




MODELOS SEMÂNTICOS PARA DADOS BIBLIOGRÁFICOS DE PUBLICAÇÕES CIENTÍFICAS DISPONIBILIZADOS COMO *LINKED DATA*

SEMANTIC MODELS FOR BIBLIOGRAPHIC DATA OF SCIENTIFIC PUBLICATIONS MADE AVAILABLE AS LINKED DATA

 Antonio Victor Wolf Tadini¹

 José Eduardo Santarem Segundo²

¹ Graduado em Biblioteconomia e Ciência da Informação pela Universidade de São Paulo (USP).


E-mail: antoniovwt@gmail.com

² Doutor em Ciência da Informação pela Universidade Estadual Paulista Júlio de Mesquita (UNESP). Professor no Departamento de Educação, Informação e Comunicação da Universidade de São Paulo (USP).

E-mail: santarem@usp.br



ACESSO ABERTO

Copyright: Esta obra está licenciada com uma Licença Creative Commons Atribuição 4.0 Internacional. 

Conflito de interesses: Os autores declaram que não há conflito de interesses.

Financiamento: Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP)

Declaração de disponibilidade dos dados: Todos os dados relevantes estão disponíveis neste artigo.

Recebido em: 07/11/2019.

Aceito em: 27/11/2019.

Revisado em: 20/12/2019.

Como citar este artigo:

TADINI, Antonio Victor Wolf; SANTAREM SEGUNDO, José Eduardo. Modelos semânticos para dados bibliográficos de publicações científicas disponibilizados como *Linked Data*. **Informação em Pauta**, Fortaleza, v. 4, n. 2, p. 32-57, jul./dez. 2019. DOI: [10.32810/2525-3468.ip.v4i2.2019.42640.32-57](https://doi.org/10.32810/2525-3468.ip.v4i2.2019.42640.32-57).

RESUMO

Esta pesquisa aborda a oportunidade de aumento do impacto de publicações científicas com a disponibilização dos dados de seus registros bibliográficos como *Linked Data*. Objetiva investigar quais são as melhores práticas para tal, bem como delinear um modelo semântico para dados bibliográficos de publicações científicas. A metodologia é exploratória e descritiva. Como resultado, apresenta um conjunto de declarações RDF, que estruturam o modelo semântico. Conclui que a representação bibliográfica das publicações científicas no *Linked Data* valoriza a coexistência de vários vocabulários para descrição em um mesmo registro.

Palavras-chave: Web Semântica. *Linked Data*. Vocabulários. Metadados. Dados Bibliográficos. Comunicação Científica.

ABSTRACT

This research addresses the opportunity to increase the impact of scientific publications with the availability of data from their bibliographic records as *Linked Data*. It aims to investigate what are the best practices for this, as well as to outline a semantic model for bibliographic data of scientific publications. The methodology is exploratory and descriptive. As a result, it presents a set of RDF statements, which structure the semantic model. It concludes that the bibliographic representation of scientific publications in *Linked Data* values the coexistence of several vocabularies for description in the same record.

Keywords: Semantic Web. *Linked Data*. Vocabularies. Metadata. Bibliographic Data. Scholarly Communication.

1 INTRODUÇÃO

Disponibilizar os metadados das publicações científicas, de modo aberto, já implica em importante resultado para a visibilidade de tais documentos. Não obstante, é possível ir além: a conexão desses metadados, conforme propõem as técnicas de *Linked Data*, ao “único espaço de dados global” referido por Heath e Bizer (2011), é uma ação capaz de potencializar exponencialmente o impacto do que é publicado. Afinal, os metadados conduzem à publicação e, conforme a Web Semântica planejou e, atualmente, permite, esses dados podem ser processados automaticamente (BERNERS-LEE, 2006).

Proceder a percepções desse tipo, isto é, da conjuntura histórica em que a comunidade científica se insere, bem como das tecnologias disponíveis e das fronteiras de todo tipo que podem ser transpostas em dada conjuntura, é atribuição da Ciência da Informação. Com isso, ela se torna mais capaz de estabelecer condições para boas práticas de comunicação científica, organizando conhecimento e informação para que as publicações possam ser acessadas de modo produtor para o avanço da ciência e da sociedade. Vale notar que, nesse contexto, o impacto das publicações é cada vez mais uma questão central para a comunidade científica – entre outros motivos, por ser um parâmetro indispensável para que se dimensione a influência de um pesquisador.

Este artigo é, inevitavelmente, uma sintetização de uma pesquisa de grande porte e capaz de expandir fronteiras de conhecimento sob a perspectiva da Ciência da Informação. Estrutura-se da seguinte maneira: o referencial teórico esclarece o que significa publicar como *Linked Data* e as tecnologias necessárias, destacando-se o RDF; em seguida, são descritos os procedimentos metodológicos da pesquisa, que são divididos em três tópicos; os resultados são apresentados fundamentalmente por meio de quadros e de uma figura; a discussão, então, analisa os resultados obtidos e tece reflexões a partir deles; e, por fim, a conclusão alinha de modo mais abstrato as questões de pesquisa com o ponto a que se chegou com ela, avaliando seu êxito com base no objetivo estabelecido.

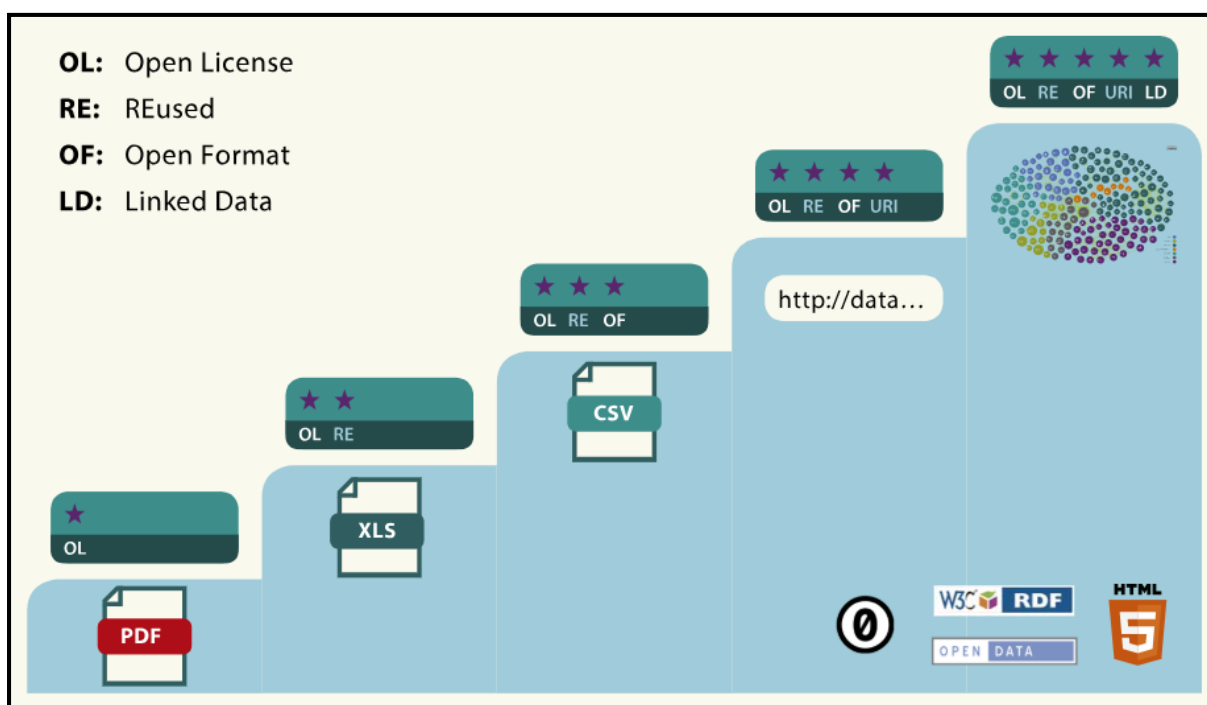
O objetivo desta pesquisa consiste em investigar quais são, para comunidades que lidam com metadados de registros bibliográficos referentes a documentos tidos como resultados de comunicação científica, as melhores práticas em uso no âmbito do *Linked Data*, para que, a partir disso, seja possível delinear um modelo semântico para registros

bibliográficos de publicações científicas, estabelecendo vocabulários e declarações RDF que nele poderiam figurar.

2 REFERENCIAL TEÓRICO

As iniciativas que lidam com metadados de registros bibliográficos e os publicam de modo aberto, mas não ainda de acordo com as práticas que caracterizam o *Linked Data*, encontram-se na posição de três estrelas na escalada proposta por Tim Berners-Lee, conforme apresentado na Figura 1.

