

XVIII ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2017

GT-02 – Organização e Representação do Conhecimento

REPRESENTANDO ÍNDICES CIENTOMÉTRICOS COMO DADOS ABERTOS CONECTADOS

Sandro Rautenberg - Universidade Estadual do Centro-Oeste (UNICENTRO)

Lucélia de Souza - Universidade Estadual do Centro-Oeste (UNICENTRO)

Josiane M. H. Dall’Agnol - Universidade Estadual do Centro-Oeste (UNICENTRO)

Tony Alexander Hild - Universidade Estadual do Centro-Oeste (UNICENTRO)

Gisane Michelin - Universidade Estadual do Centro-Oeste (UNICENTRO)

Alessandra Burda - Universidade Estadual do Centro-Oeste (UNICENTRO)

REPRESENTING SCIENTOMETRIC INDEXES AS LINKED OPEN DATA

Modalidade da Apresentação: Comunicação Oral

Resumo: A *Web de Dados* é uma plataforma global para socialização de recursos de dados em vários domínios e aplicações. Representar, preservar e curar dados nesta plataforma são atividades importantes à luz da Ciência da Informação. Neste sentido, relata-se a experiência em representar alguns elementos cientométricos na *Web de Dados* conforme as práticas da *Web Semântica*. Discute-se as decisões tomadas na modelagem dos índices *Qualis*, *SCImago Journal & Country Rank* e *Source Normalized Impact per Paper* como *Dados Abertos Conectados*. Como uma pesquisa aplicada, o desenvolvimento do trabalho fundamenta-se nas *Melhores Práticas para publicação de dados na Web de Dados* sugeridas pelo *World Wide Web Consortium* e no ciclo de vida *Linked Data Lifecycle*. Como resultado, uma representação é estabelecida, a qual contribui para o reuso, a compreensão, a interligação e a descoberta de recursos cientométricos em escala global. Desta maneira, os esforços despendidos também colaboram na interoperabilidade e na processabilidade automatizada de dados abertos em sistemas computacionais voltados à aquisição de informação cientométrica.

Palavras-Chave: Representação da Informação. Metadados. *Web Semântica*. Preservação Digital.

Abstract: The *Web of Data* is a global platform for socializing data resources across multiple domains and applications. Representing, preserving and curating data on this platform are

important activities for Information Science. The present work reports the experience in representing scientometric elements in the Web of Data according to the practices of the Semantic Web. We discuss the decisions taken in modeling the indexes Qualis, SCImago Journal & Country Rank and Source Normalized Impact per Paper as Linked Open Data. As an applied research, the development of this work is based on the Best Practices suggested by World Wide Web Consortium and the Linked Data Lifecycle. As a result, a representation is established which contributes to the reuse, comprehension, linkability and the discoverability of scientometric resources on a global scale. In this way, the expended efforts also collaborate on the interoperability and the processability of open data in scientific information systems.

Keywords: Information Representation, Metadata. Semantic Web. Digital Preservation.

1 INTRODUÇÃO

O contexto que circunscreve este trabalho baseia-se no fato que, atualmente, a humanidade tem produzido imensas bases de dados/informação para os mais variados domínios. Para Bugnion, Manivannan e Nicolas (2017), 90% dos dados produzidos pela humanidade foram gerados nos últimos 18 meses, muito em função das Tecnologias de Informação e Comunicação. Gigantes como Google, Twitter, Amazon, Netflix, dentre outros atores, têm contribuído significativamente na geração e na disseminação de grandes volumes de dados na Internet. Parte dessa expansão informacional tem suporte nos modelos da *Web Semântica* (BERNERS-LEE; HENDLER; LASSILA, 2001), os quais promovem o que se entende como a *Web de Dados*, uma imensa base informacional a ser compreendida e também fomentada pela Ciência da Informação. Neste contexto, considera-se que a *Web de Dados* é uma infraestrutura de vanguarda para socializar os elementos primários úteis em estudos da Ciência da Informação.

Segundo W3C (2017), utilizar a *Web de Dados* como estrutura global de publicação de dados abertos permite que usuários possam reusar, compreender, interligar, descobrir e acessar os dados na *Web*, dispondo aos sistemas de informação uma infraestrutura à interoperabilidade e à processabilidade automatizada dos recursos de dados. Neste sentido, insere-se a necessidade de preservar os dados para, por exemplo, privilegiar: a encontrabilidade dos elementos ao longo do tempo na Internet; a navegabilidade entre os elementos distribuídos na *Web*; ou a exploração informacional do relacionamento entre os recursos disponibilizados. Uma das formas de cooperar a essa nova realidade é o atendimento

às Melhores Práticas (MPs) recomendadas pelo Consórcio W3C (*World Wide Web Consortium*) na publicação e no uso de Dados Abertos Conectados.

Considerando o domínio da Ciência da Informação, pressupõe-se que suas disciplinas também podem se beneficiar da *Web* de Dados. Especificamente, pontua-se que os Dados Abertos Conectados subsidiam as pesquisas bibliométricas ou cientométricas, minimizando os esforços na coleta e na transformação dos dados primários. Entretanto, alguns cuidados em relação à representação dos dados na *Web* devem ser tomados, de modo que os dados possam ser preservados e curados, privilegiando a replicação de resultados ao longo do tempo.

Neste sentido, objetiva-se representar os índices Qualis, *SCImago Journal & Country Rank* (SJR) e *Source Normalized Impact per Paper* (SNIP) como Dados Abertos Conectados em consonância às MPs recomendadas pela W3C, maximizando as possibilidades de integração e de exploração desses dados primários em contextos ampliados.

Para tanto, o presente trabalho relata a experiência de uma pesquisa aplicada para representar os recursos anteriormente citados na *Web* de Dados. Além da presente seção, este artigo compreende: (i) a fundamentação teórica que discute brevemente os Dados Abertos Conectados, sua classificação e apresenta as MPs consideradas na pesquisa; (ii) os materiais e métodos utilizados na representação dos dados na *Web* de Dados; (iii) de acordo com um subconjunto de MPs, o relato dos esforços despendidos para formalizar o modelo de representação dos Dados Abertos Conectados para os índices Qualis, SJR e SNIP; (iv) a apresentação dos estudos de caso utilizados como forma de verificação da representação formalizada; (v) a discussão sobre os resultados alcançados; e (vi) as considerações finais e os trabalhos futuros.

2 FUNDAMENTAÇÃO TEÓRICA

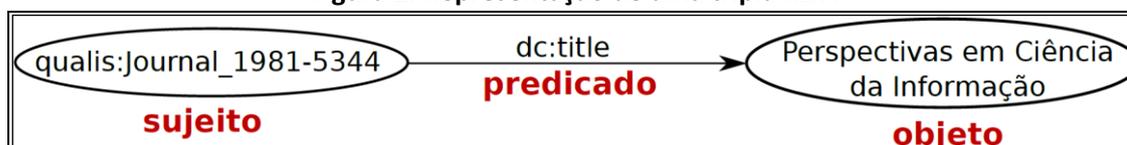
A base constitutiva deste trabalho é alinhada ao que se entende por Dados Abertos Conectados, os benefícios e as MPs para a publicação e o uso desse tipo de recurso na *Web* de Dados. Os Dados Abertos Conectados são aqueles publicados de acordo com licenças abertas, possibilitando que sejam reutilizados sem restrições, por pessoas ou aplicações e em diversos contextos. Seu entendimento é vinculado a duas indagações: (i) o que são dados abertos?; e (ii) como os dados são conectados?

