

Apoiando estudos cientométricos com Linked Open Data

Sandro Rautenberg

**Professor do Depto de Ciência da
Computação/UNICENTRO. Doutor em Engenharia
e Gestão do Conhecimento - EGC/UFSC**

Edgard Marx

**Doutorando em Ciência da Computação
Institut für Angewandte Informatik - Universität
Leipzig (Alemanha)**

Antonio Costa Gomes Filho

**Professor do Depto de
Administração/UNICENTRO. Doutor em
Engenharia e Gestão do Conhecimento -
EGC/UFSC**

Sören Auer

**Professor do Instituto de Informática
Universidade de Bonn - Alemanha
Doutor em Ciência da Computação
Institut für Angewandte Informatik - Universität
Leipzig (Alemanha)**

<http://dx.doi.org/10.1590/1981-5344/2817>

Na Cientometria, mensurar indicadores é uma tarefa complexa devido aos desafios em coletar, organizar e relacionar dados, sobretudo, na web, onde os dados são distribuídos em várias fontes e formatos incompatíveis. Esses problemas podem ser resolvidos com o emprego de tecnologias e metodologias baseadas nos princípios Linked Open Data. Tais princípios são fundamentados num conjunto de melhores práticas de Web Semântica e Dados Abertos para organizar, publicar e conectar dados na web. Por meio destes, os dados são acessados e consumidos sem restrições, em diversas aplicações. No presente trabalho, relata-se a experiência na disponibilização do histórico do índice Qualis conforme o Linked Open Data. Pressupõe-se que tal empreendimento é importante nas atividades de coleta de dados primários em pesquisas bibliométricas/cientométricas, servindo para: mensurar a evolução dos periódicos científicos; auxiliar na aferição de medidas qualitativas e

quantitativas de publicações científicas; ou obter informações relevantes a partir do cruzamento com outros indicadores cientométricos. A disponibilização do índice Qualis é verificada em três estudos de caso. Como resultado, têm-se o compartilhamento do índice Qualis (série histórica 2005-2013) por uma interface web para: (i) facilitar o reuso e a integração dos dados; e (ii) suportar a interoperabilidade e a processabilidade computacional dos recursos disponibilizados.

Palavras-chave: *Linked Open Data; Cientometria; Índice Qualis; Estudos Cientométricos; Web Semântica.*

Supporting Scientometric Studies with Linked Open Data

In Scientometric Studies, measuring scientific indicators is a complex task due to the challenges associated with data collection, organization and linking, especially in the web, where data is distributed in various sources and incompatible formats. These problems can be tackled with the technological and methodological techniques based on the Linked Open Data principles. These principles cover a set of the best practices from the fields of Semantic Web and Open Data to organize, publish and interlink the data on the Web. With the use of those best practices, the data can be accessed and consumed without restrictions, in many applications. This paper addresses the availability of a Qualis historical dataset, according to the mentioned principles. In Scientometric studies, this effort is important for data reuse, taking into the account: measuring an evolution of scientific journals; assisting production of qualitative and quantitative measures of scientific publications; or obtaining relevant information by interlinking and exploring other scientific indicators. The availability of the Qualis dataset is verified by the three use cases. As a result, the Qualis index (historical series 2005-2013) is shared by a web interface for: (i) furthering the data reuse and integration; and (ii) supporting the interoperability and computational processability of the available resources.

Keywords: *Linked Open Data; Scientometrics; Qualis Index; Scientometric Studies; Semantic Web.*

Recebido em 25.05.2016 Aceito em 13.07.2017

1 Introdução

Este artigo tem como base a relação tecnológica de dois conceitos que são úteis para organizar, formalizar e compartilhar dados sobre a pesquisa científica: a Cientometria e o *Linked Open Data*.

Para este trabalho, o entendimento do conceito Cientometria é aderente à “disciplina que tem por objetivo medir as atividades de pesquisa científica e tecnológica mediante insumos (mão-de-obra, investimentos) e produtos (equipamentos, produtos, publicações)” (CUNHA; CAVALCANTI, 2008, p. 81). Perante essa definição revelam-se alguns desafios. Principalmente, advindos da necessidade em superar os obstáculos na obtenção e no tratamento de dados primários objetivando mensurar a informação científica. Para mitigar esses desafios, comumente, os esforços despendidos são complexos (SANTOS; KOBASHI, 2009) e custosos. Por exemplo, para sobrepujar os entraves técnicos, conhecimentos diversos são exigidos, em especial de computação, para coletar, organizar, avaliar e armazenar os dados pertinentes. Pressupõe-se que a complexidade e o custo do tratamento dos dados são majorados em função dos grandes volumes de dados provenientes da Internet. Adicione-se a isso, que a disposição desses volumes é distribuída em várias fontes representadas em formatos incompatíveis, o que dificulta a manipulação.

Perante o cenário descrito, exemplarmente, considera-se duas fontes de dados abertas difundidas entre pesquisadores brasileiros e disponibilizadas na Internet no início dessa década: a Plataforma Lattes (CONSELHO NACIONAL DE DESENVOLVIMENTO CIENTÍFICO E TECNOLÓGICO, 2017) e o Sistema WebQualis¹ (COMISSÃO DE APERFEIÇOAMENTO DE PESSOAL DO NÍVEL SUPERIOR, 2013). Na Plataforma Lattes encontram-se dados sobre as comunicações científicas produzidas por pesquisadores. Já no Sistema WebQualis, os índices de qualidade de algumas dessas comunicações estavam estratificados. Interligando essas fontes, os pesquisadores podiam quantificar e qualificar algumas medidas da produção científica nacional. Mas, aferir informação a partir das referidas fontes não era uma atividade trivial, visto que os dados estavam disponíveis em formatos distintos (os currículos Lattes se encontram em páginas de Internet e o índice Qualis dos periódicos estava no formato PDF²). Este fato dificulta o cruzamento de dados e a exploração de informação de forma automatizada.

¹ Atualmente, o Sistema WebQualis se encontra indisponível, sendo substituído pela Plataforma Sucupira (Disponível em: <<https://sucupira.capes.gov.br/sucupira/public/consultas/coleta/veiculoPublicacaoQualis/listaConsultaGeralPeriodicos.jsf>>. Acesso em: 19 out. 2017). Salienta-se que a indisponibilidade do WebQualis repercutiu na perda de parte do histórico do índice Qualis. Pontualmente, o período 2005 a 2009 do referido índice não está publicado na atual plataforma.

² *Portable Document Format*: formato proprietário usado para minimizar problemas na visualização de dados.

