

Resumo: A literatura sobre dados e ontologias bibliográficas na *Web* Semântica identifica problemas, não ao nível das instâncias de dados ou da sua publicação em conjuntos isolados, mas sim relativamente às ontologias que descrevem os conceitos que lhe estão subjacentes, com reflexo na qualidade da interoperabilidade semântica e na partilha de ontologias entre sistemas.

Aborda-se a adequação à *Web* Semântica dos modelos conceptuais e limitações de FRBR - Functional Requirements for Bibliographic Records (IFLA 1998, 2018)¹; a ausência de enquadramento conceptual comum; a insuficiência de mecanismos semânticos; a baixa e deficiente reutilização de vocabulários externos; e a inadequação das metodologias de mapeamento aplicadas.

Apresenta-se um projeto de investigação que propõe uma solução para os problemas semânticos de partilha de ontologias, através da criação de um modelo de referência conceptualmente enquadrador e de uma ontologia de referência que funcione como instrumento de relacionamento semântico de alto nível e de validação de dados, recorrendo à linguagem SHACL - Shapes Constraint Language (KNUBLAUCH e KONTOKOSTAS, 2017).

Palavras-chave: Dados bibliográficos ligados; Normativos bibliográficos; Ontologias; *Web* Semântica.

Abstract: The literature on bibliographic data and ontologies on the Semantic Web identifies problems, not in terms of data instances or their publication in isolated sets, but regarding the ontologies that describe the underlying concepts, impacting on the quality of semantic interoperability and in sharing ontologies between systems.

This paper elaborates on the adequacy of conceptual models and the limitations of FRBR - Functional Requirements for Bibliographic Records (IFLA, 1998, 2018)¹ to the Semantic Web; the absence of a common conceptual framework; the insufficiency of semantic mechanisms; the low and deficient reuse of external vocabularies; and the inadequacy of mapping methodologies being applied.

A research project is presented proposing a solution to the semantic problems in sharing ontologies, through the creation of a conceptual reference model and a reference ontology as a high level mechanism for semantic relations and data validation using SHACL - Shapes Constraint Language (KNUBLAUCH e KONTOKOSTAS, 2017).

Keywords: Linked bibliographic data; Bibliographic standards; Ontologies; Semantic Web.

Introdução

Este artigo faz uma revisão dos principais problemas encontrados na literatura sobre ontologias bibliográficas na *Web* Semântica, para avançar como uma proposta de investigação sobre possíveis soluções. As principais limitações encontradas residem nas

¹ Para a evolução dos modelos baseados em FRBR ver IFLA (2018). Para a descrição dos elementos de dados FRBR em RDF, ver <https://www.iflstandards.info/fr>.

ontologias ou descrições dos conceitos, propriedades e relações subjacentes aos conjuntos de dados bibliográficos e não nas instâncias desses conjuntos de dados. Com efeito, as iniciativas de transformação de dados bibliográficos para a *Web Semântica* têm respeitado os princípios básicos de dados ligados, assegurando bons níveis de interoperabilidade técnica. Os problemas não surgem na publicação isolada de conjuntos de dados, mas sim relativamente à interoperabilidade semântica, *i.e.*, na partilha de ontologias entre sistemas.

Conforme afirma Dutta (2017), a ênfase tem recaído na publicação de dados seguindo os princípios de dados ligados, mas pouca atenção tem sido prestada à descrição desses dados em termos de conceitos, propriedades e relações entre conjuntos de dados. Este facto indicia um problema de expressividade nos conjuntos de dados ligados, que se limitam a ser uma mera coleção de triplos, que pouca utilização fazem do potencial das linguagens de ontologias como o RDFS – Resource Description Framework Schema (BRICKLEY e GUHA, 2014) e a OWL – Ontology Web Language (HITZLER *et al.*, 2012). No caso das ontologias bibliográficas, estas limitações semânticas fazem-se sentir não só ao nível do modelo conceptual que lhes está subjacente, mas também nas estruturas de descrição de dados implementadas e na forma como diferentes vocabulários de elementos são relacionados entre si. A baixa interoperabilidade semântica dos conjuntos de dados bibliográficos é agravada pelos problemas de interoperabilidade sintática das linguagens RDFS/OWL, cuja permissividade não garante a validação de dados e a deteção de erros ou inconsistências. Esta circunstância determina a necessidade de se utilizarem outras linguagens que permitam, no contexto do paradigma de mundo aberto da *Web Semântica*, detetar erros e inconsistências nas estruturas de dados bibliográficos.

Visando ultrapassar estes problemas, apresenta-se um projeto de investigação de soluções que envolvem a criação de um modelo de referência capaz de enquadrar conceptualmente as diferentes iniciativas de transformação de dados bibliográficos e de uma ontologia de referência que seja um instrumento de relacionamento semântico de alto nível e de validação de dados, recorrendo à linguagem SHACL que se diferencia, assim, dos sistemas tradicionais de controlo bibliográfico e de soluções de mapeamento próprias das bases de dados como *crosswalks* e perfis de aplicação.

Para o entendimento quer das questões em foco, quer da natureza e potencial alcance do tipo de soluções a investigar, torna-se importante, antes de mais, abordar os conceitos subjacentes à avaliação da qualidade dos dados na *Web Semântica*.

1. Qualidade dos dados na Web Semântica: conceitos básicos e avaliação

A dimensão fundamental da qualidade dos dados na *Web Semântica* é a da interoperabilidade que, em geral, consiste na capacidade de dois ou mais sistemas partilharem informação, de modo a poderem utilizá-la. Mas essa dimensão não tem uma só medida, antes se desenvolve em vários níveis: a interoperabilidade técnica, que respeita aos protocolos de troca de dados, que no caso dos dados ligados correspondem aos protocolos da *Web*; a interoperabilidade sintática, que está relacionada com os formatos de dados, neste caso as linguagens da *Web Semântica* como o RDF – Resource Description Framework (CYGANIAK, WOOD e LANTHALER, 2014), o RDFS e a OWL; e a interoperabilidade semântica, que consiste não apenas na simples troca de dados, mas na

partilha da sua significação pretendida. Para se entender essa significação, os dados trocados são descritos de forma não ambígua e interpretável por máquina, através da utilização de ontologias. Por esta razão se pode afirmar que a interoperabilidade semântica consiste na capacidade de dois ou mais sistemas trocarem e usarem ontologias.

Neste artigo, utilizaremos o termo “ontologia” no sentido T-Box (DE GIACOMO e LENZERINI, 1996), enquanto representação de nível mais elevado, que inclui conceitos, propriedades e restrições. Não utilizaremos a expressão “esquema de metadados”, pois as ontologias são sistemas mais complexos que fornecem regras de inferência e lógica descritiva para o raciocínio computacional (TALLERAS, 2018). Quanto às denominadas ontologias A-Box, que consistem nos dados ou instâncias geradas de acordo com o T-Box, preferimos usar a expressão “conjuntos de dados”.

Na análise das avaliações de ontologias bibliográficas, considerámos os trabalhos de autores que adaptaram aos dados ligados critérios de qualidade de dados originalmente definidos para o mundo fechado das bases de dados, a saber, Hogan (2012), Zaveri (2012), Kontokostas (2014), Schmachtenberg (2014), Farber (2016) e Dutta (2017).

Partilhamos com Kontokostas (2014) o entendimento de que a qualidade dos dados não tem uma medida absoluta, devendo resultar de uma avaliação da adequação dos dados face a um determinado caso de uso, consistindo, assim, na correspondência com as necessidades do utilizador e não na conformidade com uma especificação (JURAN e GODFREY, 1998). Mas no contexto da *Web Semântica* esse critério de adequação já não se limita às necessidades das organizações individuais, antes surge intimamente ligado ao relacionamento com outras comunidades na *Web de Dados*, consubstanciando-se em funções como o acesso integrado aos dados, através de ligação de diferentes conjuntos de dados na *Web* e do seu enriquecimento pelo interrelacionamento e contextualização (HOGAN, 2012; DUTTA, 2017; TALLERAS, 2018). Para satisfazer estas necessidades, a publicação de dados bibliográficos como dados ligados não se pode limitar ao cumprimento de requisitos de interoperabilidade técnica e sintática, mas terá de assentar na interoperabilidade semântica, sendo, por esse motivo, esta a categoria mais importante para a análise da qualidade dos dados neste artigo.

Para a navegação entre diferentes fontes de dados bibliográficos no contexto da *Web Semântica*, recorre-se a aplicações como navegadores e rastreadores de dados ligados que, através de *links* ou *queries*, agregam dados de conjuntos heterogéneos, como se tratasse de uma única base de dados, propiciando mais informação e respostas mais completas, à medida que novas fontes de dados vão aparecendo na *Web* (HOGAN, 2012). Para possibilitar esta exploração semântica, os dados devem ter, ao nível das instâncias, um grau de qualidade que permita que essas aplicações os possam localizar, processar e consumir. Ao nível das ontologias, a qualidade que possibilita a navegação e pesquisa semânticas avalia-se pela forma como esses vocabulários de elementos de dados se interligam e mapeiam entre si, como são reutilizados externamente, pelos padrões de modelação e pelo raciocínio permitido pela semântica que cada ontologia define (HOGAN, 2012).

Para cada nível de análise da interoperabilidade – técnica, sintática e semântica – foram consideradas as dimensões de avaliação de qualidade e critérios sugeridos pela literatura, conforme se sintetiza na Fig. 1:

Fig. 1 – Dimensões de qualidade por nível de interoperabilidade

Fonte: PATRÍCIO, CORDEIRO e RAMOS, 2019.

2. Avaliação de ontologias bibliográficas

A identificação dos problemas de interoperabilidade das ontologias bibliográficas que se apresenta em seguida resume a análise de literatura realizada em Patrício, Cordeiro e Ramos (2019 e 2020) e abrange não apenas implementações de dados ligados no domínio bibliográfico, mas em todas as categorias de conjuntos de dados publicados na nuvem de dados ligados (MCCRAE ed., 2020).

