

Web Semântica e Bibliografia: um Estudo para a Publicação de Registros Bibliográficos como Dados Abertos Conectados

Sandro Rautenberg

Doutor; Universidade Estadual do Centro-Oeste, Guarapuava, PR, Brasil;
srautenberg@unicentro.br

Resumo: Neste artigo é apresentado um estudo interdisciplinar, envolvendo elementos da Web Semântica e Bibliografia. Pontualmente, tem-se como alicerce os conceitos de Dados Abertos Conectados e registros bibliográficos, considerando que: i) à Bibliografia reserva-se o tratamento dos diversos modos de registro do conhecimento em livros, artigos, jornais, documentos digitais, entre outros; ii) em face da expansão informacional, a Bibliografia necessita abarcar novas Tecnologias de Informação e Comunicação para a manutenção dos registros bibliográficos; e iii) um dos conceitos contemporâneos da Web Semântica difundidos são os Dados Abertos Conectados, os quais podem utilizados para organizar e representar unidades do conhecimento e sua semântica. Assim, objetiva-se investigar o alinhamento dos preceitos dos Dados Abertos Conectados para a organização, a representação e o compartilhamento de recursos de dados aderentes aos estudos da Bibliografia. Para tanto, realizou-se uma busca por modelos comumente utilizados para representação de elementos bibliográficos na Web de Dados. A referida pesquisa foi realizada a partir do repositório de vocabulários abertos conectados denominado *Linked Open Vocabularies*. Como resultado, identificou-se uma ontologia para representar recursos bibliográficos na Web de Dados, a *Bibliographic Ontology*. Semanticamente, essa ontologia compreende 58 classes e 67 propriedades inspiradas em diversos metadados de descrição de documentos. Em sua essência, a *Bibliographic Ontology* pode ser usada como uma ontologia de citação e de classificação de documentos. Neste sentido, conclui-se que a *Bibliographic Ontology* pode facilitar a interoperabilidade dos dados e metadados, repercutindo nos benefícios de processabilidade, interoperabilidade, compreensão e reutilização de registros e sua semântica para com os estudos bibliográficos.

Palavras-chave: Registros Bibliográficos. Ontologia. Vocabulário. Web Semântica. Dados Abertos Conectados.

1 Introdução

É notório que, nos últimos tempos, a expansão científica e tecnológica tem influenciado no rompimento das fronteiras do compartilhamento e do (re)uso de

recursos digitais pela Sociedade da Informação. A velocidade, a transformação e a disseminação de dados e de informações tornaram-se forças importantes que potencializaram a construção social de novos conhecimentos. Neste cenário, reserva-se à web o papel de um ambiente estruturante de redes virtuais, o que permite relacionar conteúdos, pessoas e artefatos - Internet das Coisas - (van der AALST, 2014).

Neste cenário, Tim Bernes-Lee vislumbrou a evolução da web, conceitualizando a Web Semântica (BERNERS-LEE; HENDLER; LASSILA, 2001). Nesta evolução, a web, que antes previa somente a utilização de hiperlinks entre conteúdos, passa a abarcar os hiperlinks entre dados estruturados. Como resultado, tem-se a expectativa da formação de uma grande base de dados global com facilidades de busca, recuperação e exploração de informação pertinente por parte de agentes humanos e computacionais (SHADBOLT; HALL; BERNERS-LEE, 2006).

Diante disso, considerando o domínio da Ciência da Informação, vislumbra-se que a Web Semântica tem se tornado um objeto de estudo e/ou de apoio para com as pesquisas do referido campo científico. Ao se tratar a Web Semântica como um objeto de estudo, a Ciência da Informação pode contribuir conceitualmente com pesquisas devotadas à: **i)** semântica inerente às unidades de informação, considerando o significado e o significante; **ii)** estrutura, pontuando as formas de organização e representação dos elementos informacionais; e **iii)** sintaxe, considerando o conteúdo das mensagens quando da comunicação entre emissores e receptores. Já, ao considerar a Web Semântica como um objeto de apoio, esta é a estrutura basilar para pesquisas aplicadas, contribuindo para um ambiente de compartilhamento de dados e metadados nos diversos sistemas de informação. Ressalta-se que este é o contexto deste trabalho, o qual é melhor discutido na sequência.

Mediante o avanço da Internet, admite-se que a humanidade vem produzindo cada vez mais dados nas mais variadas plataformas digitais. Vários dispositivos interconectados (sensores, computadores, câmeras, dentre outros) e aplicativos relacionam uma miríade de eventos na web (van der AALST, 2014), coletando e armazenando enormes quantidades de registros, sinais, imagens,

vídeos e *posts*. Ou seja, a conexão de pessoas e tecnologias com os ambientes virtuais tem produzido enormes quantidades de dados, as quais necessitam ser organizadas e representadas em consonância a domínios especializados.