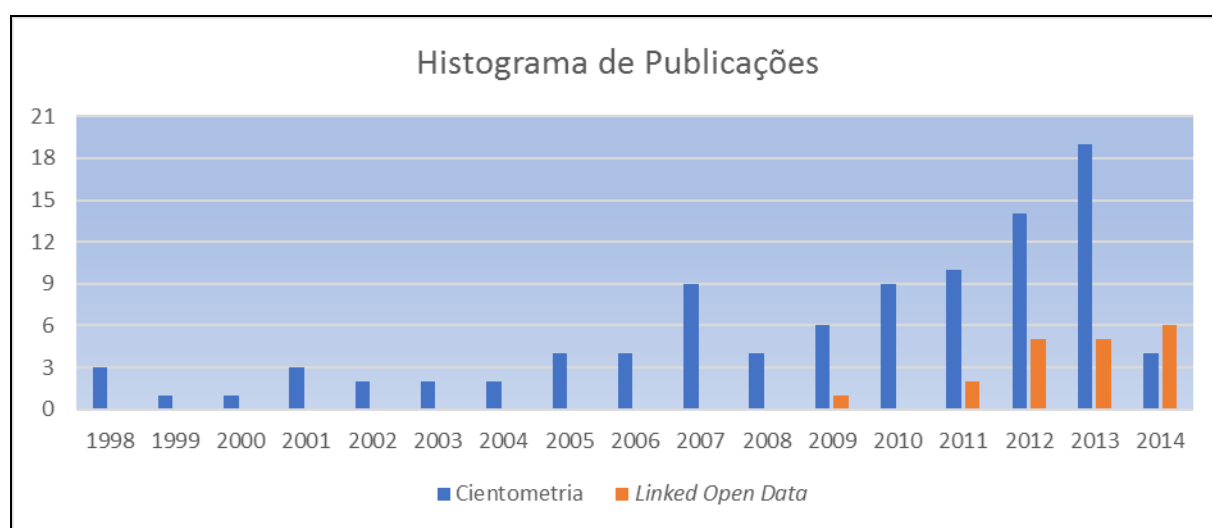
Uma das formas de contornar os desafios de integração de dados seria o uso da Web Semântica, que oferece o suporte metodológico do *Linked Open Data*. O *Linked Open Data* se baseia em um conjunto de melhores práticas (WORLD WIDE WEB CONSORTIUM, 2017) para organizar, publicar, interligar, descobrir, acessar e reusar dados abertos na *web*. Em suma, seguindo os princípios *Linked Open Data*, os dados são disponibilizados sob o uso de licenças abertas, possibilitando que sejam reutilizados sem restrições, por pessoas ou aplicações e em diversos contextos.

Contudo, estudos que envolvam a interdisciplinaridade da Cientometria e do *Linked Open Data* ainda são raros na literatura. Buscas por trabalhos correlatos que relacionem os referidos constructos foram realizadas em bases de dados científicas. Em junho de 2014 foram consultados os portais ScienceDirect (SCIENCEDIRECT, 2014) e Scielo (SCIELO, 2014), conforme descrito a seguir.

No portal ScienceDirect, realizou-se uma busca por artigos que abordam os termos "*Scientometric*" e/ou "*Linked Open Data*" em títulos, palavras-chave e resumos, considerando os últimos quinze anos. Foram recuperados 73 artigos, sendo 55 abordando Cientometria e 18 discorrendo sobre *Linked Open Data*. Como fato importante identificado, nenhum artigo recuperado aborda os termos da pesquisa simultaneamente, constando-se uma oportunidade a uma pesquisa interdisciplinar e aplicada.

No contexto nacional, a oportunidade de um estudo interdisciplinar também se confirmou. A partir do portal Scielo, foram recuperados 42 artigos que abordam o tema Cientometria e um estudo relacionado com *Linked Open Data*. Novamente, não foram localizados artigos que relacionam os termos, simultaneamente.

Figura 1 - Histograma dos artigos científicos sobre Cientometria e *Linked Open Data*



Fonte: Adaptado de SCIENCEDIRECT (2014) e SCIELO (2014).

Com o auxílio de um histograma sobre as publicações recuperadas (Figura 1), percebe-se que os estudos envolvendo Cientometria e *Linked Open Data* têm se intensificado, corroborando a percepção da contemporaneidade dos conceitos.

Diante dessas considerações, encontra-se espaço para o desenvolvimento de tecnologias e de soluções para o compartilhamento de *Linked Open Data* para o suporte aos estudos bibliométricos ou cientométricos.

Em face a isso, este trabalho relata a experiência da publicação do histórico do índice Qualis conforme os princípios *Linked Open Data*, evoluindo um trabalho preliminar (RAUTENBERG *et al.*, 2014), ao propor uma interface *web* para suportar o consumo de dados. Pressupõe-se que tal empreendimento é importante à coleta de dados (entenda-se reutilização de dados) em vários estudos cientométricos no âmbito nacional, podendo servir de base para: (i) a mensuração da evolução dos periódicos científicos; (ii) o auxílio na aferição de medidas qualitativas ou quantitativas de publicações científicas em universidades ou institutos de pesquisa brasileiros; ou (iii) a obtenção de informações relevantes a partir do cruzamento com outros indicadores cientométricos.

Para discutir a disponibilização do referido histórico, além desta seção introdutória, este artigo compreende: (i) a conceitualização de Cientometria e *Linked Open Data*; (ii) o relato dos materiais utilizados e do procedimento metodológico adotado para a disponibilização do índice Qualis; (iii) a verificação do processo de disponibilização frente a alguns cenários no consumo do índice Qualis, exemplificando a reutilização dos dados em estudos cientométricos; (iv) a discussão dos resultados alcançados perante os cenários apresentados; e (v) as considerações finais e os apontamentos aos trabalhos futuros, seguidos pelos agradecimentos e referências bibliográficas.

2 Revisão de literatura

No âmbito deste trabalho, dois conceitos se relacionam: a Cientometria e o *Linked Open Data*. Nessa seção, estes conceitos são brevemente discutidos, considerando sua interdisciplinaridade quanto ao objeto de estudo, a publicação do histórico do índice Qualis como *Linked Open Data* na *Web* de Dados.

2.1 A Cientometria e os eixos para estudo

Historicamente, o termo Cientometria é relativamente novo ao se comparar sua origem com a de conceitos em outras áreas do conhecimento. Segundo Garfield (2009), o surgimento do referido conceito é atribuído aos autores Nalimov e Mul'chenko, que em 1969 publicaram a obra russa intitulada "*Scientometrics. The Study of Science as an Information Process*".

Desde sua infância, interdisciplinarmente, a Cientometria encontrou suporte na Estatística e na Ciência da Computação. Seu entendimento é

reservado à pesquisa quantitativa da ciência e da tecnologia (van RAAN, 1997). De acordo com essa definição, a Cientometria visa mensurar a informação científica/tecnológica, a partir de fontes de dados primários como os documentos científicos (resumos, artigos, anais de conferência, periódicos ou livros), os autores de comunicações científicas, ou as instituições científicas envolvidas no fomento da ciência (STOCK; STOCK, 2015).