Da revisão da bibliografia, destacam-se as avaliações de Papadakis *et al.* (2015), relativa a iniciativas de transformação em dados ligados de sete bibliotecas nacionais; de Jett *et al.* (2016), que analisaram representações de normativos bibliográficos e da ontologia Schema.org (SCHEMA.ORG, 2021) como dados ligados; e de Talleras (2018) relativamente à publicação de dados bibliográficos como dados ligados pela Bibliothèque Nationale de France (BNF), pela Deutsche Nationalbibliothek (DNB), pela British Library (BNB – British National Bibliography) e pela Biblioteca Nacional de España (BNE). Foram ainda tidos em conta os resultados do inquérito da OCLC, realizado entre abril e maio de 2018, a 81 instituições e em que se incluem 13 bibliotecas nacionais (SMITH-YOSHIMURA, 2018).

Do conjunto da bibliografia analisada resulta que, em regra, os conjuntos de dados bibliográficos apresentam conformidade com os princípios de dados ligados (interoperabilidade técnica) havendo, no entanto, um elevado nível de divergência nas implementações que pode impedir a interoperabilidade semântica e conduzir a problemas significativos quanto aos critérios da completude e consistência. Estas limitações são agravadas pelos problemas de qualidade sintática das linguagens atualmente utilizadas para o desenvolvimento de ontologias, que não permitem a deteção desses erros, nem podem ser utilizadas na validação da qualidade dos dados a esse nível.

Nos pontos que se seguem, exploram-se em maior detalhe as limitações identificadas na bibliografia, seguindo as dimensões de análise de qualidade de dados sintetizadas na Fig. 1.

2.1. Interoperabilidade técnica

Os princípios e boas práticas de dados ligados correspondem, genericamente, à regra de cinco estrelas definida por Tim Berners-Lee (2006) e mais recentemente desenvolvida em recomendações e notas técnicas do W3C (HYLAND *et al.*, 2014; LÓSCIO e BURLE, 2017), que visam assegurar que os conjuntos de dados ligados observam os protocolos da *Web* e que, portanto, são tecnicamente interoperáveis.

Embora a maioria dos autores considere que os conjuntos de dados bibliográficos estão, em geral, conformes com os princípios essenciais de dados ligados, Talleras (2018) chama a atenção para o facto de tal não se verificar no que respeita às normas bibliográficas, uma vez que os seus elementos não estão, em regra, descritos de forma processável por máquina.

Existem, ainda, outros problemas pontuais identificados na bibliografia relativamente aos conjuntos de dados bibliográficos e que podem ser agrupados nas seguintes categorias de princípios enunciadas por Hogan (2012): nomeação com URI – Uniform Resource Identifier (URI PLANNING INTEREST GROUP, 2001), desreferenciação, utilização de RDF e de URI como *links* externos.

Quanto à nomeação inequívoca de entidades através de URI, verificam-se algumas falhas como a utilização de valores literais, e outras desconformidades como a utilização de nós brancos (HOGAN, 2012; TALLERAS, 2018).

No que respeita à desreferenciação, que consiste na utilização do protocolo HTTP na nomeação de entidades, ao contrário do que sucede na generalidade dos domínios, em que os conjuntos de dados são altamente conformes com esta recomendação (HOGAN, 2012), no caso dos dados bibliográficos ligados verifica-se a existência de alguns problemas. Com efeito, na análise de Papadakis (2015), verifica-se que os conjuntos de dados da BL, BNE, Biblioteca Nacional da Suécia (LIBRIS) e Biblioteca Nacional da Hungria não fazem desreferenciação. Por outro lado, os URI HTTP devem devolver representações RDF, pelo menos no formato RDF/XML, mas Hogan (2012) regista que 30% dos conjuntos de dados analisados estavam desconformes com este princípio.

No que se refere ao princípio que recomenda a utilização de URI para identificar entidades em conjuntos de dados externos, Schmachtenberg (2014) conclui que nos conjuntos de dados por ele analisados há um número relativamente baixo de *links* externos (8%, em média). Esta baixa percentagem de *links* externos verifica-se sobretudo nos conjuntos de dados bibliográficos, em que prevalecem ligações a entidades das mesmas comunidades (PAPADAKIS, 2015). Com efeito, Talleras (2018) detetou que os 4 conjuntos bibliográficos analisados (BNF, DNB, BNE e BNB) ligam para apenas 28 conjuntos de dados externos, dos quais 11 são bibliográficos e apenas 4 são de outras comunidades, destacando-se neste

último caso a DBPedia² e o Geonames³. Considerando as ligações entre elementos de ontologias, na análise de ontologias normativas efetuada por Jett *et al.* (2016) verifica-se que, com exceção do BIBFRAME – Bibliographic Framework Initiative (LIBRARY OF CONGRESS, 2016), todas utilizam preferencialmente valores literais em vez de URI nas ligações a elementos de outras ontologias.

2.1.1. Limitações sintáticas

Relativamente às linguagens de representação, Yee (2009) considera inadequada a utilização do RDF/RDFS e OWL para dados bibliográficos, devido à granularidade e atonicidade dos seus elementos. Outros autores chamam a atenção para a necessidade de se evitar constructos mais complexos do RDF como a reificação, as coleções e contentores (HOGAN, 2012; FARBER, 2016). Há, ainda, referência a problemas de incompatibilidade na utilização conjunta da OWL com o RDF/RDFS (FEENEY, BRENNAN e GLEASON, 2018).

Contudo, o maior problema sintático consiste na “permissividade” das linguagens da *Web Semântica* (FEENEY, BRENNAN e GLEASON, 2018), *i.e.*, na sua incapacidade para a definição de restrições e validação de estruturas de dados. De facto, os constructos de restrição do RDFS/OWL apenas constroem a utilização de um elemento para efeitos de inferência de nova informação, não podendo ser usados para validação (FEENEY, BRENNAN e GLEASON, 2018). Esta limitação tem um grande impacto na implementação de ontologias bibliográficas, pois é seu requisito a validação de estruturas de dados relativamente a determinados constrangimentos ou restrições (BAKER, COYLE e PETIYA, 2014).

Em suma, os problemas e incapacidades das linguagens da *Web Semântica* para a representação de dados bibliográficos consistem a) na impossibilidade de validação de estruturas de dados e b) na inadequação do RDF, conforme se explica nos pontos seguintes.

a) Impossibilidade de validação de estruturas de dados

O FRBR aplica o modelo ER (Entidade-Relação), tendo cada entidade determinado atributo ou propriedade. No entanto, no RDF a utilização de uma propriedade não pode ser limitada a determinada classe, pois a utilização de restrições RDF, como *range* e *domain*, para constroem a utilização de uma propriedade em função de uma dada classe apenas permite inferir nova informação, mas não valida a aplicação dessas propriedades. A especificação de estruturas de dados validáveis, necessária a modelos multi-entidade como o FRBR terá, pois, de ser feita com outras linguagens que não as da família RDF (BAKER, COYLE e PETIYA, 2014).

A mesma limitação surge no contexto da hierarquia, pois o RDF permite expressar transitividade de propriedades e classes, mas apenas para efeitos de inferência. Ou seja, não se pode recorrer ao RDF para “impor” uma determinada hierarquia. Quer isto dizer que o RDF não resolve os problemas históricos de falta de transitividades dos modelos de

² <https://www.dbpedia.org/>.

³ <https://www.geonames.org/>.

dados bibliográficos, pois, por exemplo, permite ligar entidades WEMI (*Work, Expression, Manifestation, Item*) mas não de forma hierárquica. Com efeito, a estrutura de dados em grafo (não hierárquica) do RDF permite conectar virtualmente tudo e em qualquer direção (YEE, 2009).

A especificação de muitas restrições em RDF, que ocorre na maior parte das ontologias bibliográficas, é também pouco adequada do ponto de vista da própria estrutura de grafos. Com efeito, os vocabulários RDF são mais reutilizáveis se tiverem menos restrições de dados. Esses constrangimentos isolam muito as ontologias, devendo ser expressos em “perfis” para controlo de qualidade de dados, de forma autónoma das ontologias a que se aplicam (BAKER, COYLE e PETIYA, 2014).

b) Inadequação do RDF

A *Web* semântica surge no contexto das “representações do conhecimento” da inteligência artificial. Contudo, nas bibliotecas não se pretende codificar conhecimento sobre o mundo, com o objetivo de os computadores raciocinarem sobre a realidade. Segundo Yee (2009), o que se pretende no âmbito da descrição bibliográfica é “representar informação” que descreva e indexe os recursos de modo a facilitar a sua recuperação. Por isso, considera conveniente analisar se o RDF pode ser usado para registar factos (não conhecimento) sobre recursos e sobre os seus assuntos (não sobre a “realidade”). Outra crítica que a mesma autora apresenta à adequação do RDF para a expressão de dados bibliográficos consiste na excessiva atomização dos dados, que se reflete em normativos como o RDA – Resource Description and Access (RDA STEERING COMMITTEE, 2020)⁴, demasiado granular para poder suportar RDF. Defende Yee (2009), que a atomização tem como objetivo a eficiência no processamento automático dos dados, mas pode prejudicar a sua utilidade para seres humanos, nomeadamente em tarefas como o seu agrupamento ou ordenação para apresentação e indexação.

Por último, Yee (2009) coloca a questão de saber se alguma vez a Internet será suficientemente rápida para agregar a informação equivalente a um recurso que se encontra dispersa em centenas ou até milhares de triplos RDF. Com efeito quanto mais granulares forem os dados, mais ligações são necessárias entre os elementos. Acresce que no RDF as ligações são unidirecionais, pelo que é sempre necessário multiplicar ligações recíprocas.

2.2. Problemas de interoperabilidade semântica

Ao nível interoperabilidade semântica ou da integração potencial de dados de diferentes conjuntos, as limitações das ontologias bibliográficas referenciadas pela literatura podem ser agrupadas em três dimensões de análise: a) adequação dos modelos conceptuais ao paradigma da *Web Semântica*; b) implementação de mecanismos semânticos pelas estruturas de dados; e c) reutilização de vocabulários externos.