Os dados são considerados abertos quando podem ser livremente (re)utilizados e (re)distribuídos, com as exigências de atribuição da fonte e de compartilhamento pelas regras de sua fonte original (OPEN KNOWLEDGE INTERNATIONAL, 2017).

Por sua vez, os dados conectados referem-se a um conjunto de práticas para publicar e interligar dados estruturados na *Web* (BIZER; HEATH; BERNES-LEE, 2009; HEATH; BIZER, 2011). Porém, isto somente é possível pela utilização de um modelo padrão para relacionar dados de diversas origens. Sob este prisma, tem-se o *Framework* de Descrição de Recursos (*Resource Description Framework* - RDF), uma linguagem padrão para conectar dados na *Web*. Em poucas palavras, o RDF descreve os recursos em três partes (sujeito – predicado – objeto), relacionando um sujeito a um objeto por meio de um predicado.

Como exemplo, a Figura 1 evidencia uma tripla RDF no domínio do índice Qualis. Neste exemplo, um sujeito identificado por “qualis:Journal_1981-5344” tem um predicado (dc:title do Vocabulário Dublin Core) que aponta ao nome “Perspectivas em Ciência da Informação”. Ressalta-se que um conjunto de triplas RDF restrito a um assunto gera uma grande estrutura de ligações, formando um grafo RDF. Neste sentido, o exemplo expresso na Figura 1 faz parte de um grafo, denominado QualisBrasil, o qual é discutido na Seção 4.

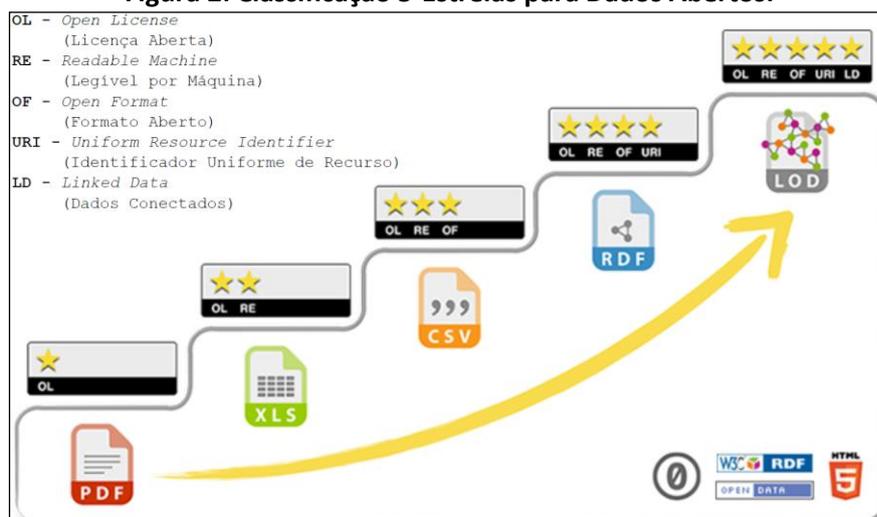
Figura 1: Representação de uma tripla RDF.



Fonte: Elaborada pelos autores - 2017.

Em relação aos dados abertos, estes são classificados conforme seu nível de acesso e de ligação a outros dados. Essa classificação é representada na Figura 2, a qual é denominada 5-Estrelas (*5-Star*). Proposta por Tim Bernes-Lee, a referida classificação é organizada incrementalmente como segue (5-STAR, 2017):

Figura 2: Classificação 5-Estrelas para Dados Abertos.

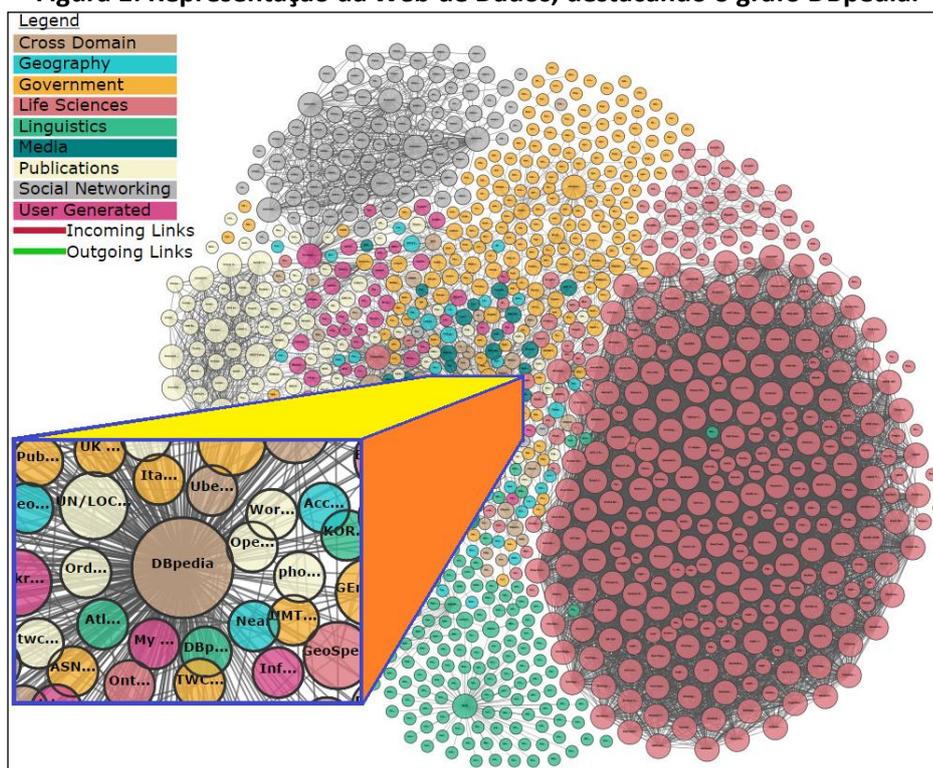


Fonte: Adaptada de (5-STAR, 2017).

- 1ª **Estrela** - é atribuída aos dados que são publicados sob uma licença aberta (*Open License* - OL), entretanto, em um formato proprietário. Os dados somente podem ser manipulados (lidos, visualizados ou impressos) por determinados *softwares*.
- 2ª **Estrela** - é conferida à publicação de dados estruturados legíveis por máquinas (*Readable Machine* - RE). Os dados são processados por *softwares* proprietários e podem ser exportados em outros formatos.
- 3ª **Estrela** - é concedida aos dados que são publicados em formato aberto (*Open Format* - OF). A manipulação dos dados não necessita de um *software* proprietário.
- 4ª **Estrela** - é designada à utilização de Identificador Uniforme de Recursos (*Uniform Resource Identifier* - URI) para rotular os dados, permitindo que outros usuários criem ligações e façam reuso de dados disponibilizados em ambientes distribuídos.
- 5ª **Estrela** - é atribuída aos dados conectados a outros dados (*Linked Data* - LD). Isso permite a navegação entre os dados e a descoberta de informação. Dessa forma, acrescenta-se valor aos dados, ao facilitar uma contextualização mais ampliada.