Ao se considerar o presente, é notório que o meio acadêmico presencia os avanços na massificação de dados e informação e, por conseguinte, experimenta maior apropriação da comunicação científica. Muito em decorrência do atual estágio das Tecnologias de Informação e Comunicação, principalmente, da Internet. Inserindo-se neste contexto, a Cientometria mostra grande potencial de aplicabilidade (SILVA; BIANCHI, 2001) para entender as dinâmicas do relacionamento da ciência e da tecnologia frente a Sociedade da Informação e do Conhecimento. Neste sentido, as pesquisas cientométricas são desenvolvidas para: (i) propor o uso de novos métodos ou tecnologias para avançar a própria disciplina; (ii) conjunturalmente, entender as políticas de ciência e de tecnologia; (iii) mapear o conhecimento em instituições de pesquisa; ou (iv) conhecer um objeto de pesquisa em particular. Sob esse prisma, o Quadro 1 lista alguns trabalhos encontrados na literatura, conforme os quatro eixos enumerados.

Quadro 1 - Exemplos de pesquisas cientométricas relatadas na literatura especializada

Eixo	Autor(es)	Sinopse
Propor o uso de novos métodos ou tecnologias	(FRENKEN; HARDEMAN; HOEKMAN, 2009)	discorre sobre os aspectos espaciais dos sistemas de ciência. Revisam a distribuição geográfica das pesquisas e citações, a existência do viés de colaboração, a mobilidade e o impacto das citações nacionais <i>versus</i> colaboração internacional. Propõe um <i>framework</i> analítico baseado no conceito de proximidade.
	(GLÄNZEL, 2010)	apresenta um possível modelo estocástico para a Cientometria. Fundamenta a confiabilidade e a robustez do modelo perante quatro dimensões: (i) a correlação entre variáveis; (ii) a confiabilidade estatística para comparações; (iii) o ranqueamento; e (iv) os aspectos preditivos.
	(de la PENA, 2011)	fundamenta matematicamente uma função de impacto para qualquer conjunto de artigos científicos. A função tem seu valor definido a partir da ponderação das arestas do grafo formado pelo relacionamento dos artigos e suas citações.
Entender as políticas de ciência e de tecnologia	(GEORGIU; THOMAIDIS, 2001)	avalia as publicações referentes ao tema Química Analítica dos países da União Europeia, no período 1993-1999. São discutidos o número de publicações de cada país, o número de publicações <i>per capita</i> e o fator de impacto médio para cada país.
	(LEYDESDORFF; WAGNER, 2009)	apresenta alguns indicadores macroeconômicos de nações, tendo como base a perspectiva cientométrica. Cita o caso da China, em que o governo incentivou a expansão do número de publicações científicas. Discute a necessidade de investimento em Pesquisa e Desenvolvimento por parte dos diversos países do mundo.
	(ALHAIDER; MUEEN AHMED; GUPTA, 2013)	analisa a contribuição da Arábia Saudita na pesquisa farmacêutica, nos anos de 2001 a 2010. Considera a produção científica dos 15 países mais produtivos e a posição da Arábia Saudita, ponderando o fator de impacto e a colaboração com os países parceiros. Apesar de haver uma produção significativa no triênio 2008-2010, afirma a necessidade de um conjunto maior de artigos para estabelecer um quadro comparativo.

Mapear conhecimento em instituições de pesquisa	(SERRANO <i>et al.</i> , 2007)	apresenta a estrutura da rede de colaboração entre centros médicos e o compartilhamento de relações de autoria em artigos científicos sobre fumo. O estudo envolveu os centros médicos da Espanha entre os anos de 1999 e 2003. São considerados 35 centros médicos. Desses, 21 centros médicos possuem projetos colaborativos.
	(ZHOU; LEYDESDORFF, 2011)	mapeia e normaliza o fator de impacto de 27 Departamentos da Tsinghua <i>University</i> de Pequim entre os anos de 2005-2009. Observa a evolução científica do Departamento de Língua Chinesa e de Linguística.
	(MENEHINI, 2011)	faz uma análise cientométrica das citações de artigos de 12 Institutos de Física (10 do Brasil e dois estrangeiros) para ranqueá-los. Argumenta que a análise de citações é mais apropriada para realizar o <i>ranking</i> , visto que consegue relacionar produtividade e impacto das publicações no meio científico.
Conhecer um objeto de pesquisa	(RUSHTON, 2001)	descreve a vida científica de Hans Eysenck, um dos psicólogos mais citados após Sigmund Freud e Jean Piaget. Eysenck iniciou os trabalhos da Teoria da Personalidade, incluindo a genética e a base de comportamento para a discussão das doenças mentais.
	(MEYER <i>et al.</i> , 2004)	avalia as contribuições científicas, tecnológicas e de inovação de Keith Pavitt. Perante as citações, identifica que o referido pesquisador é um elo interdisciplinar e importante em dois domínios: Cientometria e Políticas para Pesquisa.
	(ARBOIT; BUFREM; GONZÁLEZ, 2011)	discute o conceito de institucionalização científica e suas relações com a produção de livros, artigos de periódicos, capítulos de livro e trabalhos em eventos no exterior. Considera o universo dos pesquisadores brasileiros do campo da Ciência da Informação. Pontua que o processo de institucionalização da Ciência da Informação encontra-se em construção, frente a realidade expressa nos indicadores aferidos.

Fonte: Elaborado pelos autores.

Van Raan (1997) também considera que existem vários eixos de estudos em Cientometria. Em uma visão atemporal, o autor citado pontua que parte das pesquisas focam os sistemas de informação voltados à ciência e tecnologia, envolvendo o tratamento e a estruturação de dados, assim como a adoção de novas tecnologias. Perante essa visão, o presente trabalho pretende avançar na discussão da Cientometria amparada tecnologicamente. Para tanto, propõem-se a estruturação de um índice cientométrico como *Linked Open Data*, compartilhando recursos na *web* para demais pesquisas bibliométricas/cientométricas.

2.2 Linked Open Data, sua classificação e benefícios

No campo da *Web Semântica*, o *Linked Open Data* tem se difundido como um conjunto de melhores práticas para publicar, conectar e socializar dados estruturados na *web* (LINKED DATA, 2012a). Constitutivamente, a percepção do que é *Linked Open Data* está relacionado a dois entendimentos: (a) o que é *open data* (dados abertos); e (b) o que é *linked data* (dados conectados).

Open data são os dados regidos por licenças que não impedem o (re)uso, desde que obedecidas as regras pertinentes de distribuição (LINKED DATA, 2012b).

Por sua vez, *linked data* refere-se a um conjunto de práticas para publicar e interligar dados estruturados usando a *web* como uma plataforma global (BIZER; HEATH; BERNES-LEE, 2009; HEATH; BIZER, 2011). Isto somente é possível pela utilização de um modelo padrão para relacionar os dados de forma descentralizada. Sob este prisma, tem-se o

Resource Description Framework (RDF), uma linguagem para conectar os dados usando tecnologias da *web*. Em suma, o RDF representa recursos na Internet na forma de triplas (sujeito → predicado → objeto), relacionando um sujeito a um objeto através de um predicado.