⁴ A norma RDA, da responsabilidade do RDA Steering Committee, encontra-se publicada no RDA Toolkit, disponível em <https://www.rdatoolkit.org/subscribe>. Os elementos, classes, propriedades e vocabulários RDA estão acessíveis num *registry*, disponível em <http://www.rdaregistry.info/>.

a) Adequação dos modelos conceptuais

Ao nível conceptual é muito importante compreender se os modelos conceptuais bibliográficos estão bem alinhados ou se refletem o paradigma da *Web Semântica*, uma vez que os paradigmas que estão na base do desenvolvimento das linguagens de modelação constroem a sua aplicabilidade (CORDEIRO, 2005).

Conforme se explica em Patrício, Cordeiro e Ramos (2019, 2020), a não adequação dos modelos bibliográficos ao paradigma da *Web Semântica* resulta tanto de limitações próprias do modelo FRBR, como da ausência de um modelo de referência que funcione como enquadramento conceptual para as ontologias bibliográficas.

Limitações do FRBR

Em Patrício, Cordeiro e Ramos (2019 e 2020) foram resumidas as principais críticas ao modelo FRBR, que decorrem de os seus elementos terem sido derivados de normativos que assentam em paradigmas anteriores à *Web Semântica* (WILLER e DUNSIRE, 2013; MURRAY, 2008). Estas limitações consistem essencialmente na incapacidade de expressar hierarquia, com a conseqüente falta de transitividade e mecanismos de inferência; e no facto de o FRBR não ser verdadeiramente um modelo multi-entidade, uma vez que as suas entidades não são estruturas de dados desenhadas para serem conectadas (MURRAY, 2008), o que torna difícil a coexistência com outras comunidades de dados.

Com efeito, apesar de o FRBR ser o primeiro modelo conceptual explicitamente definido para dados bibliográficos e ter sido impulsionado pelo surgimento da Internet e das tecnologias World Wide Web, importa analisar até que ponto está alinhado com o paradigma da *Web Semântica*. O facto de os seus elementos terem derivado de normativos anteriores à *Web Semântica*, reforça a necessidade de se repensarem os modelos mais abstratos, e não apenas definir um enquadramento novo para elementos antigos (WILLER e DUNSIRE, 2013). Murray (2008) já identificara vários problemas decorrentes do facto de o modelo ter como requisitos funções e estruturas de dados característicos de sistemas antigos como o catálogo bibliográfico de fichas manuais ou o próprio formato MARC (Machine Readable Cataloging). Para além das entidades, atributos e relações do FRBR derivarem de normativos pré-existentes, o modelo não contextualiza esses elementos em ambientes mais abrangentes, quer sejam bibliográficos, quer de outros domínios, o que dificulta a coexistência das descrições FRBR com descrições produzidas por diferentes organizações, levando mesmo a que as entidades FRBR sejam muitas vezes vistas como puramente teóricas.

No que se refere às representações do FRBR em linguagens da *Web Semântica*, de que se destacam as ontologias FRBRer⁵, FRBR Core (DAVIS e NEWMAN, 2005) e FRBRoo (IFLA Working Group on FRBR/CRM Dialogue, 2016)⁶, alguns autores, como Karen Coyle (2016) e Peponakis (2012), referem que essas modelações não estão, em geral, bem alinhadas com os princípios e técnicas de dados ligados, pois não permitem a hierarquia de classes, impossibilitando a transitividade e mecanismos básicos de inferência. Consequentemente,

⁵ Elementos de dados em RDF disponíveis em <http://iflstandards.info/ns/fr/frbr/frbrer/>.

⁶ Elementos de dados em RDF disponíveis em <https://www.iflstandards.info/fr/frbr/frbroo.html>.

as entidades de nível inferior à sequência WEMI não conseguem utilizar os atributos das entidades superiores. Para Coyle (2016), esta intransitividade deriva do modelo subjacente ao FRBR, o modelo ER (Entidade-Relação), que não suporta hierarquias.

Nesta mesma linha, autores como Zapounidou, Sfakakis e Papatheodorou (2016) explicam que não existe hierarquia nas entidades WEMI por terem sido modeladas como classes disjuntas e implementadas com restrições de cardinalidade que determinam não a transitividade própria da hierarquia, mas uma sequência de instanciação de entidades WEMI. Com efeito, a disjunção de classes WEMI impede a partilha de propriedades ou relações e a ligação para dados similares (COYLE, 2015) e determina que cada instância apenas possa pertencer a uma das classes WEMI, quando na vida real os recursos podem ser instâncias de mais do que uma classe (BAKER, COYLE e PETIYA, 2014).

As divergências mais notórias do FRBR com a *Web Semântica* decorrem, precisamente, do modelo ER ser uma técnica própria do “mundo fechado” das bases de dados (COYLE, 2015), o que pode justificar o facto de, em muitos casos de publicação do modelo FRBR para a *Web Semântica*, ter havido uma mera “transcrição sintática” para RDF (WILLER e DUNSIRE, 2013). Neste mesmo sentido apontam Baker, Coyle e Petiya (2014) ao referir que a aplicação do modelo ER ao FRBR determinou que as entidades WEMI têm determinados atributos (restrições *domain* e *range*) e estão ligadas entre si por relações de dependência (restrições de cardinalidade). Segundo o modelo ER, os dados deveriam, portanto, poder ser validados contra estes constrangimentos. Sucede, no entanto, que estes conceitos ER não se adequam à *Web Semântica*, pois nem as restrições de domínio RDF, nem os axiomas OWL de cardinalidade validam dados. Em ambos os casos, são constrangimentos que apenas permitem inferir nova informação, que pode ser errada. Este é um problema das linguagens da *Web Semântica*, que se abordará mais à frente; mas traduz também o carácter superficial das publicações do FRBR como ontologias RDFS/OWL. Por exemplo, na ontologia FRBRer especificou-se a restrição de cardinalidade que determina que uma *Expression* só pode ser realização de 1 só obra. Neste caso, se uma expressão estiver relacionada com mais do que uma obra, os raciocinadores semânticos, ao invés de assinalarem essas declarações como erro, vão inferir que ambas as obras com que a expressão se relaciona são a mesma obra com diferentes URI, podendo suceder que esta nova informação inferida seja incorreta (BAKER, COYLE e PETIYA, 2014).

Sendo um modelo “multi-entidade”, o FRBR diverge conceptualmente dos modelos de estruturação de dados em “registos bibliográficos”, em que cada registo descreve uma única entidade, com os respetivos atributos e de que são exemplo o formato MARC ou a representação de registos em linguagens XML, como o Dublin Core (BAKER, COYLE e PETIYA, 2014). O FRBR mudou o foco do registo como um todo, para componentes desagregadas de dados, em que os elementos e atributos deixam de ser partes de um registo e passam a estar ligados a entidades específicas (HOWARTH, 2012). Este modelo está perfeitamente adequado às regras subjacentes aos dados ligados, pois ficando libertos dos registos, os dados podem mais facilmente ligar-se a outra informação. Trata-se de uma estrutura de dados mais flexível e apropriada para o ambiente digital, que permite um processamento mais eficiente por computador, o enriquecimento de dados com ligações externas, as agregações de dados a pedido e a criação de serviços adicionais para além do catálogo bibliográfico (PEPONAKIS, 2012). Por último, trata-se de um modelo especialmente adequado à *Web Semântica*, pois possibilita a expressão de múltiplos pontos de vista sobre determinado recurso, tanto pela comunidade bibliográfica, como por outras comunidades (BAKER, COYLE e PETIYA, 2014). A este propósito, Murray e Tillet (2011)

defendem, até, que cada entidade FRBR não é propriamente uma entidade autónoma, mas antes um ponto de vista sobre o recurso, que numa perspetiva multi-entidade pode ser expresso em grafos de dados bibliográficos que agrupam múltiplas declarações ou pontos de vista sobre determinado recurso. Cada grupo de declarações pode ter uma identidade própria, podendo ser reutilizada separadamente. Poderá, então, concluir-se que está ultrapassado objetivo de criação de um modelo bibliográfico único, subjacente à criação do FRBR nos anos 90, uma vez que o RDF está preparado para otimizar a fusão de dados de múltiplas fontes, através de estruturas de grafos que agrupam múltiplas descrições ou pontos de vista (BAKER, COYLE e PETIYA, 2014; MARTIN e MUNDLE, 2014).

Apesar da adequação à *Web Semântica* acima descrita, o modelo FRBR é criticado por não ter ido mais além, prevendo a criação de “uma super-entidade” que agrupe as entidades WEMI e por não lhes atribuir propriedades que permitam tratá-las como um todo. Por outro lado, o FRBR não surge como o modelo mais adequado a esta perspetiva das entidades como pontos de vista, pois efetua uma demarcação rígida das entidades WEMI e específica com pouca clareza as relações entre as mesmas, o que leva alguns autores a referir a necessidade de se repensar o FRBR (BAKER, COYLE e PETIYA, 2014), (MARTIN e MUNDLE, 2014) e ponderar a hipótese de criar um modelo multi-entidade diferente do FRBR, pois a distinção entre as entidades WEMI não pode ser tão rígida, dado não ter carácter universal e variar culturalmente (MURRAY e TILLET, 2011).

Em 2016 a IFLA colocou à discussão pública um novo modelo, visando não apenas uma consolidação editorial dos vários modelos anteriores da família FRBR, mas também a construção de um modelo único e coerente, apto a estruturar os dados bibliográficos com maior clareza, adequando-os à *Web Semântica* (RIVA, 2016). O novo modelo, designado IFLA Library Reference Model (IFLA-LRM) foi aprovado em agosto de 2017 (RIVA, LE BOEUF e ZUMER, 2017) e revisto em dezembro desse mesmo ano, mantendo o enquadramento ER, pelo que nesse aspeto se lhe aplicam as críticas acima apresentadas. No que respeita aos restantes problemas inerentes ao FRBR, tudo indica terem sido ultrapassados (RIVA, 2016) com a nova versão do modelo. Este novo modelo conceptual corresponderia à necessidade de combinar os diferentes ângulos de análise dos vários membros da “família FRBR”, utilizando um modelo e uma terminologia comuns (PEPONAKIS, 2016), sendo necessário analisar, em trabalhos futuros, se o modelo IFLA-LRM responde adequadamente a este objetivo.