Figura 1 – Percurso de requisitos para as cinco estrelas de Berners-Lee

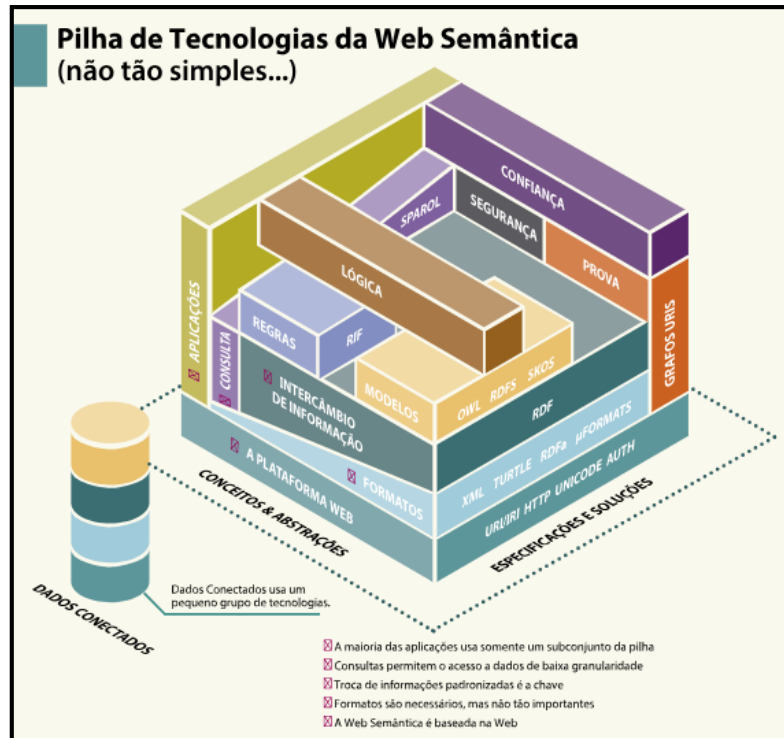


Fonte: Isotani e Bittencourt (2015, item 2.2.1.).

Estar na posição de três estrelas significa que: os dados estão publicados na Web sob uma licença aberta (estrela 1); estão estruturados de modo a serem legíveis por máquina (estrela 2); e a estruturação que confere a estrela 2 é feita utilizando um formato não proprietário (estrela 3). No entanto, os dados com três estrelas, para atingir as estrelas 4 e 5, necessitam estar sob a utilização de identificadores únicos e conectados a outros *datasets*, isto é, conjuntos de dados de outras iniciativas.

Para tanto, convém observar que basta ser utilizada apenas uma parte do *layer cake* que estrutura a Web Semântica. Corresponde, na Figura 2, à relativamente pequena pilha à esquerda, identificada como Dados Conectados e composta pelas camadas denominadas: Plataforma Web, Formatos, Intercâmbio de Informação e Modelos.

Figura 2 – Tecnologias da Web Semântica (*layer cake*)



Fonte: Isotani e Bittencourt (2015, item 1.3.).

Em termos gerais, esse “*layer cake* mínimo” se constitui da seguinte maneira: as entidades e relacionamentos devem estar identificados por um URI (ex: cada artigo científico é uma entidade com um URI); um formato fornece a sintaxe; o RDF estrutura os dados em triplas (recurso-propriedade-valor, ou sujeito-predicado-objeto), conectando esses dados entre si e a outros conjuntos de dados; e, por fim, são acoplados vocabulários genéricos, como OWL e RDFS, permitindo que vocabulários mais específicos a cada comunidade sejam posteriormente também acoplados.

O uso do RDF é fundamental para o *Linked Data*: promove sua estruturação ao mesmo tempo em que garante interoperabilidade, pois consegue ocupar, em termos funcionais e ainda que operando em um nível mais abstrato, o espaço que a padronização dos dados costumava ocupar. Nesse sentido, sabe-se que:

No domínio bibliográfico, em especial, nos processos descritivos da catalogação, sempre se buscou a utilização de padrões de metadados universais, de modo a facilitar o intercâmbio e o reuso de dados. **No âmbito da Web Semântica, a utilização de um mesmo padrão de metadados, ou melhor, de um mesmo vocabulário de classes e de propriedades, deixa de ser o principal requisito para o intercâmbio e o reuso dos dados. Ocupando essa posição, encontra-se o modelo de dados RDF** que, por ter como unidade básica uma tripla, possibilita a descrição de recursos utilizando, simultaneamente, classes e propriedades de diferentes vocabulários, assim como possibilita, a partir de recursos do RDF Schema (RDFS) e da Web Ontology Language (OWL), um “caminhar” por entre essas classes e propriedades para a realização de inferências acerca dos dados (ASSUMPCÃO; SANTOS, 2016, p. 526, grifo nosso).

O RDF, desse modo, faz com que exista hoje uma tendência à “[...] coexistência de vários formatos de metadados, pois cada um apresenta uma característica específica que atende à necessidade de descrição de determinados tipos de recursos informacionais e comunidades de interesses distintos” (SANTOS; ALVES, 2009 apud ASSUMPCÃO; SANTOS, 2016, p. 526). Assim, é fundamental para esta pesquisa, a fim de atingir seu objetivo, desenvolver subsídios para que se obtenha conhecimento amplo e qualificado sobre como é a necessidade de descrição das publicações no atual contexto da comunidade científica.

3 METODOLOGIA

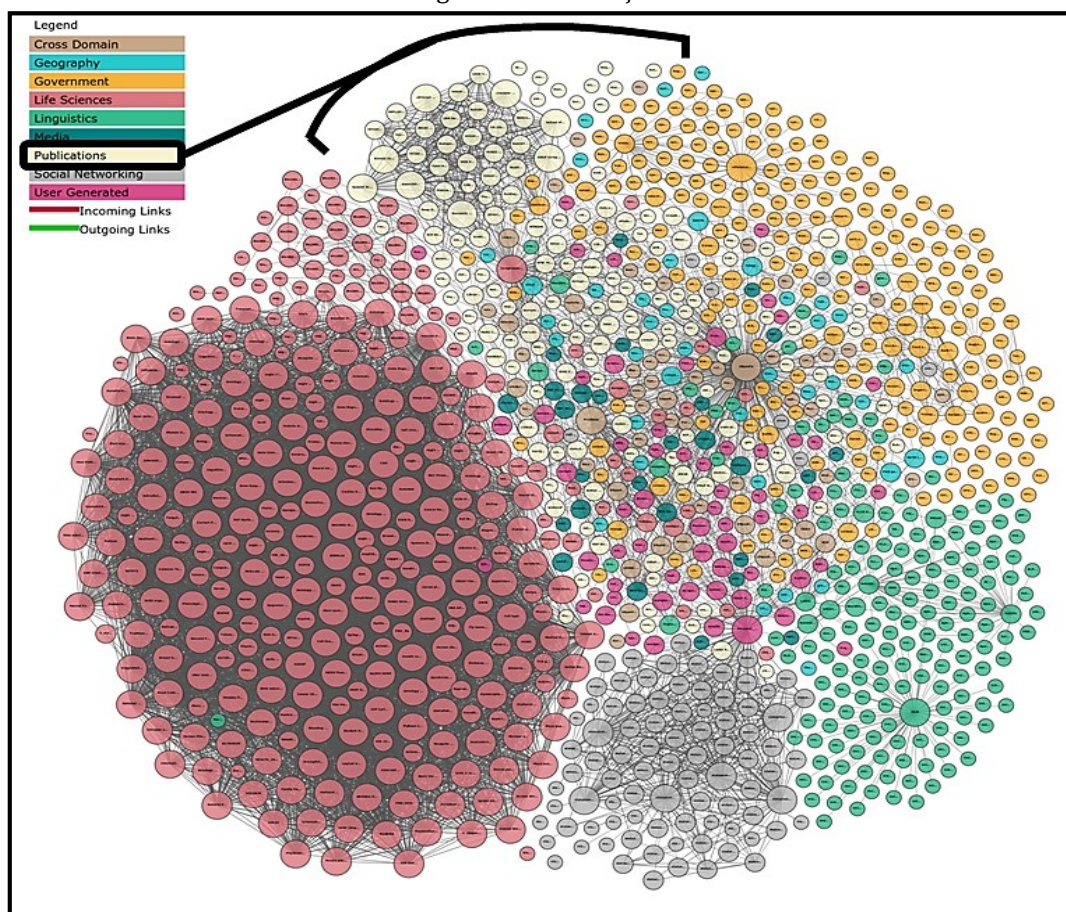
Trata-se de uma pesquisa de caráter exploratório e descritivo, que consiste em realizar um diagnóstico de *datasets* publicados no Linked Open Data ¹ (diagrama publicado em 22/08/2017) dentro do segmento temático “Publicações”, identificando as principais estruturas utilizadas para formalização dos registros bibliográficos referentes a publicações científicas. Além disso, destacou-se a partir das etapas iniciais desse diagnóstico um modelo semântico específico, da British Library, para registros bibliográficos referentes a publicações seriadas, para ser descrito neste estudo.