Considerando a classificação anterior, a união de dados abertos e de dados conectados é estabelecida na 5ª Estrela. Ressalta-se que essa combinação constitui a base informacional do imenso grafo RDF denominado *Web* de Dados (Figura 1). Neste sentido, a publicação e o uso de Dados Abertos Conectados têm como objetivo usar a arquitetura da *Web* para compartilhar dados em escala global (BIZER; HEATH; BERNES-LEE, 2009).

Figura 1: Representação da Web de Dados, destacando o grafo DBpedia.



Fonte: Adaptada de (ABELE *et al.*, 2017).

Para publicar e usar os Dados Abertos Conectados, o Consórcio W3C (2017) sugere um conjunto de MPs. Num ambiente colaborativo, onde os publicadores e os consumidores de dados compartilham e usam Dados Abertos Conectados, a incorporação de tais práticas repercute nos seguintes benefícios:

1. **Reuso** – é o aumento das chances de reutilização dos dados por diferentes grupos de consumidores.
2. **Compreensão** – propicia aos seres humanos um melhor entendimento sobre a natureza, a estrutura, o significado e os metadados dos dados disponibilizados.
3. **Interligação** – possibilita a criação dos relacionamentos entre recursos de dados, desde os conjuntos de dados até os itens de dados nesses conjuntos.
4. **Descoberta** – habilita os computadores a descobrirem automaticamente um conjunto de dados e os dados aninhados neste conjunto.
5. **Confiança** – para os consumidores de dados, certifica que o conjunto de dados é curado ao longo do tempo.
6. **Acesso** – facilita o acesso aos dados atualizados em diversas formas, tanto para os usuários como para os computadores.

7. **Interoperabilidade** – permite que os dados publicados sejam consumidos por diferentes sistemas, nos mais variados formatos.

8. **Processabilidade** – habilita os computadores a processar e a manipular automaticamente os dados contidos na *Web* de Dados.

Para tanto, o Consórcio W3C descreve 35 MPs (W3C, 2017). O documento que descreve tais práticas é atualizado constantemente, sendo que sua versão mais recente é acessada em <<https://www.w3.org/TR/dwbp/>>. Considerando que o objetivo deste trabalho é a discussão de um modelo de representação para alguns conjuntos de índices cientométricos na *Web* de Dados, das 35 MPs, destaca-se que um subconjunto de MPs tem maior aderência a esse propósito. Este subconjunto é formado por quatro MPs, sendo elas:

- **Forneça metadados estruturais (MP03)** – para capacitar os seres humanos na interpretação do esquema de representação dos dados. Para que os agentes de *software* possam processar automaticamente as distribuições do conjunto de dados.
- **Use URIs persistentes como identificadores de conjuntos de dados (MP09)** – para permitir a descoberta e a citação de conjuntos de dados ao longo do tempo, considerando a disponibilidade ou o formato em que os dados são apresentados.
- **Use URIs persistentes como identificadores dentro de conjuntos de dados (MP10)** – para permitir o mapeamento de itens de dados na *Web*, criando um espaço global de informações acessíveis por seres humanos e computadores.
- **Reuse vocabulários, dando preferência aos padronizados (MP15)** – para melhorar a interoperabilidade e o consenso entre os publicadores e os consumidores de dados.

Salienta-se que, ao publicar os índices cientométricos Qualis, SNIP e SJR, outras MPs também são atendidas. Porém, estas MPs não abordadas neste artigo, por tratarem outros requisitos diferentes da representação de recursos na *Web* de Dados.

3 MATERIAIS E MÉTODOS

Nesta seção, são descritos: (i) o processo de aquisição das bases de dados primários utilizadas no desenvolvimento do trabalho; e (ii) o procedimento metodológico que guiou a representação dos índices cientométricos na *Web* de Dados.

3.1 Bases de Dados utilizadas

Neste trabalho são considerados três conjuntos de dados abertos: 1) o histórico do índice Qualis; 2) o histórico do índice SJR; e 3) o histórico do índice SNIP.

3.1.1 Qualis

Segundo WebQualis (2013), “Qualis é o conjunto de procedimentos utilizados pela CAPES para estratificação da qualidade da produção intelectual dos programas de pós-graduação”. O Qualis afere a qualidade de produções científicas a partir da análise da qualidade dos periódicos científicos. Sua classificação compreende oito estratos em ordem decrescente de valor: A1, A2, B1, B2, B3, B4, B5 e C.

O índice Qualis foi coletado ao longo dos últimos doze anos, por meio de pesquisas anteriores, a partir do Sistema WebQualis (WEBQUALIS, 2013) e da Plataforma Sucupira (SUCUPIRA, 2017). A preservação do índice Qualis como Dados Abertos Conectados é discutida em Rautenberg e Burda (2016) e Rautenberg *et al.* (2016).

3.1.2 SNIP

O índice SNIP é uma métrica que mede o impacto de citação contextual de uma comunicação científica, normalizando a distância interna das citações das comunicações de um periódico perante o universo das citações em uma área de conhecimento (JOURNAL METRICS, 2017). Em outras palavras, o SNIP é definido como a razão do impacto bruto de um jornal/revista por publicação e o potencial de citação nas áreas de conhecimento. Isto permite, por exemplo, a avaliação de uma revista em comparação com seus pares e fornece informações mais contextualizadas, dando uma melhor imagem do impacto em determinado domínio.

O SNIP foi coletado nos anos 2015 e 2017. A partir do Portal *Journal Metrics*¹, os dados primários são extraídos em formato XLS², com o período de referência de 2005 a 2015.

3.1.3 SJR

O *Journal SCImago & Country Rank*³ é um portal que disponibiliza informações cientométricas a partir de dados contidos na base de dados Scopus. Dentre as informações, está o índice SJR, o qual pode ser utilizado para avaliar a qualidade e a reputação de periódicos científicos (JOURNAL METRICS, 2017). Este índice foi coletado a partir do Portal *Journal*

¹ URL: <<https://www.journalmetrics.com/>>.

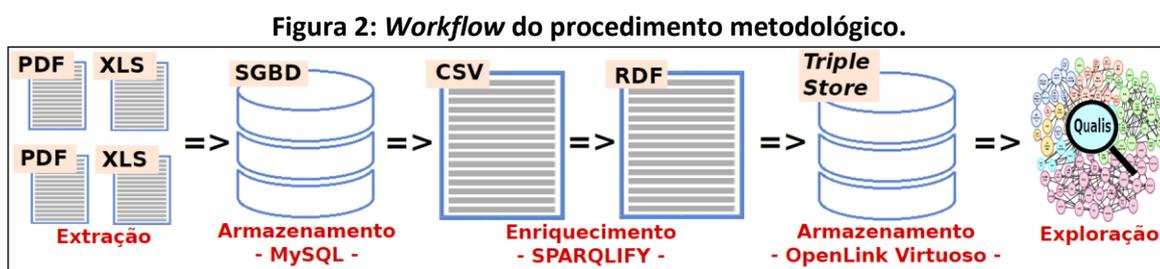
² XLS - eXcel Spreadsheet (formato de planilha eletrônica da Microsoft).

³ URL: <<http://www.scimagojr.com/journalrank.php>>.

SCImago & Country Rank, em formato XLS, também, com o período de referência de 2005 a 2015.