Ausência de enquadramento conceptual

Tal como referido em Patrício, Cordeiro e Ramos (2019), a dispersão de normativos bibliográficos foi replicada na sua publicação em linguagens RDF, não estando garantida a consistência e a qualidade do seu interrelacionamento. A necessidade de um modelo enquadrador revela-se, em primeiro lugar, ao nível do relacionamento entre os normativos, já que a multiplicidade de normativos bibliográficos pode conduzir facilmente a contradições e a dificuldades na sua aplicação conjunta (YEE, 2009). Também Sprochi (2016) aponta a necessidade de um modelo de referência que enquadre os diferentes níveis normativos bibliográficos, pois normativos como o FRBR, RDA e BIBFRAME relacionam-se estreitamente e dependem fortemente uns dos outros, para implementação. Por outro lado, antes do RDA não existia propriamente um modelo lógico subjacente às práticas catalográficas; pois apesar de as AACR e outros códigos de catalogação propiciaram o nível de normalização necessário para a partilha e a disseminação de dados bibliográficos, baseavam-se mais nas práticas tradicionais do que numa análise dos elementos realmente

necessários para o utilizador identificar, seleccionar e obter os recursos de que necessita (SPROCHI, 2016).

No que respeita à publicação da ISBD - International Standard Bibliographic Description (IFLA, 2011) como dados ligados⁷, as principais críticas consistem no facto de não assentar num modelo explícito baseado em entidades e relações, como é próprio da *Web Semântica*, mas antes ser baseado no modelo ‘plano’ do registo bibliográfico como um texto (SVENSSON, 2013), sendo alguns autores de opinião que a ISBD deveria ser substituída por outra linguagem de descrição mais adequada aos novos paradigmas (WILLER e DUNSIRE, 2013).

O RDA, por sua vez, é um normativo que já nasce na geração do FRBR sendo, na opinião da generalidade dos autores (por ex., SZETO, 2013, e COYLE, 2016), completamente compatível com a *Web Semântica* porque implementa o FRBR enquanto modelo conceptual multi-entidade e, na sua representação RDF, tem ligação a ontologias como a DCMI Metadata Terms (DCMI, 2020), utilizando relações de subclasse com os seus elementos, facto que facilita a compreensão dos mais de 900 elementos do RDA. Por exemplo, ligando um sem número de tipos de “Título” RDA ao elemento “*Title*” de um vocabulário muito conhecido na *Web*, como é o Dublin Core (DCMI, 2012), os elementos RDA não ficam presos nos silos de dados bibliográficos (COYLE, 2016), o que representa uma mudança radical nos métodos tradicionais de catalogação (SPROCHI, 2016). Há, no entanto, autores como Peponakis (2016) que apontam divergências significativas entre o modelo FRBR e o RDA, o que justifica uma análise mais profunda desta norma bibliográfica.

Ao nível das normas de codificação, os formatos MARC constituem os esquemas de metadados mais utilizados para a codificação estruturada e a partilha de informação bibliográfica entre sistemas automatizados de bibliotecas (ORTIZ-REPISO JIMÉNEZ, 2002; RILLEY, 2017). A introdução de modelos conceptuais baseados em grafos e em modelos de árvore, como o FRBR, tornou os formatos MARC inadequados pois a sua estrutura de “modelo de registo”, correspondendo a ficheiros “planos”, originalmente pensados para serem acedidos de forma sequencial, implica uma redundância considerável de metadados e assenta em dados textuais (valores textuais) em vez de URI. À limitação do MARC em não mostrar facilmente relacionamentos em aberto contrapõe-se, na medida inversa, a grande vantagem do RDF tanto ao nível da modelação de dados como da escalabilidade dos sistemas: o RDF, tem a capacidade de representar relações múltiplas e em múltiplos sentidos, pois os predicados e os objetos podem eles próprios ser recursos e, assim, ter os seus próprios predicados e objetos. Ou seja, com o RDF a partir de um simples recurso podemos recuperar de forma rigorosa relações complexas com recursos externos. A vantagem do RDF face ao MARC está, pois, em ser um modelo de dados de grafos e utilizar URI, em vez de texto, apontando para outros recursos que têm mais informação sobre a propriedade ou sobre o objeto da declaração (MITCHELL, 2013). Apesar de já terem representações em RDF, os formatos MARC têm limitações estruturais de adaptação ao ambiente *Web*, que decorrem de se tratar de normas de codificação com mais de 40

⁷ Publicação disponível em <https://www.iflstandards.info/isbd>.

Por último, verifica-se que no panorama dos dados bibliográficos ligados existe uma tensão, que pode ser prejudicial do ponto de vista da consistência conceptual, entre diferentes abordagens. Por um lado, há as abordagens *top-down*, realizadas em processos de representação RDF de normativos bibliográficos como o RDA e o BIBFRAME, que consistem na criação de ontologias holísticas, com nomes únicos para classes e propriedades (TALLERAS, 2018), que visam instanciações exclusivas (VRANDECIC, 2010); por outro, as abordagens *bottom-up*, levadas a cabo pelas iniciativas de transformação realizadas por bibliotecas, que aplicam diferentes ontologias misturando elementos locais com elementos de vocabulários externos. A existência de um modelo conceptual enquadrador aliviaria esta tensão, garantindo um relacionamento consistente entre os diferentes tipos de ontologias bibliográficas.

b) Implementação de mecanismos semânticos

Mecanismos semânticos básicos, como a classificação e a hierarquia, não estão a ser utilizados pelas ontologias bibliográficas, não se tirando assim total partido das potencialidades de dados ligados. Com efeito, muitas ontologias aplicam classes externas diretamente ao nível das instâncias, não as reutilizando, porém, ao nível do vocabulário de elementos. Este facto impede, por exemplo, a inferência de todas as instâncias de uma classe local como instâncias de uma classe externa. Sem o mecanismo de classificação, torna-se necessário classificar cada instância ao nível dos dados.

Outro exemplo de subutilização de mecanismos semânticos é apontado por Coyle (2016) ao referir a pouca utilização da hierarquia pelo RDA, obrigando a que essas relações sejam definidas ao nível das instâncias.

Por último, os constrangimentos ou restrições de dados quando expressos em RDFS/OWL possibilitam a inferência de novas declarações, não sendo, no entanto, formalizadas pelos normativos bibliográficos com recurso a essas linguagens, antes constando apenas de notas textuais.

c) Reutilização de vocabulários externos

Tanto ao nível dos normativos, como das iniciativas locais de transformação de dados, existe um número excessivo de ontologias bibliográficas, cuja heterogeneidade, sobreposição e ausência de interligação tornam difícil a pesquisa, integração e reutilização de dados (JAIN *et al.*, 2010a), WILLER e DUNSIRE, 2013, TALLERAS, 2018). A proliferação de ontologias bibliográficas locais é bem evidenciada tanto pelos resultados do inquérito da OCLC, em que 22% das organizações reportaram utilizar vocabulários locais (SMITH-YOSHIMURA, 2018), quer pela análise de Talleras (2018), em que se conclui que a utilização de vocabulários locais nos quatro conjuntos de dados bibliográficos por si analisados foi, em média, de 70,4%, com cada conjunto de dados a utilizar elementos diferentes e exclusivos para expressar as mesmas entidades FRBR.

Para além dos baixos níveis de reutilização de ontologias, a bibliografia aponta ainda outros dois problemas na reutilização de vocabulários externos: a má utilização de ontologias externas e os mapeamentos ponto-a-ponto. Estes problemas foram já referenciados em Patrício, Cordeiro e Ramos (2020), apresentando-se em seguida um resumo das principais limitações semânticas reportadas na literatura.

Ausência de links externos

No desenvolvimento de ontologias bibliográficas tem sido seguida uma metodologia “*cherry-picking*” (GOSDBY, 2016), que consiste na utilização de elementos de vocabulários externos, misturados com classes e propriedades locais. Trata-se de uma boa prática que visa reduzir a heterogeneidade dos conjuntos de dados e aumentar a sua visibilidade em comunidades externas, sendo preferível à criação de termos “insulares” sem qualquer ligação a ontologias externas (HOGAN, 2012). No entanto, esta metodologia só é útil se incluir *links* para os elementos externos, o que não tem sido a prática na maior parte dos casos. Com efeito, análises preliminares referenciadas em Patrício, Cordeiro e Ramos (2019) chamam a atenção para níveis baixos de ligações para elementos externos no FRBRer, na ontologia ISBD e no RDA, por exemplo.

No que respeita às ontologias *bottom-up* desenvolvidas por bibliotecas, nos conjuntos de dados bibliográficos da BNE, BNF, DNB e BNB analisados por Talleras (2018) só 28 ontologias externas são referenciadas, das quais apenas 8 são partilhadas por pelo menos dois conjuntos de dados. Por outro lado, relativamente a vocabulários de elementos de dados, apenas 3 propriedades (owl:sameAs; rdf:type e dct:language) são partilhadas pelos quatro conjuntos de dados, num universo global de 1.141 propriedades (TALLERAS, 2018). No mesmo sentido, a análise de Jett *et al.* (2016) conclui que as ontologias analisadas não contêm declarações explícitas de equivalência entre classes, espelhando uma tendência no desenvolvimento de ontologias que poderá dever-se à incerteza semântica causada por ontologias pouco documentadas.

Por outro lado, verifica-se que não são os normativos bibliográficos as ontologias mais referenciadas pelos conjuntos de dados bibliográficos (SCHMACHTENBERG, 2014), sendo o BIBFRAME, ainda assim, a ontologia mais reutilizada (SMITH-YOSHIMURA, 2018). Quanto à reutilização de ontologias locais, o inquérito da OCLC demonstra que apenas a ontologia British Library Terms (BLTerms) é referenciada por mais do que um conjunto de dados (SMITH-YOSHIMURA, 2018).