3.1 Seleção dos *datasets* do segmento “Publicações”

¹ A partir do conjunto de melhores práticas que caracterizam o *Linked Data*, e por iniciativa de um grupo formalizado junto ao W3C, foi criado o Linked Open Data (LOD), um projeto que teve como objetivo estimular a publicação de dados abertos em formato semântico. O grupo foi finalizado, no entanto algumas dessas pessoas resolveram manter o projeto, que tem crescido significativamente nos últimos anos. Os dados do LOD estão organizados em *datasets*, que são classificados em domínios que apontam a temática de tais dados. Essa divisão em domínios temáticos é feita desde o diagrama publicado em março de 2009. Dentre esses domínios, destaca-se o denominado “Publicações”, que abarca conteúdos acerca de publicações científicas e conferências científicas, bem como sistemas de organização do conhecimento, listas de leitura de universidades, *datasets* de bibliotecas e identificadores de autoridade.

Inicialmente, enquanto pesquisa exploratória, cumpria a esta etapa enfrentar um importante volume de dados, que demandou operações de seleção de modo a reduzi-lo progressivamente. Assim, a exploração iniciou-se com a contabilização, por meio do diagrama presente na Figura 3, dos *datasets* do segmento “Publicações” do objeto de estudo, que foram dispostos em uma planilha, totalizando 156 *datasets*. O seu processo de seleção envolve, a priori, o exame dos recursos disponibilizados na plataforma DataHub², catálogo no qual cada *dataset* possui uma página a ele relativa. Tais recursos são disponibilizados em diversos formatos, muitos deles como exemplos concretos, e isso aumenta a dificuldade da exploração, visto que o objeto de estudo é bastante heterogêneo. Além disso, investiga-se a fonte dos dados, com interface para humanos, cuja referência é dada pela plataforma.

Figura 3 – Diagrama do Linked Open Data publicado em 22/08/2017, com destaque para os *datasets* do segmento “Publicações”



² <https://old.datahub.io/>

Fonte: Adaptado de Abele *et al.* (2017).

Tadini, Coneglian e Santarem Segundo (2017) apresentaram pesquisa que caracteriza este segmento temático, “Publicações”, também por estudo exploratório, com base no diagrama Linked Open Data de 2014. A categorização dos *datasets* realizada pelos autores é baseada em Schmachtenberg, Bizer e Paulheim (2014), de modo a estabelecer as seguintes categorias: *datasets* de biblioteca; *datasets* de informações sobre publicações científicas; *datasets* de informações sobre conferências científicas; sem acesso; não classificados; sistemas de organização do conhecimento (SOCs); listas de leitura de universidades; e identificadores de autoridade.

Note-se que muitos dos *datasets* permanecem no diagrama atual, com ou sem atualização. Assim, foi realizada uma comparação cautelosa com o estudo de Tadini, Coneglian e Santarem Segundo (2017), para aproveitar as informações já delineadas acerca dos *datasets* que permaneceram, sobretudo quanto à sua categorização, tendo sido consideradas as atualizações.

Foram confrontados os 156 *datasets* do segmento “Publicações” do diagrama de 22/08/2017 com os 133 *datasets* presentes no de 2014. Uma primeira análise buscou perceber se o *dataset* se repetia com base no seu nome, 47 foram detectados, preliminarmente. Para aqueles em que houve dúvida, uma segunda análise comparou os registros dos *datasets* nos catálogos do *DataHub* de 2014 e de agosto de 2017. Foram detectados, então, mais 39, totalizando 86.

Para que fosse executada a seleção, primeiramente foram adotadas como “naturezas” interessantes ao escopo desta pesquisa: *datasets* de biblioteca, *datasets* de informações sobre publicações científicas, e *datasets* de informações sobre conferências científicas. Consequentemente, foram descartados aqueles que, sendo repetidos de 2014, fossem enquadrados nas demais “naturezas”. Assim, foram descartados 30 *datasets*, de modo que restaram 56 repetidos de 2014, e mais 70 novos, sem qualquer descrição até então.

Desse modo, o próximo passo foi descrevê-los – minimamente, neste momento. Com o mesmo critério do descarte realizado anteriormente, mais 46 *datasets* foram desprezados, de modo que foram selecionados 24, que se somaram aos 56 anteriormente selecionados. No total, foram selecionados, preliminarmente, 80 *datasets* de interesse.

Desses 80, 24 foram enquadrados como de biblioteca, 54 como de informações sobre publicações científicas, e apenas 2 como de informações sobre conferências científicas.

3.2 Descrição dos *datasets* selecionados quanto ao uso de vocabulários

O estudo descritivo que se inicia neste ponto propõe-se, então, a descrever os *datasets* de acordo com o escopo da pesquisa. Para que isso fosse feito, aprimorou-se a seleção qualitativa operada na pesquisa exploratória. Descartaram-se 7 dos 80 *datasets* selecionados anteriormente, com base no seguinte critério: a descrição almejada requiritava que fosse possível examinar se os chamados vocabulários – tecnologias recorrentes no âmbito da Web Semântica e determinantes na formalização de metadados – estavam sendo utilizados ou não em cada conjunto de dados. Nos 7 *datasets* descartados não foi possível tal exame.

Restaram, assim, 73 (19 de biblioteca, 53 de informações sobre publicações científicas e 1 de informações sobre conferências científicas), fechando-se o conjunto de *datasets* a serem descritos.

A identificação de vocabulários se compromete a ser rica, mas não exaustiva. Para que houvesse método, utilizou-se o Linked Open Vocabularies (LOV)³, ferramenta adotada por Rozsa, Dutra e Nhacuongue (2017), bem como por Freitas Junior e Jacynto (2016). Registraram-se apenas vocabulários que nele estivessem presentes, com poucas exceções, que foram marcadas quando aplicadas. Segundo as referências supracitadas, o LOV é adotado pelo W3C – organização que administra e fomenta a Web Semântica – como ponto central para informações sobre vocabulários.

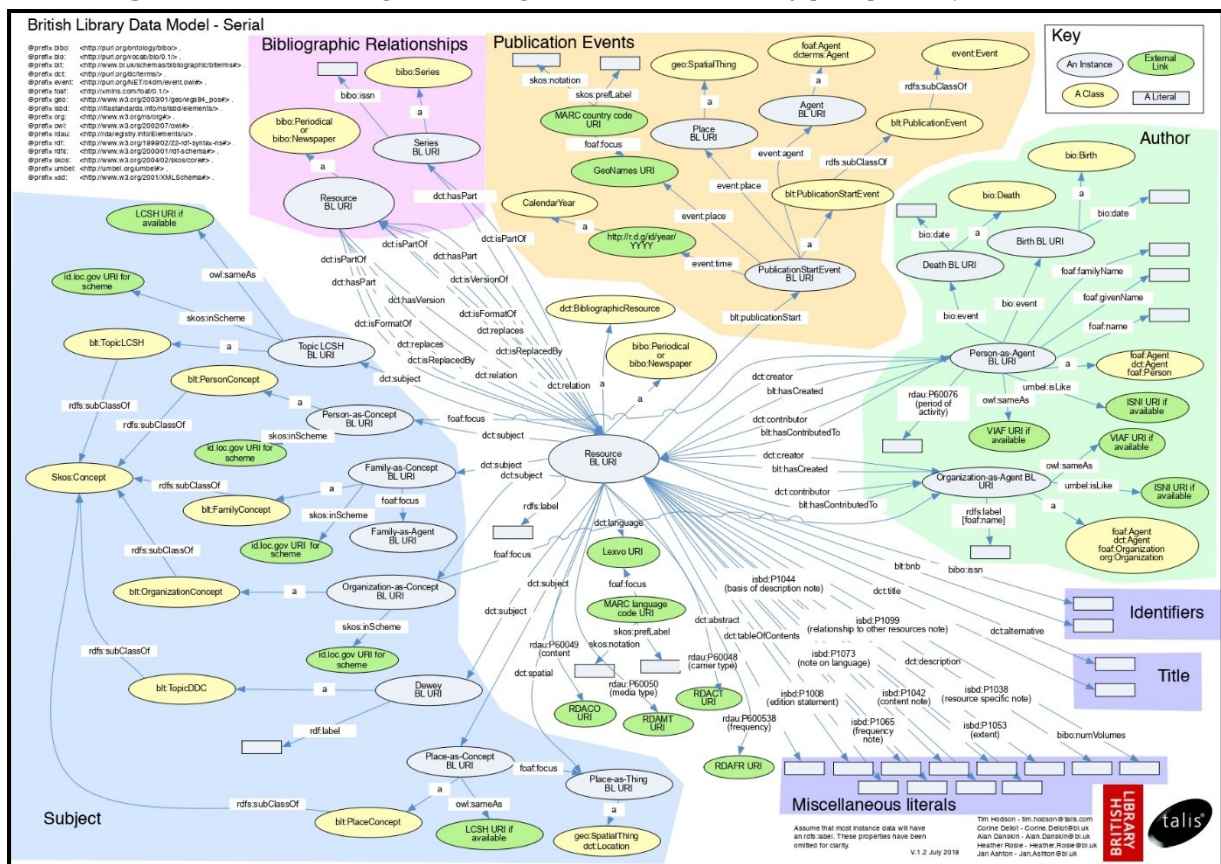
Os vocabulários foram, ao longo da descrição, separados por serem tipicamente bibliográficos ou não. Pontua-se que os vocabulários da iniciativa *Dublin Core* (DCTERMS e DCE), apesar de serem amplamente utilizados no contexto bibliográfico, foram categorizados como não tipicamente bibliográficos, visto que sua aplicação transpõe esse contexto alcançando os mais variados tipos de objetos digitais.