3.2 Procedimento Metodológico e Ferramental Tecnológico

Originalmente, os índices Qualis, SNIP e SJR são compartilhados na *Web* em formatos proprietários. Considerando a Classificação 5-Estrelas, destaca-se que: a) os dados do índice Qualis capturados do Sistema WebQualis estavam na 1ª Estrela, no formato PDF⁴; b) os dados dos índices SNIP e SJR são disponibilizados conforme a 2ª Estrela, no formato XLS; e c) a partir da Plataforma Sucupira, o índice Qualis é consumido no formato XLS. Para compatibilizar os formatos e elevar os dados à 5ª Estrela, cinco atividades baseadas no ciclo de vida *Linked Data Lifecycle* (AUER, 2014) fazem parte do procedimento metodológico, conforme a Figura 4:



Fonte: Elaborada pelos dos autores - 2017.

- **Atividade 01 - Extração** – os arquivos em formato original foram convertidos para arquivos texto. Alguns *scripts* de pré-processamento foram desenvolvidos na linguagem de programação PHP⁵ para organizar e criticar os dados, facilitando o armazenamento dos dados em uma base de dados legada.
- **Atividade 02 - Armazenamento** – os dados são armazenados em um Sistema Gerenciador de Banco de Dados MySQL⁶ para serem usados em sistemas legados.
- **Atividade 03 – Enriquecimento** – os dados são extraídos de uma base de dados legada para arquivos no formato CSV (3ª Estrela) e mapeados para RDF com o auxílio da ferramenta Sparqlify⁷. Neste ponto, os dados alcançam a 4ª Estrela.

⁴ PDF - Portable Document Format - Formato Portátil de Documento.

⁵ É uma linguagem de uso geral, especialmente adequada para o desenvolvimento de aplicações *Web*. Disponível em: <<http://www.php.net/>>

⁶ É um Sistema Gerenciador de Banco de Dados relacional *open-source* que pode ser usado em aplicações para gerir bases de dados. Disponível em: <<https://www.mysql.com/>>.

⁷ É uma ferramenta *open-source* do Instituto *Agile Knowledge and Semantic Web* que enriquece os dados primários, convertendo os dados em triplas RDF. Disponível em: <<http://aksw.org/Projects/Sparqlify.html>>.

- **Atividade 04 - Armazenamento** – para possibilitar o (re)uso de recursos na *Web* de Dados, os dados em RDF são armazenados no servidor de dados Open Link Virtuoso⁸.
- **Atividade 05 – Exploração** - na *Web* de Dados, consulta-se os Dados Abertos Conectados objetivando a aquisição de informação contextualizada. Ao se relacionar recursos oriundos dos vários grafos RDF disponibilizados, alcança-se a 5ª Estrela.

Vale destacar que, conforme o objetivo traçado, a Atividade 3 – Enriquecimento – compreende os esforços despendidos na representação dos índices cientométricos como Dados Abertos Conectados. Para tanto, a seção a seguir é reservada à discussão pormenorizada das escolhas realizadas nesta atividade de acordo com as quatro MPs para publicação de dados na *Web* de Dados pontuadas anteriormente.

4 REPRESENTANDO OS ÍNDICES CIENTOMÉTRICOS COMO DADOS ABERTOS CONECTADOS

Para discutir a representação dos índices cientométricos na *Web* de Dados, esta seção é organizada em quatro subseções, conforme o atendimento às MPs apresentadas na Seção 2.

4.1 MP09 - Use URIs persistentes como identificadores de conjuntos de dados

A adoção de um sistema comum de identificação é um requisito essencial para o gerenciamento e o reuso de dados na *Web* de Dados. Os publicadores de dados devem criar os URIs dos conjuntos de dados e de domínios de modo que a persistência e a recuperação dos recursos na *Web* de Dados sejam facilitadas ao longo do tempo (W3C, 2017).

Atendendo à MP09, criou-se o domínio <<http://lod.unicentro.br/>> (Figura 5a). Neste domínio, tem-se acesso à interface do servidor Open Link Virtuoso, onde os grafos dos índices cientométricos podem ser dereferenciados⁹ conforme os prefixos (Figura 5b):

⁸ É considerado um sistema universal para acesso, integração e gerenciamento de dados baseados no modelo RDF. Disponível em: <<http://virtuoso.openlinksw.com/>>.

⁹ *Dereferencing* é o processo de procurar um URI na *web* para acessar as informações do recurso referenciado.

Figura 3: Representação dos identificadores de conjuntos de dados.



Fonte: Elaborada pelos autores – 2017.

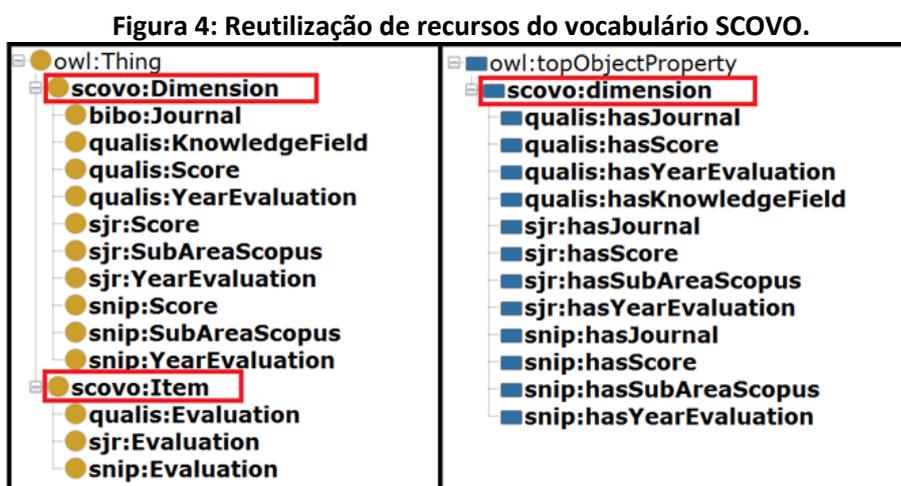
- **PREFIX qualis:**<<http://lod.unicentro.br/QualisBrasil/>> - representa o grafo dos recursos do índice Qualis.
- **PREFIX sjr:**<<http://lod.unicentro.br/SJR/>> - faz a referência ao grafo dos recursos do fator de impacto SJR.
- **PREFIX snip:**<<http://lod.unicentro.br/SNIP/>> - analogamente, aponta aos recursos do fator de impacto SNIP.

4.2 MP15 - Reuse vocabulários, dando preferência aos padronizados

Para melhorar a interoperabilidade entre sistemas, permitindo o consumo facilitado de recursos de dados, é aconselhável utilizar vocabulários e/ou ontologias já utilizados na *Web* de Dados (W3C, 2017). Neste sentido, alguns vocabulários/ontologias foram investigados no sítio *Linked Open Vocabularies* (LOV, 2017), sendo reutilizados:

- **SCOVO - The Statistical COre VOcabulary** - é um vocabulário simples para representar dados estatísticos na *Web* (SCOVO, 2017). Neste trabalho, é usado para organizar os índices Qualis, SJR e SNIP de forma multidimensional. Os elementos utilizados são:
 - **scovo:Item** – classe que representa um item de dado estatístico. Na representação dos índices cientométricos, também representa os sujeitos qualis:Evaluation, snip:Evaluation e sjr:Evaluation.
 - **scovo:Dimension** – classe que representa uma dimensão associada a um scovo:Item. Para cada prefixo, quatro scovo:Dimension são representadas:
 - **qualis** – bibo:Journal, qualis:KnowledgeField, qualis:Score e qualis:YearEvaluation;
 - **sjr** - bibo:Journal, sjr:Score, sjr:SubAreaScopus e sjr:YearEvaluation;

- **snip** - bibo:Journal, snip:Score, snip:SubAreaScopus e snip:YearEvaluation.
- **scovo:dimension** – relação que mapeia as dimensões scovo:Dimension de um scovo:Item. Para cada prefixo, quatro scovo:dimension são representadas:
 - **qualis** – qualis:hasJournal, qualis:hasKnowledgeField, qualis:hasScore e qualis:hasYearEvaluation;
 - **sjr** – sjr:hasJournal, sjr:hasScore, sjr:hasSubAreaScopus e sjr:hasYearEvaluation; e
 - **snip** - snip:hasJournal, snip:hasScore, snip:hasSubAreaScopus e snip:hasYearEvaluation.
- **BIBO - Bibliographic Ontology Specification** – é uma ontologia que modela os conceitos e as propriedades de referências bibliográficas (BIBO, 2017). Desta ontologia, são reutilizadas a classe bibo:Journal e as propriedades bibo:issn e foaf:name. Ressalta-se que esta classe representa uma scovo:Dimension comum em todos os índices cientométricos.