Em suma, pelo reduzido número de ligações para elementos externos as ontologias bibliográficas não beneficiam da reutilização de vocabulários conhecidos, que suportem a interoperabilidade, aumentem a utilização de uma semântica comum e a usabilidade por terceiros.

Outro problema consiste na utilização de elementos externos de forma direta, ao nível das instâncias de dados, ou seja, sem que haja um alinhamento explícito com outros vocabulários, através de declarações de equivalência. A ausência deste tipo de ligações não permite que as ontologias bibliográficas beneficiem de vantagens como a criação de instâncias de estruturação ontológica muito específicas ou a ligação a vocabulários mais genéricos (JETT *et al.*, 2016), perdendo-se, assim, a decorrente otimização de pesquisa por motores de busca (Search Engine Optimization - SEO).

A situação mais grave provocada pela falta de ligação entre ontologias é a que ocorre nas denominadas “ontologias proprietárias” (SCHMACHTENBERG, 2014), que não são reutilizadas por nenhum outro vocabulário externo, sendo utilizadas apenas por um único conjunto de dados. Da análise de conjuntos de dados de vários domínios efetuada por Schmachtenberg (2014), conclui-se estarem nesta situação 59% dos vocabulários

representados na nuvem de dados ligados, descendo esta percentagem para 34% quando analisada apenas a categoria “Publicações”.

“Pilhagem” de ontologias

A definição formal e explícita de elementos realizada pelas ontologias não significa que essas definições sejam respeitadas, podendo haver erros nas ligações a elementos de outras ontologias externas se forem utilizadas de forma grosseira (HOGAN, 2012; TALLERAS, 2018). Com efeito, é difícil assegurar a consistência na reutilização de ontologias, sobretudo se forem desenvolvidas de forma independente e as suas componentes forem combinadas em momento posterior. Nesses casos pode ocorrer o que os autores Feeney, Brennan e Gleason (2018) denominam por “*ontology hijacking*” ou “pilhagem” de ontologias. Este problema é especialmente grave dada a permissividade das linguagens da Web Semântica, que não permitem identificar este tipo de problemas, nem impor mecanismos de validação.

A “pilhagem” de ontologias pode provocar erros de “interoperabilidade não coordenada” (FEENEY, BRENNAN e GLEASON, 2018), que se verifica quando cada ontologia faz referências externas na sua própria perspectiva e no seu âmbito. Com efeito, não sendo estas referências modulares, quando combinadas tornam-se inconsistentes no seu conjunto. Pode também acontecer que declarações liguem de forma errada duas ontologias diferentes e que esses erros não sejam reportados, dado essas relações estarem conformes o especificado nas declarações que ligam as ontologias. Erros como, por exemplo, referenciar determinada propriedade na ontologia A como uma classe na ontologia B, podem levar à inconsistência dos modelos (FEENEY, BRENNAN e GLEASON, 2018).

Outro problema da “pilhagem” de ontologias, é a denominada degenerescência das ontologias. Esta questão decorre do ciclo de vida das ontologias e consiste no facto de as ontologias referenciadas poderem ficar indisponíveis ou ser alteradas de tal modo que se tornem incompatíveis com as ontologias que as referenciam (FEENEY, BRENNAN e GLEASON, 2018). No que respeita à indisponibilidade de ontologias externas, por exemplo, 12% das ontologias analisadas no artigo de Feeney (2018) referenciam ontologias que deixaram de estar disponíveis. Por exemplo, as ontologias RDA utilizam termos de ontologias que já não existem⁸; e no FRBR Core, as classes `frbr:Work` e `frbr:Event` são definidas como subclasses de elementos de ontologias externas também já inexistentes.

Mas o problema mais grave da “pilhagem” ocorre quando há ontologias que explicitamente alteram outras ontologias, o que sucede, na maior parte dos casos, pelo recurso a relações de equivalência. Para evitar estas situações, é sempre preferível usar relações de hierarquia ou simplesmente aplicar diretamente o elemento da ontologia externa. Na análise feita por Feeney *et al.* (2018), as ontologias que mais violam ontologias terceiras são o FRBR Core (32 violações detetadas, alterações nas ontologias FOAF, DC, CC, GEO, RDF, entre outras) e a BIBO (50 violações no DC, FOAF, RDF E RDFS).

Relativamente a este último aspeto, na análise do conjunto de dados da BNE, Kontokostas *et al.* (2014) realizaram testes de correção face a todos os esquemas externos referenciados,

⁸Por exemplo, Open Metadata Registry Application Profile Properties (<http://metadataregistry.org/schema/show/id/98.html>)

verificando 11 mil erros face ao FRBRer, 37 mil face ao DCTerms e 28 milhões face à representação RDF da ISBD. A maior parte destes erros ocorreu por violação das propriedades *rdfs:range* e *rdfs:domain* e, bem assim, de propriedades de disjunção.

Em conclusão, torna-se urgente um mecanismo que permita a validação de conjuntos de dados, de modo a assegurar a respetiva correção ao nível dos esquemas ou ontologias que os mesmos aplicam.

Mapeamentos ponto-a-ponto

Sendo a reutilização de elementos muito pouco frequente e havendo um baixo nível de *links* para vocabulários externos, o mapeamento entre elementos de diferentes ontologias torna-se ainda mais relevante para a interoperabilidade. Apesar de existirem vários alinhamentos entre ontologias bibliográficas, não temos, contudo, conhecimento de nenhuma ontologia criada a um nível de abstração mais elevado para exprimir relações entre vocabulários bibliográficos.

Tal como se referiu em Patrício, Cordeiro e Ramos (2019), os mapeamentos entre ontologias bibliográficas têm-se realizado através de ligações ponto-a-ponto. Esta metodologia, resulta bem em relações 1:1, mas não assegura interoperabilidade semântica em relações 1:* ou *:1, assim como também não resolve situações de falta de correspondência (HOWARTH, 2012). A criação de uma ontologia central que representasse, a um nível de abstração superior, estas conexões semânticas entre vocabulários, preveniria estas situações.

No que respeita às ontologias locais de bibliotecas, os problemas de interoperabilidade dos conjuntos de dados identificados por Talleras (2018) resultam da aplicação dos princípios de dados ligados, seguindo uma metodologia de interoperabilidade baseada em perfis de aplicação, misturando elementos de diferentes normativos. Efetivamente, a utilização de perfis de aplicação e de outras tecnologias próprias do mundo fechado das bases de dados e dos esquemas XML, como "*crosswalks*" ou mapeamentos entre esquemas, facilita a troca de dados entre conjuntos com esquemas diferentes, mas não resolve questões de compatibilidade semântica (HOWARTH, 2012). De facto, estes conceitos de mapeamento diferem dos mecanismos de mapeamento semântico utilizados na interligação de ontologias (DOERR, RIVA e ZUMER, 2012), (DUNSIRE, HILLMAN e PHIPPS, 2012).

3. Modelo de referência e ontologia de referência em SHACL

A criação de modelos de referência e o desenvolvimento de normativos da *Web Semântica*, como língua franca para resolver problemas de heterogeneidade entre domínios e conjuntos de dados (TALLERAS, 2018), parece-nos ser o contexto em que deve inscrever-se a solução que nos propomos investigar.

Contudo, tal como foi já acima referido, uma simples adaptação ou combinação de ontologias em RDFS/OWL não é suficiente e pode revelar-se, até, inadequado em virtude das limitações dessas linguagens para determinados requisitos de interoperabilidade.

Neste contexto, importa investigar a possibilidade de, através de mecanismos de abstração mais elevada, potenciar a interoperabilidade semântica, a dois níveis:

- i) A criação de um modelo de referência, que funcione como uma estrutura de enquadramento conceptual de alto nível para a informação bibliográfica. Este modelo de referência permitiria melhorar a consistência dos normativos bibliográficos e a qualidade dos seus inter-relacionamentos, correspondendo à necessidade de repensar modelos a um nível mais abstrato, não nos limitando a meros reenquadramentos de elementos de dados antigos (WILLER e DUNSIRE, 2013).
- ii) A especificação de uma ontologia de referência baseada no modelo de referência, i.e., uma ontologia de nível de abstração mais elevado do que o das ontologias normativas e locais.

Seguindo o entendimento de autores como Jain *et al.* (2010b), Jett *et al.* (2016) e Feeney *et al.* (2018), consideramos que os problemas derivados da ausência de *links* semânticos entre ontologias podem ser resolvidos através de uma ontologia de alto nível, para aumentar a descoberta de conhecimento, a capacidade de raciocínio computacional e propiciar a verificações de consistência (JAIN *et al.*, 2010b).

Neste mesmo sentido, Jett *et al.* (2016) defendem o desenvolvimento de ontologias que se sobreponham a outras ontologias, fazendo “pontes” entre as suas classes. Feeney *et al.* (2018) sugerem a criação de um modelo unificado de ontologias, capaz de combinar vocabulários de dados ligados heterogêneos num modelo lógico consistente.

Esta ontologia de referência seria um instrumento para relacionar semanticamente elementos de diferentes vocabulários bibliográficos e para especificar mecanismos de restrição ou constrangimento de dados.

Para apoiar a resolução dos problemas de interoperabilidade semântica, a investigação que acima se propõe pretende explorar a possibilidade de: a) criação de mecanismos de enriquecimento semântico, através da ontologia de referência; e b) especificar mecanismos de restrição ou constrangimento de dados, recorrendo à linguagem SHACL.

a) Mecanismos de enriquecimento semântico

A ontologia de referência visa a representação de conceitos comuns e a formalização de relacionamentos e restrições transversais a todo o domínio do conhecimento bibliográfico, que poderão ser aplicados ao nível das ontologias locais ou de domínio, para potenciar as ligações entre os elementos de vocabulário.

