>	<i>ONIX for books/ 15 UntilDate</i>
Data remoção	<i>Thing/ Intangible/ Offer/ availabilityEnds</i>	<i>ONIX for books/ 32 AssociatedEndDate</i>
Biblioteca (identificada pelo código do cliente)	<i>Thing/ Organization/ schemacustomer</i>	<i>Sales/ H.101 BuyerParty</i>
Data licenciamento (relacionado com código do cliente)	<i>Thing/ Product/ purchaseDate</i>	<i>Sales/ B.123 IssueDateTime</i>
Data vigência licenciamento (relacionado com código do cliente)	<i>Thing/ Action/ TradeAction/ BuyAction/ startTime</i>	<i>Best practices/ P.24.11 StartDate</i>
Data término licenciamento (relacionado com código do cliente)	<i>Thing/ Action/ TradeAction/ BuyAction/ endTime</i>	<i>Best practices/ P.24.12 EndDate</i>
Situação livro	<i>Thing/ Intangible/ Offer/ availability</i>	<i>ONIX-PL/ 34 LicenseStatus</i>



Data atualização dos dados no <i>dataset</i>	<i>Thing/ Intangible/ DataFeedItem/ dateModified</i>	<i>ONIX for books/ 17 LastUpdated</i>
Fornecedor	<i>Thing/ Action/ TradeAction/ BuyAction/ price</i>	<i>Sales/ H.77 SupplierParty</i>
Pedido	<i>Thing/ Intangible/ Order/ orderedItem</i>	<i>Book Trade Order/ Header/ 1 OrderNumber</i>
Data pedido	<i>Thing/ Intangible/ Order/ orderDate</i>	<i>Book Trade Order/ Header/ 2 IssueDateTime</i>

Fonte: Elaborado pela autora a partir de SCHEMA, 2018, EDITEUR, 2008, 2012, 2014, 2018, 2018a.

Após os mapeamentos realizados, a conversão ferramental é encerrada. A **Tabela 11** reúne as especificações definidas para o modelo nesta etapa.

**Tabela 11:** Definições para conversão ferramental

Atividades	Definições
Conversão ferramental	Fornecedores com dados bibliográficos em ONIX: mapeamento ONIX- <i>SchemaBook</i>
	Fornecedores com dados bibliográficos em planilha de dados: modelo de planilha de dados - <i>SchemaBook</i>
	Mapeamento dos dados de licenciamento para a planilha de dados, <i>SchemaBook</i> e ONIX
	Bibliotecas: mapeamento <i>SchemaBook</i> – MARC 21. Além do MARC 21, os dados podem ser disponibilizados em outros formatos como MARC XML, ISO2709 ou ainda em <i>SchemaBook</i> ou RDF, caso a biblioteca já utilize estes modelos, dispensando a conversão

Fonte: elaborado pela autora.

A **Tabela 12** apresenta as boas práticas da W3C (LÓSCIO; BURLE; CALEGARI, 2017) que foram identificadas nesta etapa.

**Tabela 12:** Boas práticas da W3C na etapa de Conversão ferramental

No. Boas práticas	Descrição
14	Forneça os dados em múltiplos formatos
17	Forneça o <i>download</i> do conjunto completo
18	Forneça o <i>download</i> de subconjuntos para grandes conjuntos de dados
19	Use a negociação de conteúdo para oferecer dados disponíveis em múltiplos formatos

Fonte: elaborado pela autora a partir de LÓSCIO; BURLE; CALEGARI, 2017.

#### 5.4 Processo de recuperação

Uma vez que os dados estão aptos para serem publicados na Web são definidas as formas de distribuição para consumo de humanos e máquinas. O SPARQL *EndPoint* é uma ferramenta de busca utilizada com frequência em projetos de Web Semântica, resolvendo a distribuição dos dados para máquinas (SANTARÉM SEGUNDO, 2018).

Para humanos pode ser disponibilizada uma interface para consulta, ou ainda para a visualização de todos os dados. Os dados do *dataset* devem ser passíveis de exportação em outros formatos, com possibilidade de exportar a quantidade desejada de dados, ou ainda sua totalidade. Para o modelo são necessárias duas interfaces para humanos: uma para os fornecedores e outra para as bibliotecas, visto que possuem funcionalidades diferentes. Os *datasets* devem ser disponibilizados de forma que alguns livros digitais ou todo o conjunto sejam consumidos. A cada publicação ou atualização do *dataset* a versão do mesmo deve ser identificada, indicando a data da última alteração.

A interface deve permitir a realização de buscas por estratégias diversas, facilitando a navegação do bibliotecário e do fornecedor por entre os livros digitais disponibilizados. Critérios como datas (alteração do livro digital, disponibilização no OPAC, data do último acesso, data de vencimento do licenciamento etc.), nome do fornecedor, biblioteca, forma de contratação, dados bibliográficos, preço etc. podem ser utilizados para filtrar somente os livros digitais desejados. As notificações também são passíveis de buscas, refinando a informação por tipo de notificação recebida, seu conteúdo ou data.

É possível encaminhar pedidos pelo modelo, o que pode contribuir com agilidade na contratação de livros digitais. Além de poder licenciar um livro digital pelo modelo, também podem ser encaminhadas solicitações de alterações dos modelos de licenciamento, visando a contratação por meio de estratégias mais econômicas quando comparados os valores do livro por aquisição perpétua ou assinatura e os custos por modelos transitórios de livros digitais que possuem alta procura dos usuários. A **Tabela 13** reúne as especificações definidas para o modelo nesta etapa. Na **Figura 10** corresponde ao passo 5.

**Tabela 13:** Definições para a publicação e recuperação dos dados

<b>Atividades</b>	<b>Definições</b>
Ferramenta para recuperação	SPARQL <i>EndPoint</i>
Apresentação dos dados	Legíveis para máquinas e pessoas
Disponibilização dos dados	Disponibilizar dados para <i>download</i> e em mais de um formato
Envio de pedidos	Possibilidade de solicitar pedidos aos fornecedores diretamente pelo modelo

Fonte: elaborado pela autora.

Para esta etapa foram identificadas as boas práticas da W3C (LÓSCIO; BURLE; CALEGARI, 2017) atendidas (**Tabela 14**):

**Tabela 14:** Boas práticas da W3C na etapa de Processo de recuperação

No. Boas práticas	Descrição
11	Atribua URIs às versões e séries dos conjuntos de dados
17	Forneça o <i>download</i> do conjunto completo
18	Forneça o <i>download</i> de subconjuntos para grandes conjuntos de dados
23	Torne os dados acessíveis por meio de API
24	Use padrões da Web como base para as APIs
25	Forneça a documentação completa para a API
26	Evite alterações que quebrem a API

Fonte: elaborado pela autora a partir de LÓSCIO; BURLE; CALEGARI, 2017.

### 5.5 Marketing e feedback

Após a conversão ferramental os dados estão aptos para serem publicados e disponibilizados na Web. O modelo não define qual ferramenta deve ser utilizada para a publicação, cabendo ao responsável pelo projeto a seleção. Os dados do licenciamento devem ser armazenados em ambiente que proporcione segurança para fornecedores e bibliotecas.

Uma vez os dados publicados, a atualização dos mesmos deve ser constante e realizada de forma automática, identificando os livros digitais que foram alterados, acrescentados ou removidos. No caso dos livros digitais removidos, para que a biblioteca não perca controle dos livros digitais que contratou (mesmo que o licenciamento tenha terminado), os mesmos devem ser sinalizados como ‘Removidos’, orientando a biblioteca para inibi-los no OPAC. Ao remover um livro digital de seu *dataset* o fornecedor deve informar o motivo da remoção. Um dos pontos principais que motivaram esta pesquisa é exatamente permitir que o fornecedor disponibilize dados atualizados a seus clientes (bibliotecas). Esta ação proporcionará às bibliotecas a identificação de livros digitais que foram incluídos na plataforma de cada fornecedor, assim como aqueles que foram removidos ou que serão removidos, visto que alguns fornecedores informam com antecedência sobre a retirada dos livros digitais de um pacote.

Os dados disponibilizados no *dataset* devem ser atualizados e com ocorrência de comunicação entre fornecedores e bibliotecas em tempo real. Esta comunicação é estabelecida por API, conectando os *datasets* ao modelo.

As notificações estabelecem a comunicação entre fornecedores e bibliotecas. As principais notificações a serem implementadas são:

- a) Alterações realizadas pelo fornecedor informando os livros digitais que foram incluídos, alterados, removidos ou que ficarão indisponíveis em breve;

- b) Informar fornecedor que os dados foram alterados pela biblioteca, com possibilidade de visualizar e atualizar os dados presentes no *dataset* aprovando ou não a alteração realizada;
- c) Informar alterações sobre os licenciamentos.

As notificações foram concebidas para utilizar o protocolo *Data Notifications* (LDN), que está em desenvolvimento pelo LD4P2. É necessário acompanhar as iniciativas sobre esta funcionalidade para avaliar se sua implementação será viável. Os dados de notificações (*NotificationType*) presentes no ONIX podem fornecer informações para sinalização de alterações nos livros digitais do fornecedor que será visto adiante, na etapa Marketing e *feedback*. Os tipos de notificações presentes no ONIX são:

- 01 -- early notification*
- 02 -- advance notification (confirmed)*
- 03 -- notification confirmed*
- 04 -- update (partial)*
- 05 -- delete*
- 08 -- notice of sale*
- 09 -- notice of acquisition*

Estas notificações podem ser utilizadas para comunicar no modelo quais livros digitais são novos (01 ou 09), quais serão removidos (02) e quais foram removidos (05). É preciso avaliar o emprego das notificações presentes no ONIX e acompanhar o processo para que a LDN seja uma recomendação da W3C. Na **Figura 10** esta etapa é a 9.

O *Data Catalog Vocabulary* (DCAT) é uma recomendação da W3C e trata-se de um vocabulário RDF construído para facilitar a interoperabilidade entre catálogos publicados na Web. Ao utilizar este vocabulário editores aumentam as possibilidades de encontrabilidade de seus livros digitais, facilitando a consulta a múltiplos catálogos. Esta ferramenta pode ser utilizada por fornecedores para aumentar a descoberta de publicações e o consumo de dados. Novos estudos podem ser realizados para avaliar a possibilidade de utilização deste vocabulário nos *datasets* dos fornecedores.

Como um dos objetivos do modelo é proporcionar a geração de dados estatísticos e indicadores, a utilização do protocolo SUSHI e do Código de práticas COUNTER permitirá que bibliotecas e fornecedores tenham autonomia para coletar e utilizar os dados de uso dos livros digitais. O SUSHI não permite que a coleta de dados seja realizada de diversos

provedores de uma única vez. Entretanto, o modelo pode mandar requisições aos *datasets* dos provedores com os quais a biblioteca trabalha para coletar e reunir os dados, facilitando a geração de estatísticas (NISO, 2009).

O *dataset* deve ser disponibilizado em mais de um formato de dados, proporcionando flexibilidade aos desenvolvedores caso sejam feitas integrações com outros sistemas. Os formatos de serialização foram definidos no item 5.2.

A **Tabela 15** reúne as especificações definidas para o modelo nesta etapa.

**Tabela 15:** Definições das estratégias para marketing e *feedback*

<b>Atividades</b>	<b>Definições</b>
Publicação dos <i>datasets</i> para consumo de bibliotecas	Ferramenta para publicação de <i>datasets</i>
Retroalimentação	APIs desenvolvidas para atualização dos dados em tempo real <i>Data Catalog Vocabulary</i> (DCAT)
Notificações	Protocolo <i>Data Notifications</i> (LDN)
Dados estatísticos de uso	SUSHI, Código de práticas COUNTER

Fonte: elaborado pela autora.

Na **Tabela 16** constam as boas práticas da W3C (LÓSCIO; BURLE; CALEGARI, 2017) atendidas nesta etapa.

**Tabela 16:** Boas práticas da W3C na etapa de Marketing e *feedback*

<b>No. Boas práticas</b>	<b>Descrição</b>
14	Forneça os dados em múltiplos formatos
19	Use a negociação de conteúdo para oferecer dados disponíveis em múltiplos formatos
20	Forneça acesso em tempo real
21	Forneça dados atualizados
22	Forneça uma explicação para os dados que não estão disponíveis
29	Obtenha o <i>feedback</i> dos consumidores
30	Torne o <i>feedback</i> disponível
32	Forneça apresentações complementares
33	Forneça <i>feedback</i> para o publicador original

Fonte: elaborado pela autora a partir de LÓSCIO; BURLE; CALEGARI, 2017.

## 5.6 *Qualidade e preservação*

A disponibilização de dados em ambientes abertos permite expansão do conhecimento com o compartilhamento de fontes de dados de forma gratuita. Porém a não adoção de padrões e aderência a boas práticas pode resultar em obstáculos para compreensão dos dados

disponibilizados (WU et al., 2017). Quando se fala em qualidade é necessário especificar quais critérios nortearão o nível esperado e que aspectos dos dados devem ser analisados e controlados. De acordo com Melo, Botega e Santarém Segundo:

Deste modo, tratar e detectar problemas de qualidade é crucial para o sucesso das aplicações baseadas em Web Semântica, visto que quando não avaliada, os problemas são descobertos apenas quando as aplicações falham ou as buscas nos conjuntos de dados retornam resultados incorretos (MELO; BOTEAGA; SANTARÉM SEGUNDO, 2017, p. 81).

Mesmo com a atividade de *Data cleaning* realizada na primeira etapa para publicação, são necessárias ações constantes para assegurar a integridade e qualidade dos *datasets*. Dados imprecisos podem ser causados por diversos fatores, como utilização de dados desatualizados, modelagem ruim e até mesmo a presença de erros criados de forma proposital (WU et al., 2017). Beek et al. (2014) discorrem que mesmo *datasets* que são limpos em sua criação podem ficar sujeitos com o decorrer do tempo. Aspectos da qualidade dos dados é analisada por Rietveld:

A qualidade baixa de nuvens LOD exige considerável esforço de cientistas da Web Semântica e desenvolvedores para consumo. Na prática, isso significa que o LOD é menos efetivamente (re)utilizado do que poderia e deveria (RIETVELD, 2016, p. 19, tradução nossa)<sup>40</sup>.

Um *dataset* deve ser constantemente monitorado e corrigido. Os problemas que podem ser identificados são dados com conteúdos em padrões diferentes, redundância e duplicidade de registros, nodos em branco, conferência das hierarquias de classes, dados imprecisos, erros sintáticos, URIs quebradas etc. O RDF permite a publicação de dados legíveis para pessoas e máquinas. Entretanto, ele pode apresentar problemas, como a desordenação de triplas, o que exige maior processamento dos computadores para compreender o que foi publicado, além de poder resultar em ilegibilidade para humanos (WU et al., 2017).