³ <https://lov.linkeddata.es/dataset/lov>

3.3 Modelo de registro bibliográfico da British Library para publicações seriadas: Estudo Descritivo

Nesta fase, desenvolve-se um estudo descritivo de uma estrutura em específico identificada no diagnóstico decorrente das atividades conforme itens anteriores, que consiste em um dos documentos presentes no registro do *dataset* da British Library no *DataHub*. É um modelo semântico de registro bibliográfico utilizado pela biblioteca para publicações seriadas, como periódicos científicos, que corresponde à Figura 4.

Figura 4 – Modelo de registro bibliográfico da British Library para publicações seriadas



Fonte: British Library e Talis (2018).

Como é possível verificar, o diagrama é dividido em 7 regiões (mais 1 espaço sem denominação); são elas: Autor, Identificadores, Título, Relacionamentos Bibliográficos, Eventos de Publicação, Assunto e Miscelânea. Os vocabulários utilizados no modelo são: BIBO, BIO, BLT (BLTERMS), DCTERMS, EVENT, FOAF, GEO, ISBD, ORG, OWL, RDAU, RDF, RDFS, SKOS, UMBEL e XSD.

4 RESULTADOS

No primeiro estudo descritivo (correspondente ao item 3.2, da Metodologia), identificou-se pelo menos um vocabulário tipicamente bibliográfico em 28 (38,4%) do total de *datasets* selecionados. Foram identificados os seguintes vocabulários tipicamente bibliográficos: BIBO, vocabulários RDA, vocabulários MARC, ISBD, FRBR e FRBRER, BLT (BLTERMS), GNDO, BIBTEX, BF (BIBFRAME), LIBRIS, BNF, SUDOC, vocabulários DNB e TEF.

A frequência de ocorrência ou não desses vocabulários nos *datasets* está disposta no Quadro 1. Como opção metodológica, os números não inteiros servem para resolver o problema de um *dataset* ser entendido como pertencente a mais de uma “natureza”, ou seja, se um *dataset* é de biblioteca, mas também de informações sobre publicações científicas, e possui o vocabulário BIBO na sua formalização, por exemplo, confere-se 0,5 ocorrência para uma “natureza” e 0,5 para a outra. As porcentagens servem para reconhecer o total de *datasets* de cada “natureza”, e também o total de *datasets* selecionados.

Quadro 1 – Frequência de vocabulários considerados tipicamente bibliográficos nos *datasets* selecionados

VOCABULÁRIO	Está no LOV?	FREQUÊNCIA							
		Datsets de biblioteca		Datsets de informação sobre publicações científicas		Datsets de informação sobre conferências científicas		TOTAL	
		nº	%	nº	%	nº	%	nº	%
BIBO	Sim	6,5	34,2%	13,5	25,5%	0	0%	20	27,4%
Vocabulários RDA	Parte	8	42,1%	1	1,9%	0	0%	9	12,3%
Vocabulários MARC	Parte	7,5	39,5%	1,5	2,8%	0	0%	9	12,3%
ISBD	Sim	5,5	28,9%	2,5	4,7%	0	0%	8	11%
FRBR e FRBRER	Sim	4,5	23,7%	1,5	2,8%	0	0%	6	8,2%
BLT (BLTERMS)	Sim	2	10,5%	0	0%	0	0%	2	2,7%
GNDO	Sim	1	5,3%	1	1,9%	0	0%	2	2,7%
BIBTEX	Sim	0	0%	1	1,9%	1	100%	2	2,7%
BF (BIBFRAME)	Sim	1	5,3%	0	0%	0	0%	1	1,4%
LIBRIS	Não	1	5,3%	0	0%	0	0%	1	1,4%
BNF	Não	1	5,3%	0	0%	0	0%	1	1,4%
SUDOC	Não	1	5,3%	0	0%	0	0%	1	1,4%
Vocabulários DNB	Parte	0,5	2,6%	0,5	0,9%	0	0%	1	1,4%

TEF	Não	0	0%	1	1,9%	0	0%	1	1,4%
-----	-----	---	----	---	------	---	----	---	------

Fonte: Elaborado pelos autores.

Parte-se, então, para a descrição dos *datasets* quanto a vocabulários não tipicamente bibliográficos, o que consiste em desafio maior. A frequência considerando os 73 *datasets* selecionados está disposta no Quadro 2 (a seguir).

Quadro 2 – Frequência de vocabulários considerados não tipicamente bibliográficos nos *datasets* selecionados

VOCABULÁRIO	Está no LOV?	FREQUÊNCIA							
		Datasets de biblioteca		Datasets de informação sobre publicações científicas		Datasets de informação sobre conferências científicas		TOTAL	
		nº	%	nº	%	nº	%	nº	%
FOAF	Sim	15	78,9%	43	81,1%	0	0%	58	79,5%
VOID	Sim	7	36,8%	41	77,4%	0	0%	48	65,8%
DCE	Sim	13	68,4%	34	64,2%	0	0%	47	64,4%
DCTERMS	Sim	10	52,6%	36	67,9%	0	0%	46	63%
RDFS	Sim	15	78,9%	20	37,7%	1	100%	36	49,3%
AKT	Sim	0	0%	29	54,7%	0	0%	29	39,7%
OWL	Sim	12	63,2%	11	20,8%	0	0%	23	31,5%
SKOS	Sim	9,5	50%	8,5	16%	0	0%	18	24,7%
XSD	Sim	7,5	39,5%	8,5	16%	0	0%	16	21,9%
GEO	Sim	5	26,3%	4	7,5%	0	0%	9	12,3%
SCHEMA	Sim	2,5	13,2%	4,5	8,5%	0	0%	7	9,6%
SCOVO	Sim	1	5,3%	6	11,3%	0	0%	7	9,6%
VIVO	Sim	0,5	2,6%	5,5	10,4%	0	0%	6	8,2%
DCMITYPE	Sim	3,5	18,4%	1,5	2,8%	0	0%	5	6,8%
PRV	Sim	2	10,5%	3	5,7%	0	0%	5	6,8%
GEONAMES	Sim	3	15,8%	1	1,9%	0	0%	4	5,5%
DBPEDIA	Sim	2	10,5%	2	3,8%	0	0%	4	5,5%
EVENT	Sim	2	10,5%	2	3,8%	0	0%	4	5,5%
DOAP	Sim	1	5,3%	3	5,7%	0	0%	4	5,5%
PROV	Sim	1	5,3%	3	5,7%	0	0%	4	5,5%
SD	Sim	1	5,3%	3	5,7%	0	0%	4	5,5%
UMBEL	Sim	2,5	13,2%	0,5	0,9%	0	0%	3	4,1%
WDRS	Sim	2,5	13,2%	0,5	0,9%	0	0%	3	4,1%
ORG	Sim	1	5,3%	2	3,8%	0	0%	3	4,1%
PRVTYPES	Sim	1	5,3%	2	3,8%	0	0%	3	4,1%
CRSW	Sim	0	0%	3	5,7%	0	0%	3	4,1%
DCAT	Sim	0	0%	3	5,7%	0	0%	3	4,1%
SIO	Sim	0	0%	3	5,7%	0	0%	3	4,1%
SWRC	Sim	0	0%	3	5,7%	0	0%	3	4,1%

Vocabulários com 2 ocorrências: BIO, GR, IRW, LIDO, META, OBORO, OPM, ORE, OV, PML, PMLJ, PMLP, RESEX, SIOC, TAG, UNITS, VANN, VSR e WV.
Vocabulários com 1 ocorrência: API, B3KAT, BFLC, BTE, CC, CON, CONTACT, CONV, COREF, D2R, DATAFAQS, DATE, DM2E, EBU, ECS, EDITEUR, EDM, EP, EXTN, FORMATS, FREQ, GBV, GRAFFLE, GRDDL, IR, ISWC, JISC, KISTI, LV, MEDIA, MO, MUSIM, MVCB, NFO, NIF, OCD, ODS, OO, P20, PAV, PRONOM, QB, RE, REL, SF, SUJ, SWIVT, URANAI, V, VCARD, VOAG, WGS, XLINK e YAGO.

Fonte: Elaborado pelos autores.

Uma informação necessária é que 33 entre os 73 *datasets* pertencem à iniciativa RKB Explorer⁴, e, por apresentarem muitas semelhanças entre si, acabam tendo impacto significativo nas estatísticas. Para resolver esse problema, procedeu-se à descrição desses *datasets* em separado. Todos os *datasets* RKB Explorer foram categorizados como de informação sobre publicações científicas. Em seguida, foram desconsiderados, e se procedeu à elaboração das estatísticas para o grupo dos 40 *datasets* restantes, como se pode verificar no Quadro 3.