Tendo em vista a resolução dos problemas de interoperabilidade das ontologias bibliográficas locais, há autores que defendem a adoção de um modelo único comum (SVENSSON, 2013), mas esse tipo de solução não nos parece conforme aos princípios fundamentais da *Web Semântica*. Com efeito, a *Web Semântica* assenta num modelo distribuído de conhecimento que se caracteriza pelo princípio de um mundo aberto (OWA – Open World Assumption) (GRIMM e MOTIK, 2005), em que cada um pode dizer qualquer coisa sobre tudo (AAA – Anyone can say Anything about anything) e pelo princípio da múltipla identificação (NUNA – Non Unique Name Assumption).

Segundo o princípio OWA, não saber se uma declaração é verdadeira não significa que a mesma seja falsa. Ou seja, a nova informação é cumulativa, pode ser contraditória com informação anterior, mas não a pode remover. Quer isto dizer que só é possível fazer

raciocínios sobre informação conhecida e que a ausência de uma informação não pode ser usada para inferir outra informação. Num mundo fechado, passa-se o contrário: podemos presumir como sendo falso tudo o que não sabemos ser verdadeiro. O princípio AAA decorre das próprias estruturas de dados da *Web Semântica*, que foram desenhadas precisamente para acomodar múltiplas fontes e pontos de vista, podendo haver declarações sem sentido ou inconsistentes (BAKER, COYLE e PETIYA, 2014). Por último, de acordo com o princípio NUNA, o mesmo recurso pode ser identificado por múltiplos URI, pois é descrito de forma distribuída, por múltiplos utilizadores, em múltiplas localizações, podendo cada um deles usar o seu próprio URI para descrever o recurso. O princípio dos nomes múltiplos estabelece que não se pode inferir que recursos identificados por múltiplos URI correspondam a diferentes recursos. A redundância e ambiguidade dos dados é assumida e, assim, podemos fazer equivaler recursos sem destruir informação que lhes está associada. No mundo fechado das bases de dados, passa-se o contrário, toda a informação é conhecida e, portanto, pode-se atribuir um identificador único, em cada sistema, para cada recurso (HEBELER *et al.*, 2009).

Em suma, a solução para a interoperabilidade semântica de ontologias bibliográficas tem de enquadrar-se nos princípios da *Web Semântica* e não em pressupostos do “mundo fechado” característico das bases de dados, pelo que não é possível adotar um modelo único como propõe Svensson (2013).

Justifica-se, assim, investigar a hipótese de uma ontologia de nível superior ao das ontologias de domínio (vocabulários normativos e ontologias locais), que permita maximizar a semântica e a inferência propiciada pelas linguagens OWL e RDFS, claramente subutilizadas pelas ontologias do domínio bibliográfico. Tal ontologia não corresponderia a um modelo conceptual único, pois no ambiente atual de metadados recombinaíveis já não é necessário (nem possível) impor um consenso através da conformidade obrigatória com formatos complexos e monolíticos, nem com restrições típicas de um “mundo fechado”. Os dados podem ser customizados para necessidades e contextos específicos, sendo a interoperabilidade alcançada na medida em que os grafos de dados partilhem vocabulários e restrições coincidentes, i.e., que se sobrepõem. Em apoio à teoria de Murray e Tillet (2011), é necessária uma mudança de perspetiva nos modelos bibliográficos, passando as entidades bibliográficas a ser vistas como grafos (BAKER, COYLE e PETIYA, 2014).

Se a conformidade com restrições não é necessária para a interoperabilidade semântica, verifica-se, contudo, que ao nível da qualidade dos dados é importante e útil definir e implementar constrangimentos de dados, sendo este aspeto tratado no ponto que se segue.

b) Utilização da linguagem SHACL

O recurso a linguagens da *Web Semântica* para melhorar a interoperabilidade nas ontologias locais não resolve a ausência de mecanismos de restrição ou constrangimento de dados (WILLER e DUNSIRE, 2013). Para este efeito, não se pode recorrer apenas ao RDFS e à OWL, pois foram desenhadas de acordo com os princípios OWA, AAA e NUNA tendo em vista a inferência e não para assegurar a conformidade dos dados com determinado esquema (BAKER, COYLE e PETIYA, 2014). São, aliás, os princípios OWA e NUNA que impedem que as linguagens de modelação tradicionais da *Web Semântica* possam ser usadas para validação de dados (ALLEMANG e HENDLER, 2011; KNUBLAUCH, 2017), uma vez que a semântica de inferência do RDFS/OWL tem natureza

informativa, *i.e.*, adiciona contexto e sentido mas não restringe a utilização (BAKER, COYLE e PETIYA, 2014).

A falta de normativos para expressar constrangimentos de dados e esquemas num contexto de “mundo fechado” conduziu à criação da SHACL, uma linguagem de esquema para RDF que permite a especificação de restrições (denominadas “*shapes*”) para a validação de grafos RDF (KNUBLAUCH e KONTOKOSTAS, 2017).

Aprovada como Recomendação do W3C, em julho de 2017, a SHACL vai mais além da OWL, pois não só permite validar dados, como também possibilita a inferência de novos factos, podendo ainda ser genericamente usada para inferências baseadas em regras (KNUBLAUCH, 2017). Recorreremos, assim, a esta linguagem para a construção da ontologia de referência já que é mais poderosa do que a OWL quanto aos mecanismos de inferência, permite a especificação de mecanismos de restrição para a validação de ontologias e é um normativo formal para a implementação de restrições de dados, permitindo que sejam processáveis por máquina.

Adicionalmente, investigaremos a hipótese de utilizar o vocabulário W3C PROF - *The Profiles Vocabulary* (ATKINSON e CAR, 2019), para descrever a nossa ontologia de referência como um recurso SHACL que define e implementa especificações e mecanismos de restrição aplicáveis quer a vocabulários mais gerais, como os normativos bibliográficos, quer a ontologias desenvolvidas localmente por bibliotecas, para aumentar a sua interoperabilidade semântica.

A utilização da SHACL e do vocabulário PROF no desenvolvimento da ontologia de referência, é uma solução que difere da criação de perfis de aplicação para esquemas de metadados (ATKINSON e CAR, 2019), distinguindo-se também das extensões de ontologias RDFS/OWL, que utilizam mecanismos de restrição apenas para inferência. Por outro lado, a nossa ontologia de referência pretende ser mais do que um perfil de mecanismos de restrição, distinguindo-se de perfis SHACL como o ARM SHACL (Art and Rare Materials BIBFRAME profile) (KOVARY *et al.*, 2018).

Contrariamente ao entendimento de que a SHACL apenas se aplica a instâncias de dados (DEBATTISTA, 2018) e de que não existe forma de definir, numa linguagem *standard*, esquemas de metadados com mecanismos de restrição (COYLE, 2019), defendemos que a SHACL é uma linguagem de validação que se aplica tanto a instâncias como a ontologias RDFS/OWL como, por exemplo, na validação da ontologia DBPedia (GAYO *et al.*, 2018).

Em suma, o nosso objetivo de investigação é testar e demonstrar a possibilidade de especificar uma ontologia de alto-nível para a descrição e validação de vocabulários bibliográficos, usando SCHACL como linguagem e PROF como vocabulário formal para perfis.

4. Conclusão

Neste artigo analisámos as questões que se colocam à interoperabilidade de dados e ontologias do universo bibliográfico no âmbito da *Web*, recolhendo, da literatura sobre o tema, os principais fatores que impedem a sua otimização. Constatou-se que esses fatores não se verificam ao nível da interoperabilidade técnica, mas antes decorrem, ao nível sintático, de limitações de linguagens como o RDF/RDFS e OWL quer para a representação de dados com a granularidade dos dados bibliográficos quer para a sua validação, um aspeto essencial para a qualidade da interoperabilidade no contexto da *Web Semântica*.

Outros fatores concorrem para as dificuldades em termos de interoperabilidade semântica: a falta de alinhamento conceptual entre os modelos do universo bibliográfico, como o FRBR, e a natureza de mundo aberto da *Web*; a mera tradução para linguagens da *Web Semântica* de normativos bibliográficos de outra geração, próprios do mundo fechado das bases de dados; e a proliferação de vocabulários normativos e ontologias bibliográficas locais pouco reutilizáveis, isto é, que fazem pouco ou deficiente uso de relacionamentos entre si e não permitem, assim, o raciocínio computacional inerente à *Web Semântica*.

Para a resolução destes problemas, e partindo do princípio da *Web Semântica* como “mundo aberto” (OWA), não se preconiza um modelo único ou a “normalização” de vocabulários e ontologias, mas antes a criação de mecanismos formais de enquadramento conceptual e de relacionamento de ontologias que permitam a coexistência de heterogeneidades. Ou seja, a criação de mecanismos integradores e processáveis na *Web Semântica*, que permitam maior eficiência no processamento de dados, ultrapassem a escassez de ligações externas a outros vocabulários e os mapeamentos ponto a ponto, que não tiram partido do mecanismo URI.

Neste sentido, propõe-se a investigação de um modelo de referência para enquadramento conceptual comum e de uma ontologia de alto nível que enriqueça e integre a semântica tanto de normativos bibliográficos, como de ontologias desenvolvidas localmente por bibliotecas. As soluções a investigar não terão a natureza de perfis de aplicação, nem de vocabulários de extensão ou de outros mecanismos próprios das linguagens de esquemas de metadados, uma vez que nenhum desses mecanismos, usualmente expressos em RDFS/OWL, tem efeitos de validação na *Web Semântica*. Em alternativa, propõe-se a utilização da linguagem SHACL, que permite a implementação de mecanismos de restrição que podem viabilizar essa validação e, assim, assegurar a qualidade semântica dos dados.

Referências bibliográficas

ALLEMANG, D.; HENDLER, J.

2011 *Semantic Web for the working ontologist: effective modeling in RDFS and OWL*. 2nd ed. Waltham (MA): Morgan Kaufmann, 2011.

ATKINSON, R.; CAR, N. J., ed.

2019 *The Profiles vocabulary: W3C working draft 2 April 2019, W3C*. [Em linha]. 2019. [Consult. 7 mar. 2021]. Disponível em: <https://www.w3.org/TR/dx-prof/>.