Fonte: Elaborada pelos autores - 2017.

Estruturalmente, a Figura 6 apresenta a base constitutiva de reutilização do vocabulário SCOVO frente a modelagem dos recursos e propriedades dos índices cientométricos. Quanto às propriedades de dados, salienta-se que também são utilizadas as propriedades dc:identifier e dc:title do vocabulário Dublin Core (DCMI, 2017) e rdf:value do framework RDF.

4.3 MP03 - Forneça metadados estruturais

Disponibilizar informação sobre a modelagem dos grafos é essencial para que os consumidores de dados consigam explorar os dados compartilhados na *Web* de Dados. Compartilhar os metadados estruturais também auxilia o entendimento do significado dos dados. Segundo W3C (2017), ao fornecer os metadados estruturais, os seres humanos podem melhor interpretar o esquema dos dados e os agentes de *software* conseguem processar automaticamente os recursos disponibilizados. Neste sentido, o Quadro 1 apresenta os metadados e suas definições organizados segundo cada índice cientométrico compartilhado.

Quadro 1: Descrição dos recursos que representam índices cientométricos na *Web* de Dados.

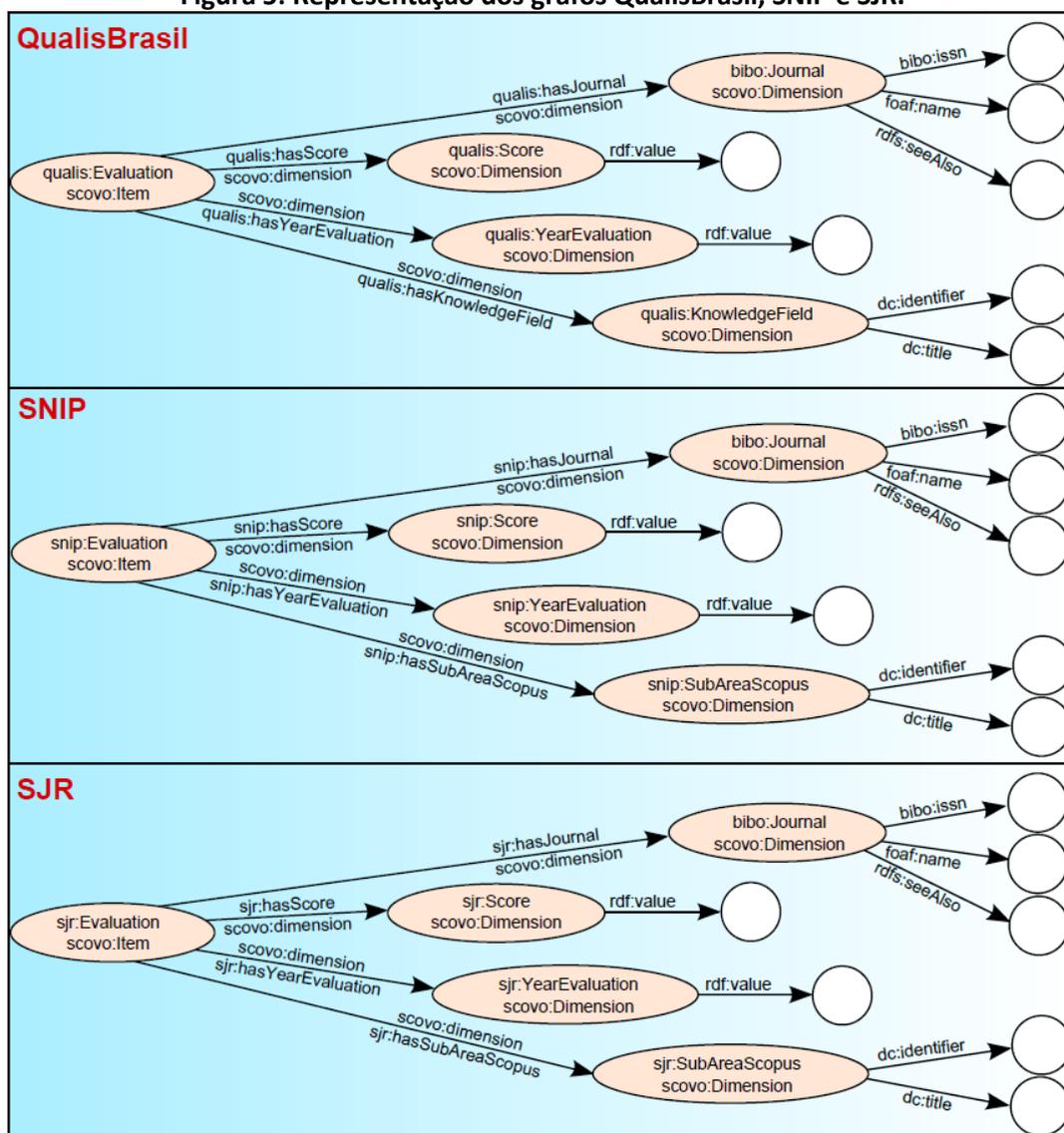
Grafo	Tipo Recurso	Definição
QualisBrasil	Classe	qualis:Evaluation – classe que representa uma avaliação Qualis, relacionando uma revista, a área de conhecimento, o ano e o estrato alcançado.
		qualis:KnowledgeField – classe que representa uma área de avaliação, relacionando o código de identificação da área e o nome, segundo o Conselho Nacional de Pesquisa - CNPq.
		bibo:Journal – classe reutilizada da <i>Bibliografic Ontology Specification</i> .
		qualis:Score – classe que representa os estratos Qualis.
		qualis:YearEvaluation – classe que representa os anos de avaliação do índice Qualis.
	Relação	qualis:hasKnowledgeField – relação funcional que determina a área de conhecimento de uma avaliação.
		qualis:hasJournal – relação funcional que determina a qual jornal/revista a avaliação é relacionada.
		qualis:hasScore – relação funcional que determina qual é o estrato Qualis de uma avaliação.
		qualis:hasYearEvaluation – relação funcional que determina qual é o ano associado a uma avaliação.
	Propriedade	bibo:issn – propriedade de dados da <i>Bibliografic Ontology Specification</i> .
		foaf:name – propriedade de dados da <i>Bibliografic Ontology Specification</i> .
		dc:identifier – propriedade de dados utilizada para identificar a área de conhecimento segundo o CNPq.
		dc:title – propriedade de dados utilizada para nomear a área de conhecimento segundo o CNPq.
		rdf:value – propriedade de dados para atribuir valores para o estrato Qualis e o ano de avaliação.
		rdfs:seeAlso – propriedade de dados para vincular recursos da <i>Web</i> de Dados aos jornais/revistas.
SNIP	Classe	snip:Evaluation – classe que representa uma avaliação SNIP, relacionando uma revista, a área de conhecimento, o ano e o estrato alcançado.
		snip:SubAreaScopus – classe que representa uma área de avaliação, relacionando o código de identificação da área e o nome, segundo as subáreas definidas pelo Scopus.
		bibo:Journal – classe da <i>Bibliografic Ontology Specification</i> .
		snip:Score – classe que representa os índices de impacto das avaliações.
		snip:YearEvaluation – classe que representa os anos de avaliação do índice SNIP.
	Relação	snip:hasSubAreaScopus – relação funcional que determina a área de conhecimento de uma avaliação.
		snip:hasJournal – relação funcional que determina a que jornal/revista a avaliação é relacionada.