Pontualmente, este artigo está circunscrito ao domínio especializado da Bibliografia. Conforme a definição encontrada em (MALM, 1950, p. 7-8 *apud* CUNHA; CAVALCANTI, 2008, p. 46), a Bibliografia “é o ramo da Bibliologia – ou ciência do livro – que consiste na pesquisa de textos impressos ou multigrafados para indicá-los, descrevê-los e classificá-los com a finalidade de estabelecer instrumentos (de busca) e organizar serviços apropriados a facilitar o trabalho intelectual [...]”. Diante essa definição, pontua-se que a Bibliografia, como campo de estudo, pode apropriar-se de tecnologias da Web Semântica para sustentar parte de suas pesquisas científicas quando da organização, da representação, da descrição e da classificação de obras bibliográficas. Em poucas palavras, pressupõe-se que a Web Semântica pode contribuir com seus modelos de organização e representação do conhecimento para publicar e compartilhar recursos digitais de registros bibliográficos entre seus pesquisadores. Em consonância, a Figura 1 melhor ilustra tal pressuposto. A publicação de dados e da semântica inerente aos livros ocorre com o uso de modelos de organização e representação (vocabulários ou ontologias).



Fonte: Adaptado de Lóscio, Burle e Calegari (2019).

Neste sentido, o esforço de publicar e compartilhar os dados de livros e seus metadados em conformidade aos preceitos da Web Semântica ocorre em dois momentos distintos. Primeiramente, quando um conhecimento a respeito de algum objeto é formalmente representado via um modelo (ontologias ou vocabulários), está-se relacionando os metadados inerentes do recurso em

estudo aos padrões e as Melhores Práticas de Publicação na Web de Dados (LÓSCIO; BURLE; CALEGARI, 2019). No segundo momento, quando os dados de um objeto são materializados em consonância ao modelo de representação escolhido, os dados são organizados e estruturados, tornando-se Dados Conectados. E, se os Dados Conectados advêm de bases de dados livremente acessíveis por pessoas ou sistemas, estes são classificados como Dados Abertos Conectados. A Figura 2 representa esse relacionamento.

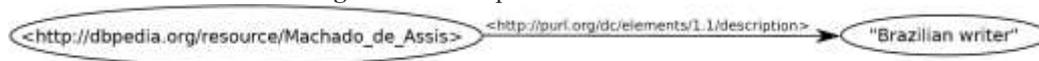
Figura 2 – Dados Abertos Conectados e seus benefícios



Fonte: Elaborada pelo autor (2019).

Em outras palavras, o conceito de Dados Abertos Conectados é configurado por duas questões norteadoras: O que são Dados Abertos? E o que são Dados Conectados? Resumidamente, na Web Semântica, os Dados Abertos são regidos por licenças que definem as regras de distribuição e (re)utilização que as pessoas ou aplicações computacionais devem seguir quando do uso desses recursos (LINKED DATA, 2019). Alguns exemplos de regras podem ser: **a)** a necessidade de citação da fonte original; **b)** o uso não comercial dos dados; **c)** a não derivação ou não adaptação dos dados; **d)** o compartilhamento dos dados em mesmo formato. Por sua vez, os Dados Conectados referem-se aos dados que são representados mediante os protocolos de acesso da Internet e que usam constitutivamente os formalismos da Web Semântica como infraestrutura para promover sua publicidade (BIZER; HEATH; BERNES-LEE, 2009). Para melhor entender o que é Dado Conectado, considere a Figura 3.

Figura 3 – Exemplo de Dado Conectado



Fonte: Elaborada pelo autor (2019).

A Figura 3 expressa alguns Dados Conectados do escritor “Machado de Assis”. De acordo com os protocolos de acesso da Internet, o recurso que representa o escritor “Machado de Assis” é codificado pelo URL¹ <http://dbpedia.org/resource/Machado_de_Assis>. A partir desse URL, pode-se navegar pelo URL <http://purl.org/dc/elements/1.1/description> e descobrir um Dado Conectado que informa que o recurso em evidência é um escritor brasileiro. Diante desse exemplo, pode-se conjecturar o potencial da Web Semântica, o qual permite aferir informações por meio da ligação/navegação entre os milhares de recursos já disponibilizados na Web de Dados.

Ainda prospectando com o exemplo anterior, uma vez publicado um conjunto de Dados Abertos Conectados em consonância às Melhores Práticas de Publicação na Web de Dados, os publicadores e consumidores desses dados são beneficiados pelo (LÓSCIO; BURLE; CALEGARI, 2019): **a) reúso** – aumenta as chances de reutilização de dados por consumidores, independentemente do domínio de aplicação; **b) compreensão** – com a publicação de conjuntos de dados e seus respectivos metadados, os seres humanos têm melhor entendimento das natureza, estrutura e semântica dos dados compartilhados; **c) ligação** – por publicar os dados em consonância com os protocolos da Internet, possibilita-se a criação de relações (*hiperlinks*) entre recursos de dados (conjuntos de dados e itens de dados); **d) descoberta** – mediante os *hiperlinks*, autonomamente, os computadores descobrem os conjuntos de dados e navegam entre os recursos digitais, aferindo conhecimento relacional entre os itens compartilhados; **e) confiança** – com a publicação de metadados de proveniência, os publicadores expressam como os conjuntos de recursos digitais são custodiados ao longo do tempo; **f) acesso** – utilizando os protocolos de acesso da Internet, promove-se a encontrabilidade e a exploração de recursos digitais pelos agentes (humanos e de *software*); **g) interoperabilidade** – permite que os recursos digitais sejam automaticamente convertidos em formatos abertos diversos (CSV², JSON³,

RDF⁴, XML⁵, entre outros); **h) processabilidade** – adicionalmente à interoperabilidade, a processabilidade permite que as aplicações computacionais manipulem os recursos digitais contidos na Web de Dados de forma automática.