A redundância de dados pode ser explicada pela ampla oferta de fontes de dados na Web, que permite que um mesmo dado seja coletado de *datasets* diferentes. Além de incluírem dados duplicados, redundâncias semânticas também são observadas, com informações excedentes para representar um mesmo conceito ou valor (WU et al., 2017). Critérios para

---

<sup>40</sup> *The poor data quality of the LOD clouds requires considerable effort from SW scientists and developers to consume. In practice, this means that LOD is less effectively (re)used than it should and could be.* (RIETVELD, 2016, p. 19).

identificar com segurança casos de duplicidades são complexos e sua aplicação pode comprometer o *dataset* caso dados sejam removidos erroneamente.

Diversas ferramentas podem ser empregadas para limpar *datasets* como *LOC Washing Machine*, *LOD Laundromat*, *TRank*, *TripleRank* etc. A *LOD Laundromat*<sup>41</sup> informa em seu sítio que permite que todos os LODs publicados no mundo sejam acessados por meio do *crawling LOD WashingMachine*, disponibilizado de forma gratuita. A *TRank* analisa as hierarquias de classes de entidades em RDF, porém não as triplas. Essas, por sua vez, podem ser analisadas com a *TripleRank* (RIETVELD, 2016). O *Shapes Constraint Language* (SHACL) é uma ferramenta que pode ser utilizada para o exemplo de validação de grafos e é uma recomendação da W3C (KNUBLAUCH; KONTOKOSTAS, 2017).

Outro aspecto destacado por Wu et al. (2017) é relacionado com o enorme volume de dados disponíveis, o que pode tornar complexo o controle de qualidade. Com o consumo de dados provenientes de diversos *datasets*, caso estes estejam modelados de forma diferente, podem ocorrer as redundâncias e duplicidades.

De acordo com Melo, Botega e Santarém Segundo (2017) a partir de modelo proposto para avaliação de qualidade para dados conectados, as dimensões a serem analisadas são: interligação (*interlinking*), consistência, completude, licenciamento, avaliação temporal, precisão sintática e semântica. Estas dimensões são analisadas de acordo com métricas, descritas na **Tabela 17**.

**Tabela 17:** Dimensões e métricas de qualidade para dados conectados

Dimensões	Métricas	Características
Interligação	Detectar <i>links</i> de boa qualidade	Não utilizar detalhes de implementação, não utilizar identificadores numéricos para recursos
	Detectar existência de <i>links</i> para fontes externas	Ter ao menos 50 <i>links</i> para fontes externas
Licenciamento	Identificar licença na documentação do conjunto de dados	Incluir documentação sobre a licença para utilização dos dados junto com o conjunto disponível
	Especificar a licença	Definir o modelo de licença adotado e possibilidade de conferir sua aplicação
	Licença aberta para o conjunto de dados	Permitir que os registros sejam publicados sob licença aberta

<sup>41</sup> <http://lodlaundromat.org/>

Consistência	Identificar alinhamento dos dados com a ontologia adotada	Utilizada principalmente em <i>datasets</i> que utilizam diversos vocabulários para identificar a presença de dados não compatíveis
	Verificar tipo de dado permitido para a propriedade	Identificar se os atributos das propriedades que possuem tipo pré-determinado de dados estão aderentes com os valores permitidos
Precisão sintática	Identificar uso de regras sintáticas	Validar os dados permitidos por tipo de campo
Precisão semântica	Identificar o uso de propriedades inexistentes ou não definidas	Confirmar se as classes e propriedades utilizadas são as presentes nos vocabulários em uso. Verificar a descrição de recursos com valores literais
Compleitude	Esquema	Analisar se a ontologia utiliza propriedades e classes suficientes para o domínio do <i>dataset</i>
	Propriedade	Realizar cálculo do número de propriedades utilizadas da ontologia adotada
	População	Realizar cálculo do número de classes utilizadas em relação a quantidade existente
	No. de metadados necessários	Conferir os dados disponibilizados nos registros
	Atributos prioritários	De acordo com o domínio, identificar quais são os metadados de preenchimento mandatório ou essenciais e a proporção em que foram preenchidos
Avaliação temporal	Analisar atualização dos dados	Conferir se os dados são atuais em relação ao contexto que será utilizado e analisar a frequência com que são atualizados

Fonte: MELO; BOTEGA; SANTARÉM SEGUNDO, 2017.

A preservação dos dados é compreendida como parte da qualidade e visa garantir que os metadados presentes nos *datasets* estejam disponíveis para consulta e utilização em longo prazo. Para a preservação de *datasets* deve ser traçada uma política de preservação, considerando as especificações e documentações do *Schema* adotado, ontologias empregadas, RDF etc. Todas as linguagens utilizadas devem possuir especificação e estes dados não devem ficar restritos ao *namespace* presente nos arquivos (BOUNTOURI; GRATZ; SANMARTIN, 2017).

Outro aspecto a ser considerado para a preservação é a indicação da origem dos dados, assim como a sua versão, com os dados representados de acordo com padrões apropriados, com a “coleta e a atribuição de informações de representação necessárias para o entendimento do dado e para sua apresentação [...]” (SAYÃO; SALES, 2012, p. 185). De acordo com Bountouri,



Gratz e SanMartin (2017, p. 370, tradução nossa), “uma das informações semânticas mais significativas para documentar o ciclo de vida do recurso digital é a procedência do metadado”<sup>42</sup>.

A ontologia PREMIS<sup>43</sup>, é um padrão internacional para preservação (LIBRARY OF CONGRESS, 2018b) e será adotada para o modelo.

Na **Figura 10** esta etapa está representada no item 10, com a centralização dos dados na biblioteca. O envio de dados selecionados (licenciados) ao catálogo é o item 11, enquanto o item 12 simboliza a possibilidade de envio de pedidos aos fornecedores.

A **Tabela 18** reúne as especificações definidas para o modelo nesta etapa.

**Tabela 18:** Definições para controle de qualidade e preservação

<b>Atividades</b>	<b>Definições</b>
Qualidade dos dados	Aplicação de dimensões e métricas para aferir a qualidade, de acordo com o modelo proposto para avaliação de qualidade para dados conectados descrito por Melo, Botega e Santarém Segundo (2017). Como existem diversas ferramentas para aferir a qualidade de <i>datasets</i> , a seleção da ferramenta adequada será realizada pelo responsável pelo projeto
Preservação	PREMIS

Fonte: elaborado pela autora.

A **Tabela 19** identifica as boas práticas da W3C (LÓSCIO; BURLE; CALEGARI, 2017) que foram atendidas.

**Tabela 19:** Boas práticas da W3C na etapa de Qualidade e preservação

<b>No. Boas práticas</b>	<b>Descrição</b>
6	Forneça informações sobre a qualidade dos dados
7	Forneça um indicador da versão
8	Forneça o histórico de versões
11	Atribua URIs às versões e séries dos conjuntos de dados
27	Preserve indicadores
32	Forneça apresentações complementares
34	Siga os termos do licenciamento
35	Cite a publicação original

Fonte: elaborado pela autora a partir de LÓSCIO; BURLE; CALEGARI, 2017.

<sup>42</sup> *One of the most significant parts of the semantic information needed to document the life cycle of a digital resource is the provenance metadata* (BOUNTOURI; GRATZ; SANMARTIN, 2017, p. 370).

<sup>43</sup> <http://id.loc.gov/ontologies/premis-3-0-0.html>

Com a definição do modelo inicia-se o exemplo de validação do mesmo, verificando se sua aplicação é factível. O exemplo de validação será realizado a seguir, no Capítulo 6.

## 6 Exemplo de validação do modelo

---

### Preâmbulo

Este capítulo discorre sobre o exemplo de validação do modelo proposto, visando identificar se o processo foi completamente mapeado, garantindo a troca de registros entre fornecedores e bibliotecas.

Para o exemplo de validação do modelo foi realizado um processo de simulação a partir da seleção de um registro de um fornecedor de livro digital licenciado com atuação no Brasil, recebido em planilha de dados. São mostrados os dados originais recebidos e após tratamento para padronização. Primeiramente foram identificadas as informações incompletas ou preenchidas fora de padrão de acordo com o tipo de dado presente na coluna na planilha de dados. A ferramenta *OpenRefine* foi utilizada para o *data cleaning*.

Em seguida foi feito o mapeamento para o *Schema* e para o MARC 21, com inclusão de URIs identificadas em *datasets*. Para o *data lifting* foram selecionados os *datasets* VIAF e *GeoNames*, com o intuito de padronização e enriquecimento de autoridades do tipo Pessoa e informações do Local de publicação e País. Outras informações foram obtidas por meio do enriquecimento de dados, como a capa da publicação, correção da URL, relação entre os autores e o registro etc.

As transformações dos dados originais ao padrão RDF são sinalizadas em tabela de acordo com suas propriedades, classes e subclasses buscando simular o processo de tratamento e conversão dos dados desejados com o modelo.

O exemplo de validação do modelo visa verificar se o mesmo atende aos objetivos desejados e se as funcionalidades necessárias que foram identificadas estão contempladas, atendendo aos principais interessados (*stakeholders*): fornecedores e bibliotecas. Para as bibliotecas, os principais afetados são os responsáveis pela aquisição e os catalogadores. Aos fornecedores o objetivo é proporcionar maior qualidade aos registros bibliográficos. Também busca o estabelecimento de comunicação entre os *stakeholders*, de forma a tornar as operações transparentes, notificando sobre alterações dos livros digitais e formas de licenciamento.

Para simular a aplicação do modelo foi convertido um registro apresentando os dados originais e, após o processo de publicação finalizado, as partes tratadas, de acordo com o dado que é representado.

Os dados originais, recebidos em planilha de dados, foram obtidos junto a um fornecedor de conteúdo digital de origem internacional e com atuação no Brasil. Nem todos os campos foram preenchidos pelo fornecedor. Conforme explicado no Capítulo 5, alguns campos em planilhas de dados, sinalizados com asterisco, necessitam de tratamento específico. Estes campos estão relacionados abaixo.

- a) Autorias: autores Pessoa para entrada principal e secundárias (**Tabela 20**).

**Tabela 20:** Regras de descrição de Autorias

Regras de descrição	Diversas ocorrências podem ser incluídas nesta coluna, separando os autores com ponto e vírgula (;), resultando em uma <i>tag</i> para cada nome. Caso o autor exerça uma função específica na obra, é identificado como secundário (700  a) e a função desempenhada deve ser preenchida entre parênteses, após o nome (700  e). O primeiro nome é cadastrado na <i>tag</i> 100 caso não exerça função específica ou que estejam presentes até três autores exercendo a mesma função. Caso existam mais de 3 autores exercendo a mesma função, a <i>tag</i> 100 fica em branco e todos os autores são informados em <i>tags</i> 700. Os demais autores são cadastrados em <i>tags</i> 700 com a indicação da função no  e. Caso não exista indicação de função os autores são identificados como 'autor'. O nome dos autores deve ser incluído na <i>tag</i> 245 c em ordem direta, seguidos do nome das funções desempenhadas na obra.
Dados originais	Silva, José João da; Almeida, Luciana de; Campos, Arthur; Pacheco, Renato de Souza (tradutor); Travassos, Regina de Pontes (ilustrador)
MARC 21 após tratamento	100 1_  a Silva, José João da  e autor 245 10  c José João da Silva, Luciana de Almeida, Arthur Campos (editor), Renato de Souza Pacheco (organizador), Regina de Pontes Travassos (ilustrador) 700 1_  a Almeida, Luciana de  e autor 700 1_  a Campos, Arthur  e autor 700 1_  a Pacheco, Renato de Souza  e tradutor 700 1_  a Travassos, Regina de Pontes  e ilustrador

Fonte: elaborado pela autora.

b) Edição: detalhamento dos dados de edição (**Tabela 21**).

**Tabela 21:** Regras de descrição de Edição

Regras de descrição	Informações complementares da edição são informados separados por ponto e vírgula.
Dados originais	5 ed. ; rev. e ampl.
MARC 21 após tratamento	250 __  a 5. ed.  b rev. e ampl.

Fonte: elaborado pela autora.

c) Número de páginas (extensão): descrição de livros digitais possui diferenças em relação às publicações impressas (**Tabela 22**).

**Tabela 22:** Regras de descrição de Número de páginas (extensão)

Regras de descrição	A indicação do número de páginas, quando conhecida, é mostrada entre parêntesis após a expressão “1 recurso eletrônico”. Caso o número de páginas não seja informado, somente a expressão “1 recurso eletrônico” é utilizada.
Dados originais	234 p.
MARC 21 após tratamento	300 __  a 1 recurso eletrônico (234 p.)

Fonte: elaborado pela autora.

d) ISBN da versão impressa: preenchimento de mais de um campo (**Tabela 23**).

**Tabela 23:** Regras de descrição de ISBN da versão impressa

Regras de descrição	Caso esteja preenchido o campo “ISBN versão impressa” deve ser gerada <i>tag</i> 530 com o texto “Obra disponível em formato impresso sob o ISBN <ISBN versão impressa>.” Também é gerada uma <i>tag</i> 776  z com o ISBN, com inclusão do  i com a expressão “Versão impressa”
Dados originais	9788598309823
MARC 21 após tratamento	530 __  a Obra disponível em formato impresso sob o ISBN 9788598309823. 776 __  i Versão impressa  z 9788598309823.

Fonte: elaborado pela autora.

e) Local de publicação: preenchimento de mais de um campo ou ausência do dado (**Tabela 24**).

**Tabela 24:** Regras de descrição de Local de publicação

Regras de descrição	Caso exista mais de um local de publicação, as ocorrências devem ser separadas por ponto e vírgula, gerando dois subcampos __ 260  a (exemplo 1).
---------------------	---

	Caso o campo esteja em branco, é acrescentada a expressão [S. 1.] de forma automática (exemplo 2). Caso o campo da editora esteja vinculado a um <i>dataset</i> que identifique sua localização geográfica, o local de publicação pode ser completado no enriquecimento dos dados.
Dados originais	Exemplo 1: São Paulo; Rio de Janeiro Exemplo 2:
MARC 21 após tratamento	Exemplo 1: 260 __  a São Paulo;  a Rio de Janeiro Exemplo 2: 260 __  a [S. 1.]

Fonte: elaborado pela autora.

f) Editora: preenchimento de mais de um campo ou ausência do dado (**Tabela 25**).

**Tabela 25:** Regras de descrição de Editora

Regras de descrição	Caso exista mais de uma editora, as ocorrências devem ser separadas por ponto e vírgula, gerando dois subcampos __ 260  b (exemplo 1). Caso o campo esteja em branco, é acrescentada a expressão [s. n.] de forma automática (exemplo 2).
Dados originais	Exemplo 1: Record ; FGV Exemplo 2:
MARC 21 após tratamento	Exemplo 1: __ 260  b Record :  b FGV Exemplo 2: __ 260  b [s. n.]

Fonte: elaborado pela autora.

g) Assuntos: descritores de assuntos (**Tabela 26**).