XVIII ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2017
23 a 27 de outubro de 2017 – Marília – SP

		snip:hasScore – relação funcional que determina qual é o fator de impacto de uma avaliação.
		snip:hasYearEvaluation – relação funcional que determina qual é o ano associado a uma avaliação.
	Propriedade	As mesmas propriedades do grafo QualisBrasil, atreladas às classes análogas.
SJR	Classe	Dada a verossimilhança estrutural, têm o mesmo conjunto de classes do grafo SNIP.
	Relação	Dada a verossimilhança estrutural, têm o mesmo conjunto de relações do grafo SNIP.
	Propriedade	Conforme o grafo SNIP, as mesmas propriedades do grafo QualisBrasil, atreladas às classes análogas.

Fonte: Elaborado pelos autores – 2017.

Figura 5: Representação dos grafos QualisBrasil, SNIP e SJR.



Fonte: Elaborada pelos dos autores - 2017.

Para facilitar o entendimento, a Figura 7 representa graficamente a organização dos metadados estruturais discutidos no Quadro 1.

4.4 MP10 - Use URIs persistentes como identificadores dentro de conjuntos de dados

O potencial da *Web* está no efeito de complementariedade da informação que a rede proporciona. Os dados se tornam mais valiosos ao serem relacionados com outros dados, incrementando o conhecimento sobre os dados (por exemplo, vincular lugares e pessoas aos eventos históricos). Para tornar os vínculos entre os dados perenes, seus identificadores devem ser consistentes, preservando os requisitos de encontrabilidade. Deste modo, as regras de nomeação de URIs dos dados devem ser consistentes, mantendo a inter-relação dos dados representados ao longo do tempo. Para W3C (2017), isso permite que os itens de dados sejam relacionados na *Web*, em um espaço global de informação acessível por humanos e máquinas.

Para preservar a navegabilidade entre os itens de dados cientométricos, neste trabalho, a nomeação de URIs dos dados segue duas regras generalizadas:

1. **<classe_do_recurso>_<id_recurso>** - para URIs que representam um recurso isolado. Deve-se evidenciar: (i) a classe à qual o recurso é categorizado para diferenciar de outros tipos em um grafo; (ii) o atributo que atua como identificador interno para itens de dados de mesma categoria. Concatenando tais itens, computacionalmente, esta regra preserva a identificação única de um recurso em um grafo. E, para o ser humano, evidencia um entendimento literal sobre qual recurso é discutido. Como exemplo, dado o item de dado “qualis:Journal_0020-0255” é perceptível que este se refere a um periódico, especificamente, aquele identificado pelo ISSN “0020-0225”.
2. **(<URI>_)*<URI>** - para URIs que representam mais de um recurso, deve-se concatenar todos URIs dos recursos isolados. Como um exemplo, dado o item “QualisEvaluation_Journal_0020-0255_KnowledgeField_31_YearEvaluation_2016_Score_A1”, literalmente, identifica-se que o item refere-se uma avaliação Qualis, do periódico cujo ISSN é “0020-0255”. Essa avaliação considera a área de conhecimento identificada pelo CNPq com o código “31”. No ano de 2016, o periódico em questão tem como estrato Qualis associado o valor “A1”.

Considerando as generalizações apresentadas, o Quadro 2 lista as regras de nomeação de URIs dos dados para cada um dos índices cientométricos considerados neste trabalho.

XVIII ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2017
23 a 27 de outubro de 2017 – Marília – SP

Quadro 2: Regras de nomeação das URIs dos recursos de índices cientométricos e seus exemplos.

Grafo	Regra	Prefixo do grafo + Exemplo
QualisBrasil	qualis:Score - 'Score_' + estrato	qualis:Score_A1
	qualis:YearEvaluation - 'YearEvaluation_' + ano	qualis:YearEvaluation_2016
	qualis:KnowledgeField - 'KnowledgeField_' + idCNPq	qualis:KnowledgeField_31
	bibo:Journal - 'Journal_' + issn	qualis:Journal_0020-0255
	qualis:Evaluation - 'QualisEvaluation_' + bibo:Journal + '_' + qualis:KnowledgeField + '_' + qualis:YearEvaluation + '_' + qualis:Score	qualis:QualisEvaluation_Journal_0020-0255_KnowledgeField_31_YearEvaluation_2016_Score_A1
SNIP	snip:Score - 'Score_' + (score * 1000)	snip:Score_2489
	snip:YearEvaluation - idem a regra do grafo QualisBrasil	snip:YearEvaluation_2015
	snip:SubAreaScopus - 'SubAreaScopus_' + idSubareaScopus	snip:SubAreaScopus_1802
	bibo:Journal - idem a regra do grafo QualisBrasil	snip:Journal_0020-0255
	snip:Evaluation - 'SNIPEvaluation_' + bibo:Journal + '_' + snip:SubAreaScopus + '_' + snip:YearEvaluation + '_' + snip:Score	snip:SNIPEvaluation_Journal_0020-0255_SubAreaScopus_1802_YearEvaluation_2015_Score_2489
SJR	sjr:Score - idem a regra do grafo SNIP	sjr:Score_2513
	sjr:YearEvaluation - idem a regra do grafo SNIP	sjr:Year_2015
	sjr:SubArea - idem a regra do grafo SNIP	sjr:SubAreaScopus_1802
	bibo:Journal - idem a regra do grafo SNIP	sjr:Journal_0020-0255
	sjr:Evaluation - 'SJREvaluation_' + bibo:Journal + '_' + sjr:SubAreaScopus + '_' + sjr:YearEvaluation + '_' + sjr:Score	sjr:SJREvaluation_Journal_0020-0255_SubAreaScopus_1802_Year_2015_Score_2513

Fonte: Elaborada pelos autores - 2017.

Ressalta-se que as quatro MPs discutidas nesta seção garantem que os índices Qualis, SJR e SNIP sejam representados conforme a 4ª Estrela na classificação 5-Estrelas. Para alcançar a 5ª Estrela, os contextos de utilização dos recursos disponibilizados devem ser implementados. Neste sentido, a próxima seção exemplifica dois cenários em que os grafos são explorados simultaneamente, ampliando a capacidade informacional dos índices compartilhados.