Figura 2 - Representação de uma tripla RDF



Fonte: Elaborado pelos autores.

Como exemplo, a Figura 2 evidencia uma tripla descrita em RDF e sua representação gráfica. Cabe ressaltar que o referido exemplo faz parte do grafo³ <http://lod.unicentro.br/QualisBrasil/> (prefixo = qualis), o qual é discutido com maior propriedade na próxima seção. No exemplo, um sujeito <http://lod.unicentro.br/QualisBrasil/Periodico_1413-9936> é identificado por "qualis:Journal_1981-5344" e tem um predicado <http://purl.org/dc/elements/1.1/title> (a propriedade dc:title do Vocabulário Dublin Core) que aponta ao nome "Perspectivas em Ciência da Informação".

Ao unir os conceitos *open data* e *linked data* têm-se o *Linked Open Data*. Mediante essa integração conceitual, os dados podem ser publicados na *web* de acordo com licenças abertas, possibilitando que os mesmos sejam (re)utilizados em diversos domínios. Ressalta-se que os dados disponibilizados conforme os princípios Linked Open Data são classificados conforme seu nível de abertura e de conexão a outros dados. Representada na Figura 3, essa classificação é denominada 5-Estrelas e é organizada como segue (5-STAR, 2016):

1ª Estrela - é atribuída aos dados que são publicados sob uma licença aberta (*Open License* - OL), não importando o formato. Assim, a manipulação e o processamento dos dados abertos são dificultados (Exemplo: a planilha das avaliações qualis no ano de 2009 foi publicada como *open data*, porém, disponibilizada em formato PDF);

2ª Estrela - é conferida à publicação de dados estruturados legíveis por máquinas (*Readable Machine* - RE). Os dados podem ser diretamente processados por *softwares* proprietários e podem ser convertidos para outros formatos (Exemplo: a planilha das avaliações qualis no ano de 2007, em XLS⁴, a qual é processável pelo *software* proprietário Microsoft Excel®);

3ª Estrela - é concedida aos dados que são publicados em formato aberto não proprietário (*Open Format* - OF). Neste

³ Estrutura que comporta um conjunto de triplas RDF acerca de um mesmo assunto, organizado segundo um mesmo modelo.

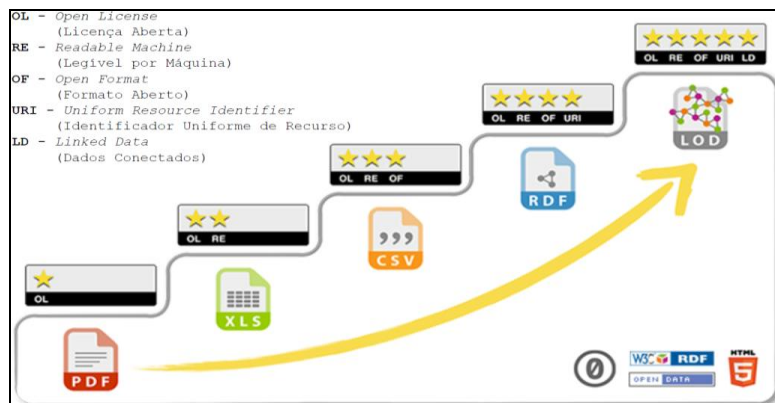
⁴ Planilha eletrônica: formato proprietário, possível de ser manipulado por alguns aplicativos.

patamar, a manipulação dos dados é realizada sem a necessidade do uso de um *software* proprietário (Exemplo: os dados de uma planilha XLS salvos em formato CSV⁵);

4ª **Estrela** - é designada à utilização de Identificadores Universais de Recursos (*Universal Resource Identifier* - URI) para nomear os dados na *web*, permitindo a criação de ligações entre os dados (Exemplo: o conteúdo de um arquivo CSV convertido para recursos no formato RDF);

5ª **Estrela** - é atribuída aos dados que são conectados (*Linked Data* - LD) a outros dados. Isso permite a navegação entre dados e a descoberta de informação relacionada. Dessa forma, acrescenta-se valor aos dados ao fornecer uma contextualização mais ampliada (Exemplo: vincular recursos RDF locais a outros recursos na *web*).

Figura 3 - Classificação 5-Estrelas



Fonte: Adaptado de 5-STAR (2016).

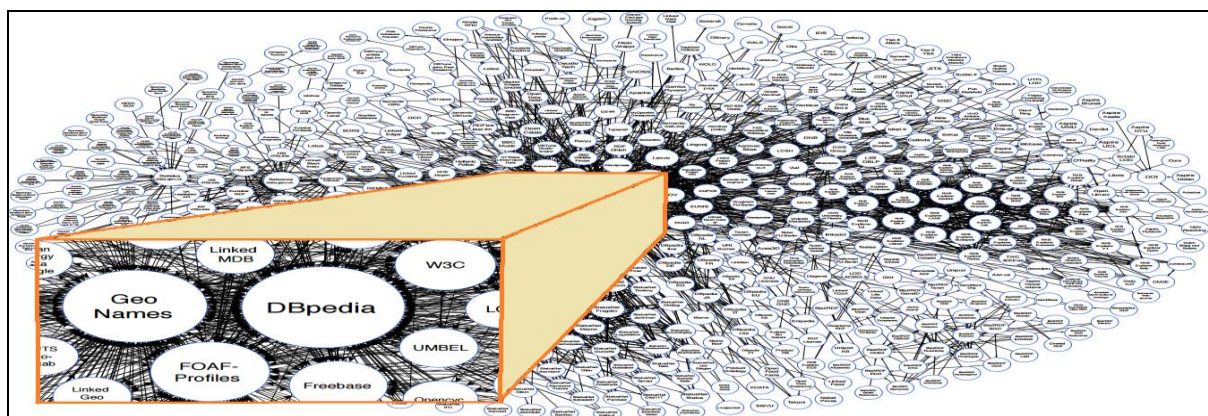
Considerando a classificação anterior, a união de *open data* e *linked data* é potencializada ao atingir-se a 5ª Estrela. Salienta-se que a referida junção é a base informacional de um imenso grafo RDF denominado *Web de Dados* (Figura 4), uma estrutura global que suporta a publicação e o consumo de vários conjuntos de dados abertos. Em face dessa dinâmica de compartilhamento, publicadores e consumidores de *Linked Open Data* aproveitam-se de oito benefícios (WORLD WIDE WEB CONSORTIUM, 2017):

1. **Reuso** – o aumento das chances de reutilização de dados por diferentes grupos de consumidores de dados;
2. **Compreensão** – a estruturação dos elementos para entender a natureza e o significado dos dados disponibilizados;
3. **Interligação** – a possibilidade de criação de relacionamentos entre os recursos de dados geograficamente distribuídos;

⁵ *Comma Separated Values*: formato amplamente utilizado para exportação de dados entre aplicações.