**Tabela 26:** Regras de descrição de Assuntos

Regras de descrição	As ocorrências são separadas por ponto e vírgula, sendo uma <i>tag</i> para cada ocorrência. Caso exista traço separando dois termos, será criada uma subdivisão geral. Apesar de existirem outros tipos de subdivisões para assuntos e não somente a geral, optou-se por uma descrição simplificada no MARC 21, evitando complexidade no preenchimento dos dados na planilha. Eventuais ajustes podem ser realizados pelos bibliotecários.
Dados originais	Medicina – Anatomia; Músculos; Força; Força – Prova de função
MARC 21 após tratamento	650 _4  a Medicina  x Anatomia 650 _4  a Músculos 650 _4  a Força 650 _4  a Força  x Prova de função

Fonte: elaborado pela autora.

h) URL para acesso: inclusão de informações adicionais (**Tabela 27**).

**Tabela 27:** Regras de descrição de URLs para acesso

Regras de descrição	O subcampo 856  y é inserido com texto padrão “Acesso restrito” em todas as URLs. Verificação se a URL possui o <http://>. Se ausente, o dado é incluído.
Dados originais	www.fornec1.com.br/9788598309823

MARC 21 após tratamento	856 41  u <a href="http://www.fornec1.com.br/9788598309823">http://www.fornec1.com.br/9788598309823</a>  y Acesso restrito
-------------------------------	--

Fonte: elaborado pela autora.

A **Tabela 28** apresenta os dados originais do fornecedor (Valores originais) e após a primeira etapa de tratamento dos dados (Valores tratados), antes da conversão ao *Schema*.

**Tabela 28:** Dados originais recebidos de fornecedor em planilha de dados

Campos	Valores originais	Valores tratados
Idioma		Por
ISBN	978-85-3529142-1	9788535291421
ISBN versão impressa	9788535291179	9788535291179
DOI		
Autorias (pessoa)	Mark Miller; James Browne; Brian Cole; Andrew Cosgarea e Brett Owens	Miller, Mark; Browne, James; Cole, Brian; Cosgarea, Andrew; Owens, Brett
Entidade coletiva		
Dados sobre eventos		
Título	Cirurgia do Joelho 2ED Série CIRURGIAS ORTOPÉDICAS	Cirurgia do joelho
Subtítulo		
Edição	2	2
Obra volumada?		
Volume		
Local de publicação		Rio de Janeiro
Editora	Elsevier Editora	Elsevier
Data de publicação	1/15/19	2019
Data de copyright	2019	© 2019
País de publicação		Brasil
Número de páginas	408	1 recurso online (408 p.)
Formato do arquivo	EPUB	ePub
Série		Cirurgias ortopédicas
Volume da série		
Nota geral		
Nota de conteúdo		
Nota de resumo	Ideal para residentes ortopedistas, colegas e cirurgiões praticantes, oferece todas as orientações passo a passo que você precisa para	Ideal para residentes ortopedistas, colegas e cirurgiões praticantes, oferece todas as orientações passo a passo que você precisa para realizar as técnicas mais recentes na

	realizar as técnicas mais recentes na cirurgia do joelho.  Como parte da série de Técnicas Operatórias que são altamente visuais, possui descrições breves em bullets e um layout limpo para facilitar o uso, enquanto as dicas clínicas juntamente com observações práticas dos principais especialistas em cada uma das técnicas, e as radiografias de alta qualidade mostram problemas de apresentação e resultados pós-cirúrgicos.	cirurgia do joelho. Como parte da série de Técnicas Operatórias que são altamente visuais, possui descrições breves em bullets e um layout limpo para facilitar o uso, enquanto as dicas clínicas juntamente com observações práticas dos principais especialistas em cada uma das técnicas, e as radiografias de alta qualidade mostram problemas de apresentação e resultados pós-cirúrgicos.
Dados biográficos		
Nota de premiação		
Nota de edição impressa		
Público alvo		
Assuntos	SAÚDE. MEDICINA. ORTOPEDIA CIRÚRGICA	Saúde; Medicina; Ortopedia cirúrgica
Local como assunto		
URL para acesso	<a href="http://www.evolution.com.br/product/9788535291421">www.evolution.com.br/product/9788535291421</a>	<a href="http://www.evolution.com.br/product/9788535291421">http://www.evolution.com.br/product/9788535291421</a>
Capa		<a href="https://www.evolution.com.br/viewinside/9788535291421">https://www.evolution.com.br/viewinside/9788535291421</a>
Moeda	R\$	R\$
Preço	325,00	325
Modelo de negócio	Assinatura	Assinatura
Uso concorrente?	Não	Não
No. acessos simultâneos	1	1
Disponibilidade	Disponível na plataforma	Disponível
Data inclusão	12 de janeiro 2019	12/01/2019
Data última alteração (plataforma)	31 JAN. 2019	31/01/2019
Data prevista remoção	31-01-2020	31/01/2020
Data remoção		
Situação livro	Ativo	Ativo

Fonte: dados originais de fornecedor, recebidos em planilha de dados. Dados tratados pela autora.

Na **Tabela 28** são observados nos dados originais com informações faltantes ou preenchidas de forma incompleta ou fora dos padrões da Representação descritiva. Para o *data*



*cleaning* foi utilizada a ferramenta *OpenRefine*<sup>44</sup> para correção dos problemas. Destaque para os campos de datas que usualmente apresentam diversas formas de grafia. Nesta etapa foram aplicadas as regras definidas no item 5.3.1 desta pesquisa. A descrição de ISBN impresso, autorias, assuntos, edição, extensão e *link* para acesso receberam o tratamento descrito para, após esta normalização, serem convertidos para o RDF. Os dados de edição e série foram removidos do título e transferidos para os respectivos campos. O local de publicação foi incluído na planilha por ser relacionado com a editora. A partir do local de publicação foi possível incluir o país. O idioma foi incluído por tratar-se de um livro publicado no idioma do local de publicação. Foi aplicada capitalização nos caracteres para transformar letras do maiúsculo para o minúsculo. Foram removidas quebras de linhas do campo da nota Resumo. O *link* para acesso da obra foi complementado com o `http://` e a URL da imagem da capa do livro foi obtida no sítio da própria editora.

O enriquecimento dos dados foi realizado por meio do VIAF para o campo ‘Autorias’. No exemplo adotado foram incluídas datas de nascimento de alguns autores e seus identificadores. O *dataset GeoNames*<sup>45</sup> foi utilizado para identificar os dados nos registros dos campos Local de publicação e País, completando informações faltantes.

A **Figura 13** apresenta os dados tratados convertidos para o modelo *Schema*, com utilização dos *vocabulários Book, Place, Offer, Intangible e Creative Work*.

Figura 13: Registro convertido ao *Schema*

```
<?xml version="1.0" encoding="utf-8" ?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:schema="http://schema.org/">
<schema:Book>
<schema:isbn>9780030426599</schema:isbn>
  <schema:MediaObject>
    <schema:encodingFormat>
      <schema:printEdition>print</schema:printEdition>
    <schema:isbn>9788535291179</schema:isbn>
    </schema:encodingFormat>
  </schema:MediaObject>
<schema:inLanguage>por</schema:inLanguage>
<schema:name>Cirurgia do joelho</schema:name>
  <schema:author>
    <schema:Person>
      <schema:name>Miller, Mark D.</schema:name>
      <schema:birthdate>1957</schema:birthdate>
      <schema:sameAs>href=http://viaf.org/viaf/44532190</schema:sameAs>
    </schema:Person>
  </schema:author>
</schema:author>
```

<sup>44</sup> <http://openrefine.org/>

<sup>45</sup> <https://www.geonames.org/>

```

    <schema:Person>
      <schema:name>Browne, James A.</schema:name>
      <schema:sameAs>href=http://viaf.org/viaf/93922076</schema:sameAs>
    </schema:Person>
  </schema:author>
  <schema:author>
    <schema:Person>
      <schema:name>Cole, Brian J.</schema:name>
      <schema:birthDate>1962</schema:birthDate>
<schema:sameAs>href=http://viaf.org/viaf/64253230</schema:sameAs>
    </schema:Person>
  </schema:author>
  <schema:author>
    <schema:Person>
      <schema:name>Cosgarea, Andrew J.</schema:name>
<schema:sameAs>href=http://viaf.org/viaf/4247148209309500460007</schema:sameAs>
    </schema:Person>
  </schema:author>
  <schema:author>
    <schema:Person>
      <schema:name>Owens, Brett D.</schema:name>
      <schema:birthDate>1972</schema:birthDate>
      <schema:sameAs>href=http://viaf.org/viaf/159998615</schema:sameAs>
    </schema:Person>
  </schema:author>
  <schema:bookEdition>2</schema:bookEdition>
  <schema:pubPlace>Rio de
Janeiro</schema:pubPlace><schema:sameAs>href=http://www.geonames.org/3451190</schema:sameAs>
  <schema:publisher>Elsevier</schema:publisher>
  <schema:copyrightYear>2019</schema:copyrightYear>
  <schema:datePublished>2019</schema:datePublished>
</schema:Book>
<schema:Place>
  <schema:sameAs>href=http://www.geonames.org/countries/BR/</schema:sameAs>
  <schema:addressCountry>BR</schema:addressCountry>
</schema:Place>
<schema:Book>
  <schema:numberOfPages>1 recurso online (408 p.)</schema:numberOfPages>
  <schema:bookFormat>ePub</schema:bookFormat>
  <schema:isPartOf>Cirurgias ortopédicas</schema:isPartOf>
  <schema:description>Ideal para residentes ortopedistas, colegas e cirurgiões praticantes,
oferece todas as orientações passo a passo que você precisa para realizar as técnicas mais recentes
na cirurgia do joelho.Como parte da série de Técnicas Operatórias que são altamente visuais, possui
descrições breves em bullets e um layout limpo para facilitar o uso, enquanto as dicas clínicas
juntamente com observações práticas dos principais especialistas em cada uma das técnicas, e as
radiografias de alta qualidade mostram problemas de apresentação e resultados pós-
cirúrgicos.</schema:description>
  <schema:about>Saúde</schema:about>
  <schema:about>Medicina</schema:about>
  <schema:about>Ortopedia cirúrgica</schema:about>
  <schema:resource>href="https://www.evolution.com.br/product/978535291421"</schema:r
esource>

```

```

    <schema:image
rdf:src="9788535291421.jpg">href="http://www.evolution.comn.br/viewinside/9788535291421"</
schema:image>
</schema:Book>
<schema:Offer>
  <schema:price>325,00</schema:price>
  <schema:priceCurrency>BRL</schema:priceCurrency>
  <schema:additionalType>Assinatura</schema:additionalType>
</schema:Offer>
<schema:Intangible>
  <schema:eligibleTransactionVolume>1</schema:eligibleTransactionVolume>
  <schema:eligibleQuantity>1</schema:eligibleQuantity>
</schema:Intangible>
<schema>Action>
  <schema:actionStatus>Disponível na plataforma</schema:actionStatus>
</schema>Action>
<schema:Book>
  <schema:dateCreated rdf:content="20190112">19 jan. 2019</schema:dateCreated>
  <schema:dateModified rdf:content="20190131">31 jan. 2019</schema:dateModified>
</schema:Book>
<schema:Intangible>
  <schema:eligibleDuration rdf:content="20200131">31 jan. 2020</schema:eligibleDuration>
</schema:Intangible>
<schema:CreativeWork>
  <schema:name>Creative Commons BY-NC-ND Licence</schema:name>
  <schema:license>All content on this record is licensed under
http://creativecommons.org/ns#license</schema:license>
  <schema:URL>http://creativecommons.org/licences/by/3.0/</schema:URL>
</schema:CreativeWork>
</rdf:RDF>

```

Fonte: elaborado pela autora a partir dos dados presentes em *Schema.org*.

Ao ter os dados no *Schema* é possível convertê-lo para outros formatos. Para esta pesquisa os dados foram convertidos para o formato JSON-LD a partir da ferramenta *JSON-LD Playground*<sup>46</sup> (**Figura 14**).

Figura 14: Registro convertido do *Schema* ao formato JSON-LD

```

{
  "@graph": [
    {
      "@type": "http://schema.org/Book",
      "http://schema.org/MediaObject": {
        "@id": "_:b1"
      },
      "http://schema.org/author": [
        {
          "@id": "_:b2"
        },
        {
          "@id": "_:b3"
        }
      ]
    }
  ]
}

```

<sup>46</sup> <https://json-ld.org/playground/>

```

    "@id": "_:b4"
  },
  {
    "@id": "_:b5"
  },
  {
    "@id": "_:b6"
  }
],
"http://schema.org/bookEdition": "2",
"http://schema.org/copyrightYear": "2019",
"http://schema.org/datePublished": "2019",
"http://schema.org/inLanguage": "por",
"http://schema.org/isbn": "9780030426599",
"http://schema.org/name": "Cirurgia do joelho",
"http://schema.org/pubPlace": "Rio de Janeiro",
"http://schema.org/publisher": "Elsevier",
"http://schema.org/sameAs": "href=http://www.geonames.org/3451190"
},
{
  "@id": "_:b1",
  "@type": "http://schema.org/encodingFormat",
  "http://schema.org/isbn": "9788535291179",
  "http://schema.org/printEdition": "print"
},
{
  "@type": "http://schema.org/Intangible",
  "http://schema.org/elegibleQuantity": "1",
  "http://schema.org/elegibleTransactionVolume": "1"
},
{
  "@type": "http://schema.org/Action",
  "http://schema.org/actionStatus": "Disponível na plataforma"
},
{
  "@type": "http://schema.org/Book",
  "http://schema.org/dateCreated": "19 jan. 2019",
  "http://schema.org/dateModified": "31 jan. 2019"
},
{
  "@type": "http://schema.org/Intangible",
  "http://schema.org/eligibleDuration": "31 jan. 2020"
},
{
  "@type": "http://schema.org/CreativeWork",
  "http://schema.org/URL": "http://creativecommons.org/licences/by/3.0/",
  "http://schema.org/license": "All content on this record is licensed under
http://creativecommons.org/ns#license",
  "http://schema.org/name": "Creative Commons BY-NC-ND Licence"
},
{
  "@id": "_:b2",
  "@type": "http://schema.org/Person",
  "http://schema.org/birthdate": "1957",
  "http://schema.org/name": "Miller, Mark D.",