5 VERIFICANDO A REPRESENTAÇÃO PERANTE ALGUNS CENÁRIOS

Nesta seção verifica-se a representação dos índices cientométricos perante as MPs reportadas na seção anterior. Para tanto, dois estudos de caso exploram a integração dos grafos disponibilizados. Dessa forma, acrescenta-se valor aos recursos em contextos mais ampliados.

**XVIII ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2017
23 a 27 de outubro de 2017 – Marília – SP**

Listagem 1: Consulta SPARQL para relacionar as áreas de conhecimento do CNPq com as áreas de conhecimento do Scopus usando o ISSN dos periódicos como o item de relacionamento.

```

01 PREFIX qualis: <http://lod.unicentro.br/QualisBrasil/>
02 PREFIX sjr: <http://lod.unicentro.br/SJR/>
03 PREFIX dc: <http://purl.org/dc/elements/1.1/>
04 PREFIX foaf: <http://xmlns.com/foaf/0.1/>
05 PREFIX bibo: <http://purl.org/ontology/bibo/>
06
07 SELECT DISTINCT ?issn ?nameJournal ?nameAreaQualis ?nameAreaScopus WHERE {
08   ?EvaluationQualis qualis:hasJournal ?JournalQualis .
09   ?JournalQualis bibo:issn ?issn .
10   ?JournalQualis foaf:name ?nameJournal .
11   ?EvaluationQualis qualis:hasKnowledgeField ?KnowledgeField .
12   ?KnowledgeField dc:title ?nameAreaQualis .
13
14   ?EvaluationSJR sjr:hasJournal ?JournalSJR .
15   ?JournalSJR bibo:issn ?issn .
16   ?EvaluationSJR sjr:hasSubAreaScopus ?SubAreaScopus .
17   ?SubAreaScopus dc:title ?nameAreaScopus .
18   FILTER (?issn = "1588-2861")
19 }

```

Fonte: Elaborada pelos autores – 2017.

No primeiro estudo de caso, uma consulta SPARQL (Listagem 1) relaciona os grafos <http://lod.unicentro.br/QualisBrasil/> e <http://lod.unicentro.br/SJR/> ao mapear o relacionamento entre as áreas de conhecimento do CNPq e do Scopus, considerando um periódico. A Tabela 1 enumera alguns recursos recuperados ao processar a referida consulta.

Tabela 1: Resultado parcial do processamento da consulta da Listagem 1.

?issn	?nameJournal	?nameAreaQualis	?nameAreaScopus
1588-2861	Scientometrics	Ciência da Computação	Library and Information Sciences
1588-2861	Scientometrics	Ciência da Computação	General Social Sciences
1588-2861	Scientometrics	Comunicação e Informação	Computer Science Applications
1588-2861	Scientometrics	Comunicação e Informação	Law
[...]	[...]	[...]	[...]

Fonte: Elaborada pelos autores - 2017.

No segundo estudo de caso, objetiva-se relacionar os índices cientométricos associados a um periódico. Neste sentido, a Listagem 2 recupera os fatores e áreas de conhecimento para os índices Qualis, SJR e SNIP do periódico “Information Sciences” no ano 2015. A Tabela 2 enumera o resultado parcial ao processar a consulta da Listagem 2.

Listagem 1: Exemplo de consulta SPARQL que relaciona o periódico *Information Sciences* e seus índices cientométricos no ano 2015.

```

01 PREFIX qualis: <http://lod.unicentro.br/QualisBrasil/>
02 PREFIX sjr: <http://lod.unicentro.br/SJR/>
03 PREFIX snip: <http://lod.unicentro.br/SNIP/>
04 PREFIX dc: <http://purl.org/dc/elements/1.1/>
05 PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
06 PREFIX foaf: <http://xmlns.com/foaf/0.1/>
07 PREFIX bibo: <http://purl.org/ontology/bibo/>
08
09 SELECT ?issn ?name ?year ?nameAreaCNPq ?qualis ?nameAreaScopus ?snip ?sjr WHERE {
10   ?EvaluationQualis qualis:hasJournal ?JournalQualis .

```

XVIII ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2017
23 a 27 de outubro de 2017 – Marília – SP

11	?EvaluationQualis	qualis:hasKnowledgeField	?KnowledgeField .
12	?EvaluationQualis	qualis:hasScore	?ScoreQualis .
13	?EvaluationQualis	qualis:hasYearEvaluation	?YearEvaluationQualis.
14	?JournalQualis	bibo:issn	?issn .
15	?JournalQualis	foaf:name	?name .
16	?KnowledgeField	dc:title	?nameAreaCNPq .
17	?ScoreQualis	rdf:value	?qualis.
18	?YearEvaluationQualis	rdf:value	?year .
19			
20	?EvaluationSJR	sjr:hasJournal	?JournalSJR .
21	?EvaluationSJR	sjr:hasScore	?ScoreSJR .
22	?EvaluationSJR	sjr:hasYearEvaluation	?YearEvaluationSJR.
23	?JournalSJR	bibo:issn	?issn .
24	?YearEvaluationSJR	rdf:value	?year .
25	?ScoreSJR	rdf:value	?sjr.
26			
27	?EvaluationSNIP	snip:hasJournal	?JournalSNIP .
28	?EvaluationSNIP	snip:hasScore	?ScoreSNIP .
29	?EvaluationSNIP	snip:hasYearEvaluation	?YearEvaluationSNIP.
30	?EvaluationSNIP	snip:hasSubAreaScopus	?SubAreaScopus .
31	?SubAreaScopus	dc:title	?nameAreaScopus .
32	?JournalSNIP	bibo:issn	?issn .
33	?YearEvaluationSNIP	rdf:value	?year .
34	?ScoreSNIP	rdf:value	?snip.
35	FILTER (?year = "2015" && ?issn = "0020-0255")		
36	}		

Fonte: Elaborada pelos autores - 2017.

Tabela 2: Resultado parcial do processamento da consulta da Listagem 2.

?issn	?name	?year	?nameAreaCNPq	?qualis	?nameArea-Scopus	?snip	?sjr
0020-0255	Information Sciences	2015	Ciência da Computação	A1	Control and Systems Engineering	2.489	2.513
0020-0255	Information Sciences	2015	Ciências Biológicas I	A2	Artificial Intelligence	2.489	2.513
[...]	[...]	[...]	[...]	[...]	[...]	[...]	[...]

Fonte: Elaborada pelos autores – 2017.

Com a execução das consultas desenvolvidas nos estudos de caso, é possível verificar a capacidade representacional dos grafos modelados e a possibilidade do reuso dos recursos de dados cientométricos na *Web* de Dados em contextos informacionais ampliados.

6 RESULTADOS E DISCUSSÃO

Diante dos estudos de caso relatados anteriormente, encontra-se suporte na discussão de alguns resultados. Argumenta-se que as MP03, MP09, MP10 e MP15 impactam positivamente nas atividades de enriquecimento, armazenamento e exploração dos recursos cientométricos como Dados Abertos Conectados na *Web* de Dados.

Com a MP03, as definições essenciais dos recursos de dados são estabelecidas, contribuindo à organização dos dados cientométricos como Dados Abertos Conectados.

A MP09 estabelece o domínio e os URIs <<http://lod.unicentro.br/SJR/>>, <<http://lod.unicentro.br/QualisBrasil/>> e <<http://lod.unicentro.br/SNIP/>>. Com essas

definições, os conjuntos podem ser dereferenciados na *Web*, possibilitando a obtenção de recursos cientométricos pelos agentes (humanos ou de *software*) nos mais variados contextos.