Quadro 3 – Frequência de vocabulários considerados não tipicamente bibliográficos nos *datasets* selecionados, excluindo-se os *datasets* RKB Explorer

VOCABULÁRIO	Está no LOV?	FREQUÊNCIA							
		Datasets de biblioteca		Datasets de informação sobre publicações científicas		Datasets de informação sobre conferências científicas		TOTAL (sem RKB Explorer)	
		nº	%	nº	%	nº	%	nº	%
RDFS	Sim	15	78,9%	15	75%	1	100%	31	77,5%
FOAF	Sim	15	78,9%	15	75%	0	0%	30	75%
DCE	Sim	13	68,4%	9	45%	0	0%	22	55%
OWL	Sim	12	63,2%	10	50%	0	0%	22	55%
DCTERMS	Sim	10	52,6%	10	50%	0	0%	20	50%
SKOS	Sim	9,5	50%	8,5	42,5%	0	0%	18	45%
VOID	Sim	7	36,8%	8	40%	0	0%	15	37,5%
XSD	Sim	7,5	39,5%	5,5	27%	0	0%	13	32,5%
GEO	Sim	5	26,3%	4	20%	0	0%	9	22,5%
SCHEMA	Sim	2,5	13,2%	4,5	22,5%	0	0%	7	17,5%
VIVO	Sim	0,5	2,6%	5,5	27,5%	0	0%	6	15%
DCMITYPE	Sim	3,5	18,4%	1,5	7,5%	0	0%	5	12,5%
PRV	Sim	2	10,5%	3	15%	0	0%	5	12,5%
GEONAMES	Sim	3	15,8%	1	5%	0	0%	4	10%
DBPEDIA	Sim	2	10,5%	2	10%	0	0%	4	10%

⁴ “RKB Explorer é uma aplicação da Web Semântica que é capaz de apresentar visualizações unificadas de um número significativo de fontes de dados heterogêneas com relação a um determinado domínio” (GLASER; MILLARD, 2007, p.1, tradução nossa).

EVENT	Sim	2	10,5%	2	10%	0	0%	4	10%
DOAP	Sim	1	5,3%	3	15%	0	0%	4	10%
PROV	Sim	1	5,3%	3	15%	0	0%	4	10%
SD	Sim	1	5,3%	3	15%	0	0%	4	10%
UMBEL	Sim	2,5	13,2%	0,5	2,5%	0	0%	3	7,5%
WDRS	Sim	2,5	13,2%	0,5	2,5%	0	0%	3	7,5%
ORG	Sim	1	5,3%	2	10%	0	0%	3	7,5%
PRVYPES	Sim	1	5,3%	2	10%	0	0%	3	7,5%
DCAT	Sim	0	0%	3	15%	0	0%	3	7,5%
SIO	Sim	0	0%	3	15%	0	0%	3	7,5%
SWRC	Sim	0	0%	3	15%	0	0%	3	7,5%
Vocabulários com 2 ocorrências: BIO, GR, IRW, LIDO, META, OBORO, OPM, ORE, OV, PML, PMLJ, PMLP, SIOC, TAG, UNITS, VANN, VSR e WV.									
Vocabulários com 1 ocorrência: API, B3KAT, BFLC, BTE, CC, CON, CONTACT, CONV, CRSW, D2R, DATAFAQS, DATE, DM2E, EBU, ECS, EDITEUR, EDM, FORMATS, FREQ, GBV, GRAFFLE, GRDDL, IR, ISWC, LV, MEDIA, MO, MUSIM, MVCB, NFO, NIF, OCD, ODS, OO, P20, PAV, PRONOM, QB, RE, REL, SCOVO, SF, SUJ, SWIVT, URANAI, V, VCARD, VOAG, WGS, XLINK e YAGO.									

Fonte: Elaborado pelos autores.

Por meio do estudo empírico de descrição quanto ao uso de vocabulários, tipicamente bibliográficos ou não, da amostra de *datasets* preestabelecida, obtém-se como resultado um panorama quanto à utilização de vocabulários para metadados referentes a publicações científicas disponibilizados como *Linked Data*. Assim, encerram-se os resultados do primeiro estudo descritivo.

Os resultados do segundo estudo descritivo (correspondente ao item 3.3, da Metodologia) estão organizados nos quadros de número 4 a 11. Neles, elencam-se as declarações RDF, presentes no modelo da British Library, de cada vocabulário que foi considerado interessante para a pesquisa. Optou-se por não se especificarem as declarações da região Assunto, pois é demasiadamente detalhada e restrita à lógica do contexto das bibliotecas.

O Quadro 4 estabelece as declarações RDF que envolvem uso do vocabulário BIBO, nome simplificado para The Bibliographic Ontology. Segundo o Linked Open Vocabularies (LOV), o BIBO “provê os principais conceitos e propriedades para descrição de citações e referências bibliográficas [...] na Web Semântica” (ONTOLOGY ENGINEERING GROUP, 2018, tradução nossa). Pode-se afirmar que se trata de um vocabulário praticamente tão aplicado no contexto da comunidade científica quanto no das bibliotecas – o que se pode verificar no Quadro 1, apresentado anteriormente.

Quadro 4 – Declarações RDF com uso do vocabulário BIBO no modelo da British Library para publicações seriadas

Região	Recurso/ Sujeito	Chave	Propriedade/ Predicado	Chave	Valor/ Objeto	Chave
Identificadores	Resource BL URI	Instância	bibo:issn	-	-	Literal
Relacionamentos Bibliográficos	Resource BL URI	Instância	-	-	bibo:periodical ou bibo:newspaper	Classe
	Series BL URI	Instância	bibo:issn	-	-	Literal
	Series BL URI	Instância	-	-	bibo:series	Classe
Miscelânea	Resource BL URI	Instância	bibo:numVolumes	-	-	Literal
Espaço sem denominação	Resource BL URI	Instância	-	-	bibo:periodical ou bibo:newspaper	Classe

Fonte: Elaborado pelos autores.

No Quadro 5, encontram-se as triplas RDF que utilizam o vocabulário BIO, descrito pelo LOV como “um vocabulário para descrição de informação biográfica sobre pessoas, tanto vivas quanto mortas” (ONTOLOGY ENGINEERING GROUP, 2018, tradução nossa).

Quadro 5 – Declarações RDF com uso do vocabulário BIO no modelo da British Library para publicações seriadas

Região	Recurso/ Sujeito	Chave	Propriedade/ Predicado	Chave	Valor/ Objeto	Chave
Autor	Person-as- Agent BL URI	Instância	bio:event	-	Birth BL URI	Instância
	Birth BL URI	Instância	-	-	bio:Birth	Classe
	Birth BL URI	Instância	bio:date	-	-	Literal
	Person-as- Agent BL URI	Instância	bio:event	-	Death BL URI	Instância
	Death BL URI	Instância	-	-	bio:Death	Classe
	Death BL URI	Instância	bio:date	-	-	Literal

Fonte: Elaborado pelos autores.

O BLT, ou BLTERMS, ou ainda British Library Terms RDF Schema, é o vocabulário desenvolvido pela própria British Library. Conforme descrito no LOV, trata-se de “alguns

termos úteis para descrição de recursos bibliográficos que outros modelos não incluem” (ONTOLOGY ENGINEERING GROUP, 2018, tradução nossa). As declarações RDF com o BLT estão dispostas no Quadro 6 (a seguir).

Quadro 6 – Declarações RDF com uso do vocabulário BLT (BLTERMS) no modelo da British Library para publicações seriadas

Região	Recurso/ Sujeito	Chave	Propriedade/ Predicado	Chave	Valor/ Objeto	Chave
Autor	Person-as-Agent BL URI	Instância	blt:hasCreated	-	Resource BL URI	Instância
	Person-as-Agent BL URI	Instância	blt:hasContributedTo	-	Resource BL URI	Instância
	Organization-as-Agent BL URI	Instância	blt:hasCreated	-	Resource BL URI	Instância
	Organization-as-Agent BL URI	Instância	blt:hasContributedTo	-	Resource BL URI	Instância
Identificadores	Resource BL URI	Instância	blt:bnb	-	-	Literal
Eventos de Publicação	Resource BL URI	Instância	blt:publicationStart	-	PublicationStart Event BL URI	Instância
	PublicationStart Event BL URI	Instância	-	-	blt:PublicationStartEvent	Classe
	blt:PublicationStartEvent	Classe	rdfs:subClassOf	-	blt:PublicationEvent	Classe
	blt:PublicationEvent	Classe	rdfs:subClassOf	-	event:Event	Classe
Assunto						

Fonte: Elaborado pelos autores.

A Dublin Core Metadata Initiative (DCMI) é uma iniciativa que mantém tecnologias para a descrição de objetos informacionais no meio digital. Entre elas, estão dois vocabulários principais: o Dublin Core Metadata Element Set (DCE), composto de apenas 15 propriedades para descrição, e o DCMI Metadata Terms (DCTERMS), que veio para sofisticar o primeiro.

Segundo o LOV, ambos estão entre os vocabulários mais populares da Web Semântica, sendo que o líder é o DCTERMS. O modelo da British Library faz uso dele, como se verifica no Quadro 7.