4. **Descoberta** – o suporte aos computadores para descobrir automaticamente um conjunto de dados;
5. **Confiança** – a certificação que um conjunto de dados é melhorado ao longo do tempo;
6. **Acesso** – a facilidade do acesso aos dados atualizados e disponibilizados em múltiplos formatos;
7. **Interoperabilidade** – o apoio à troca de dados entre os sistemas; e
8. **Processabilidade** – o suporte para que os sistemas computacionais processem os (conjuntos de) dados, automaticamente.

Figura 4 - Representação dos grafos na *Web* de Dados em 30 de agosto de 2014, destacando a centralidade da DBpedia



Fonte: Adaptado de SCHMACHTENBERG; BIZER; PAULHEIM (2017).

Observando os benefícios citados e considerando as universidades como polos de produção e de consumo de dados, adotar o *Linked Open Data* se torna cada vez mais pertinente no meio científico. Por exemplo, interdisciplinarmente, pode-se relacionar o *Linked Open Data* à Cientometria, visto que os estudos cientométricos podem se basear em volumes de dados primários e que, por natureza, podem ser *open data*. Em suma, ao considerar que na Internet permeia-se um grande repositório de dados, a adoção dos princípios *Linked Open Data*, em consonância aos instrumentos metodológicos de estudos cientométricos, facilitaria as atividades de coleta de dados primários. Neste sentido, este trabalho discorre sobre as experiências em publicar o histórico do índice Qualis como *Linked Open Data* e como utilizar os recursos compartilhados na *Web* de Dados.

3 Materiais e métodos

Conforme um trabalho anterior (RAUTENBERG *et al.*, 2014), para disponibilizar o índice Qualis como *Linked Open Data*, são considerados os seguintes materiais:

- a. a tabela da classificação do índice Qualis (formato de planilha eletrônica - XLS) acessada no ano de 2007 e denominada neste trabalho de Qualis_2007;
- b. a tabela da classificação do índice Qualis (formato *Portable Document File* - PDF) acessada no ano de 2009 e denominada neste trabalho de Qualis_2009;
- c. a tabela da classificação do índice Qualis (formato *Portable Document File* - PDF) acessada no ano de 2013 e denominada neste trabalho de Qualis_2013; e
- d. o metadados Dublin Core (DUBLIN CORE METADATA INITIATIVE, 2014) para organização e representação dos dados conforme um vocabulário da *Web Semântica*.

Para o estabelecimento de uma linha temporal do referido índice, adotou-se o critério de replicação dos dados para representar três triênios, como segue:

- a. Qualis_2007 - representa os índices Qualis para os anos 2005 a 2007;
- b. Qualis_2009 - representa os índices Qualis para os anos 2008 a 2010; e
- c. Qualis_2013 - representa os índices Qualis para os anos 2011 a 2013.

Ressalta-se que os arquivos anteriormente enumerados foram pré-processados, removendo as tuplas que não constavam, ou o ISSN de uma avaliação, ou uma classificação qualis válida. A Tabela 1 apresenta a quantidade de tuplas processadas, evidenciando a baixa porcentagem de perda de dados, considerando as restrições de consistência relatadas.

Tabela 1 - Resumo do pré-processamento dos arquivos do índice Qualis

Arquivo	Tuplas processadas	Tuplas validadas	% de perda
Qualis_2007	35.190	35.020	0,48
Qualis_2009	54.493	54.233	0,48
Qualis_2013	108.562	107.429	1,04

Fonte: Elaborado pelos autores.

Listagem 1 - código-fonte em RDF – exemplo de descrição de recursos

01	<?xml version="1.0" encoding="UTF-8"?>
02	<rdf:RDF
03	xmlns:owl="http://www.w3.org/2002/07/owl#"
04	xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
05	xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
06	xmlns:dc="http://purl.org/dc/elements/1.1/"
07	xmlns:qualis="http://lod.unicentro.br/QualisBrasil/">

```

08
09 <rdf:Description rdf:about="qualis:Area_31">
10 <dc:identifier>31</dc:identifier>
11 <dc:title>CIÊNCIAS SOCIAIS APLICADAS I</dc:title>
12 </rdf:Description>
13
14 <rdf:Description rdf:about="qualis:Ano_2013">
15 <rdf:value>2013</rdf:value>
16 </rdf:Description>
17
18 <rdf:Description rdf:about="qualis:Qualis_A1">
19 <dc:identifier>A1</dc:identifier>
20 </rdf:Description>
21
22 <rdf:Description rdf:about="qualis:Periodico_1413-9936">
23 <dc:identifier>1413-9936</dc:identifier>
24 <dc:title>Perspectivas em Ciência da Informação</dc:title>
25 </rdf:Description>
26
27 <rdf:Description rdf:about="qualis:umaAvaliacao">
28 <temPeriodico xmlns="qualis:" rdf:resource="qualis:Periodico_1413-9936"/>
29 <temArea xmlns="qualis:" rdf:resource="qualis:Area_31"/>
30 <temAno xmlns="qualis:" rdf:resource="qualis:Ano_2013"/>
31 <temQualis xmlns="qualis:" rdf:resource="qualis:Qualis_A1"/>
32 </rdf:Description>
33 </rdf:RDF>

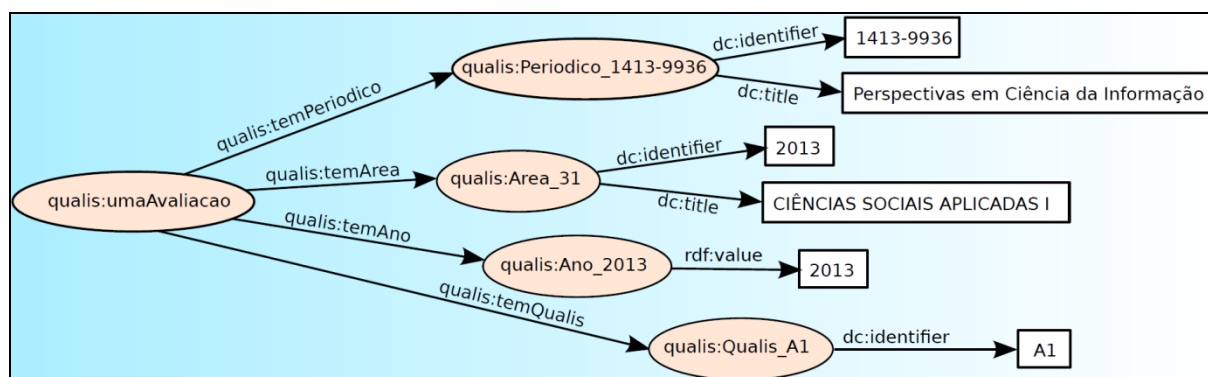
```

Fonte: Elaborado pelos autores.