Ao seguir a MP10, garante-se que um recurso de dados cientométrico tenha sua identificação única estabelecida, com um URI permanente ao longo do tempo na *Web* de Dados.

Ao atender a MP15, reusando os vocabulários estabelecidos, tende-se a facilitar a interoperabilidade entre sistemas e o entendimento dos dados pelos humanos, visto que se padroniza a forma de acesso aos recursos de dados.

Diante a conformidade às MPs, admite-se que os usuários são beneficiados quanto ao reuso, a interligação, a compreensão e a confiança dos recursos de dados compartilhados, tornando possível a interoperabilidade e a processabilidade dos índices Qualis, SJR e SNIP.

7 CONSIDERAÇÕES FINAIS

Este artigo relata a experiência em representar dados cientométricos conforme algumas MPs recomendadas pelo W3C para a publicação e o uso de dados na *Web* de Dados. Pontualmente, são consideradas as MP03, MP09, MP10 e MP15 na publicação dos índices cientométricos Qualis, SJR e SNIP como Dados Abertos Conectados. Advoga-se que ao seguir as referidas MPs, contribui-se para o reuso, a compreensão, a interligação e a descoberta de recursos cientométricos em escala global. Desta maneira, os esforços despendidos colaboram na interoperabilidade e na consequente processabilidade de dados em sistemas computacionais para aquisição de informação cientométrica mais contextualizada.

Por isso, considerando o domínio da Ciência da Informação, a contribuição deste trabalho reside, inicialmente, no estabelecimento de um modelo de representação útil no compartilhamento de dados primários para demais pesquisas bibliométricas ou cientométricas. Futuras pesquisas podem se beneficiar deste modelo para: (i) minimizar os esforços na compreensão e na compatibilização dos formatos de representação de dados primários; e (ii) maximizar as possibilidades de integração e de exploração de dados cientométricos em contextos ampliados, onde os índices Qualis, SJR e SNIP são necessários. Cientificamente, o trabalho também contribui à discussão do uso interdisciplinar de Dados Abertos Conectados como arcabouço tecnológico na Ciência da Informação. Ou seja, tem-se um exemplo profícuo no que tange as novas demandas de preservação e de curadoria de

**XVIII ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2017
23 a 27 de outubro de 2017 – Marília – SP**

dados em ambientes digitais baseados na *Web* de Dados. Diante dessas considerações, como trabalhos futuros são traçados:

- atender as demais MPs recomendadas pelo W3C na publicação e uso de recursos cientométricos dos índices Qualis, SJR e SNIP na *Web* de Dados;
- atualizar os grafos <<http://lod.unicentro.br/SJR/>>, <<http://lod.unicentro.br/SNIP/>> e <<http://lod.unicentro.br/QualisBrasil/>> ao longo do tempo, atuando na preservação digital de seus recursos; e
- modelar os mecanismos de gestão de dados cientométricos no escopo das universidades brasileiras, mediante um “Modelo Tecnológico ao Compartilhamento de Dados para Estudos Cientométricos baseado em *Linked Open Data*”.

AGRADECIMENTOS

O autor principal agradece à Fundação Araucária pelo suporte financeiro (Projeto nº 601/2014 - Modelo para Compartilhamento de Informações sobre Pesquisas baseado em *Linked Open Data* para Estudos Cientométricos).

REFERÊNCIAS

5-STAR. **5-Star OPEN DATA**. Disponível em: <<http://5stardata.info/en>>. Acesso em: 16 de Abril de 2016 09:00.

ABELE, A.; MCCRAE, J. P.; BUITELAAR, P. JENTZSCH, A. CYGANIAK, R. **Linking Open Data cloud diagram 2017**. Disponível em: <<http://lod-cloud.net>>. Acesso em: 13 de abr de 2017.

AUER, S. Introduction to lod2. In AUER, S.; BRYL, V.; TRAMP, C (ed). **Linked Open Data – Creating Knowledge Out of Interlinked Data**. Springer-Verlag, 2014. 215p.

BERNERS-LEE, T.; HENDLER, J.; LASSILA, O. The Semantic Web. **Scientific American Magazine**, v. 284, n. 5, p. 34-43, 2001.

BIBO. **Bibliographic Ontology Specification | The Bibliographic Ontology**. Disponível em: <<http://bibliontology.com/>>. Acesso em: 16 de Abril de 2017.

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.

BUGNION, P.; MANIVANNAN, A.; NICOLAS, P. R. **Scala: Guide for Data Science Professionals**. Birmingham: Packt Publishing, 2017. 1077p.

DCMI. **DCMI Metadata Terms**. Disponível em: <<http://dublincore.org/documents/dcmi-terms/>>. Acesso em: 27 de Julho de 2017 14:00.

HEATH, T.; BIZER, C. **Linked Data Evolving the Web into a Global Data Space**. Londres: Morgan & Claypool, 2011. 136p.

XVIII ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2017
23 a 27 de outubro de 2017 – Marília – SP

JOURNAL METRICS. **Journal Metrics - Scopus.com**. Disponível em:
<<https://www.journalmetrics.com/>>. Acesso em: 16 de Abril de 2017.

LOV. **Linked Open Vocabularies**. Disponível em: <<http://lov.okfn.org/dataset/lov>>. Acesso em:
16 de abr de 2017.

OPEN KNOWLEDGE INTERNATIONAL. O que são Dados Abertos? Disponível em:
<http://opendatahandbook.org/guide/pt_BR/what-is-open-data/>. Acesso em: 14 de Junho
de 2017 21:00.

RAUTENBERG, S.; BURDA, A. C. Linked Open Data para Cientometria: Compartilhando e
Mantendo o índice Qualis na *Web* de Dados In: ENCONTRO BRASILEIRO DE BIBLIOMETRIA E
CIENTOMETRIA, 5., 2016, São Paulo. **Anais...** São Paulo: USP, 2016. p. A34.

RAUTENBERG, S.; *et al.* Linked Data Workflow Project Ontology: uma Ontologia de Domínio
para Publicação e Preservação de Dados Conectados. **Tendências da Pesquisa Brasileira em
Ciência da Informação**, v. 9, p. 1-19, 2016.

SCOVO. **The Statistical Core Vocabulary (SCOVO) | DERI Vocabularies**. Disponível em:
<<http://vocab.deri.ie/scovo>>. Acesso em: 16 de Abril de 2017.

SUCUPIRA. **Plataforma Sucupira**. Disponível em:
<[https://sucupira.capes.gov.br/sucupira/public/consultas/coleta/veiculoPublicacaoQualis/lista
ConsultaGeralPeriodicos.jsf](https://sucupira.capes.gov.br/sucupira/public/consultas/coleta/veiculoPublicacaoQualis/listaConsultaGeralPeriodicos.jsf)>. Acesso em: 03 de Abril de 2017 21:00.

W3C. **Data on the Web Best Practices - W3C Recommendation 31 January 2017**. Disponível
em <<https://www.w3.org/TR/2017/REC-dwbp-20170131/>>. Acesso em 22 mar 2017 22:00.

WEBQUALIS. **Sistema WebQualis - Portal Capes**. Disponível em:
<<http://qualis.capes.gov.br/webqualis/principal.seam>>. Acesso em: 25 de Agosto de 2013
10:00.