Vale pontuar que essa vasta aplicabilidade do Dublin Core deve ser atribuída a sua premissa de ser capaz de descrever os mais variados objetos digitais e atender diversos tipos de comunidades. Os vocabulários são flexíveis, isto é, possuem propriedades genéricas – a ponto de serem eventualmente criticados por isso. São notadamente

aderentes à comunidade científica (SILVA; RIBEIRO; LOPES, 2016; SIMEK *et al.*, 2017), ao domínio bibliográfico (BAKER, 2012), e ao *Linked Data* (CATARINO; SOUZA, 2012).

Quadro 7 – Declarações RDF com uso do vocabulário DCTERMS no modelo da British Library para publicações seriadas

Região	Recurso/ Sujeito	Chave	Propriedade/ Predicado	Chave	Valor/ Objeto	Chave
Autor	Resource BL URI	Instância	dct:creator	-	Person-as-Agent BL URI	Instância
	Resource BL URI	Instância	dct:contributor	-	Person-as-Agent BL URI	Instância
	Resource BL URI	Instância	dct:creator	-	Organization-as- Agent BL URI	Instância
	Resource BL URI	Instância	dct:contributor	-	Organization-as- Agent BL URI	Instância
	Person-as- Agent BL URI	Instância	-	-	dct:Agent	Classe
	Organization- as-Agent BL URI	Instância	-	-	dct:Agent	Classe
Título	Resource BL URI	Instância	dct:title	-	-	Literal
	Resource BL URI	Instância	dct:alternative	-	-	Literal
Relacionamentos Bibliográficos	Resource BL URI	Instância	dct:isPartOf	-	Series BL URI	Instância
	Series BL URI	Instância	dct:hasPart	-	Resource BL URI	Instância
	Resource BL URI	Instância	dct:isPartOf	-	Resource BL URI	Instância
	Resource BL URI	Instância	dct:hasPart	-	Resource BL URI	Instância
	Resource BL URI	Instância	dct:isFormatOf	-	Resource BL URI	Instância
	Resource BL URI	Instância	dct:hasVersion	-	Resource BL URI	Instância
	Resource BL URI	Instância	dct:replaces	-	Resource BL URI	Instância
	Resource BL URI	Instância	dct:isReplacedBy	-	Resource BL URI	Instância
	Resource BL URI	Instância	dct:relation	-	Resource BL URI	Instância
Eventos de Publicação	Agent BL URI	Instância	-	-	dct:Agent	Classe
Assunto						
Miscelânea	Resource BL URI	Instância	dct:abstract	-	-	Literal
	Resource BL URI	Instância	dct:tableOf Contents	-	-	Literal
	Resource BL URI	Instância	dct:description	-	-	Literal
Espaço sem denominação	Resource BL URI	Instância	-	-	dct:Bibliographic Resource	Classe
	Resource BL URI	Instância	dct:language	-	Lexvo URI	Link externo

Fonte: Elaborado pelos autores.

O Quadro 8 é composto pelas declarações RDF que utilizam o vocabulário EVENT, ou The Event Ontology. O LOV o descreve da seguinte maneira: “Essa ontologia trabalha com a noção de eventos reificados – eventos vistos como objetos de primeira-classe” (ONTOLOGY ENGINEERING GROUP, 2018, tradução nossa). Em outras palavras, o vocabulário EVENT se propõe, basicamente, a fornecer termos para descrição de eventos.

Quadro 8 – Declarações RDF com uso do vocabulário EVENT no modelo da British Library para publicações seriadas

Região	Recurso/ Sujeito	Chave	Propriedade/ Predicado	Chave	Valor/ Objeto	Chave
Eventos de Publicação	Publication StartEvent BL URI	Instância	event:time	-	http://r.d.g/ id/year/YYYY	Link externo
	Publication StartEvent BL URI	Instância	event:place	-	GeoNames URI	Link externo
	Publication StartEvent BL URI	Instância	event:place	-	Place BL URI	Instância
	Publication StartEvent BL URI	Instância	event:agent	-	Agent BL URI	Instância
	blt:Publication Event	Classe	rdfs:subClassOf	-	event:Event	Classe

Fonte: Elaborado pelos autores.

FOAF (Friend of a Friend Vocabulary) é um vocabulário que se insere em um “projeto destinado a conectar pessoas e informação usando a Web” (ONTOLOGY ENGINEERING GROUP, 2018, tradução nossa). Vale notar que a necessidade de especificar relações entre as pessoas, contemplada por esse vocabulário, é intrínseca à comunidade científica. As triplas RDF com o FOAF estão dispostas no Quadro 9.

Quadro 9 – Declarações RDF com uso do vocabulário FOAF no modelo da British Library para publicações seriadas

Região	Recurso/ Sujeito	Chave	Propriedade/ Predicado	Chave	Valor/ Objeto	Chave
Autor	Person-as-Agent BL URI	Instância	foaf:familyName	-	-	Literal
	Person-as-Agent BL URI	Instância	foaf:givenName	-	-	Literal
	Person-as-Agent BL URI	Instância	foaf:name	-	-	Literal
	Person-as-Agent BL URI	Instância	-	-	foaf:Agent foaf:Person	Classe
	Organization-as-Agent BL URI	Instância	rdfs:label [foaf:name]	-	-	Literal
	Organization-as-Agent BL URI	Instância	-	-	foaf:Agent foaf:Organization	Classe
Eventos de Publicação	Agent BL URI	Instância	-	-	foaf:Agent	Classe
	MARC country code URI	Link externo	foaf:focus	-	Geonames URI	Link externo
Assunto						
Espaço sem denominação	MARC language code URI	Link externo	foaf:focus	-	Lexvo URI	Link externo

Fonte: Elaborado pelos autores.

O Quadro 10 apresenta as declarações RDF que envolvem o uso do vocabulário GEO. Segundo o LOV, trata-se de “um vocabulário para representar informação de latitude, longitude e altitude nos dados de referência geodésica WGS84” (ONTOLOGY ENGINEERING GROUP, 2018, tradução nossa).

Quadro 10 – Declarações RDF com uso do vocabulário GEO no modelo da British Library para publicações seriadas

Região	Recurso/ Sujeito	Chave	Propriedade/ Predicado	Chave	Valor/ Objeto	Chave
Eventos de Publicação	Place BL URI	Instância	-	-	geo:SpatialThing	Classe
Assunto						

Fonte: Elaborado pelos autores.

O vocabulário ORG, nome mais simples para Core Organization Ontology, é referido em apenas uma declaração RDF do modelo, como se pode observar no Quadro 11. O LOV o define como “vocabulário para descrever estruturas organizacionais, especializado para uma ampla variedade de tipos de organização” (ONTOLOGY ENGINEERING GROUP, 2018, tradução nossa).

Quadro 11 – Declarações RDF com uso do vocabulário ORG no modelo da British Library para publicações seriadas

Região	Recurso/ Sujeito	Chave	Propriedade/ Predicado	Chave	Valor/ Objeto	Chave
Autor	Organization- as-Agent BL URI	Instância	-	-	org:Organization	Classe

Fonte: Elaborado pelos autores.

Cabe justificar o descarte de determinados vocabulários presentes no modelo, quanto ao estudo das declarações que envolvem seu uso. RDF, OWL, XSD (XML Schema) e UMBEL são vocabulários muito genéricos e abstratos, isto é, aplicam-se a qualquer matéria. ISBD e RDAU, em oposição, são vocabulários demasiadamente específicos, mas específicos ao contexto da descrição em bibliotecas, distinto do contexto da descrição de artigos de periódicos para comunicação científica, estudado nesta pesquisa.

4.1 Modelo semântico para publicações científicas: uma proposição

Com base nos estudos desenvolvidos, é possível propor um modelo semântico estruturado por declarações RDF, conforme objetivado. Tal proposição se constrói ao longo de um percurso composto sucessivamente pelos seguintes passos:

- (1) Selecionar propriedades no vocabulário DCTERMS, uma vez que, entre outros motivos, possui alta aderência à comunidade científica (SILVA; RIBEIRO; LOPES, 2016; SIMEK *et al.*, 2017), bem como por ter sido utilizado significativamente no protótipo para catalogação semântica de publicações proposto por Freitas Junior e Jacynto (2016);

- (2) Estipular declarações RDF para que os recursos sejam relacionados a identificadores bibliográficos de interesse (identificadores de autoridade, ISSN e DOI);
- (3) Verificar se existe algum outro vocabulário observado no estudo descritivo dos *datasets* selecionados que suscite atenção específica, para ser utilizado na composição das declarações;
- (4) Buscar contemplar as 7 regiões do modelo da British Library complementando nesse sentido o conjunto de declarações propostas nos passos anteriores, e dando preferência a declarações que utilizem vocabulários com bons índices no estudo descritivo de *datasets*.

No Quadro 12, faz-se uma analogia com o modelo de dados da British Library destinado a publicações seriadas, por meio da organização de todas as declarações RDF obtidas durante o percurso exposto acima de modo a encaixá-las nas regiões em que se divide o referido modelo.