BAKER, T.; COYLE, K.; PETIYA, S.

2014 Multi-entity models of resource description in the Semantic Web: a comparison of FRBR, RDA and BIBFRAME. *Library Hi Tech*. [Em linha].32:4 (2014) 562-582. [Consult. 7 mar. 2021]. DOI: [10.1108/LHT-08-2014-0081](https://doi.org/10.1108/LHT-08-2014-0081).

BERNERS-LEE, T.

2006 *Linked data: W3C*. [Em linha]. 2006. [Consult. 7 mar. 2021]. Disponível em: <https://www.w3.org/DesignIssues/LinkedData.html>.

BRINKLEY, J. F.; GUHA, R.

2014 *RDF Schema 1.1: W3C Recommendation 25 February 2014*. [Em linha]. Cambridge (MA): W3C, 2014. [Consult. 7 mar. 2021]. Disponível em: <https://www.w3.org/TR/rdf-schema>.

BRINKLEY, J. F. [et al.]

2006 A Framework for using reference ontologies as a foundation for the semantic web. In AMIA ANNUAL SYMPOSIUM, Washington, 2006 – *Proceedings*. [Em linha]. Washington: AMIA, 2006, p. 96-100. [Consult. 7 mar. 2021]. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1839690>

CORDEIRO, M. I.

2005 *Information technology frameworks in LIS: exploring IT constructs as sources of conceptual alignment*. 2005. Tese de doutoramento - University College of London.

COYLE, K.

2019 I'd like to buy a VOWEL. *Coyle's InFormation*. [Em linha]. 2019. [Consult. 7 mar. 2021]. Disponível em: <http://kcoyle.blogspot.com/2019/03/id-like-to-buy-vowel.html>.

COYLE, K.

2016 *FRBR, before and after: a look at our bibliographic models*. [Em linha]. Chicago: American Library Association, 2016. [Consult. 7 mar. 2021]. Disponível em: <http://www.kcoyle.net/beforeAndAfter/978-0-8389-1364-2.pdf>.

COYLE, K.

2015 FRBR, twenty years on. *Cataloging and Classification Quarterly*. [Em linha]. 53:3/4 (2015) 265-285. [Consult. 7 mar. 2021]. DOI: <https://doi.org/10.1080/01639374.2014.943446>.

CYGANIAK, R., WOOD, D.; MCBRIDE, B.

2014 *RDF 1.1 concepts and abstract syntax: W3C Recommendation 25 February 2014*. [Em linha]. Cambridge (MA): W3C., 2014. [Consult. 7 mar. 2021]. Disponível em: <https://www.w3.org/TR/rdf11-concepts/>.

DAVIS, I.; NEWMAN, R.

2005 *Expression of core FRBR concepts in RDF*. [Em linha]. 2005. [Consult. 7 mar. 2021]. Disponível em: <https://vocab.org/frbr/core>.

DCMI

2020 *DCMI metadata terms*. [Em linha]. 2020. [Consult. 7 mar. 2021]. Disponível em: <https://www.dublincore.org/specifications/dublin-core/dcmi-terms/>.

DCMI

2012 *Dublin Core™ metadata element set*. [Em linha]. 2012. [Consult. 7 mar. 2021]. Disponível em: <https://www.dublincore.org/specifications/dublin-core/dces/>.

DE GIACOMO, G.; LENZERINI, M.

1996 TBox and ABox reasoning in expressive description logics. In INTERNATIONAL WORKSHOP ON DESCRIPTION LOGICS, Cambridge (MA) USA, 1996 – *Proceedings*. [Em linha]. Cambridge, MA, USA: AAAI Press, 1996. [Consult. 7 mar. 2021]. Disponível em: <https://www.aaai.org/Papers/Workshops/1996/WS-96-05/WS96-05-004.pdf>.

DEBATTISTA, J. [et al.]

2018 Evaluating the quality of the LOD cloud: an empirical investigation. *Semantic Web*. [Em linha]. 9:6 (2018) 859-901. [Consult. 7 mar. 2021]. Disponível em: <http://www.semantic-web-journal.net/system/files/swj1757.pdf>.

DOERR, M.; RIVA, P.; ZUMER, M.

2012 FRBR entities: identity and identification. *Cataloging & Classification Quarterly*. [Em linha]. 50:5-7 (2012) 517-541. [Consult. 7 mar. 2021]. DOI: <https://doi.org/10.1080/01639374.2012.681252>.

DUNSIRE, G., HILLMAN, D.; PHIPPS, J.

2012 Reconsidering Universal Bibliographic Control in light of the Semantic Web. *Journal of Library Metadata*. 12:2/3 (2012) 164-176.

DUTTA, B.

2017 Examining the interrelatedness between ontologies and Linked Data. *Library High Tech*. 35:2 (2017) 312-329.

FARBER, M. [et al.]

2016 Linked Data quality of DBpedia, Freebase, OpenCyc, Wikidata, and YAGO. *Semantic Web*. 1:1/5 (2016) 1-45.

FEENEY, K. C., BRENNAN, R.; GLEASON, G. M.

2018 Linked data schemata: fixing unsound foundations. *Semantic Web*. 9:1 (2018) 53-75.

GAYO, J. E. L. [et al.]

2018 *Validating RDF data*. San Rafael, CA: Morgan & Claypool, 2018.

GODBY, C. J.

2016 A Division of labor: the role of Schema.org in a Semantic Web model of library resources. In JONES, Ed, ed. - *Linked data for the cultural heritage*. Chicago: ALCTS, 2016, p. 73-101.

GRIMM, S.; MOTIK, B.

2005 Closed world reasoning in the Semantic Web through epistemic operators. In OWLED'05 WORKSHOP ON OWL: EXPERIENCES AND DIRECTIONS, Galway, Ireland, 2005 – *Proceedings*. [Em linha]. 2005. [Consult. 7 mar. 2021]. Disponível em: <http://ceur-ws.org/Vol-188/sub12.pdf>.

GUERRINI, M.; POSSEMATO, T.

2016 From record management to data management: RDA and new applications models BIBFRAME, RIMMF, and Olsuite/WeCat. *Cataloging & Classification Quarterly*. 54 (feb. 2016) 179-199.

HEBELER, J. [et al.]

2009 *Semantic Web programming*. Indianapolis: Willey, 2009.

HITZLER, P. [et al.]

2012 *OWL 2 Web Ontology Language Primer: W3C Recommendation 11 December 2012*. [Em linha]. 2nd ed. Cambridge (MA): W3C, 2012. [Consult. 7 mar. 2021]. Disponível em: <https://www.w3.org/TR/owl2-primer/>.

HOGAN, A. [et al.]

2012 An Empirical survey of Linked Data conformance. *Journal of Web Semantics*. [Em linha]. 14 (2012) 14-44. [Consult. 7 mar. 2021]. DOI: <https://doi.org/10.1016/j.websem.2012.02.001>.

HOWARTH, L. C.

2012 FRBR and linked data: connecting FRBR and linked data. *Cataloging & Classification Quarterly*. [Em linha]. 50:5/7 (2012) 763-776. [Consult. 7 mar. 2021]. DOI: <https://doi.org/10.1080/01639374.2012.680835>.

HYLAND, B. [et al.]

2014 *Best practices for publishing Linked Data*. W3C Working Group Note 9 January 2014. [Em linha]. Cambridge (MA): W3C, 2014. [Consult. 7 mar. 2021]. Disponível em: <https://www.w3.org/TR/ld-bp>.

IFLA

2018 *IFLA's bibliographic conceptual models*. [Em linha]. 2018. [Consult. 7 mar. 2021]. Disponível em: <https://www.ifla.org/node/2016>.

IFLA

2011 *ISBD: International Standard Bibliographic Description*. Consolidated ed. Berlin; München: De Gruyter Saur, 2011. (IFLA series on bibliographic control; vol. 44).

IFLA

1998 *Functional Requirements for Bibliographic Records*. [Em linha]. Munich: K.G. Saur Verlag, 1998. [Consult. 7 mar. 2021]. Disponível em: <https://www.ifla.org/publications/functional-requirements-for-bibliographic-records>.

IFLA. Working Group on FRBR/CRM Dialogue

2016 *Definition of FRBROO: a conceptual model for bibliographic information object-oriented formalism*. [Em linha]. 2016. [Consult. 7 mar. 2021]. Disponível em https://www.ifla.org/files/assets/cataloguing/FRBRoo/frbroo_v_2.4.pdf.

JETT, J. [et al.]

2016 Enhancing scholarly use of digital libraries: a comparative survey and review of bibliographic metadata ontologies. In *2016 IEEE/ACM Joint Conference on Digital Libraries (JCDL)*. 2016, p. 35-44.

JAIN, P. [et al.]

2010a *Linked data is merely more data*. Stanford, CA: Association for the Advancement of Artificial Intelligence, 2010.

JAIN, P. [et al.]

2010b *Ontology alignment for Linked Open Data*. Shangai: Springer, 2010.

JURAN, J. M.; GODFREY, A. B.

1998 *Juran's quality handbook*. 5th ed. New York: McGraw-Hill, 1998.

KNUBLAUCH, H.

2017 *SHACL and OWL compared*. *spinrdf.org*. [Em linha]. 2017. [Consult. 7 mar. 2021]. Disponível em: <http://spinrdf.org/shacl-and-owl.html>.

KNUBLAUCH, H.; KONTOKOSTAS, D., ed.

2017 *Shapes Constraint Language (SHACL): W3C Recommendation 20 July 2017*. [Em linha]. 2017. [Consult. 7 mar. 2021]. Disponível em: <https://www.w3.org/TR/shacl/>.

KONTOKOSTAS, D. [et al.]

2014 Test-driven evaluation of Linked Data quality. In *WWW '14 Proceedings of the 23rd international conference on World wide web*. New York: ACM, 2014, p. 747-757.

KOVARI, J. [et al.]