Os arquivos tiveram seu formato migrado para CSV, permitindo a transformação dos dados para o formato RDF. A Listagem 1 exemplifica a codificação, em RDF, para a avaliação qualis do periódico *Perspectivas em Ciência da Informação*, na área de *Ciências Sociais Aplicadas I*, no ano 2013. Na referida listagem ressalta-se: (a) **linha 7** - a padronização da URI para os dados a partir de `<http://lod.unicentro.br/QualisBrasil/>`⁶, com o prefixo *qualis*; (b) **linhas 9 a 12** - a criação de um recurso de representação da área *Ciências Sociais Aplicadas I*, considerando o vocabulário Dublin Core; (c) **linhas 14 a 16** - a descrição de um recurso para vinculação do ano 2013; (d) **linhas 18 a 20** - a criação do recurso para representação do estrato A1; (e) **linhas 22 a 25** - a descrição do recurso para representar o periódico *Perspectivas em Ciência da Informação*; e (f) **linhas 27 a 32** - a criação de um recurso que agrega os recursos anteriores.

⁶ O endereço `<http://lod.unicentro.br>` encontra-se em construção. Entretanto, os recursos do histórico do índice Qualis podem ser acessados a partir do sítio `<http://lodunicentro.aks.org>`, conforme exemplificado nos casos de uso.

Figura 5 - grafo de representação da Listagem 1



Fonte: Elaborado pelos autores.

A Figura 5 traz a representação gráfica e o relacionamento dos recursos citados na Listagem 1. Ressalta-se que, para o período de 2005-2013, o histórico do índice Qualis é compartilhado, com um total de 590.046 avaliações. Esse conjunto de recursos forma um grafo RDF denominado <<http://lod.unicentro.br/QualisBrasil/>>, o qual pode ser explorado conforme os cenários descritos a seguir.

4 Cenários para o reuso de recursos do grafo <<http://lod.unicentro.br/QualisBrasil/>>

Para verificar o reuso dos recursos do grafo <<http://lod.unicentro.br/QualisBrasil/>> foram projetados três estudos de caso, com os cenários prototipados e acessados a partir de uma interface *web* no sítio <<http://lodunicentro.aksw.org>>.

4.1 Consumo de dados

Considerando os princípios *Linked Open Data*, deve-se disponibilizar os recursos de dados de forma estruturada ao utilizar formatos não proprietários. Neste trabalho, o histórico do índice Qualis pode ser consumido de duas maneiras: uma de caráter mais computacional e outra de forma mais intuitiva.

Computacionalmente, uma consulta em linguagem SPARQL pode ser codificada para recuperar os dados ([formatos abertos CSV, RDF, XML⁷, por exemplo](#)) diretamente do *endpoint*⁸ <<http://space.sina.aksw.org/sparql>>. A Listagem 2 exemplifica uma consulta em SPARQL para o consumo de um subconjunto dos recursos disponibilizados. Cabe ressaltar que consultas similares podem ser baseadas neste exemplo e integradas em diversas aplicações, alcançando os benefícios do reuso, da interligação e do acesso aos dados; assim como da interoperabilidade entre sistemas e da processabilidade automatizada. Na listagem, destaca-se: (a) linha 01 - a padronização da nomenclatura dos recursos a partir de <<http://lod.unicentro.br/QualisBrasil/>>, com o prefixo *qualis*; (b) linhas 06 - as

⁷ *Extensible Markup Language*: formato amplamente utilizado para exportação de dados entre aplicações.

⁸ Sítio da Internet que hospeda os recursos semânticos.

dimensões dos recursos a serem recuperadas; (c) linhas 08 a 11 - os recursos considerados; (d) linhas 13 a 18 - as dimensões das triplas que caracterizam as dimensões recuperadas; e (e) linhas 20 a 22 - o filtro da consulta (neste caso, todos os recursos disponíveis para a Área 31 – Ciências Sociais Aplicadas I, com o estrato qualis A1 do ano de 2013).

Listagem 2 - exemplo de consulta de dados em linguagem SPARQL

```

01 PREFIX qualis: <http://lod.unicentro.br/QualisBrasil/>
02 PREFIX dc: <http://purl.org/dc/elements/1.1/>
03 PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
04 PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
05
06 SELECT DISTINCT ?ISSN ?nomePeriodico ?idCNPq ?nomeArea ?anoQualis ?qualisPeriodico
07 WHERE {
08   ?Avaliacao qualis:temPeriodico ?Periodico .
09   ?Avaliacao qualis:temArea ?Area .
10   ?Avaliacao qualis:temAno ?Ano .
11   ?Avaliacao qualis:temQualis ?Qualis .
12
13   ?Periodico dc:identifier ?ISSN .
14   ?Periodico dc:title ?nomePeriodico .
15   ?Area dc:identifier ?idCNPq .
16   ?Area dc:title ?nomeArea .
17   ?Ano rdf:value ?anoQualis .
18   ?Qualis rdf:value ?qualisPeriodico .
19   FILTER (
20     ?Area = qualis:Area_31 &&
21     ?Ano = qualis:Ano_2013 &&
22     ?Qualis = qualis:Qualis_A1)
23 } ORDER BY ?Ano ?Qualis ?nomePeriodico

```

Fonte: Elaborado pelos autores.

Figura 6 - Consumo de dados via interface

ISSN	Revista	Código Área	Área	Ano	Qualis
1981-5344	Perspectivas em Ciência da Informação (Online)	31	CIÊNCIAS SOCIAIS APLICADAS I	2013	A1
1527-8018	Public Culture	31	CIÊNCIAS SOCIAIS APLICADAS I	2013	A1
1588-2861	Scientometrics	31	CIÊNCIAS SOCIAIS APLICADAS I	2013	A1
0138-9130	Scientometrics (Print)	31	CIÊNCIAS SOCIAIS APLICADAS I	2013	A1

Fonte: Elaborado pelos autores.

A consulta anterior também pode ser realizada de forma intuitiva, sendo que os dados são acessados através da interface ilustrada na Figura 6. Conforme os critérios selecionados (Figura 6a), o consumidor de dados tem a possibilidade de exportar (Figura 6b) os dados apresentados (Figura

6c). Na interface, os dados podem ser exportados em quatro formatos (CSV, XML, XLS e PDF). Em especial, CSV e XML são os formatos não proprietários e seu uso é aconselhado segundo os preceitos do *Linked Open Data*.