```

```

"http://schema.org/sameAs": "href=http://viaf.org/viaf/44532190"
},
{
"@id": "._:b3",
"@type": "http://schema.org/Person",
"http://schema.org/name": "Browne, James A.",
"http://schema.org/sameAs": "href=http://viaf.org/viaf/93922076"
},
{
"@id": "._:b4",
"@type": "http://schema.org/Person",
"http://schema.org/birthdate": "1962",
"http://schema.org/name": "Cole, Brian J.",
"http://schema.org/sameAs": "href=http://viaf.org/viaf/64253230"
},
{
"@id": "._:b5",
"@type": "http://schema.org/Person",
"http://schema.org/name": "Cosgarea, Andrew J.",
"http://schema.org/sameAs": "href=http://viaf.org/viaf/4247148209309500460007"
},
{
"@id": "._:b6",
"@type": "http://schema.org/Person",
"http://schema.org/birthdate": "1972",
"http://schema.org/name": "Owens, Brett D.",
"http://schema.org/sameAs": "href=http://viaf.org/viaf/159998615"
},
{
"@type": "http://schema.org/Place",
"http://schema.org/addressCountry": "BR",
"http://schema.org/sameAs": "href=http://www.geonames.org/countries/BR/"
},
{
"@type": "http://schema.org/Book",
"http://schema.org/about": [
"Saúde",
"Medicina",
"Ortopedia cirúrgica"
],
"http://schema.org/bookFormat": "ePub",
"http://schema.org/description": "Ideal para residentes ortopedistas, colegas e cirurgiões praticantes, oferece todas as orientações passo a passo que você precisa para realizar as técnicas mais recentes na cirurgia do joelho. Como parte da série de Técnicas Operatórias que são altamente visuais, possui descrições breves em bullets e um layout limpo para facilitar o uso, enquanto as dicas clínicas juntamente com observações práticas dos principais especialistas em cada uma das técnicas, e as radiografias de alta qualidade mostram problemas de apresentação e resultados pós-cirúrgicos.",
"http://schema.org/image":
"href=http://www.evolution.comn.br/viewinside/9788535291421\"",
"http://schema.org/isPartOf": "Cirurgias ortopédicas",
"http://schema.org/numberOfPages": "1 recurso online (408 p.)",
"http://schema.org/resource": "href=http://www.evolution.com.br/product/978535291421\""
},
{

```

```

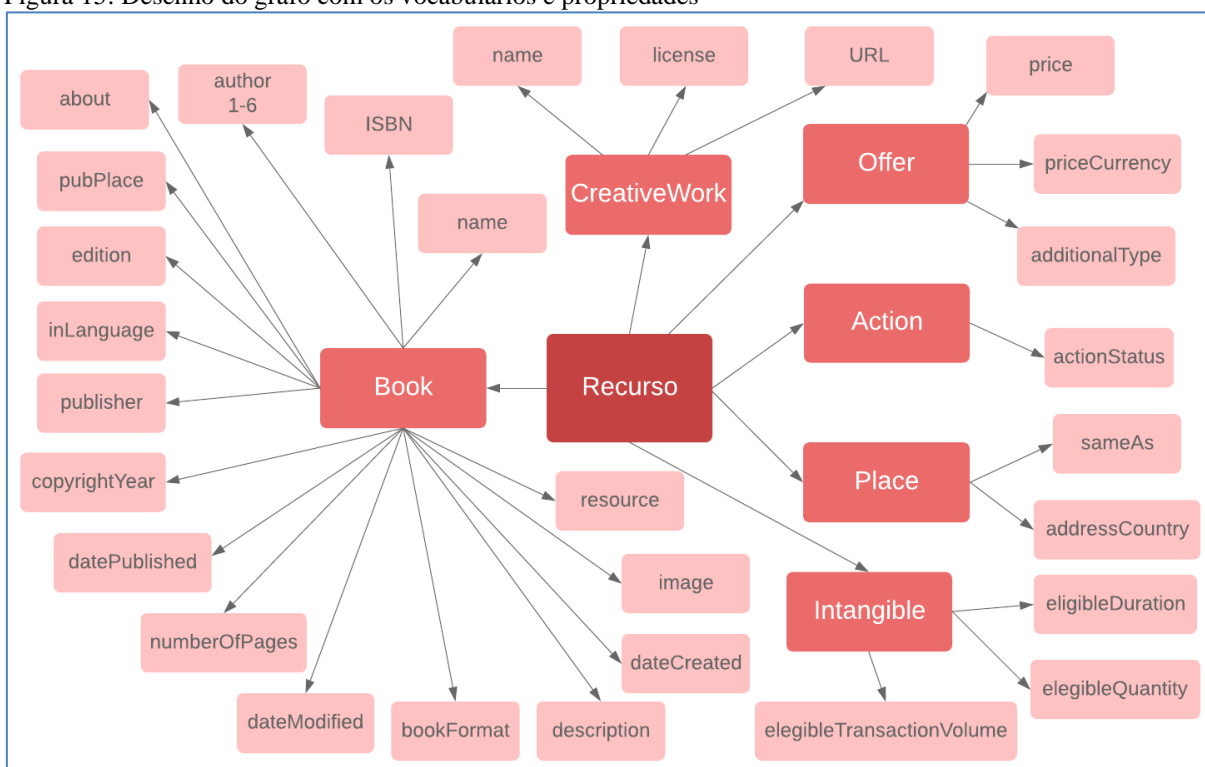
"@type": "http://schema.org/Offer",
"http://schema.org/additionalType": "Assinatura",
"http://schema.org/price": "325,00",
"http://schema.org/priceCurrency": "BRL"
}
]
}

```

Fonte: elaborado pela autora a partir dos dados em *Schema* e convertidos pela ferramenta *JSON-LD Playground*

O recurso está vinculado aos vocabulários *Book*, *Offer*, *Action*, *Intangible*, *CreativeWork* e *Place* e gerou um grafo com 68 triplas. Uma única imagem foi gerada pela ferramenta *EasyRDF*<sup>47</sup>. Entretanto, como não seria legível colocar o grafo completo para visualização em uma única página, optou-se por separá-los por vocabulário. Uma representação do recurso com todos os vocabulários vinculados e o primeiro nível de dados pode ser conferido na **Figura 15**.

Figura 15: Desenho do grafo com os vocabulários e propriedades



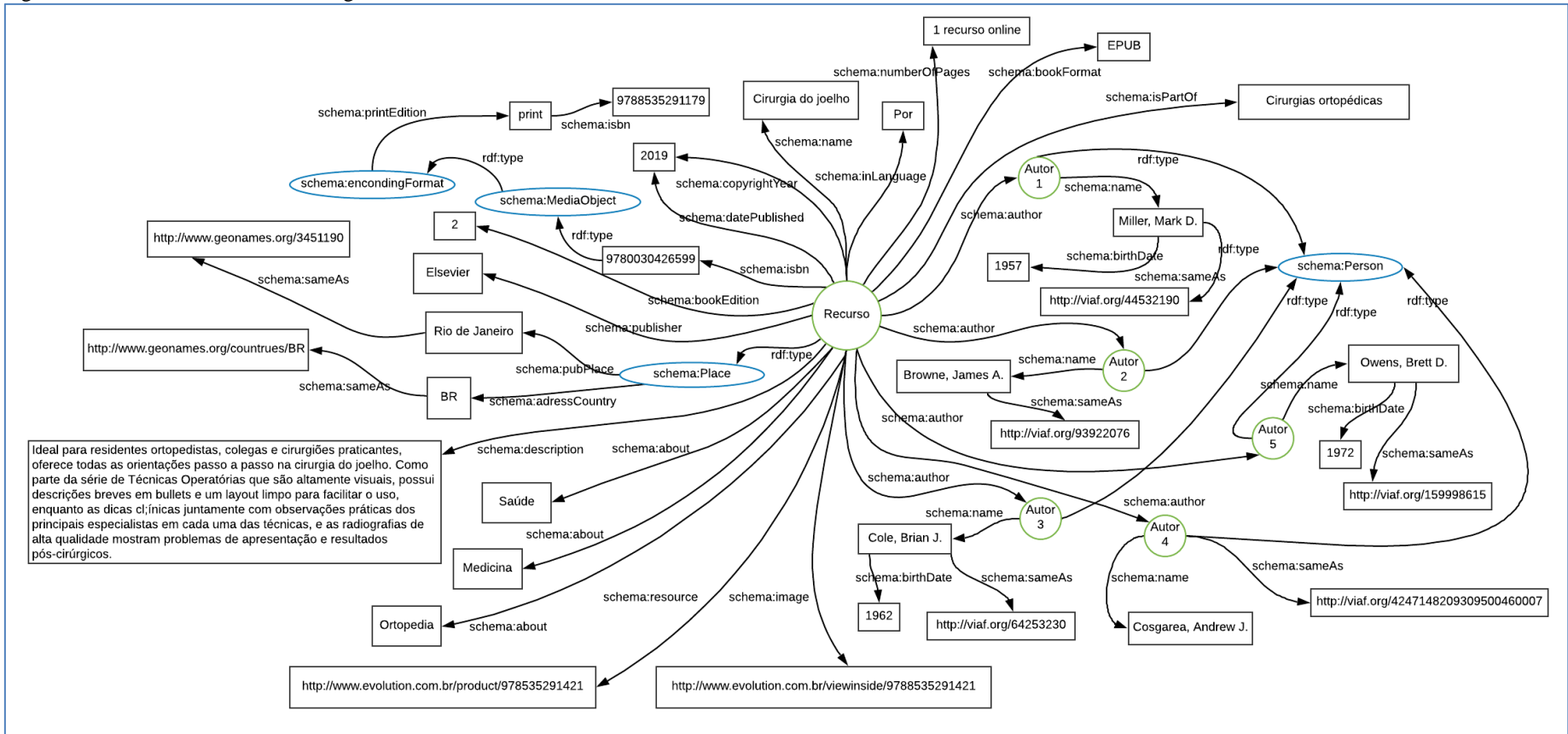
Fonte: elaborado pela autora com a ferramenta *Lucidchart*

Os grafos por vocabulário estão representados da **Figura 16** até a **Figura 20**.

A **Figura 16** reúne os dados bibliográficos do recurso, compreendidos pelo *SchemaBook*. Como este conjunto de dados é extenso em relação aos grupos dos demais vocabulários utilizados neste exemplo de validação (*CreativeWork*, *Offer*, *Action* e *Intangible*), alguns dados foram explicitados no grafo para melhor identificação dos elementos.

<sup>47</sup> <http://www.easyrdf.org/converter>

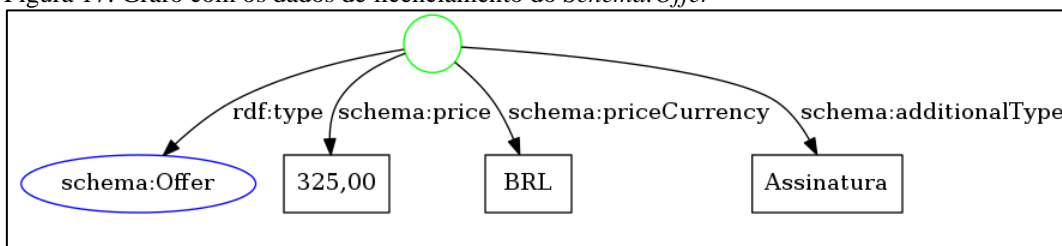
Figura 16: Grafo com os dados bibliográficos do *Schema:Book*



Fonte: elaborado pela autora a partir dos dados em *Schema* e editados no *Lucidchart*

Os dados de licenciamento estão ilustrados na **Figura 17**.

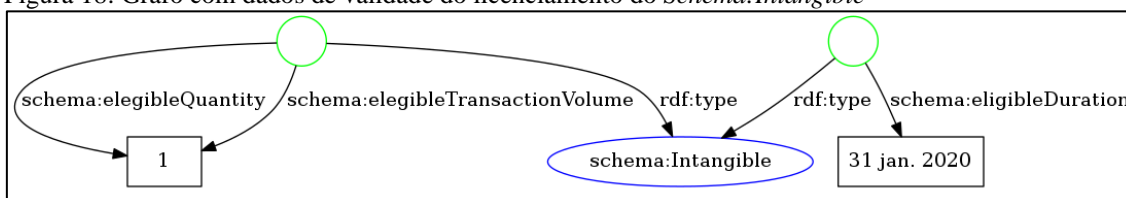
Figura 17: Grafo com os dados de licenciamento do *Schema:Offer*



Fonte: elaborado pela autora a partir dos dados em *Schema* convertidos pela ferramenta *EasyRDF*

A **Figura 18** apresenta grafo com os dados sobre a validade do licenciamento.

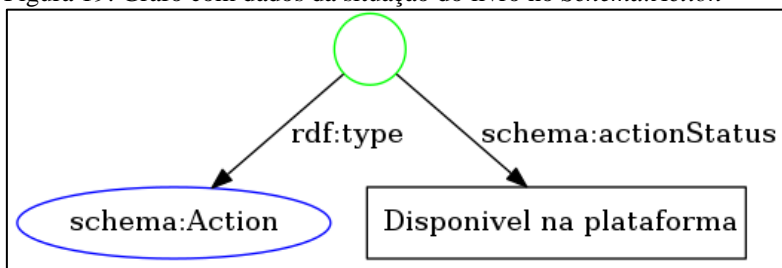
Figura 18: Grafo com dados de validade do licenciamento do *Schema:Intangible*



Fonte: elaborado pela autora a partir dos dados em *Schema* convertidos pela ferramenta *EasyRDF*

A **Figura 19** ilustra o grafo com os dados sobre a situação do livro.

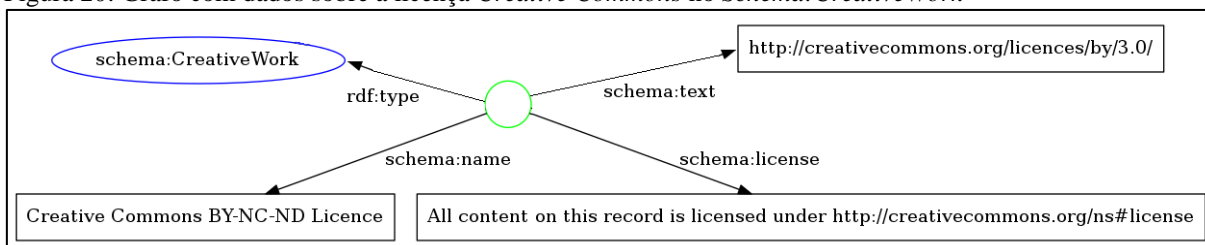
Figura 19: Grafo com dados da situação do livro no *Schema:Action*



Fonte: elaborado pela autora a partir dos dados em *Schema* convertidos pela ferramenta *EasyRDF*

O grafo com os dados da licença estão na **Figura 20**.

Figura 20: Grafo com dados sobre a licença *Creative Commons* no *Schema:CreativeWork*



Fonte: elaborado pela autora a partir dos dados em *Schema* convertidos pela ferramenta *EasyRDF*

Para permitir futuras importações de registros para o catálogo da biblioteca, os elementos bibliográficos foram mapeados e convertidos do *Schema* para o formato MARC 21, conforme pode ser verificado na **Tabela 29**.