Diante o exposto, o objetivo deste trabalho é investigar o alinhamento dos preceitos da Web Semântica como forma de organização, representação e compartilhamento de recursos digitais do domínio da Ciência da Informação. Especificamente, visa-se contribuir na adoção dos Dados Abertos Conectados como forma de compreender, reutilizar, interoperar e processar os dados e metadados de recursos aderentes aos estudos da Bibliografia. Para tanto, a próxima seção é reservada a discussão do procedimento metodológico adotado.

2 Percurso Metodológico

Em consonância às Melhores Práticas da Publicação na Web de Dados, Lóscio, Burle e Calegari (2019) pontuam que deve-se reutilizar modelos padronizados (ontologias ou vocabulários) para representar dados e metadados. Para os referidos autores, o reuso de modelos existentes especificados por especialistas ou organizações reconhecidos facilita o consenso e a comunicação de recursos digitais por comunidades de interesse. Em face disso, semanticamente, o uso de modelos já estabelecidos aumenta a interoperabilidade e reduz redundâncias, encorajando o reuso dos dados compartilhados. Em poucas palavras, a adoção de modelos compartilhados beneficia os publicadores e consumidores de Dados Abertos Conectados quanto aos benefícios de reuso, compreensão, confiança, interoperabilidade e processabilidade.

Segundo Hyland, Ateazing e Villazón-Terrazas (2019) existem várias ferramentas que coletam, analisam e indexam ontologias, vocabulários e dados semânticos. Neste sentido, os autores citados pontuam a utilização do repositório denominado *Linked Open Vocabularies* – LOV – (VANDENBUSSCHE *et. al.*, 2017). Representado na Figura 4, o LOV é um catálogo indexado de todos os modelos de organização e representação presentes na Web de Dados. Tal qual as páginas amarelas, esse catálogo facilita a identificação de vocabulários compartilhados em consonância aos termos comumente utilizados em algum domínio de interesse.

de modo a garantir sua reutilização para representar semanticamente diversos conjuntos de dados. Quanto mais um modelo for utilizado na representação de diversos conjuntos de dados, mais estabelecido o modelo se torna na Web de Dados. E, portanto, este se torna a melhor opção para adoção e reutilização, em um ciclo virtuoso. **Requisito 5)** Um modelo de representação deve ser acessível por um longo período, fornecendo alguma garantia de manutenção pelo grupo ou pela organização promotora. Idealmente, o acesso ao modelo deve ser de forma perene. **Requisito 6)** Um modelo de representação deve ser publicado e mantido por um grupo ou organização confiável. Embora qualquer indivíduo possa criar um modelo de representação na Web de Dados, é prudente verificar se existem políticas de responsabilidade quanto à manutenção do modelo. **Requisito 7)** um modelo de representação deve ter URLs persistentes na Web de Dados. O acesso persistente ao servidor que hospeda as definições dos elementos do modelo facilita a compreensão das unidades constituintes (rótulos, definições e comentários), aumentando as chances de reutilização. **Requisito 8)** Para um modelo de representação, é prudente fornecer uma política de versionamento. Os grupos ou organizações promotoras do modelo devem explicitar a compatibilidade entre versões e manter as versões do modelo ao longo do tempo. Ademais, as alterações significantes devem ser mencionadas em documentação apropriada.

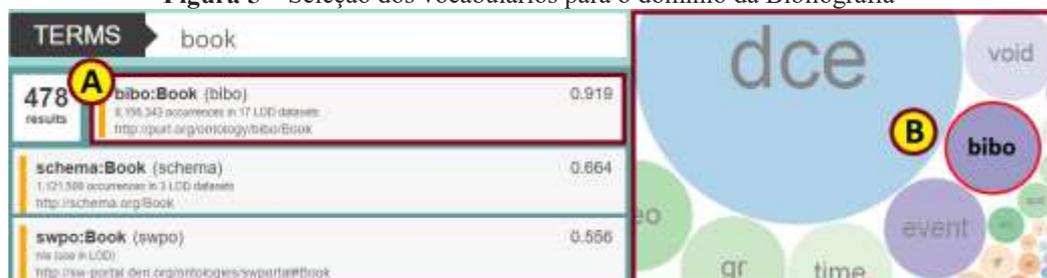
Definidos o percurso metodológico e os requisitos de adoção de um modelo de representação, a próxima seção discute os esforços despendidos na definição de um vocabulário ou de uma ontologia com vistas a organizar e a formalizar registros bibliográficos como Dados Abertos Conectados.