Quadro 12 – Conjunto de declarações RDF propostas, conforme regiões

Região	Recurso/ Sujeito	Propriedade/ Predicado	Valor/ Objeto	Passo
Autor	RECURSO URI	dct:creator	AUTOR URI	1
	AUTOR URI	owl:sameAs	IDENTIFICADOR DE AUTORIDADE URI	2
	AUTOR URI	foaf:familyName	Literal	3
	AUTOR URI	foaf:givenName	Literal	3
	AUTOR URI	foaf:name	Literal	3
Identificadores	RECURSO URI	bibo:issn	ISSN (Literal)	2
	RECURSO URI	bf:issn	ISSN (Literal)	2
	RECURSO URI	dct:identifier	RECURSO DOI	2
Título	RECURSO URI	dct:title	Literal	1
Relacionamentos Bibliográficos	RECURSO A URI	dct:bibliographicCitation	RECURSO B URI	1
	RECURSO A URI	dct:bibliographicCitation	RECURSO B DOI	1
	RECURSO A URI	dct:isReferencedBy	RECURSO B URI	1

Região	Recurso/ Sujeito	Propriedade/ Predicado	Valor/ Objeto	Passo
	RECURSO A URI	dct:isReferencedBy	RECURSO B DOI	1
	SÉRIE URI	bibo:issn	Literal	3
	RECURSO URI	dct:isPartOf	SÉRIE URI	4
	SÉRIE URI	dct:hasPart	RECURSO URI	4
	RECURSO A URI	dct:isPartOf	RECURSO B URI	4
	RECURSO A URI	dct:hasPart	RECURSO B URI	4
	RECURSO A URI	dct:isFormatOf	RECURSO B URI	4
	RECURSO A URI	dct:hasVersion	RECURSO B URI	4
	RECURSO A URI	dct:replaces	RECURSO B URI	4
	RECURSO A URI	dct:isReplacedBy	RECURSO B URI	4
	RECURSO A URI	dct:relation	RECURSO B URI	4
Eventos de Publicação	RECURSO URI	event:producedIn	EVENTO URI	4
	EVENTO URI	event:place	GeoNames URI	4
	EVENTO URI	event:time	Literal	4
Assunto	RECURSO URI	dct:subject	Literal	1
Miscelânea	RECURSO URI	dct:date	Literal	1
	RECURSO URI	dct:publisher	Literal	1
	RECURSO URI	dct:accessRights	Literal	1
	RECURSO URI	dct:abstract	Literal	1
	RECURSO URI	dct:tableOfContents	Literal	1
	RECURSO URI	bibo:numVolumes	Literal	3
Espaço sem denominação	RECURSO URI	dct:type	dctype:_____	1
	RECURSO URI	dct:language	Lexvo URI	1

Fonte: Elaborado pelos autores.

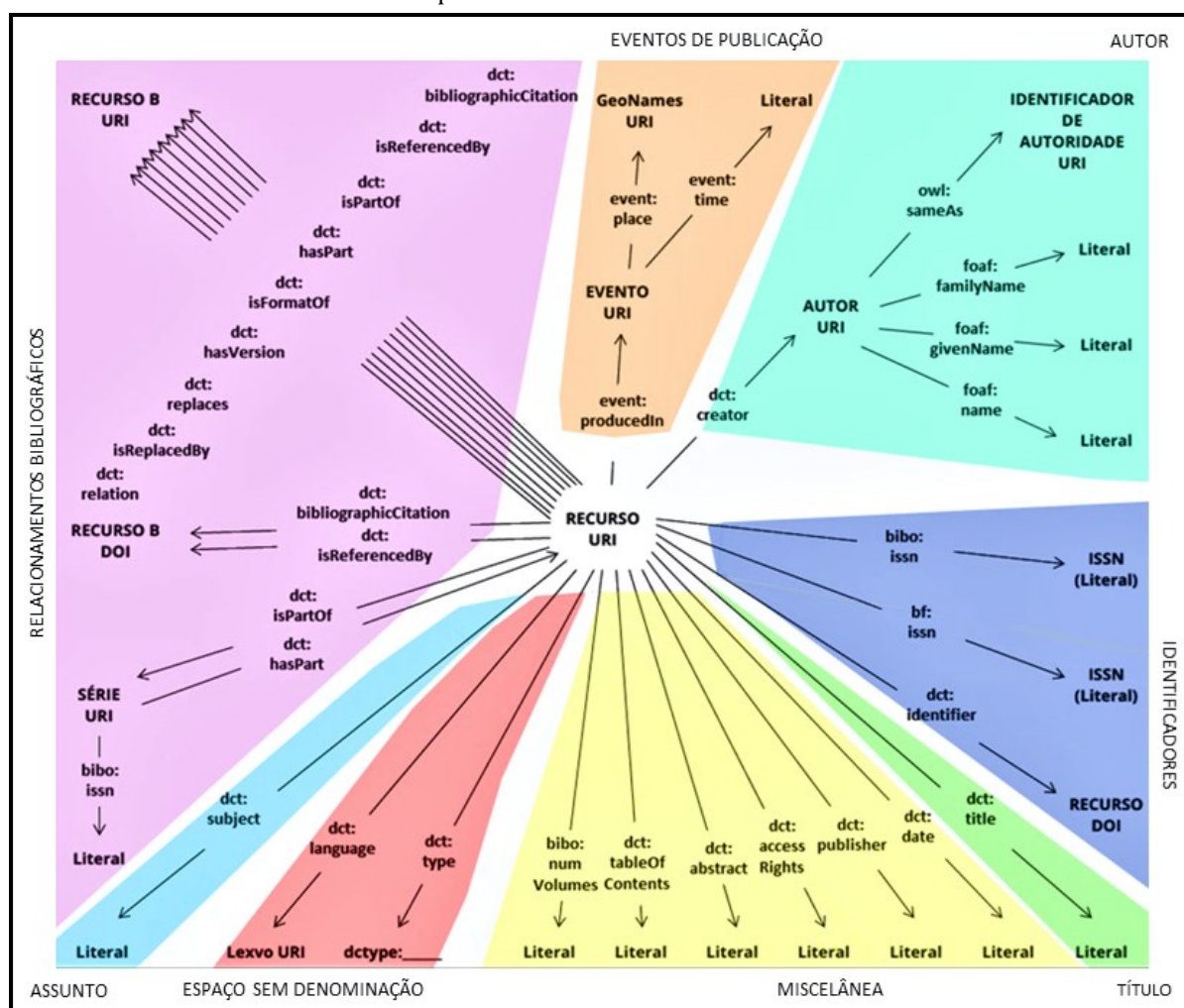
Tem-se que: são 35 declarações RDF propostas no total; são 7 vocabulários coexistindo no registro (DCTERMS, BIBO, FOAF, OWL, DCMI Type, BIBFRAME e EVENT); e é evidente a preponderância do DCTERMS, que integra 24 declarações.

No mais, vale citar o quarto princípio do *Linked Data*, segundo o qual a posição valor/objeto de uma tripla RDF deve ser preferencialmente preenchida por um URI, em

vez de um literal (BERNERS-LEE, 2006). Disso, interpreta-se que tal princípio se aplica a tudo que não é literal, seja uma instância ou um link externo, por exemplo. O conjunto apresentado, nesse sentido, pode ser considerado exitoso ao elencar 20 declarações que conseguem atender a esse princípio.

Como último resultado, apresenta-se a Figura 5, que esquematiza as declarações RDF do Quadro 13 de modo semelhante ao esquema visual utilizado e publicado pela British Library (Figura 5).

Figura 5 - Modelo semântico proposto para dados bibliográficos de publicações científicas disponibilizados como *Linked Data*



Fonte: Elaborado pelos autores.

Assim como no esquema visual que representa o modelo da British Library, cada região corresponde a uma cor, e, via de regra, irradia-se a partir do centro da imagem, onde se encontra o URI referente ao objeto informacional descrito.

5 DISCUSSÃO

Esta pesquisa concerne à investigação sobre quais são as melhores práticas para publicação em formato semântico, isto é, como *Linked Data*, de metadados de registros bibliográficos referentes a documentos tidos como resultados de comunicação científica. Isso pressupõe a utilização de tecnologias elementares da Web Semântica, e os vocabulários estão entre elas.

Schaible, Gottron e Scherp (2014) aplicaram um *survey* sobre vocabulários, respondido por publicadores e consumidores de *Linked Data*, a partir do qual se identificou a preferência, na construção de um *dataset*, por “se manter uma mistura equilibrada entre vocabulários populares e específicos de domínio, de modo a fornecer uma estrutura clara dos dados e facilitar seu consumo”.

No contexto da publicação de metadados bibliográficos, trabalhado no presente estudo, notou-se que os vocabulários entendidos como tipicamente bibliográficos correspondem aos vocabulários “específicos de domínio” mencionados. Na constituição dos *datasets*, verificou-se que eles realmente se combinam a vocabulários mais genéricos, isto é, não tipicamente bibliográficos.