**Tabela 29:** Dados bibliográficos no modelo *Schema* mapeados para o MARC 21

<i>Schema</i>	MARC 21
<schema:inLanguage>por</schema:inLanguage>	008/35-37 por
<schema:isbn>9780030426599</schema:isbn>	<b>020</b> __  a 9788535291421
<schema:MediaObject> <schema:encodingFormat> <schema:printEdition>print</schema:printEdition> </schema:MediaObject>	<b>530</b> __  a Obra disponível em formato impresso sob o ISBN 9788535291179 <b>776</b> __  i Versão impressa  z 9788535291179
<schema:author> <schema:Person> <schema:name>Miller, Mark D.</schema:name> <schema:birthdate>1957</schema:birthdate> <schema:sameAs>href=http://viaf.org/viaf/44532190</schema:sameAs> </schema:Person> </schema:author> <schema:author> <schema:Person> <schema:name>Browne, James A.</schema:name> <schema:sameAs>href=http://viaf.org/viaf/93922076</schema:sameAs> </schema:Person> </schema:author> <schema:author> <schema:Person> <schema:name>Cole, Brian J.</schema:name> <schema:birthDate>1962</schema:birthDate> <schema:sameAs>href=http://viaf.org/viaf/64253230</schema:sameAs> </schema:Person> </schema:author> <schema:author> <schema:Person> <schema:name>Cosgarea, Andrew J.</schema:name> <schema:sameAs>href=http://viaf.org/viaf/4247148209309500460007</schema:sameAs> </schema:Person> </schema:author> <schema:author> <schema:Person> <schema:name>Owens, Brett D.</schema:name> <schema:birthDate>1972</schema:birthDate> <schema:sameAs>href=http://viaf.org/viaf/159998615</schema:sameAs> </schema:Person> </schema:author>	<b>100</b> 1_  a Miller, Mark D.  d 1957-  e autor  1 <a href="http://viaf.org/viaf/44532190">http://viaf.org/viaf/44532190</a>  4 <a href="http://id.loc.gov/vocabulary/relators/aut">http://id.loc.gov/vocabulary/relators/aut</a> <b>700</b> 1_  a Browne, James A.  e autor  1 <a href="https://viaf.org/viaf/93922076/#Browne,_James_A.">https://viaf.org/viaf/93922076/#Browne,_James_A.</a>  4 <a href="http://id.loc.gov/vocabulary/relators/aut">http://id.loc.gov/vocabulary/relators/aut</a> <b>700</b> 1_  a Cole, Brian J.  d 1962  e autor  1 <a href="http://viaf.org/viaf/64253230">http://viaf.org/viaf/64253230</a>  4 <a href="http://id.loc.gov/vocabulary/relators/aut">http://id.loc.gov/vocabulary/relators/aut</a> <b>700</b> 1_  a Cosgarea, Andrew J.  e autor  1 <a href="http://viaf.org/viaf/4247148209309500460007">http://viaf.org/viaf/4247148209309500460007</a>  4 <a href="http://id.loc.gov/vocabulary/relators/aut">http://id.loc.gov/vocabulary/relators/aut</a> <b>700</b> 1_  a Owens, Brett D.  d 1972-  e autor  1 <a href="http://viaf.org/viaf/159998615">http://viaf.org/viaf/159998615</a>  4 <a href="http://id.loc.gov/vocabulary/relators/aut">http://id.loc.gov/vocabulary/relators/aut</a>

<schema:name>Cirurgia do joelho</schema:name>	<b>245</b> 10  a Cirurgia do joelho /  c Mark D. Miller, James A. Browne, Brian J. Cole, Andrew J. Cosgarea, Brett D. Owens
<schema:bookEdition>2</schema:bookEdition>	<b>250</b> __  a 2. ed.
<schema:pubPlace>Rio de Janeiro</schema:pubPlace><schema:sameAs>href=http://www.geonames.org/3451190</schema:sameAs>	<b>260</b> __  a Rio de Janeiro
<schema:publisher>Elsevier</schema:publisher>	<b>260</b> __  b Elsevier
<schema:datePublished>2019</schema:datePublished>	<b>260</b> __  c © 2019
<schema:copyrightYear>2019</schema:copyrightYear>	
<schema:Place><schema:sameAs>href=http://www.geonames.org/countries/BR/</schema:sameAs><schema:addressCountry>BR</schema:addressCountry></schema:Place>	<b>270</b> __  d Brasil
<schema:numberOfPages>1 recurso online (408 p.) </schema:numberOfPages>	<b>300</b> __  a 1 recurso online (408 p.) :  b ePub
<schema:bookFormat>ePub</schema:bookFormat>	<b>347</b> __  b ePub
<schema:isPartOf>Cirurgias ortopédicas</schema:isPartOf>	<b>490</b> 1_  a Cirurgias ortopédicas
<schema:description>Ideal para residentes ortopedistas, colegas e cirurgiões praticantes, oferece todas as orientações passo a passo que você precisa para realizar as técnicas mais recentes na cirurgia do joelho. Como parte da série de Técnicas Operatórias que são altamente visuais, possui descrições breves em bullets e um layout limpo para facilitar o uso, enquanto as dicas clínicas juntamente com observações práticas dos principais especialistas em cada uma das técnicas, e as radiografias de alta qualidade mostram problemas de apresentação e resultados pós-cirúrgicos.</schema:description>	<b>520</b> __  a Ideal para residentes ortopedistas, colegas e cirurgiões praticantes, oferece todas as orientações passo a passo que você precisa para realizar as técnicas mais recentes na cirurgia do joelho. Como parte da série de Técnicas Operatórias que são altamente visuais, possui descrições breves em bullets e um layout limpo para facilitar o uso, enquanto as dicas clínicas juntamente com observações práticas dos principais especialistas em cada uma das técnicas, e as radiografias de alta qualidade mostram problemas de apresentação e resultados pós-cirúrgicos.
<schema:about>Saúde</schema:about><schema:about>Medicina</schema:about><schema:about>Ortopedia cirúrgica</schema:about>	<b>650</b> _4  a Saúde <b>650</b> _4  a Medicina <b>650</b> _4  a Ortopedia cirúrgica
<schema:resource>href="https://www.evolution.com.br/product/978535291421"</schema:resource>	<b>856</b> 41  y Acesso restrito  u https://www.evolution.com.br/product/9788535291421

Fonte: elaborado pela autora a partir dos dados presentes em *Schema.org* ao formato MARC 21.

O exemplo validado neste capítulo apresenta as etapas pelas quais o registro recebeu tratamento para padronização, enriquecimento dos dados com informações complementares

(datas de nascimento dos autores, por exemplo), identificadores em formato de URI. Os mapeamentos sugeridos entre dados dispostos em planilha de dados para conversão para o *Schema* e MARC 21 foram testados e ilustrados neste exemplo de validação.

No próximo capítulo são apresentadas as considerações finais desta pesquisa.

## 7 Considerações finais

---

Após identificação das dificuldades enfrentadas por bibliotecas para a inclusão e gestão de livros digitais licenciados e com a análise de aplicações da Web Semântica que podem ser utilizadas em bibliotecas, esta pesquisa buscou responder o seguinte questionamento: “é possível a criação de um modelo de dados (*framework*) construído por meio da Web Semântica que permita a publicação e reuso de *datasets* favorecendo troca de registros bibliográficos entre fornecedores e bibliotecas?”. As pesquisas realizadas no decorrer do preparo desta tese proporcionaram a condições para responder a este e demais questionamentos que foram elaborados.

Para orientar a pesquisa foi identificado um objetivo geral e definidos três objetivos específicos. O **objetivo geral** é a proposta de criação de modelo de dados para gestão de livros digitais licenciados por meio da Web Semântica. Os **objetivos específicos** buscaram identificar os principais problemas observados por bibliotecas na contratação de livros digitais por fornecedores; os aspectos da Web Semântica que permitem a publicação e reutilização de dados entre fornecedores de livros digitais licenciados e bibliotecas; e, desenvolver e validar modelo que, por meio de enriquecimento, troca e reuso de dados, facilite a gestão de livros digitais licenciados.

O Capítulo 2 proporcionou subsídios para a resposta do primeiro objetivo específico (**identificar os principais problemas observados por bibliotecas na contratação de livros digitais licenciados por fornecedores**). Primeiramente foi necessário compreender o cenário do uso de livros digitais em bibliotecas. Foram discorridas explicações sobre as formas como os livros digitais podem ser contratados pelas bibliotecas (modelos de licenciamento) organizados em três tipologias: perpétuo, transitório e orientado ao usuário; a dependência da plataforma para a leitura dos conteúdos e os atributos de acesso ao conteúdo (concorrente ou não concorrente).

Neste cenário foram identificadas três principais dificuldades enfrentadas pelas bibliotecas com o uso de livros digitais licenciados: a qualidade dos registros entregues pelos fornecedores, a fragilidade do desenvolvimento da coleção, e a gestão dos metadados, com dados de uso e de licenciamento, que usualmente são obtidos de formas diversas, de acordo com os dados dos fornecedores.

Normalmente os fornecedores utilizam o formato ONIX, específico do mercado do livro, ou então em outros sistemas, proprietários ou não, normalmente com distribuição de dados em planilhas. Os fornecedores tradicionalmente não trabalham com o formato MARC 21 e prover os metadados dos livros digitais neste formato é custoso para eles. Foi apresentado um exemplo de livro digital encaminhado por fornecedor onde é exemplificada a não aderência dos dados aos padrões descritivos internacionais.

Dentre os problemas relacionados ao desenvolvimento de coleção destacam-se as alterações nos livros digitais e falhas de comunicação entre os fornecedores e as bibliotecas. A seleção e contratação dos livros digitais que serão licenciados são, na maioria das vezes, definidas pelos fornecedores, com a oferta de pacotes com diversos livros digitais. A biblioteca não tem total domínio dos livros digitais que fazem parte de sua coleção. Foram relatadas situações em que bibliotecas notaram que livros digitais haviam sido removidos da plataforma, muitas vezes sem notificação do fornecedor.

A alteração nos pacotes contratados pode ocorrer durante a vigência do licenciamento, assim como também podem existir modificações nas formas de licenciamento, o que exige atualização no catálogo. A inclusão ou remoção de livros digitais da plataforma exige que sejam atualizados os dados disponíveis no catálogo, assim como qualquer alteração realizada na forma de contratação do livro digital, como a troca do modelo de licenciamento ou dos atributos de acesso dos livros digitais. Se a biblioteca optar por não incluir os metadados dos livros digitais no catálogo, os usuários devem realizar a consulta na plataforma do fornecedor, desvinculando a oferta do recurso dentre os serviços que são oferecidos no catálogo *online*.

Outro aspecto relatado foi a presença de obras concorrentes e/ou duplicadas, o que acarreta aumento da quantidade de livros digitais efetivamente contratados, exigindo atualização dos dados no catálogo *online*. Obras concorrentes são resultados de contratação de um mesmo livro digital por mais de um fornecedor. A presença de obras duplicadas pode ocorrer devido redundância no conjunto de dados encaminhado pelo fornecedor, ou ainda a necessidade de realizar uma nova carga dos dados a cada renovação.

A gestão dos livros digitais fica fragmentada, com a biblioteca recebendo informes estatísticos e notificações de livros digitais presentes no pacote contratado sem padrões, com cada fornecedor encaminhando os dados de uma forma. Outra situação identificada é a falta de metadados do licenciamento, o que dificulta saber qual fornecedor é responsável por qual livro digital, e qual seu período de validade. Foram vistos exemplos de bibliotecas que utilizam o

formato MARC 21 para registro dos licenciamentos realizados (nome do fornecedor, data de validade, forma de contratação etc.).

Sem os dados reunidos em um único ponto as bibliotecas precisam coletar e consolidar os dados dos fornecedores, analisando a quantidade de acessos realizados (empréstimos digitais), se existe necessidade de manutenção dos livros digitais na coleção, ou ainda identificar se é preciso contratar maior quantidade de acessos simultâneos. Sem indicadores de uso e dados de licenciamento as bibliotecas encontram bastante dificuldade para realizar a gestão de seu acervo digital e ficam sem subsídios para tomadas de decisão como a renovação ou não dos livros de um determinado fornecedor.

O segundo objetivo específico foi respondido nos Capítulos 3 e 4. Este objetivo específico busca **identificar aspectos da Web Semântica que permitam a publicação e reutilização de dados entre fornecedores de livros digitais licenciados e bibliotecas**. Para responder a este objetivo primeiramente foi realizada conceituação da Web Semântica, apresentando seus aspectos como ontologias, modelos de dados e ferramentas, com a presença de triplas com sujeito, predicado e objeto onde descrições e relações podem ser representadas por URIs, tornando os dados legíveis para máquinas e também para pessoas, em modelo de dados em RDF.

Foram identificados aspectos da Web Semântica localizados na literatura consultada, com análise do conceito do *Linked Data* e as aplicações em *Linked Open Data* e *Linked Enterprise Data*. O Capítulo 3 foi finalizado com as etapas para publicação de *datasets* na Web de forma que possam ser consumidos.

Aspectos específicos da Web Semântica em bibliotecas foram apresentados no Capítulo 4, discutindo as possibilidades de utilização e algumas iniciativas e estudos que já estão em curso. Foi destacado o posicionamento que as bibliotecas podem alcançar ao disponibilizarem *datasets* com dados confiáveis e com qualidade na Web dos dados. Também foi discutida a presença de silos de informações nas bibliotecas e as possibilidades de reuso de metadados, proporcionando agilidade para bibliotecários para disponibilização de recursos contratados. Os dados das bibliotecas devem ser colocados de forma a serem consumidos abertamente na Web por outras bibliotecas e demais instituições, contribuindo com a normalização de dados bibliográficos e de autoridades.

No Capítulo 4 também foram discutidas as conversões de dados para que fiquem compatíveis com o LD e a definição de manutenção na conversão de dados para o formato MARC 21, em decorrência do desenvolvimento do BIBFRAME ainda estar em desenvolvimento. Foram apresentados os projetos LD4, o modelo de dados do *Schema.org* e o *dataset* VIAF, visando o reuso de dados e enriquecimento do catálogo. Foram citadas somente fontes que foram aplicadas na elaboração e no exemplo de validação do modelo proposto.

O terceiro objetivo específico consiste em **desenvolver e validar modelo que, por meio da troca e reuso de dados, facilite a gestão de livros digitais licenciados**. O embasamento teórico para responder a este objetivo foi alcançado com as questões abordadas no Capítulo 4 e a identificação das etapas necessárias para publicação de *datasets* dos fornecedores.

O modelo de dados foi descrito no Capítulo 5, explicando os procedimentos de cada etapa, porém sem definir ferramentas, proporcionando flexibilidade para adaptação da escolha de ontologias, formato de dados, serialização, limpeza e enriquecimento que podem ser aplicados aos dados.