3 Definindo um modelo de representação de Dados Abertos Conectados para o domínio da Bibliografia

Conforme sugerido por Hyland, Ateazing e Villazón-Terrazas (2019), para reutilizar um modelo de representação para Dados Abertos Conectados, pode-se recorrer ao repositório LOV. No referido repositório, a busca por termos já modelados e que são inerentes ao domínio de interesse configura-se como uma estratégia pertinente para recuperar os possíveis candidatos de modelos de

representação compartilhados. Neste sentido, o termo utilizado para a busca de ontologias ou vocabulários aderentes ao campo da Bibliografia foi *book*.

Figura 5 – Seleção dos vocabulários para o domínio da Bibliografia



Fonte: Adaptada de Linked Open Vocabularies (2019a) e Linked Open Vocabularies (2019b).

Como ilustrado na Figura 5a, a partir do repositório LOV são encontradas 478 ocorrências de ontologias ou vocabulários que contém o termo *book*. E como destacado na Figura 5b, percebe-se a relevância da *Bibliografic Ontology* (prefixo *bibo:*) quanto ao volume de sua utilização na representação de objetos digitais na Web de Dados. Pontualmente, a referida ontologia é utilizada para representar cerca de oito milhões de ocorrências de entidades classificadas como *bibo:Book*. Tais fatos estão em consonância aos requisitos listados na seção anterior, principalmente, ao Requisito 4. Em poucas palavras, conforme a Figura 5, a *Bibliografic Ontology* é utilizada na representação de Dados Abertos Conectados em 17 conjuntos de dados. Diante a quantidade de objetos e o número de conjuntos de dados representados, considera-se que a *Bibliografic Ontology* é uma opção de modelo de representação relevante para organizar e formalizar recursos bibliográficos como Dados Abertos Conectados.

Em sua essência, a *Bibliografic Ontology* descreve recursos bibliográficos na Web de Dados. Segundo seus idealizadores, a referida ontologia pode ser usada como (D'ARCUS; GIASSON, 2019): **a**) uma ontologia de citação; **b**) uma ontologia de classificação de documentos; ou **c**) uma maneira de descrever qualquer tipo de documento na Web de Dados. Ressalta-se que a *Bibliografic Ontology* foi inspirada em diversos metadados de descrição de documentos (*dcterms* - *Dublin Core Terms*, *skos* - *Simple Knowledge Organization System*, *foaf* - *Friend of a Friend*, por exemplo). Ademais, a

Bibliografic Ontology também atende a outros requisitos de qualidade apontados por Hyland, Ateazing e Villazón-Terrazas (2019).

Neste sentido, quanto ao requisito de documentação e auto-descrição (Requisito 1 e Requisito 2, respectivamente), a *Bibliografic Ontology* utiliza definições semânticas para seus elementos. Isto é exemplificado na Figura 6. Na referida figura, ilustra-se a descrição da classe *book* (Figura 6a). Documentalmente, são exibidos os metadados referentes ao rótulo (nome sugestivo) e a uma definição clara (Figura 6b).

Figura 6 – Representação da classe *Book* com a *Bibliografic Ontology*



Fonte: Elaborada pelo autor (2019).

Adicionalmente na Figura 6 (partes C e D), é exemplificado como na *Bibliografic Ontology* são considerados os Requisito 7 e Requisito 8. Em poucas palavras, conforme o Requisito 7, um vocabulário deve ser URLs persistentes. Na Figura 6c evidencia-se a URL <<http://purl.org/ontology/bibo/>> como endereço padrão para o acesso aos recursos da *Bibliografic Ontology*. Em relação ao Requisito 8, o qual pontua sobre a política de versionamento, a Figura 6d aponta que a classe *Book* está em uma versão estável, sendo que a utilização desta classe como parte de um modelo de representação é recomendada.

Em relação ao Requisito 5 e Requisito 6, os quais circunscrevem questões de Curadoria Digital (perenidade dos recursos e reputação da organização custodiante) cabe ressaltar que, recentemente, a renomada organização *Dublin Core Metadata Initiative* tornou-se a entidade responsável pela manutenção da *Bibliografic Ontology* (DUBLIN CORE METADATA INITIATIVE, 2019). Dada a reputação e os serviços prestados pela referida

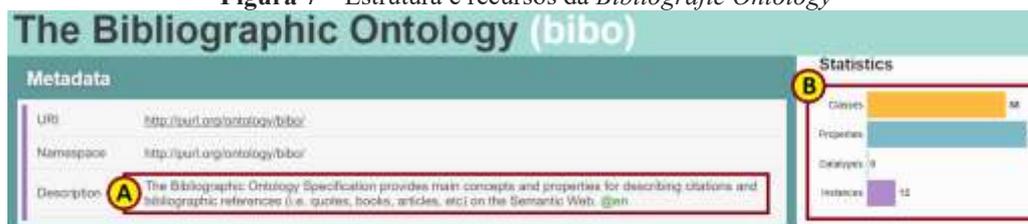
organização, pressupõe-se que tal fato indica o contínuo desenvolvimento e o uso regular da *Bibliografic Ontology* em pesquisas da Ciência da Informação.