Absolutamente, a combinação de vocabulários em um mesmo *dataset* não é, em si, um problema. Segundo Assumpção e Santos (2016), tal prática não compromete a padronização dos metadados – tão cara à descrição bibliográfica – pois isso se garante com o alicerce dado pelo RDF, que, mais abstrato do que um vocabulário, é um modelo de dados, fundamental para a interoperabilidade no contexto da Web Semântica.

A coexistência de estruturas e identificadores diversos para descrição de um mesmo objeto é um traço característico da Web Semântica. Isso costuma resultar em conjuntos de metadados complexos – talvez até demais – para a leitura a ser realizada por humanos; complexidade que, todavia, não constitui um problema para o entendimento pelas máquinas. Nesse sentido, um conjunto de metadados rico com coexistência de vocabulários relevantes é o que se almeja.

6 CONCLUSÃO

Na busca pelo impacto das publicações científicas, sabe-se que uma boa representação bibliográfica é fundamental. Isso é verdade desde as origens da

catalogação, em contexto completamente analógico. No entanto, é necessário compreender para além disso que, considerando o processamento por máquina dos metadados do registro bibliográfico, em *Linked Data*, esse potencial atinge dimensões exponencialmente maiores.

Mas, então, o que é uma boa representação bibliográfica no contexto do *Linked Data*? Como se pôde perceber neste artigo, é aquela que reconhece que a coexistência de vários vocabulários em um mesmo registro não constitui um problema.

Pelo contrário. Deve ser valorizada, em equilíbrio com outros requisitos também importantes, como as demandas de representação do tipo de objeto informacional a ser descrito e a consistência entre os registros, por exemplo. Assim, pode não ser necessário se utilizarem vocabulários a se perder de vista para que a publicação esteja bem representada, bem como pode ser indesejável descrever muito mais determinados objetos em detrimento de outros.

Em suma, considera-se que a estratégia metodológica contribuiu satisfatoriamente para a consecução dos objetivos da pesquisa: entender quais são as melhores práticas em uso no atendimento às necessidades relativas a descrição bibliográfica de publicações científicas no contexto do *Linked Data*, bem como, a partir desse entendimento, proferir um gesto no sentido de vislumbrar um modelo semântico para metadados bibliográficos de publicações científicas, estabelecendo vocabulários e declarações RDF que nele poderiam figurar. Neste estudo, o modelo se concretizou na forma de um quadro com as declarações RDF, e de uma imagem análoga à publicada pela British Library.

REFERÊNCIAS

ABELE, A. *et al.* **Linking Open Data cloud diagram**. [Galway, Ireland]: Insight Centre for Data Analytics, 2017. Disponível em: <https://web.archive.org/web/20171210161355/http://lod-cloud.net:80/>. Acesso em: 27 ago. 2019.

ASSUMPÇÃO, F. S.; SANTOS, P. L. V. A. C. *Linked Data no domínio bibliográfico: vocabulários para a publicação de dados de autoridade*. In: ENCONTRO INTERNACIONAL DE DADOS, TECNOLOGIA E INFORMAÇÃO, 3., 2016, Marília. **Perspectivas e interdisciplinaridades em Ciência da**

Informação. Marília: Unesp, 2016. v. 3, p. 507-528. Disponível em: <https://goo.gl/Lec61v>. Acesso em: 25 ago. 2019.

BAKER, T. Libraries, languages of description, and Linked Data: a Dublin Core perspective. **Library Hi Tech**, v. 30, n. 1, p. 116-133, 2012. Disponível em: <https://pdfs.semanticscholar.org/6b67/be67b58771e03eda4ca8e33c13113a522948.pdf>. Acesso em: 30 ago. 2019.

BERNERS-LEE, T. **Linked Data principles**. 2006. Disponível em: <http://www.w3.org>

/DesignIssues/LinkedData.html. Acesso em: 30 ago. 2019.

BRITISH LIBRARY; TALIS. **British Library data model**: serial. Versão 1.2. [Boston Spa, United Kingdom], jul. 2018. Disponível em: <https://www.bl.uk/bibliographic/pdfs/bldatamodelserial.pdf>. Acesso em: 27 ago. 2019.

CATARINO, M. E.; SOUZA, T. B. A representação descritiva no contexto da web semântica. **TransInformação**, v. 24, n. 2, p. 77-90, 2012. Disponível em: <http://www.scielo.br/pdf/tinf/v24n2/a01v24n2.pdf>. Acesso em: 30 ago. 2019.

FREITAS JUNIOR, N.; JACYNTO, M. D. A. Um protótipo Linked Data para catalogação semântica de publicações. **Perspectivas em Ciência da Informação**, v. 21, n. 4, p. 48-65, out./dez. 2016. Disponível em: <http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/2664/1812>. Acesso em: 03 abr. 2018.

GLASER, H.; MILLARD, I. RKB Explorer: application and infrastructure. *In*: SEMANTIC WEB CHALLENGE, 2007, Busan, Korea. **CEUR Workshop Proceedings**, v. 295, paper 13. Disponível em: <http://ceur-ws.org/Vol-295/paper13.pdf>. Acesso em: 27 ago. 2019.

HEATH, T.; BIZER, C. **Linked Data**: evolving the Web into a global data space. [S.l.]: Morgan & Claypool, 2011. Disponível em: <http://info.slis.indiana.edu/~dingying/Teaching/S604/LODBook.pdf>. Acesso em: 30 ago. 2019.

ISOTANI, S.; BITTENCOURT, I. I. **Dados Abertos Conectados**. São Paulo: Novatec, 2015. Disponível em: <https://ceweb.br/livros/dados-abertos-conectados/>. Acesso em: 30 ago. 2019.

ONTOLOGY ENGINEERING GROUP. **Linked Open Vocabularies (LOV)**. [Madri, 2018]. Disponível em: <https://lov.linkeddata.es/dataset/lov>. Acesso em: 26 nov. 2019.

ROZSA, V.; DUTRA, M. S. L.; NHACUONGUE, J. R. A. Linked Open Data no contexto acadêmico: identificação e análise de vocabulários utilizados na academia e na

pesquisa científica. **Brazilian Journal of Information Science**, v. 11, n. 3, p. 34-52, 2017. Disponível em: <http://www2.marilia.unesp.br/revistas/index.php/bjis/article/view/6780/4651>. Acesso em: 05 abr. 2018.

SCHAIBLE, J.; GOTTRON, T.; SCHERP, A. Survey on common strategies of vocabulary reuse in Linked Open Data modeling. *In*: EUROPEAN SEMANTIC WEB CONFERENCE (ESWC), 11., 2014, Anissaras, Greece. **The semantic web: trends and challenges**. Cham, Switzerland: Springer, 2014. p. 457-472. Disponível em: https://link.springer.com/content/pdf/10.1007%2F978-3-319-07443-6_31.pdf. Acesso em: 27 ago. 2019.

SCHMACHTENBERG, M.; BIZER, C.; PAULHEIM, H. Adoption of the linked data best practices in different topical domains. *In*: INTERNATIONAL SEMANTIC WEB CONFERENCE (ISWC), 13., 2014, Riva del Garda, Italy. **Lecture Notes in Computer Science**. Cham, Switzerland: Springer, 2014. v. 8796, p. 245-260. Disponível em: <http://www.planet-data.eu/sites/default/files/publications/SchmachtenbergBizerPaulheim-AdoptionOfLinkedDataBestPractices.pdf>. Acesso em: 27 ago. 2019.

SILVA, J. R.; RIBEIRO, C.; LOPES, J. C. Usage-driven Dublin Core descriptor selection: a case study using the Dendro platform for research dataset description. *In*: INTERNATIONAL CONFERENCE ON THEORY AND PRACTICE OF DIGITAL LIBRARIES (TPDL), 20., 2016, Hannover, Germany. **Lecture Notes in Computer Science**. Cham, Switzerland: Springer, 2016. v. 9819, p. 27-38. Disponível em: https://link.springer.com/chapter/10.1007%2F978-3-319-43997-6_3. Acesso em: 25 nov. 2018.

SIMEK, P. *et al.* Analysis of options and tools for semantic and effective description of data and research results in the areas of agrarian sector. *In*: INTERNATIONAL SCIENTIFIC

CONFERENCE ON AGRARIAN PERSPECTIVES, 26., 2017, Prague, Czech Republic. **Agrarian Perspectives XXVI: Competitiveness of European Agriculture and Food Sectors**. Prague, Czech Republic: Czech University Life Sciences Prague, 2017. p. 404-409. Disponível em: <https://www.cabdirect.org/cabdirect/FullTextPDF/2018/20183082664.pdf>. Acesso em: 30 ago. 2019.

TADINI, A. V. W.; CONEGLIAN, C. S.; SANTAREM SEGUNDO, J. E. Caracterização do segmento de publicações no Linking Open Data, um estudo exploratório. **Conhecimento em Ação**, v. 2, n. 2, p. 2-18, jul./dez. 2017. Disponível em: <https://revistas.ufrj.br/index.php/rca/article/view/11699/9739>. Acesso em: 25 ago. 2019.