Foi identificado que cada fornecedor deve possuir seu próprio *dataset*, de forma que possa aplicar restrições de acesso aos dados de licenciamento realizados com seus clientes, porém disponibilizando os dados bibliográficos de forma totalmente aberta. A criação de *datasets* por fornecedor também permite que as bibliotecas utilizem somente os *datasets* de fornecedores com os quais possuem contratações, evitando tráfego de informações que não serão consumidas.

Um dos objetivos do modelo é o tratamento e enriquecimento de dados, observando-se que é possível a utilização de variados *datasets* para normalização de dados utilizados na descrição de autores, assuntos, locais geográficos, editores etc.

A cada etapa para publicação do *dataset* foram identificadas as boas práticas para LOD recomendadas pela W3C, buscando com isso ampliação das possibilidades de uso do modelo de dados por diversos fornecedores e bibliotecas.

Foi definido o processo para conversão dos dados, partindo de informações originais até alcançar o nível que permite a publicação e troca de registros entre bibliotecas e fornecedores. Também foi feito o desenho da modelagem, identificando as informações que

podem ser trocadas entre fornecedores e bibliotecas, apresentando os dados que são de uso aberto (LOD) e de uso restrito (LED).

Dentre as etapas para publicação destaca-se a necessidade de mapeamentos para conversão de dados de fornecedores (ONIX, planilha de dados) e bibliotecas (MARC 21) para o modelo RDF, *Schema*, compatíveis com o LD.

Enquanto o Capítulo 5 apresentou as especificações do modelo, o Capítulo 6 buscou um exemplo de validação do mesmo para assegurar que os dados pertinentes estavam recebendo tratamento adequado para conversão, publicação e reutilização de dados.

No Capítulo 6 foi utilizado um exemplo real de livro digital de fornecedor e o tratamento dado para que o mesmo possa ser publicado em *dataset* com qualidade superior ao original. Aqui foram descritas as ferramentas que foram utilizadas e a escolha dos *datasets* VIAF e *GeoNames* para enriquecimento dos dados. Após passar por todas as etapas especificadas no Capítulo 5, é apresentado o registro enriquecido e convertido ao RDF com dados bibliográficos e de licenciamento. O registro é apresentado nos formatos RDF-XML, JSON-LD e como grafo. Como a imagem do grafo ficou muito extensa e restringia a possibilidade de visualização, foi feita uma imagem apresentando o grafo (os vocabulários utilizados e suas respectivas propriedades) e, em seguida, a imagem do grafo gerado em cada vocabulário.

A opção por fazer o exemplo de validação do modelo permitiu a identificação de dados que não tinham sido contemplados num primeiro momento, além de estressar cada etapa, certificando as possibilidades de estabelecimento do modelo. Um dos objetivos do exemplo de validação era aferir o quanto os dados poderiam ser melhorados de forma a proporcionar melhor qualidade às bibliotecas, ao mesmo tempo que as alterações realizadas pelas bibliotecas possam ser conhecidas dos fornecedores, que podem acatar ou não as sugestões dadas.

Ao alcançar o pleno atendimento dos objetivos geral e específicos no decorrer da pesquisa, foi obtida a comprovação da tese de que é possível a utilização da Web Semântica na gestão de livros digitais licenciados, contribuindo com a solução dos problemas apresentados no Capítulo 2.

Dentre as contribuições obtidas nesta pesquisa merecem destaque:

- 1) A identificação de que as bibliotecas podem (e devem) contribuir com a Web dos dados e disponibilizar *datasets* para consumo de dados e que estas contribuições são



vistas como confiáveis, em virtude da experiência e habilidade de bibliotecários em descrever dados descritivos;

- 2) A possibilidade de utilização de *datasets* externos para enriquecimento de dados da biblioteca;
- 3) Mapeamento de dados para os modelos e formatos ONIX, MARC 21 e *Schema*, permitindo a troca e reuso de informações entre fornecedores e bibliotecas;
- 4) Apresentação de regras específicas para mapeamento de dados obtidos por meio de planilhas de dados;
- 5) Diversos *datasets* podem ser utilizados para enriquecer dados do catálogo, tanto de autoridades como de informações de local de publicação, país e demais dados que estejam disponíveis na Web;
- 6) As bibliotecas podem iniciar a preparação de seus dados à Web Semântica mesmo que ainda não estejam totalmente preparados para isso, inclusive utilizando o formato MARC 21. É possível a inclusão de URIs em autoridades e em registros bibliográficos, assim como explicitar as relações existentes entre autores e registros;
- 7) Dispor de um modelo aberto de dados que pode ser aplicado para o desenvolvimento de produtos, serviços ou funcionalidades para publicação, vínculo e reuso de dados na Web dos dados, inclusive utilizando *datasets* para padronização da entrada de dados, vinculando-os a sistemas de cadastro de registros bibliográficos de fornecedores e/ou bibliotecas.

Na visão desta pesquisadora, a utilização do modelo pode trazer outras consequências às bibliotecas, como:

- 1) A partir do momento que a biblioteca possui domínio dos livros digitais que foram licenciados e que permanecem disponíveis no catálogo, evita-se que publicações removidas dos pacotes contratados sejam oferecidas aos usuários. Isto agrega credibilidade à biblioteca, disponibilizando recursos de informação que realmente podem ser acessados;
- 2) Quando ocorrer notificação de que um livro digital será removido da coleção do fornecedor, a biblioteca pode se programar para substituí-lo por outro similar ou negociar compensações em decorrência do cancelamento;
- 3) Ao trocar notificações, as bibliotecas podem contribuir com aumento da qualidade dos metadados dos fornecedores que, inclusive, pode proporcionar novas possibilidades de negociação com outras bibliotecas. Este fato também influencia

no esforço para limpeza e padronização de dados nas etapas para a publicação do *dataset*;

- 4) Ao reunir em um único ponto os livros digitais contratados por fornecedores diferentes, a biblioteca pode ter maior controle sobre o licenciamento de livros digitais duplicados e obras concorrentes, ajustando as contratações de acordo com a demanda manifestada pelos usuários, a partir da mensuração dos acessos realizados;
- 5) O domínio dos livros digitais contratados, com licenciamentos vigentes e disponibilizados aos usuários permite que a biblioteca possa avaliar o retorno no investimento realizado na contratação de recursos de informação, proporcionando clareza do cenário entre oferta dos fornecedores e demanda dos usuários;
- 6) Ao consumir *datasets* com dados bibliográficos e de licenciamento, as bibliotecas passam a ter ciência dos valores cobrados e formas de contratação oferecidos pelos fornecedores sem a necessidade de contatá-los, permitindo negociação de descontos nos valores praticados. A partir do momento que os valores para licenciamento estão públicos, as bibliotecas podem programar seus orçamentos, prevendo as quantias necessárias para atualização da coleção;
- 7) Em situações em que é necessário diminuir investimentos em licenciamento, os dados centralizados pelo modelo contribuem com a identificação dos recursos que são fundamentais, essenciais, desejados ou dispensáveis de cada fornecedor, acarretando bom uso de recursos financeiros. Também é possível identificar cortes possíveis em contratações, diminuindo os efeitos que podem causar aos usuários;
- 8) Outro aspecto a ser proporcionado pela adoção do modelo é a diminuição do controle manual e atualização de livros digitais nos catálogos, coleta e consolidação de estatísticas de uso. O tempo investido nestas tarefas pode ser direcionado para orientações aos usuários, estudos da coleção ou outras atividades bibliotecárias.

A utilização do modelo por fornecedores pode proporcionar os seguintes benefícios:

- 1) Possibilidade de publicação de seus dados, ampliando as chances de realização de licenciamentos;
- 2) Obtenção de registros no formato MARC, dispensando o investimento em ferramentas ou mão de obra para fazê-lo;
- 3) Recebimento de notificações das bibliotecas para atualização e melhoria da qualidade dos metadados oferecidos aos demais clientes;

- 4) Permitir o envio de pedidos da biblioteca diretamente ao fornecedor, contribuindo com o aumento de licenciamentos;
- 5) Permitir que as bibliotecas identifiquem as demais obras presentes em sua plataforma, porém que não foram licenciados;
- 6) Desobrigação dos fornecedores em utilizar ferramentas diferentes das utilizadas somente para permitir a publicação de dados. Os fornecedores podem manter as ferramentas que utilizam, desde que os dados sejam passíveis de exportação por meio de planilha de dados, por exemplo;
- 7) Melhoria no contato com seus clientes (bibliotecas) ao informar de forma ágil e rápida que novos títulos foram incluídos ou que serão removidos;
- 8) Favorecer a troca de dados bibliográficos com outros fornecedores e bibliotecas.

Dentre os benefícios proporcionados pela utilização do modelo proposto espera-se contribuir com dados que permitam controle de recursos financeiros e os investimentos em publicações que serão realizados, de acordo com modelos de negócios aplicados, quantidade de acessos simultâneos, sempre orientado com o uso que é feito dos livros digitais contratados. Isso permite que ajustes sejam feitos, evitando ou diminuindo os efeitos que estes cortes podem acarretar na coleção.

Conforme pode ser observado na literatura consultada, a presença de estudos no Brasil de uso da Web Semântica em bibliotecas ainda não se mostra abundante. A literatura é essencialmente estrangeira, em idioma inglês, e foi obtida em grande parte pela oportunidade de acesso às bibliotecas da UIUC e da participação em palestras, aulas e debates sobre o LD. Mesmo em nível internacional, o fato de somente um trabalho ter sido identificado com o mesmo objetivo desta pesquisa confirma o quanto o tema é novo e que precisa ser melhor estudado e explorado.

Esta pesquisa motivou a continuidade de estudos relacionados à Web Semântica e as bibliotecas, mostrando-se um tema fértil e com vastas possibilidades de abordagens, como por exemplo:

- Possibilidade de utilização de elementos da Web Semântica para inclusão de dados dinâmicos nos catálogos;
- Inclusão de vínculo entre dados de um mesmo catálogo, estabelecendo relações entre registros presentes no acervo e destes com *datasets* externos;

- Possibilidades de utilização de outras ferramentas da Web Semântica em bibliotecas, como o *Knowledge graph*, por exemplo;
- Aprofundar estudos sobre enriquecimento e reuso de dados entre bibliotecas e a Web dos dados;
- Como motivar que bibliotecas disponibilizem seus catálogos em modelo RDF para que seus dados sejam consultados pela Web dos dados.

Esta pesquisa, portanto, almeja ter contribuído com a Ciência da Informação ao discutir possibilidades de utilização da Web Semântica em bibliotecas e os benefícios que podem advir desta iniciativa, como o saneamento de fragilidades decorrentes do licenciamento de livros digitais, aproveitamento assertivo dos recursos financeiros existentes e otimização do tempo investido pelos funcionários da biblioteca na coleta, controle e gestão de dados de licenciamentos. Benefícios sociais do modelo podem ser observados ao proporcionar condições de contratação de conteúdo digital de forma enxuta, sem desperdício de recursos financeiros ou redundância de material bibliográfico, alinhados com a demanda dos usuários identificada pela quantidade de acessos e solicitações realizados. A gestão de recursos financeiros proporcionada pelo modelo pode trazer benefícios na gestão de bibliotecas, além de melhor aproveitamento de recursos humanos e financeiros.

## Referências

---

ALLEMANG, D. Semantic Web and the linked data enterprise. In: WOOD, D. (ed.). **Linking enterprise data**. New York: Springer, 2010. pp. 3-23. DOI 10.1007/978-1-4419-7665-9

ALLEMANG, D.; HENDLER, J. **Semantic Web for the working ontologist: effective modeling in RDFS and OWL**. 2. ed. London: Elsevier, 2011. DOI: 10.1016/B978-0-12-385965-5.10001-9

ARMSTRONG, C.; LONSDALE, R. Introduction. In: PRICE, K.; HAVERGAL, V. (eds.). **E-books in libraries: a practical guide**. London, UK: Facet, 2011. p. xxi-xl.

ASSUMPCÃO, F. S. **Conversão de registros em XML para MARC 21: um modelo baseado em folhas de estilo XSLT**. 2013. 135 f. Dissertação (mestrado)-Universidade Estadual Paulista Júlio de Mesquita Filho, Faculdade de Filosofia e Ciências de Marília. Disponível em:

[https://repositorio.unesp.br/bitstream/handle/11449/93658/assumpcao\\_fs\\_me\\_mar.pdf?sequen ce=1&isAllowed=y](https://repositorio.unesp.br/bitstream/handle/11449/93658/assumpcao_fs_me_mar.pdf?sequen ce=1&isAllowed=y). Acesso em 16 out. 2018.

BEALL, J. Free books: loading brief MARC records for open-access books in an academic library catalog. **Cataloging & Classification Quarterly**, v. 47, p. 452-463, 2009. DOI 10.1080/01639370902870215. Acesso em: 1 fev. 2019.

BEEK, W et al. LOD Laundromat. In: RIETVELD, L. **Publishing and consuming linked data: optimizing for the unknown**. Berlim: AKA, 2016. Disponível em: [http://lodlaundromat.org/pdf/LOD\\_Laundromat\\_-\\_A\\_Uniform\\_Way\\_of\\_Publishing\\_Other\\_Peoples\\_Dirty\\_Data.pdf](http://lodlaundromat.org/pdf/LOD_Laundromat_-_A_Uniform_Way_of_Publishing_Other_Peoples_Dirty_Data.pdf). Acesso em: 22 jan. 2019.

BERMÈS, E. Enabling your catalogue for the semantic web. In: CHAMBERS, S. (Ed.). **Catalogue 2.0: the future of library catalogue**. Chicago: Neal-Schuman, 2013. p. 117-142.

BERNERS-LEE, T. Is your Linked Open Data 5 star? In: BERNERS-LEE, T. **Linked Data**. Cambridge: W3C, 2010. Disponível em: <https://www.w3.org/DesignIssues/LinkedData.html>. Acesso em 14 out. 2018.

BERNERS-LEE, T.; HENDLER, J.; LASSILA, O. The semantic web. **Scientific American**, New York, p. 24-30, May 2001. Disponível em: <https://www.scientificamerican.com/article/the-semantic-web/>. Acesso em: 27 jul. 2016.

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, 2011. DOI: 10.4018/978-1-60960-593-3.ch008. Disponível em: <http://tomheath.com/papers/bizer-heath-berners-lee-ijswis-linked-data.pdf>. Acesso em: 17 out. 2018.

BOJARS, U.; ZOGLA, A.; EGLÏTE, E. Linking library data for quality improvement and data enrichment. In: **Proceedings of 4th International Conference on Data Management**

**Technologies and Applications** (DATA 2015), p. 184-188, 2015. DOI: 10.5220/0005554901840188.

BOOK INDUSTRY STUDY GROUP. **Best practices for product metadata**: guide for North American Data senders and receivers. [S. L.]: BISG: Booknet Canada, 2015.