Entretanto, dentre os requisitos estabelecidos, cabe ressaltar que o Requisito 3 não é devidamente atendido pela *Bibliografic Ontology*. Neste sentido, não foram observados indícios de multilinguismo para com os elementos da referida ontologia. Tal fato, entretanto, aponta para uma lacuna a ser explorada em pesquisas da Ciência da Informação.

4 Exemplificando o uso da *Bibliografic Ontology* no campo da Bibliografia

Ao considerar os registros bibliográficos, seus dados e metadados, de acordo com o entendimento abstraído de Alvares Jr e Saldanha (2017), a Bibliografia pode enveredar por duas trilhas. A primeira trilha concentra-se na representação e exploração do conteúdo de um documento. Os autores se referem a este ramo como a Bibliografia Intelectual, a qual é alcançada ao folhar, ler e entender a expressão do autor de uma obra. A segunda trilha abarca as características observáveis antes de interpretar o documento (por exemplo, dados a respeito da editora, local de publicação, preço, dentre outros). Alvares Jr e Saldanha (2017) denominam esta última trilha de Bibliografia Material. Nesta seção, pretende-se explorar um pequeno exemplo de verificação com a *Bibliografic Ontology*, observando como a referida ontologia permite representar alguns recursos digitais tanto da Bibliografia Intelectual, quanto da Bibliografia Material. Para tanto, utilizar-se-á alguns dados bibliográficos da obra “Quarto de Despejo: diário de uma favelada”. Salienta-se que o exemplo não esgota completamente a representação de todos os elementos bibliográficos da obra de Carolina Maria de Jesus.

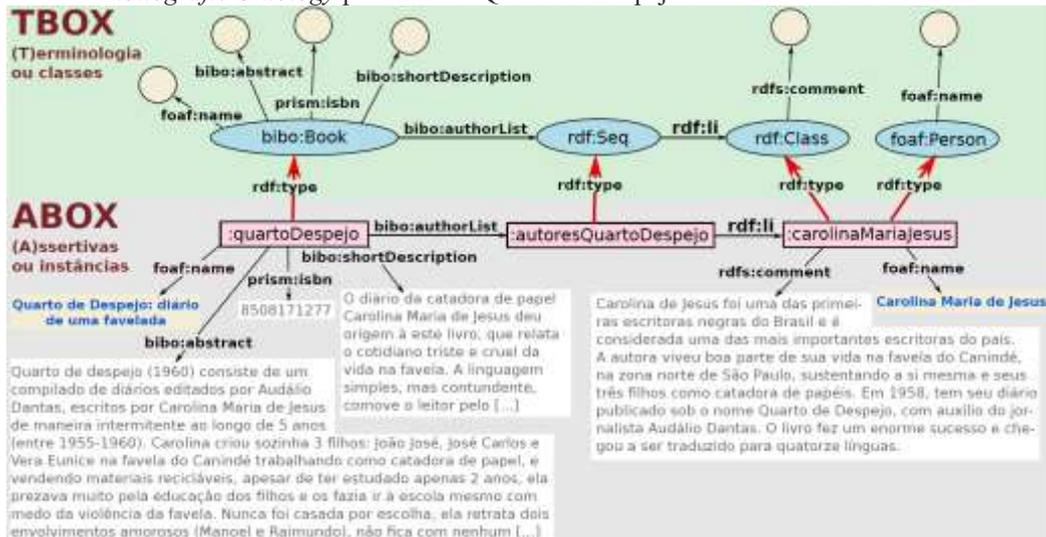
Figura 7 – Estrutura e recursos da *Bibliografic Ontology*



Fonte: Adaptada de Linked Open Vocabularies (2019c).

Conforme a Figura 7, observa-se que a *Bibliografic Ontology* compreende 58 classes e 67 propriedades, as quais podem ser utilizadas para organizar e representar semanticamente tanto os recursos digitais da Bibliografia Intelectual, quanto dos recursos digitais da Bibliografia Material.

Figura 8 – Representação gráfica de recursos digitais representados em conformidade à *Bibliografic Ontology* para a obra “Quarto de Despejo: diário de uma favelada”



Fonte: Elaborada pelo autor (2019).

De acordo com as classes e propriedades da *Bibliografic Ontology*, para o exemplo representado na Figura 8 da obra “Quarto de Despejo: diário de uma favelada”, são criados três recursos digitais (representados pelas assertivas ou instâncias na parte denominada ABOX). Um recurso é reservado à descrição semântica da obra (“:quartoDespejo”). Outro recurso é criado para organizar a lista de autores da obra em questão (“:autoresQuartoDespejo”). E finalmente, um recurso descreve a única autora da obra (“:carolinaMariaJesus”). Semanticamente, os recursos digitais criados são instâncias das classes representacionais “bibo:Book”, “rdf:Seq” e “foaf:Person”, demonstrando que a *Bibliografic Ontology* é baseada em e reutiliza outros modelos de representação já estabelecidos. Neste caso, são exemplificados o RDF (RDF WORKING GROUP, 2019), o *Publishing Requirements For Industry Standard Metadata* – PRISM (PRISM METADATA, 2019) e *Friend of a Friend* – FOAF (FOAF VOCABULARY SPECIFICATION, 2019).