2018 Modeling application profiles in the art and rare materials BIBFRAME ontology extension. In *INTERNATIONAL CONFERENCE ON DUBLIN CORE AND METADATA APPLICATIONS, 2018 – Proceedings*. [Em linha]. [Consult. 7 mar. 2021]. Disponível em: <http://dcpapers.dublincore.org/pubs/article/download/3979/2181>.

LIBRARY OF CONGRESS

2016 *Bibliographic Framework Initiative*. [Em linha]. 2016. [Consult. 7 mar. 2021]. Disponível em: <https://www.loc.gov/bibframe/>.

LÓSCIO, B. F.; BURLE, C.

2017 *Data on the Web best practices: W3C Recommendation 31 January 2017*. [Em linha]. 2017. [Consult. 7 mar. 2021]. <https://www.w3.org/TR/dwbp/#dataFormats>.

MARTIN, K. E.; MUNDLE, K.

2014 Positioning libraries for a new bibliographic universe: a review of cataloging and classification literature 2011-12. *Library Resources & Technical Service*. [Em linha]. 58:4 (2014) 233-249. [Consult. 7 mar. 2021]. Disponível em: <https://journals.ala.org/index.php/lrts/article/view/5408/6637>.

MCCRAE, J. P.

2020 *The Linked data open cloud*. [Em linha]. 2020. [Consult. 7 mar. 2021]. Disponível em: <https://lod-cloud.net>.

MITCHELL, E. T.

2013 Metadata developments in libraries and other cultural heritage institutions. *Library Technology Reports*. (Jul. 2013) 5-10.

MURRAY, R. J.

2008 The FRBR theoretic library: the role of conceptual data modeling in cultural heritage information system design. In *iPRES 2008 - INTERNATIONAL CONFERENCE ON PRESERVATION OF DIGITAL OBJECTS, 5th, 2008 - Proceedings*. London: British Library, 2008, p. 163-168.

MURRAY, R. J.; TILLET, B.

2011 Cataloging theory in search of graph theory and other ivory towers. *Information Technology and Libraries*. [Em linha]. 30:4 (2011) 170-184. [Consult. 7 mar. 2021]. Disponível em: <https://ejournals.bc.edu/ojs/index.php/ital/article/download/1868/1706>

ORTIZ-REPISO JIMÉNEZ, V.

2002 Análisis documental formal: la catalogación. In LÓPEZ YEPES, J., ed. - *Manual de Ciencias de la Documentación*. Madrid: Pirámide, 2002, p. 359-377.

PAPADAKIS, I.; KYPRIANOS, K.; STEFANIDAKIS, M.

2015 Linked Data URIs and libraries: the story so far. *D-Lib Magazine*. [Em linha]. 21:5/6 (2015). [Consult. 7 mar. 2021]. Disponível em:
<http://www.dlib.org/dlib/may15/papadakis/05papadakis.html>.

PATRÍCIO, H. S., CORDEIRO, M. I.; RAMOS, P. N.

2020 From the web of bibliographic data to the web of bibliographic meaning: structuring, interlinking and validating ontologies on the semantic web. *International Journal of Metadata, Semantics and Ontologies*. [Em linha]. 14:2 (2020) p.124-134. [Consult. 7 mar. 2021]. DOI: <https://dx.doi.org/10.1504/IJMSO.2020.108318>.

PATRÍCIO, H. S., CORDEIRO, M. I.; RAMOS, P. N.

2019 Formalizing enrichment mechanisms for bibliographic ontologies in the semantic web. In GAROUFALLOU, E. *et al.*, ed. - *Metadata and semantic research*. [Em linha]. [Heidelberg]: Springer, Cham, 2019, p. 147-158. (Communications in Computer and Information Science; vol 846. [Consult. 7 mar. 2021]. DOI: [10.1007/978-3-030-14401-2_14](https://doi.org/10.1007/978-3-030-14401-2_14).

PEPONAKIS, M.

2016 In the name of the name: RDF literals, ER attributes, and the potential to rethink the structures and visualizations of catalogs. *Information technology and libraries*. (Jun. 2016) 19-38.

PEPONAKIS, M.

2012 Conceptualizations of the cataloging object: a critique on current perceptions of FRBR Group 1 entities. *Cataloging & Classification Quarterly*. [Em linha]. 50:5/7 (2012) 587-602. Disponível em:
http://eprints.rclis.org/17461/1/Peponakis_FRBR_Group1_Conceptualizations.pdf.

RDA STEERING COMMITTEE

2020 *RDA Registry*. [Em linha]. 2020. [Consult. 7 mar. 2021]. Disponível em:
<http://www.rdaregistry.info/>.

RILEY, J.

2017 *Understanding metadata: what is metadata and what is it for?* Baltimore: National Information Standards Organization, 2017.

RIVA, P.

2016 Il Nuovo modello concettuale dell' universo bibliografico: FRBR Library Reference Model. *AIB Studi*. [Em linha]. 56:2 (2016) 265-275. [Consult. 7 mar. 2021]. Disponível em: <https://aibstudi.aib.it/article/download/11480/10740>.

RIVA, P., LE BOEUF, P.; ŽUMER, M.

2017 *IFLA Library Reference Model: a conceptual model for bibliographic information*. [Em linha]. Den Haag: IFLA, 2017. [Consult. 7 mar. 2021]. Disponível em:
https://www.ifla.org/files/assets/cataloguing/frbr-lrm/ifla-lrm-august-2017_rev201712.pdf.

SCHEMA.ORG

2021 *Schema.org: organization of schemas*. [Em linha]. 2021. [Consult. 7 mar. 2021]. Disponível em: <https://schema.org/docs/schemas.html>.

SCHMACHTENBERG, M., BIZER, C.; PAULHEIM, H.

2014 Adoption of the Linked Data best practices in different topical domains. In MIKA, P. *et al.*, ed. - *The Semantic Web*. 2014, p. 245-260. (ISWC 2014: Lecture Notes in Computer Science; vol 8.796).

SMITH-YOSHIMURA, K.

2018 Analysis of 2018 international linked data survey for implementers. *Code4Lib Journal*. [Em linha]. 42 (2018). [Consult. 7 mar. 2021]. Disponível em: <https://journal.code4lib.org/articles/13867>.

SPROCHI, A.

2016 Where are we headed? : Resource Description and Access, Bibliographic Framework, and the Functional Requirements for Bibliographic Records Library Reference Model. *International Information & Library Review*. [Em linha]. 48:2 (2016) 129-136. [Consult. 7 mar. 2021]. DOI <https://doi.org/10.1080/10572317.2016.1176455>.

SUOMINEN, O.; HYVONEN, N.

2017 From MARC silos to Linked Data silos? *O-bib: Das Offene Bibliotheksjournal*. [Em linha]. 4:2 (2017). Disponível em: <https://www.o-bib.de/article/viewFile/2017H2S1-13/5854>.

SVENSSON, L. G.

2013 Are current bibliographic models suitable for integration with the Web? *Information Standards Quarterly*. [Em linha]. 25:4 (2013) 7-13. [Consult. 7 mar. 2021]. Disponível em: http://www.niso.org/apps/group_public/download.php/11931/FE_Svensson_bibliographic_models_isqv25no4.pdf.

SZETO, K.

2013 Positioning library data for the Semantic Web: recent developments in resource description. *Journal of Web Librarianship*. [Em linha]. 7:3 (2013) 305-321. [Consult. 7 mar. 2021]. DOI: [10.1080/19322909.2013.802584](https://doi.org/10.1080/19322909.2013.802584).

TALLERAS, K.

2018 *Metadata structures of the Bibliographic Universe: transformation, interoperability, conceptualizations and quality*. [Em linha]. 2018. [Consult. 7 mar. 2021]. Disponível em: <https://oda-hioa.archive.knowledgearc.net/handle/10642/6622>.
Tese de doutoramento - Oslo Metropolitan University.

URI PLANNING INTEREST GROUP

2001 *URIs, URLs, and URNs: clarifications and recommendations 1.0: report from the joint W3C/IETF URI Planning Interest Group, W3C Note 21 September 2001*. [Em linha]. 2001. [Consult. 7 mar. 2021]. Disponível em: <https://www.w3.org/TR/uri-clarification/>.

VRANDECIC, D.

2010 *Ontology evaluation*. [Em linha]. 2010. [Consult. 7 mar. 2021]. Disponível em: http://simia.net/download/ontology_evaluation.pdf.
Tese de doutoramento - Karlsruher Institut für Technologie.

WILLER, M.; DUNSIRE, G.

2013 *Bibliographic information organization in the Semantic Web*. Oxford: Chandos, 2013.

YEE, M. M.

2009 Can bibliographic data be put directly onto the Semantic Web? *Information Technology and Libraries*. [Em linha]. (Jun. 2009) 55-80. [Consult. 7 mar. 2021]. Disponível em:

<https://ejournals.bc.edu/ojs/index.php/ital/article/viewFile/3175/2788>.

ZAPOUNIDOU, S., SFAKAKIS, M.; PAPTAEODOROU, C.

2016 Representing and integrating bibliographic information into the Semantic Web: a comparison of four conceptual models. *Journal of Information Science*. [Em linha]. (2016) 1-29. [Consult. 7 mar. 2021]. Disponível em:

<http://journals.sagepub.com/doi/pdf/10.1177/0165551516650410>

ZAVERI, A. [et al.]

2012 *Quality assessment methodologies for Linked Open Data: a systematic literature review and conceptual framework*. [Em linha]. {S. l.}, IOS Press, 2012. [Consult. 7 mar. 2021]. Disponível em:

<https://pdfs.semanticscholar.org/916a/ef4b7cb2e650628ef48b364850d209aff041.pdf>.

Helena Simões Patrício | hpatricio@bnportugal.gov.pt

Biblioteca Nacional de Portugal; ISCTE – Instituto Universitário de Lisboa

Maria Inês Cordeiro | icordeiro@bnportugal.gov.pt

Biblioteca Nacional de Portugal

Pedro Nogueira Ramos | pedro.ramos@iscte-iul.pt

ISCTE – Instituto Universitário de Lisboa