4.2 Aferição de medidas qualitativas

Uma das vantagens do *Linked Open Data* é a possibilidade em vincular os dados de diferentes grafos na *Web* de Dados, ampliando os contextos informacionais. A Listagem 3 exemplifica o consumo do histórico do índice Qualis ao atrelar os estratos às publicações de determinado grupo de pesquisadores. Na listagem, destaca-se: (a) linha 1 – um grafo que compreende a produção de artigos em periódicos, com o prefixo *pp*; (b) linha 2 - o grafo do histórico do índice Qualis, com prefixo *qualis*; (c) linha 7 - os dados recuperados; (d) linhas 8 a 10 - os artigos de um grupo de pesquisadores e suas respectivas avaliações; (e) linhas 12 a 15 - a recuperação dos recursos que compõem uma avaliação qualis; (f) linhas 17 a 21 - as dimensões dos recursos a serem recuperadas como dados da consulta; e (g) linhas 24 a 26 - os filtros da consulta que caracterizam o grupo de pesquisadores e a fatia da área/qualis a ser analisada (neste caso, a área de *Ciências Sociais Aplicadas I* e o estrato *A1*). Cabe ressaltar que os itens (b) e (e) evidenciam a reutilização de dados (o índice Qualis) no enriquecimento de um contexto (a classificação da produção científica de determinado grupo), comprovando o alcance da 5ª Estrela (5-STAR, 2016).

Listagem 3 - código-fonte em linguagem SPARQL – exemplo de consumo de dados

01	PREFIX pp: <http://lod.unicentro.br/ProducaoPeriodicos/>
02	PREFIX qualis: <http://lod.unicentro.br/QualisBrasil/>
03	PREFIX dc: <http://purl.org/dc/elements/1.1/>
04	PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
05	PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
06	
07	SELECT DISTINCT ?labelArea ?labelAno ?labelQualis ?referencia WHERE {
08	?Grupo dc:contributor ?Professor .
09	?Professor pp:temArtigo ?Artigo .
10	?Artigo pp:temAvaliacao ?Avaliacao .
11	
12	?Avaliacao qualis:temAno ?Ano .
13	?Avaliacao qualis:temArea ?Area .
14	?Avaliacao qualis:temQualis ?Qualis .
15	?Avaliacao qualis:temPeriodico ?Periodico .
16	
17	?Artigo dc:identifier ?referencia .
18	?Grupo dc:title ?labelGrupo .
19	?Area dc:title ?labelArea .
20	?Ano rdf:value ?labelAno .
21	?Qualis rdf:value ?labelQualis .
22	

23	FILTER (
24	?Grupo = pp:Grupo_3 &&
25	?Area = qualis:Area_31 &&
26	?Qualis = qualis:Qualis_A1)
27	} ORDER BY DESC(?Ano) ?Qualis

Fonte: Elaborado pelos autores.

Figura 7 - Exemplificação do enriquecimento de um contexto com o índice Qualis

Grupo	Área	Ano	Qualis	Referência
Departamento De Ciência Da Computação	CIÊNCIAS SOCIAIS APLICADAS I	2013	A1	CASTANHO, M. J. DE P.; HERNANDES, F.; RE, A. M. DE; RAUTENBERG, S.; BILLIS, A. Fuzzy expert system for predicting pathological stage of prostate cancer. Expert Systems with Applications. 2013.
Departamento De Ciência Da Computação	CIÊNCIAS SOCIAIS APLICADAS I	2011	A1	RAUTENBERG, S.; STEIL, A. V.; TODESCO, J. L. Modelo de conhecimento para mapeamento de instrumentos da gestão do conhecimento e de agentes computacionais da engenharia do conhecimento. Perspectivas em Ciência da Informação. 2011.
Departamento De Ciência Da Computação	CIÊNCIAS SOCIAIS APLICADAS I	2011	A1	RAUTENBERG, S.; TODESCO, J. L.; STEIL, A. V. Uma Ontologia para Instrumentos da Gestão do Conhecimento e Agentes da Engenharia do Conhecimento. Informação & Sociedade (UFPB, Impresso). 2011.

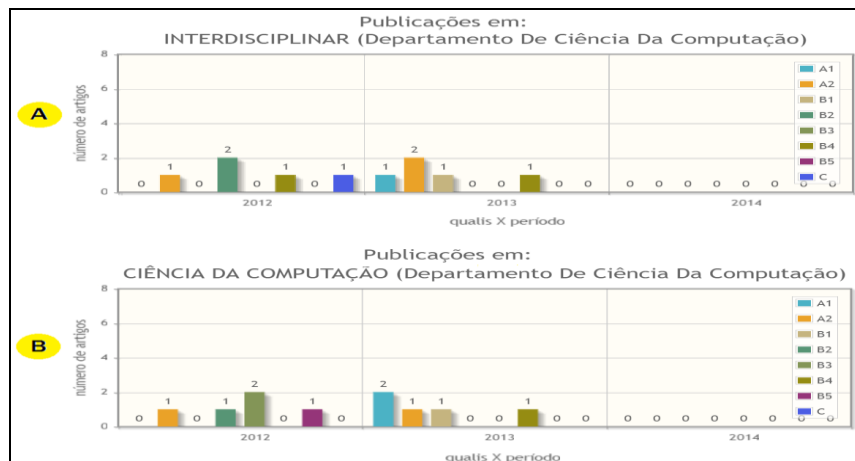
Fonte: Elaborado pelos autores.

A Figura 7 apresenta a interface *web* em conformidade com a consulta codificada na Listagem 3. Na figura, o elo de enriquecimento do contexto (5ª Estrela) é destacado e evidencia a interligação, interoperabilidade e a processabilidade dos recursos de dados compartilhados entre os grafos e <http://lod.unicentro.br/ProducaoPeriodicos/> e <http://lod.unicentro.br/QualisBrasil/>.

4.3 Obtenção de informações cientométricas

De forma geral, obter informações cientométricas auxilia: (i) no reconhecimento de competências implícitas, (ii) na definição de ações para o desenvolvimento da ciência e de tecnologias, ou (iii) simplesmente, no entendimento institucional da evolução das pesquisas (considerando os pesquisadores individuais, os grupos de pesquisa, as áreas do conhecimento, os programas de pós-graduação ou uma universidade).

Figura 8 - exemplificação de um cenário de entendimento da evolução de pesquisas



Fonte: Elaborado pelos autores.

Neste cenário também são exemplificados os benefícios de reuso, interligação, interoperabilidade e processabilidade de recursos disponibilizados. Com base no estudo de caso anterior, os recursos dos grafos <http://lod.unicentro.br/ProducaoPeriodicos/> e <http://lod.unicentro.br/QualisBrasil/> são integrados de forma gráfica para observar a evolução da produção científica nas áreas do conhecimento. Na Figura 8, um conjunto de publicações de determinado grupo de pesquisadores é classificado e sumarizado, considerando as áreas *Interdisciplinar* (Figura 8a) e *Ciência da Computação* (Figura 8b).