Ainda conforme a Figura 8, pode-se observar como a *Bibliografic Ontology* organiza e representa semanticamente os elementos da Bibliografia Material. Neste sentido, por exemplo, as propriedades sobre a obra (“foaf:name”, “prism:isbn”, “bibo:shortDescription”) descrevem parte dos dados que podem ser capturados junto à editora do livro. Já em relação à Bibliografia Intelectual, exemplarmente, a propriedade “bibo:abstract” pode descrever o resumo ou a crítica produzida por um leitor. Neste caso, o conteúdo desta propriedade foi colhido a partir de (PET LETRAS, 2019).

Figura 9 – Exemplo de recursos digitais representados em conformidade à *Bibliografic Ontology* para a obra “Quarto de Despejo: diário de uma favelada”

```

01 [...]
02
03 ### http://lod.unicentro.br/Bibliografia/quartoDespejo
04 :quartoDespejo rdf:type :Book ;
05     prism:isbn 81763315 ;
06     foaf:name "Quarto de Despejo: diário de uma favelada" ;
07     :abstract "Quarto de despejo (1960) consiste de um [...]" ;
08     :shortDescription "O diário da catadora de papel [...]" ;
09     :shortTitle "Quarto de Despejo" ;
10     :authorList :autoresQuartoDespejo .
11
12 ### http://lod.unicentro.br/Bibliografia/autoresQuartoDespejo
13 :autoresQuartoDespejo rdf:type rdf:Seq ;
14     rdfs:label "Autores de Quarto de Despejo" ;
15     rdfs:comment "Autores da obra \"Quarto de Despejo\"" .
16
17
18 ### http://lod.unicentro.br/Bibliografia/carolinaMariaJesus
19 :carolinaMariaJesus rdf:type foaf:Person ;
20     foaf:name "Maria de Jesus" ;
21     rdfs:comment "Carolina de Jesus foi uma das [...]" .
22
23 [...]

```

■ recursos criados
■ metadados bibo
■ metadados de outros modelos

Fonte: Elaborada pelo autor (2019).

A Figura 9 expressa a codificação dos elementos criados no exemplo desta seção. Em vermelho, são destacados os recursos informacionais criados. Em azul, são evidenciadas as propriedades (ou os metadados) descritas com a *Bibliografic Ontology*. Já em negrito, destacam-se os metadados oriundos de outros modelos de representação utilizados na Web de Dados.

5 Considerações Finais

Neste artigo, apresentou-se o processo de escolha e prova de conceito de um modelo de representação a recursos digitais bibliográficos como Dados Abertos Conectados. O processo foi guiado em conformidade com as Melhores Práticas de Publicação na Web de Dados (LÓSCIO; BURLE; CALEGARI, 2019) e os

requisitos para modelos de organização e representação de Dados Abertos Conectados (HYLAND; ATEMEZING; VILLAZÓN-TERRAZAS, 2019). Neste sentido, ao se utilizar o repositório LOV (VANDENBUSSCHE *et. al.*, 2017), buscou-se por modelos adequados à representação, à organização e à publicação de registros bibliográficos na Web de Dados.

Atentando aos requisitos estabelecidos, definiu-se a *Bibliografic Ontology* como o modelo, atualmente, mais adequado. A *Bibliografic Ontology* é constituída por um conjunto de conceitos e propriedades para descrever semanticamente citações e referências bibliográficas na Web de Dados (D'ARCUS; GIASSON, 2019). Neste sentido, a referida ontologia pode organizar e representar, por exemplo: os recursos referentes às críticas dos leitores; a contextualização temporal da obra perante o momento de sua escrita; as convicções dos autores; a interpretação da obra; os dados a respeito da editora, do local de publicação, do preço, dentre outros.

Cabe ressaltar que a *Bibliografic Ontology* atende a sete dos oito requisitos apresentados por Hyland; Atemezing e Villazón-Terrazas (2019) para os modelos de organização e representação de dados da Web de Dados. Pormenorizando, na *Bibliografic Ontology* observa-se a preocupação com os requisitos de documentação e auto-descrição (Requisito 1 e Requisito 2, respectivamente). Para tanto, o referido modelo utiliza definições semânticas para seus elementos constituintes.

Quanto à utilização da *Bibliografic Ontology* na Web de Dados (Requisito 4), esta representa cerca de oito milhões de ocorrências de entidades classificadas como *bibo:Book* para 17 conjuntos de dados publicados na Web de Dados, por exemplo. Isso confirma o potencial da *Bibliografic Ontology* como modelo a ser utilizado em pesquisas que envolvam a publicação e o acesso aos registros inerentes de estudos da Bibliografia.