Disponível em:

[https://static1.squarespace.com/static/550334cbe4b0e08b6885e88f/t/55d2277be4b0a2c68568aaec/1439836027770/BISG\\_Best\\_Practices\\_for\\_Product\\_Metadata\\_6.1.15.pdf](https://static1.squarespace.com/static/550334cbe4b0e08b6885e88f/t/55d2277be4b0a2c68568aaec/1439836027770/BISG_Best_Practices_for_Product_Metadata_6.1.15.pdf). Acesso em: 1 fev. 2019.

BOUNTOURI, L.; GRATZ, P.; SANMARTIN, F. Digital preservation: how to be trustworthy. **Lecture Notes in Computer Science**, v. 10605, p. 364-374, 2018. Disponível em: [https://link.springer.com/content/pdf/10.1007%2F978-3-319-75826-8\\_29.pdf](https://link.springer.com/content/pdf/10.1007%2F978-3-319-75826-8_29.pdf) . Acesso em: 25 jan. 2019.

BRANAN, B.; FUTORNICK, M. **Linked data for production (LD4P)**. 21 out. 2018.

Disponível em: <https://wiki.duraspace.org/pages/viewpage.action?pageId=74515029>. Acesso em: 12 nov. 2018.

BRAND, A.; DALY, F.; MEYERS, B. **Metadata demystified**: a guide for publishers. [S. l.]: Sheridan: NISO, 2003. Disponível em:

[https://groups.niso.org/publications/press/Metadata\\_Demystified.pdf](https://groups.niso.org/publications/press/Metadata_Demystified.pdf). Acesso em: 1 fev. 2019.

BREITMAN, K. K.; CASANOVA, M. A.; TRUSZKOWSKI, W. **Semantic web**: concepts, technologies and applications. Cham: Springer, 2007. Disponível em:

<http://dx.doi.org/10.1007/978-1-84628-710-7>. Acesso em: 23 out. 2018.

BRICKLEY, D.; MILLER, L. **FOAF vocabulary specification 0.00**. 2014. Disponível em:

<http://xmlns.com/foaf/spec/>. Acesso em: 20 dez. 2018.

BYRNE, G. ; GODDARD, L. The strongest link: libraries and linked data. **D-Lib Magazine**, v. 16, n. 11/12, nov./dez. 2010. p. 1-9. Disponível em:

<http://www.dlib.org/dlib/november10/byrne/11byrne.html>. Acesso em 19 nov. 2018.

CHEN, M.; KIM, M.; MONTGOMERY, D. Ebook record management at The University of Texas at Dallas. **Technical Services Quarterly**, v. 33, n. 3, p. 251-267, 2016. DOI: 10.1080/07317131.2016.1169781. Acesso em: 30 jan. 2019.

CHEN, Y.-N. A review of practices for transforming library legacy records into linked open data. In: GAROUFALLOU, E. et al. (ed.) **Metadata and semantic research**. Cham: Springer, 2017. p. 123 -133. (Communications in computer and information science; 755)

COLE, T. W. et al. Library MARC records into linked open data: challenges and opportunities. **Journal of Library Metadata**, v. 13, p. 163-196, 2013. DOI:

0.1080/19386389.2013.826074

CONNAWAY, L. S.; POWEL, R. R. Research methods for librarians. 5. ed. [S.l.]: **ABC-Clio**, 2010. Disponível em:

<http://ebookcentral.proquest.com/lib/uiuc/detail.action?docID=617137>. Acesso em 15 out. 2018.

COUNTING ONLINE USER NETWORKED ELECTRONIC RESOURCES. **The COUNTER Code of practice for release 4**. 2019. Disponível em:

<http://www.projectcounter.org/code-of-practice-sections/general-information/>. Acesso em: 4 fev. 2019.

COYLE, K. Library data in the Web world. **Library Technology Reports**, Chicago: ALA, feb./mar. 2010, p. 5-11.

CREATIVE COMMONS. **About the licenses**. 2018a. Disponível em:

<https://creativecommons.org/licenses/>. Acesso em 20 dez. 2018.

CREATIVE COMMONS. **Share your work**. 2018. Disponível em:

<https://creativecommons.org/share-your-work/licensing-types-examples/>. Acesso em 1 nov. 2018.

CULBERTSON, B.; MANDELSTAM, Y.; PRAGER, G. **Provider-neutral e-monograph MARC record guide**. Washington, D.C.: Program for Cooperative Cataloging, 2011.

Disponível em: <https://www.loc.gov/aba/pcc/bibco/documents/PN-Guide.pdf>. Acesso em: 9 set. 2019.

DENAUX, R. et al. Knowledge architecture for organisations. PAN, J. Z. et al. (ed.).

**Exploiting linked data and knowledge graphs in large organizations**. Cham: Springer, 2017. p. 57- 94. Disponível em: <http://dx.doi.org/10.1007/978-3-319-45654-6>. Acesso em: 05 nov. 2018.

DEWEESE, K. P.; SEGAL, D. **Libraries and the semantic web**. Williston, VT: Morgan & Claypool, 2014. (Synthesis lectures on emerging trends in librarianship). DOI: 10.2200/S00615ED1V01Y201411ETL003

DUBLIN CORE METADATA INITIATIVE. **DCMI metadata terms**. 2012. Disponível em: <http://dublincore.org/documents/dcmi-terms/>. Acesso em: 21 dez. 2018.

EDITEUR. **EDItX sales and sales tax report**: format specification, version 1.2, October 2018. 2018. Disponível em: [https://www.editeur.org/files/EDItX%20-%20sales%20report/EDItX\\_Sales\\_SaleTax\\_v1.2.pdf](https://www.editeur.org/files/EDItX%20-%20sales%20report/EDItX_Sales_SaleTax_v1.2.pdf). Acesso em: 9 fev. 2019.

EDITEUR. **General book trade order format**: version 1.3, January 2010, corrected March 2010, February 2014. 2014. Disponível em: [https://www.editeur.org/files/EDItX%20-%20trade%20book/EDItX\\_Trade\\_Order\\_V1.3\(corr3\).pdf](https://www.editeur.org/files/EDItX%20-%20trade%20book/EDItX_Trade_Order_V1.3(corr3).pdf). Acesso em: 9 fev. 2019.

EDITEUR. **Onix**: overview. 2009. Disponível em: <https://www.editeur.org/83/Overview/>. Acesso em 25 jan. 2019.

EDITEUR. **Onix for books**: implementation and best practice guide: release 3.0 rev. 5, October 2018. 2018a. Disponível em:

[https://www.editeur.org/files/ONIX%203/ONIX\\_BookProduct\\_XSD\\_schema+codes\\_Issue\\_44.zip](https://www.editeur.org/files/ONIX%203/ONIX_BookProduct_XSD_schema+codes_Issue_44.zip). Acesso em: 9 fev. 2019.

EDITEUR. **Onix for books**: product information format specification, release 3.0 rev 1, January 2012. 2012. Disponível em:

[https://www.editeur.org/files/ONIX%203/ONIX\\_for\\_Books\\_Release3-0\\_docs+codes\\_Issue\\_44.zip](https://www.editeur.org/files/ONIX%203/ONIX_for_Books_Release3-0_docs+codes_Issue_44.zip). Acesso em 31 jan. 2019.

EDITEUR. **Onix for licensing terms**: ONIX-PL publications license format, version 1.0. 2008. Disponível em: <https://www.editeur.org/files/ONIX-PL/ONIX-PL%20Format%20specification%20v1.0.pdf>. Acesso em: 10 fev. 2019.

FREDERICK, D. E. **Managing ebook metadata in academic libraries**: taming the tiger [recurso eletrônico]. Waltham, MA: Chandos Publishing, 2015. (Chandos information professional series). Disponível em:

<http://www.library.illinois.edu/proxy/go.php?url=http://www.sciencedirect.com/science/book/9780081001516>. Acesso em 15 out. 2018

GEORGAS, H. The case of the disappearing e-book: academic libraries and subscription packages. **College & Research Libraries**, v. 77, n. 11, p. 883-898, 2015.

DOI:10.5860/crl.76.7.883.

GILLILAND, A. J. Setting the stage. In: BACA, M. (Ed.). **Introduction to metadata**. 2. ed. Los Angeles: J. Paul Getty Trust, 2008. p. 1-19. Disponível em:

<http://d2aohiyo3d3idm.cloudfront.net/publications/virtuallibrary/0892368969.pdf>. Acesso em 20 out. 2018.

GODBY, C. J. **A crosswalk from ONIX version 3.0 for books to MARC 21**. Ohio: OCLC, 2012. Disponível em: <http://www.oclc.org/research/publications/library/2012/2012-04a.xls>.

Acesso em: 17 dez. 2018.

GUARINO, N. Formal ontology and information systems. In: **Proceedings of the First International Conference on Formal Ontologies in Information Systems, FOIS'98**, Trento, Italy, 1998. pp. 3–15. Disponível em:

[https://klevas.mif.vu.lt/~donatas/Vadovavimas/Temos/OntologiskaiTeisingasKonceptinisModeliavimas/papildoma/Guarino98-](https://klevas.mif.vu.lt/~donatas/Vadovavimas/Temos/OntologiskaiTeisingasKonceptinisModeliavimas/papildoma/Guarino98-Formal%20Ontology%20and%20Information%20Systems.pdf)

[Formal%20Ontology%20and%20Information%20Systems.pdf](https://klevas.mif.vu.lt/~donatas/Vadovavimas/Temos/OntologiskaiTeisingasKonceptinisModeliavimas/papildoma/Guarino98-Formal%20Ontology%20and%20Information%20Systems.pdf). Acesso em: 19 out. 2018.

GUHA, R. V.; BRINCKLEY, D.; MACBETH, S. Schema.org: evolution of structure data on the Web. **Communications of the ACM**, v. 59, n. 2, p. 44-51, Feb., 2016. DOI:

10.1145/2844544. Disponível em: <http://delivery.acm.org/10.1145/2850000/>. Acesso em: 27 nov. 2018.

HILLERBRAND, E. Semantic web and business: reaching a tipping point? In: WORKMAN, M. (ed.). **Semantic Web**. Wilmette, IL: Springer, 2016. p.213-229.



HODGE, V.; MANOFF, M.; WATSON, G. Providing access to e-books and e-book collections: struggles and solutions. **The Serials Librarian**, v. 64, n. 1-4, p. 200-205, 2013. DOI: 10.1080/0361526X.2013.760411.

HYLAND, B. Preparing for linked data enterprise. In: WOOD, D. (ed.). **Linking enterprise data**. New York: Springer, 2010. pp. 51-64. DOI 10.1007/978-1-4419-7665-9.

INTERNATIONAL FEDERATION OF INTERNATIONAL FEDERATION OF LIBRARY ASSOCIATIONS. **Statement of International Cataloguing Principles**. 2009. Disponível em: [https://www.ifla.org/files/assets/cataloguing/icp/icp\\_2009-en.pdf](https://www.ifla.org/files/assets/cataloguing/icp/icp_2009-en.pdf). Acesso em: 14 dez. 2018.

JIN, Q.; SANDBERG, J. A. Implementing RDA at the University of Illinois at Urbana-Champaign library. **Technical Services Quarterly**, v. 31, n. 3, p. 217-236, 2014. DOI: 10.1080/07317131.2014.908585

KNUBLAUCH, H.; KONTOKOSTAS, D. **Shapes constraint language (SHACL)**. 2017. Disponível em: <https://www.w3.org/TR/shacl/>. Acesso em: 20 dez. 2018.

LIBRARY OF CONGRESS. **MARC standards**. 2018a. Disponível em: <http://www.loc.gov/marc/>. Acesso em: 15 dez. 2018.

LIBRARY OF CONGRESS. **PREMIS 3 ontology**. 2018b. Disponível em: <http://id.loc.gov/ontologies/premis-3-0-0.html>. Acesso em: 25 jan. 2019.

LINKED open vocabularies (LOV). 2018. Disponível em: <https://lov.linkeddata.es/dataset/lov>. Acesso em: 20 dez. 2018.

LÓSCIO, B. F. F.; BURLE, C.; CALEGARI, N. (eds). **Data on the Web best practices: W3C recommendation**. 31 jan. 2017. Disponível em: <https://www.w3.org/TR/dwbp/>. Acesso em: 12 dez. 2018

LUTHER, J. **Streamlining book metadata workflow**. Baltimore; Dublin : NISO : OCLC, 2009. Disponível em: [https://groups.niso.org/publications/white\\_papers/StreamlineBookMetadataWorkflowWhitePaper.pdf](https://groups.niso.org/publications/white_papers/StreamlineBookMetadataWorkflowWhitePaper.pdf). Acesso em: 1 fev. 2019.

MACEWAN, A. Project InterParty: from library authority files to e-commerce. **Cataloging & Classification Quarterly**, v. 39, n. 1-2, p. 429-442, 2004. DOI: 10.1300/J104v39n01\_11

MALMSTEN, M. **Exposing library data as linked data**. IFLA, 2009. Disponível em: [http://wtlab.um.ac.ir/images/e-library/linked\\_data/other/Exposing%20Library%20Data%20as%20Linked%20Data.pdf](http://wtlab.um.ac.ir/images/e-library/linked_data/other/Exposing%20Library%20Data%20as%20Linked%20Data.pdf). Acesso em 15 nov. 2018.

MARKSCHEFFEL, M. Semantic web and libraries. In: VISAKNI, P. et al. (eds). **Dynamics of library for excellence in electronic revolution**. Delhi: Bookwell, 2017. p. 41-51.

MARTIN, K. E. Cataloging ebooks: an overview of issues and challenges. **Against the grain**, v. 19, n. 1, p. 45-47, nov. 2007. Disponível em: <http://docs.lib.purdue.edu/cgi/viewcontent.cgi?article=5233&context=atg>. Acesso em: 13 dez. 2018.

MARTIN, K. E.; MUNDLE, K. Cataloging e-books and vendor records: a case study at the University of Illinois at Chicago. **Library Resources & Technical Services**, v. 54, n. 4, 2010. Disponível em: <https://journals.ala.org/lrts/article/viewFile/5585/6883>. Acesso em: 12 dez. 2018.

MCCLURE, C. R. Increasing the usefulness of research for library managers: propositions, issues, and strategies. **Library trends**, v. 38, n. 2, p. 280-294, 1989.

MCGEE, M.; WARNER, S. **Linked Data Notification (LDN) and Linked Data Fragments (LDF)**, 2018. Disponível em: <https://wiki.duraspace.org/pages/viewpage.action?pageId=101780973>. Acesso em 21 out. 2018.

MELO, J. O. de S. F. ; BOTEGA, L. C.; SANTARÉM SEGUNDO, J. E. Metodologia de avaliação de qualidade para dados conectados. *Informação & Tecnologia*, v. 4, n. 2, p. 80-101, jul./dez. 2017. Disponível em: <http://www.periodicos.ufpb.br/ojs/index.php/itec/article/view/40539/21041>. Acesso em: 22 jan. 2019.

MINCIC-OBRAĐOVIC, K. E-books in academic libraries. Oxford: Chandos, 2011. 203 p.