Em relação às questões que circunscrevem a Curadoria Digital de Dados (perenidade dos recursos e reputação da organização custodiante do modelo de representação - Requisito 5 e Requisito 6), ressalta-se que, recentemente, a *Dublin Core Metadata Initiative* tornou-se a entidade responsável pela manutenção da *Bibliografic Ontology* (DUBLIN CORE

METADATA INITIATIVE, 2019). Dada a reputação e os serviços prestados pela referida organização, tem-se a indicação do contínuo desenvolvimento da *Bibliografic Ontology*.

Por fim, a *Bibliografic Ontology* também atende aos Requisito 7 e Requisito 8. No tocante ao Requisito 7, a referida ontologia demonstra basear-se em URLs persistentes na *web* para encontrar referências aos seus elementos constituintes, a partir do URL <<http://purl.org/ontology/bibo/>>. E em relação ao Requisito 8 (política de versionamento), internamente, a *Bibliografic Ontology* apresenta insumos informacionais a respeito da estabilidade conceitual de seus elementos constituintes (Vide a Figura 6d).

Ademais, uma prova de conceito foi realizada, ao utilizar a *Bibliografic Ontology* para organizar e representar elementos bibliográficos da obra “Quarto de Despejo: diário de uma favelada”. Com o estudo realizado, percebeu-se que o modelo de representação da *Bibliografic Ontology* permite organizar e publicar registros bibliográficos, em consonância aos estudos da Bibliografia Intelectual e da Bibliografia Material. Vale destacar que na referida prova de conceito, a organização, a representação e a publicação de registros bibliográficos seguiram os preceitos de compartilhamento de Dados Abertos Conectados. Tal fato corrobora a alguns benefícios para com os publicadores e consumidores (LÓSCIO; BURLE; CALEGARI, 2019) desse tipo especial de recurso informacional. Em especial, destaca-se os benefícios de reuso, compreensão, confiança, interoperabilidade e processabilidade de recursos informacionais na Web de Dados.

Também é oportuno destacar que a *Bibliografic Ontology* apresenta algumas limitações quanto a sua adoção. Principalmente, no que se refere o Requisito 3 apontado por Hyland, Ateazing e Villazón-Terrazas (2019). Neste sentido, não foram observados indícios de multilinguismo da referida ontologia para com os elementos constituintes. Contudo, tal fato aponta para uma lacuna a ser explorada em futuras pesquisas da Ciência da Informação, com a finalidade de promover a utilização deste modelo de representação em demais pesquisas da Ciência da Informação.

Agradecimentos

À Fundação Araucária pelo suporte financeiro ao projeto de pesquisa intitulado “Curadoria Digital e Dados Abertos Conectados: um estudo da preservação de recursos digitais na Web de Dados para estudos cientométricos” (FA - Convênio 49773/19FA).

Referências

ALVARES JR, L. S.; SALDANHA, G. S. A bibliografia como ciência [Tradução do artigo de GUILD, Reuben A. Bibliography as science. *American Library Journal*, v. 1, n. 2-3, p. 67-69, nov. 1876.]. **Perspectivas em Ciência da Informação**, v. 22, número especial, p.199-202, 2017.

BERNERS-LEE, T.; HENDLER, J.; LASSILA, O. The Semantic Web: A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities. **Scientific American**, v.1, p. 34 - 43, 2001.

BIZER, C.; HEATH, T.; BERNERS-LEE, T. Linked data: the story so far. **International Journal of Semantic Web and Information Systems**, v. 5, n. 1, p. 1-22, 2009.

CUNHA, M. B.; CAVALCANTI, C. R. O. **Dicionário de biblioteconomia e arquivologia**. Brasília: Briquet de Lemos, 2008.

D'ARCUS, B.; GIASSON, F. **Bibliographic Ontology Specification | The Bibliographic Ontology**. Disponível em: <http://bibliontology.com>. Acesso em: 20 abr. 2019.

DUBLIN CORE METADATA INITIATIVE. **DCMI to Maintain the Bibliographic Ontology (BIBO)**. Disponível em: <http://www.dublincore.org/news/2018/11-30-bibo>. Acesso em: 20 de abr. 2019.

FOAF VOCABULARY SPECIFICATION. **FOAF Vocabulary Specification 0.99**. Disponível em: <http://xmlns.com/foaf/spec>. Acesso em: 20 abr. 2019.

HYLAND, B.; ATEMEZING, G.; VILLAZÓN-TERRAZAS, B. **Best Practices for Publishing Linked Data**. Disponível em: <https://www.w3.org/TR/ld-bp>. Acesso em: 20 abr. 2019.

LINKED DATA. **Linked Data: design issues**. Disponível em: <http://www.w3.org/DesignIssues/LinkedData.html>. Acesso em: 20 abr. 2019.

LINKED OPEN VOCABULARIES. **Linked Data Vocabularies**. Disponível em: <https://lov.linkeddata.es/dataset/lov>. Acesso em: 20 abr. 2019a.

LINKED OPEN VOCABULARIES. **Linked Data Vocabularies**. Disponível em: <https://lov.linkeddata.es/dataset/lov/terms?q=book>. Acesso em: 20 abr. 2019b.

LINKED OPEN VOCABULARIES. **Linked Data Vocabularies**. Disponível em: <https://lov.linkeddata.es/dataset/lov/vocabs/bibo>. Acesso em: 20 abr. 2019c.

LÓSCIO, B. F.; BURLE, C.; CALEGARI, N. **Data on the Web Best Practices - W3C Recommendation 31 January 2017**. Disponível em: <https://www.w3.org/TR/dwbp>. Acesso em: 20 abr. 2019.

PET LETRAS. **Resumo da obra “Quarto De Despejo”, de Carolina Maria de Jesus**. Disponível em: https://www2.unicentro.br/pet-letras/2017/08/30/resumo-da-obra-quarto-de-despejo-de-carolina-maria-de-jesus/?doing_wp_cron=1555815959.8100609779357910156250. Acesso em: 20 abr. 2019.

PRISM METADATA. **Prism Publishing Requirements for Industry Standard Metadata**. Disponível em: <https://www.idealliance.org/prism-metadata>. Acesso em: 20 abr. 2019.

RDF WORKING GROUP. **Resource Description Framework (RDF)**. Disponível em: <https://www.w3.org/RDF>. Acesso em: 20 abr. 2019.

SHADBOLT, N.; HALL, W.; BERNERS-LEE, T. The Semantic Web Revisited. **IEEE Intelligent Systems**, v. 21, n. 3, p. 96-101, 2006.

van der AALST, W. Data Scientist: The Engineer of the Future. *In: INTEROPERABILITY OF ENTERPRISES SYSTEMS AND APPLICATIONS CONFERENCE (I-ESA'2014)*, 2014, Albi, France. **Anais [...]**. Albi, France, 2014.

VANDEBUSSCHE, P.-Y.; ATEMEZING, G. A.; POVEDA-VILLALÓN, M.; VATANT, B. Linked Open Vocabularies (LOV): a gateway to reusable semantic vocabularies on the Web. **Semantic Web – Interoperability, Usability, Applicability**, v. 8, n. 3, p. 437-452, 2017.

Semantic Web and Bibliography: A Study for Publishing Bibliographic Records as Linked Open Data

Abstract: This paper presents an interdisciplinary study, involving elements of the Bibliography and Semantic Web. Specifically, it is based on the concepts of bibliographic records and Linked Open Data, considering that: i) the Bibliography deals with the different ways of registering knowledge in books, articles, newspapers, digital documents, among others; ii) considering the

expansion of information across the web, the Bibliography needs to include new Information and Communication Technologies for maintaining bibliographic records; and iii) one of the contemporary concepts for sharing information on the web is Linked Open Data, which can be used to organize and represent knowledge units and their semantics. Thus, we aim to investigate the alignment of the Linked Open Data precepts for organizing, representing and sharing data resources circumscribed to Bibliography studies. Therefore, a search was conducted, seeking for vocabularies commonly used to represent bibliographic elements. This research was carried out from a repository named Linked Open Vocabularies. As a result, an ontology was identified to organize and represent bibliographic resources on the Web of Data, the Bibliographic Ontology. Semantically, this ontology comprises 58 classes and 67 properties inspired by various metadata for describing documents on the web. In few words, the Bibliographic Ontology can be used as an ontology of citation and classification of documents. In this sense, we concluded that the Bibliographic Ontology promotes the interoperability of data and metadata, increasing the benefits of processability, interoperability, comprehension and reuse of records and their semantics in bibliographic studies.

Keywords: Bibliographic Registers. Ontologies. Vocabularies. Semantic Web. Linked Open Data.

Recebido: 23/04/2019

Aceito: 18/06/2019



-
- ¹ **URL – Uniform Resource Locator** - é um termo técnico que define o endereço de rede que referencia um recurso na Internet (por exemplo: um computador, um sítio ou serviço web).
- ² **CSV – Comma Separated Values** - é um formato para armazenar dados tabulares em texto, e cada linha do arquivo representa um registro. Cada registro é composto de um ou mais campos separados por vírgula. É usado para representar dados de estruturas simples, diretamente processados por editores de planilha ou de texto.
- ³ **JSON - JavaScript Object Notation** - é um formato para troca de dados, de fácil escrita e leitura para pessoas e fácil análise e geração para máquinas. É construído sobre duas estruturas: i) uma coleção de pares nome/valor; e ii) uma lista ordenada de valores. Ambas estruturas são amplamente usadas em linguagens de programação, facilitando seu entendimento e uso.
- ⁴ **RDF - Resource Description Framework** - é uma linguagem que usa um modelo padrão para troca de dados na web, o qual possui a forma de triplas, realizando a descrição de um recurso em três partes, ligando um sujeito (recurso) a um objeto através de um predicado. Isso gera uma estrutura de ligações que, por sua vez, formam um grafo RDF.
- ⁵ **XML - Extensible Markup Language** - é uma linguagem de marcação extensível que usa um formato padrão para descrição e troca de dados na web. Os dados são organizados de forma hierárquica, permitindo a representação de todos os tipos de estruturas de dados.