MILLER, E. The Semantic Web and digital libraries. DC 2004 / SILF 2004. Shanghai, China, 13 out. 2004. Disponível em: <http://dc2004.library.sh.cn/english/prog/ppt/talk.ppt>. Acesso em: 14 dez. 2018.

MUGRIDGE, R. L.; EDMUNDS, J. Using batchloading to improve access to electronic and microform collections. **Library Resources & Technical Services**, v. 53, n. 1, p. 53-61, 2009. Disponível em: <https://journals.ala.org/index.php/lrts/article/view/5461/6700>. Acesso em: 13 dez. 2018.

NATIONAL INFORMATION STANDARDS ORGANIZATION. **Demand Driven Acquisition of monographs**. Baltimore, MD: NISO, 2014. Disponível em: [http://groups.niso.org/apps/group\\_public/download.php/13373/rp-20-2014\\_DDA.pdf](http://groups.niso.org/apps/group_public/download.php/13373/rp-20-2014_DDA.pdf). Acesso em: 5 set. 2014.

NATIONAL INFORMATION STANDARDS ORGANIZATION. **NISO launches e-book special interest group**. 2011. Disponível em: <https://www.niso.org/press-releases/2011/05/niso-launches-e-book-special-interest-group>. Acesso em: 23 nov. 2018.

NATIONAL INFORMATION STANDARDS ORGANIZATION. **SUSHI FAQs**: general questions. 2010. Disponível em: <https://groups.niso.org/workrooms/sushi/faq/general/#q1>. Acesso em: 4 fev. 2019.

NATIONAL INFORMATION STANDARDS ORGANIZATION. **SUSHI FAQs**: librarian questions. 2009. Disponível em: <https://groups.niso.org/workrooms/sushi/faq/librarian>. Acesso em: 4 fev. 2019.

OCLC. **VIAF**: connect authority data across cultures and languages to facilitate research. 2018. Disponível em: <https://www.oclc.org/en/viaf.html>. Acesso em: 27 nov. 2018.

PROGRAM FOR COOPERATIVE CATALOGING. **Formulating and obtaining URIs**: a guide to commonly used vocabularies and reference sources. 14 Feb. 2018a. Disponível em: [https://www.loc.gov/aba/pcc/bibframe/TaskGroups/formulate\\_obtain\\_URI\\_guide.pdf](https://www.loc.gov/aba/pcc/bibframe/TaskGroups/formulate_obtain_URI_guide.pdf). Acesso em: 21 nov. 2018.

PROGRAM FOR COOPERATIVE CATALOGING. **MARC record guide for monograph aggregator vendors**. 2011. Disponível em: <http://www.loc.gov/aba/pcc/sca/documents/FinalVendorGuide.pdf>. Acesso em: 21 nov. 2018.

RAMALHO, R. A. S. **Desenvolvimento e utilização de ontologias em bibliotecas digitais**: uma proposta de aplicação. 2010. 145 f. Tese (Doutorado) - Universidade Estadual Paulista, Faculdade de Filosofia e Ciências, 2010. Disponível em: <http://hdl.handle.net/11449/103376>. Acesso em: 11 dez. 2018.

RAUPP, F. M.; BEUREN, I. M. Metodologia da pesquisa aplicável às ciências sociais. In: BEUREN, I. M. (org.). **Como elaborar trabalhos monográficos em contabilidade: teoria e prática**. 3. ed. São Paulo: Atlas, 2006. p. 76-97.

RAYLE, L. **Project description**: (LD4L Labs). 9 jun. 2016. Disponível em: [https://wiki.duraspace.org/pages/viewpage.action?pageId=76843838#\\_ftn6](https://wiki.duraspace.org/pages/viewpage.action?pageId=76843838#_ftn6). Acesso em: 13 nov. 2018.

REGISTER, R.; MCLROY, T. **The metadata handbook**: a book publisher's guide to creating and distributing metadata for print and ebooks. 2. ed. Columbus, OH: Datacurate, 2015.

RIETVELD, L. **Publishing and consuming linked data**: optimizing for the unknown. Berlin: AKA, 2016.

ROSSMANN, D.; FOSTER, A.; BABBITT, E. P. Ebook MARC records: do they make the mark? **Serials**, v. 22, n. 3 supplement, p. 546-550, nov. 2009. Disponível em: <https://scholarworks.montana.edu/xmlui/bitstream/handle/1/323/ebookMARC.pdf>. Acesso em: 12 dez. 2018.

SANTARÉM SEGUNDO, J. E. Web Semântica: fluxo para publicação de dados abertos e ligados. **Informação em Pauta**, Fortaleza, v. 3, n. especial, nov. 2018. Disponível em: <https://doi.org/10.32810/2525-3468.ip.v3iEspecial.2018.39721.117-140>. Acesso em: 12 dez. 2018.

SANTOS, P. L. V. A. da C.; ALVES, R. C. V. Metadados e Web Semântica para estruturação da Web 2.0 e Web 3.0. **DataGamaZero**: revista de ciência da informação, v. 10, n. 6, p. 1-9,

dez. 2009. Disponível em: <http://www.brapci.inf.br/index.php/article/download/52958>. Acesso em: 11 nov. 2018

SAPON-WHITE, R. E-book cataloging workflow at Oregon State University. **Library Resources & Technical Services**, v. 58, n. 2, 2014. Disponível em: <https://journals.ala.org/lrts/article/view/5335/6515>. Acesso em: 12 dez. 2018.

SARACEVIC, T. Interdisciplinary nature of information science. **Ciência da Informação**, v. 24, n. 1, 1995. Disponível em: <http://revista.ibict.br/ciinf/article/view/608>. Acesso em: 15 out. 2018.

SAYÃO, L. F.; SALES, L. F. Curadoria digital: um novo patamar para preservação de dados digitais de pesquisa. **Informação & Sociedade**, v. 22, n. 3, p. 179-191, set./dez., 2012. Disponível em: <http://www.periodicos.ufpb.br/ojs/index.php/ies/article/view/12224/8586>. Acesso em: 25 jan. 2019.

SCHEMA.ORG. **Book**. 2018. Disponível em: <https://schema.org/Book>. Acesso em: 18 dez. 2018.

SCHREURS, P. E.; LORIMER, N. Linked data in libraries' technical services workflows. In: GAROUFALLOU, E. et al. (ed.) **Metadata and semantic research**. Cham: Springer, 2017. p. 224-229. (Communications in computer and information science; 755).

SERRA, L. G. Bibliotecas e livros digitais: breve história e novos desafios. In: RIBEIRO, A. C. M. L.; FERREIRA, P. C. G. (orgs). **Biblioteca do século XXI: desafios e perspectivas**. Brasília: IPEA, 2017. Disponível em: [http://www.ipea.gov.br/portal/index.php?option=com\\_content&view=article&id=29215&Itemid=419](http://www.ipea.gov.br/portal/index.php?option=com_content&view=article&id=29215&Itemid=419). Acesso em: 22 abr. 2019.

SERRA, L. G.; SANTARÉM SEGUNDO, J. E. Sobre livros digitais e modelos de negócios. In **ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO**, 19., Londrina: ANCIB, 2018.

SERRA, L. G.; SILVA, J. F. M. da. Empréstimo de dispositivos de leitura em bibliotecas: análise de experiências. **Perspectivas em Ciência da Informação**, v. 22, n.2, p. 264-76, 2017. Disponível em: <http://dx.doi.org/10.1590/1981-5344/2701>. Acesso em: 1 fev. 2019.

SERRA, L. G.; SILVA, J. F. M. da. Livros digitais licenciados e os modelos de negócios transitórios. In **INTERNATIONAL CONFERENCE ON INFORMATION SYSTEMS AND TECHNOLOGY MANAGEMENT**, 13., São Paulo, 2016. p. 878-893. Disponível em: <http://contecsi.fea.usp.br/index.php/contecsi/13CONTECSI/paper/view/3786/2508>. Acesso em: 3 fev. 2019.

SHIEH, J.; REESE, T. The importance of identifiers in the new Web environment and using the Uniform Resource Identifier (URI) in subfield zero (\$0): a small step that is actually a big step. **Journal of Library Metadata**, v. 15, p. 208-226, 2015. DOI: 10.1080/19386389.2015.1099981.

SHOTTON, D.; PERONI, S. **FaBiO, the FRBR-aligned Bibliographic ontology**. 2018. Disponível em: <https://sparontologies.github.io/fabio/current/fabio.html>. Acesso em: 20 dez. 2018.

SINGER, R. Linked library data now! **Journal of Electronic Resources Librarianship**, v. 21, p. 114-126, 2009. Disponível em: <https://www.tandfonline.com/doi/pdf/10.1080/19411260903035809?needAccess=true>. Acesso em: 14 dez. 2018

SPURGIN, K. M. Getting what we paid for: a script to verify full access to e-resources. **Code4LIB Journal**, n. 25, July 2014. Disponível em: <https://journal.code4lib.org/articles/9684>. Acesso em: 12 dez. 2018.

SVENONIUS, E. **The intellectual foundation of information organization**. London, UK: MIT Press, 2001.

SVENSSON, L. G.; ATKINSON, R.; CAR, N. J. **Content negotiation by profile**. 2018. Disponível em: <https://www.w3.org/TR/2018/WD-dx-prof-conneg-20181218/#bib-rfc3986>. Acesso em: 20 dez. 2018.

THOMPSON, K.; TRAILL, S. Leveraging Python to improve ebook metadata selection, ingest, and management. **Code4lib Journal**, n. 38, out. 2017. Disponível em: <https://journal.code4lib.org/articles/12828>. Acesso em: 12 dez. 2018.

TRAILL, S. Quality issues in vendor-provided e-monograph records. **Library Resources & Technical Services**, v. 57, n. 4, 2013. Disponível em: <https://journals.ala.org/index.php/lrts/article/view/5589>. Acesso em: 12 dez. 2018.

USE CASES for Linked Data Notifications in the library contexto. 2018. Disponível em: [https://docs.google.com/document/d/124FAYW73AXGVxFL7I1nUtBeLObSge\\_6o8he2jHdGvUo](https://docs.google.com/document/d/124FAYW73AXGVxFL7I1nUtBeLObSge_6o8he2jHdGvUo). Acesso em 22 out. 2018.

VAN KLEECK, D. et al. Managing bibliographic data quality for electronic resources. **Cataloguing & Classification Quarterly**, v. 55, n. 7-8, p. 560-577, 2017. DOI: 10.1080/01639374.2017.1350777.

VILLAZON-TERRAZAS, B. et al. Construction of enterprise knowledge graphs I. In: PAN, J. Z. et al. (ed.). **Exploiting linked data and knowledge graphs in large organizations**. Cham: Springer, 2017a. p. 87- 116. Disponível em: <http://dx.doi.org/10.1007/978-3-319-45654-6>. Acesso em: 29 out. 2018.

VILLAZON-TERRAZAS, B. et al. Knowledge graph foundations. In: PAN, J. Z. et al. (ed.). **Exploiting linked data and knowledge graphs in large organizations**. Cham: Springer, 2017. p. 17- 55. Disponível em: <http://dx.doi.org/10.1007/978-3-319-45654-6>. Acesso em: 29 out. 2018.

WALLIMAN, N. Qualitative data analysis. In: WALLIMAN, N. **Social research methods**. London: Sage, 2011. p. 129-147. DOI: <http://dx.doi.org/10.4135/9781849209939>

WALTERS, W. H. E-books in academic libraries: challenges for discovery and access. **Serials Review**, v. 39, n. 2, p. 97-104, 2013. Disponível em: <https://doi.org/10.1080/00987913.2013.10765501>. Acesso em: 12 dez. 2018.

WALTERS, W. H. E-books in academic libraries: challenges for sharing and use. **Journal of Librarianship and Information Science**, v. 46, n. 2, p. 85-95, 2014. <http://www.doi.org/10.1177/0961000612470279>. Acesso em: 30 jan. 2019.

WESSELS, B. et al. **Open data and the knowledge society**. Amsterdam: Amsterdam University Press, 2017.

WIKOFF, K. **Electronics resources management in the academic library: a professional guide**. Santa Barbara, CA: Libraries Unlimited, 2012.

WONG, E. Y. LD4L: linked data for libraries. **Technical Services Quarterly**, v. 34, n. 3, p. 332-333, 2017. DOI: 10.1080/07317131.2017.1321405. Disponível em: <https://www.tandfonline.com/doi/pdf/10.1080/07317131.2017.1321405?needAccess=true>. Acesso em: 14 out. 2018.

WOODS, A.; WARNER, S. **Linked Data for Libraries Labs (LD4L Labs)**, 2018. Disponível em: <https://wiki.duraspace.org/display/ld4LABS>. Acesso em: 22 out. 2018.

WORLD WIDE WEB CONSORTION. **Introduction to SKOS**. 2012. Disponível em: <https://www.w3.org/2004/02/skos/intro>. Acesso em: 20 dez. 2018.

WORLD WIDE WEB CONSORTION. **RDF vocabularies current status**. 2018. Disponível em: [https://www.w3.org/standards/techs/rdfvocab#w3c\\_all](https://www.w3.org/standards/techs/rdfvocab#w3c_all). Acesso em: 20 dez. 2018.

WORLD WIDE WEB CONSORTIUM. **RDF Primer**. Cambridge, 2014. Disponível em: <https://www.w3.org/TR/2014/NOTE-rdf11-primer-20140624/>. Acesso em: 22 out. 2018.

WORLD WIDE WEB INCUBATOR GROUP REPORT. **Library Linked Data Incubator Group final report**. 25 out. 2011. Disponível em: <https://www.w3.org/2005/Incubator/lld/XGR-lld-20111025/>. Acesso em: 14 nov. 2018.

WU, A.; MITCHELL, A. M. Mass management of e-book catalog records. **Library Resources & Technical Services**, v. 54, n. 3, 2010. Disponível em: <https://journals.ala.org/lrts/article/view/5558/6841>. Acesso em: 12 dez. 2018.

WU, H. et al. Understanding knowledge graphs. In: Pan, J. Z. et al. (ed.). **Exploiting linked data and knowledge graphs in large organizations**. Cham: Springer, 2017. pp. 147-180. Disponível em: <http://dx.doi.org/10.1007/978-3-319-45654-6>. Acesso em: 30 out. 2018.

ZHANG, L.; JIN, M. Cataloging e-books: dealing with vendors and various other problems. **The Serials Librarian**, v. 67, p. 76-80, 2014. DOI 10.1080/0361526X.2014.899295. Acesso em: 4 fev. 2019.

ZHANG, Y. E-books for academic libraries in the USA. In: **2018 5th International Symposium on Emerging Trends and Technologies in Libraries and Information**

**Services** (ETTLIS), Noida, India: IEEE, 2018. p. 7-11. DOI: 10.1109/ETTLIS.2018.8485208.  
Acesso em: 4 fev. 2019.