5 Resultados e discussão

Os cenários de utilização apresentados na seção anterior permitem realizar algumas discussões. Pontualmente, estes exemplificam como o *Linked Open Data* pode alavancar alguns estudos sobre a informação científica, principalmente, as fases de coleta de dados primários em pesquisas cientométricas. Neste sentido, os cenários apresentados nas subseções exemplificam o reuso de recursos de dados primários do grafo <http://lod.unicentro.br/QualisBrasil/> no contexto da classificação de recursos bibliométricos do grafo <http://lod.unicentro.br/ProducaoPeriodicos/>.

Diante os estudos de caso, atesta-se que a contribuição deste trabalho reside na disponibilização do histórico do índice Qualis seguindo os princípios *Linked Open Data*. Ou seja, antes o índice Qualis estava disponível no formato PDF na interface do sistema WebQualis (1ª Estrela, na Classificação 5-Estrelas proposta por Tim Bernes-Lee), prejudicando o uso automatizado do referido índice em estudos cientométricos. Com o trabalho, o índice Qualis tem seus recursos elevados à 5ª Estrela, minimizando os esforços na compatibilização dos formatos de dados entre aplicações computacionais. Neste sentido, adicionalmente ao trabalho de Rautenberg *et al.* (2014), uma interface *web* de consulta (acessada em

<<http://lodunicentro.aksw.org>>) foi implementada, a qual facilita o consumo dos dados disponibilizados no ambiente distribuído da Internet.

Desta forma, na perspectiva dos benefícios discutidos pela World Wide Web Consortium (2017), o grafo <<http://lod.unicentro.br/QualisBrasil/>> suporta: (i) o reuso de dados por diferentes grupos de consumidores; (ii) a interligação com outros recursos na *Web* de Dados, por exemplo, advindos da Plataforma Lattes; (iii) considerando a manutenção do índice ao longo do tempo (é considerado como trabalho futuro e permanente), o acesso aos dados atualizados; (iv) a interoperabilidade entre os sistemas de informação cientométricos por minimizar os esforços na compatibilização de formatos de representação de dados; e (v) por conseguinte, a processabilidade automática do referido índice por parte dos computadores.

6 Considerações finais

Neste artigo apresenta-se um estudo interdisciplinar e aplicado que envolve os conceitos *Linked Open Data* e Cientometria. Os autores consideram que a obtenção de informações em estudos cientométricos envolve conhecimentos diversos, em especial de computação, devido aos desafios na coleta, na organização e no relacionamento de dados pertinentes. Pressupõe-se que tais desafios existem porque os dados são provenientes da Internet e estão distribuídos em várias fontes, apresentados em formatos proprietários ou incompatíveis, dificultando os procedimentos de manipulação.

Como resultado principal, publica-se o histórico do índice Qualis (período 2005-2013) na *Web* de Dados, seguindo os princípios *Linked Open Data*. Esse conjunto de dados está armazenado no grafo <<http://lod.unicentro.br/QualisBrasil/>>, é acessível a partir no *endpoint* <<http://space.sina.aksw.org/sparql>> e, por meio de consultas SPARQL, é disponibilizado em vários formatos abertos (CSV, RDF, XML, entre outros). Ressalta-se que os dados originais estavam representados no formato proprietário PDF, o que dificultava seu reuso como dados primários em pesquisas cientométricas.

Como contribuição científica, ao advogar a adoção dos princípios *Linked Open Data* para organizar e compartilhar dados primários da Cientometria, o presente trabalho colabora na discussão interdisciplinar, ao propor o uso de tecnologias para suportar os estudos da referida disciplina. Neste sentido, destaca-se a realização de uma revisão de literatura nas bases de dados científicas ScienceDirect e Scielo. Observou-se que a aderência dos princípios do *Linked Open Data* aos estudos da Cientometria ainda é pouco explorada, oportunizando a realização desta pesquisa.

Como trabalhos futuros pretende-se:

- a. atualizar o grafo <<http://lod.unicentro.br/QualisBrasil/>> ao longo do tempo, atuando na preservação digital de seus recursos. Ressalta-se que parte da série histórica do

referido índice já se encontra indisponível nas plataformas oficiais (o sítio WebQualis foi descontinuado e a atual Plataforma Sucupira apresenta o histórico do índice Qualis somente a partir do ano 2010);

- b. relacionar o índice Qualis a outros índices de classificação de periódicos (como os fatores de impacto *Journal Citation Reports* e *SCImago Journal Rank*), baseando-se também nos princípios *Linked Open Data*; e
- c. modelar os mecanismos para a gestão de informações cientométricas no escopo das universidades brasileiras, repercutindo em 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) e a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES - Processo número: 18228/12-7).

Referências

5-STAR. *5-Star Open Data*. 2016. Disponível em: <<http://5stardata.info/en>>. Acesso em: 8 jun. 2016.

ALHAIDER, A.; MUEEN AHMED, K. K.; GUPTA, B. M. Pharmaceutical research in the Kingdom of Saudi Arabia: a scientometric analysis during 2001–2010. *Saudi Pharmaceutical Journal*, v. 23, n. 3, p. 215-222, 2013.

ARBOIT, A. E.; BUFREM, L. S.; GONZALEZ, J. A. M. A produção brasileira em Ciência da Informação no exterior como reflexo de institucionalização científica. *Perspectivas em Ciência da Informação*, v. 16, n. 3, p. 75-92, 2011.

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.

COMISSÃO DE APERFEIÇOAMENTO DE PESSOAL DO NÍVEL SUPERIOR (CAPES). *Sistema WebQualis*: Portal Capes. 2013. Disponível em: <<http://qualis.capes.gov.br/webqualis/principal.seam>>. Acesso em: 25 ago. 2013.

CONSELHO NACIONAL DE DESENVOLVIMENTO CIENTÍFICO E TECNOLÓGICO (CNPq). *Plataforma Lattes*. 2017. Disponível em: <lattes.cnpq.br>. Acesso em: 25 abr. 2017.

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

DUBLIN CORE METADATA INITIATIVE. *DCMI Metadata Terms*. 2014. Disponível em: <<http://dublincore.org/documents/dcmi-terms/>>. Acesso em: 27 jul. 2014.

de la PENA, J. A. Impact functions on the citation network of scientific articles. *Journal of Informetrics*, v. 5, n. 4, p. 565-573, 2011.

FRENKEN, K.; HARDEMAN, S.; HOEKMAN, J. Spatial scientometrics: towards a cumulative research program. *Journal of Informetrics*, v. 3, n. 3, p. 222-232, 2009.

GARFIELD, E. From the science of science to Scientometrics visualizing the history of science with HistCite software. *Journal of Informetrics*, v. 3, n. 3, p. 173-179, 2009.

GEORGIU, C. A.; THOMAIDIS, N. S. Analytical chemistry in the European Union during 1993-1999: an appraisal on the basis of papers abstracted in Analytical Abstracts. *TrAC Trends in Analytical Chemistry*, v. 20, n. 9, p. 462-466, 2001.

GLÄNZE, W. On reliability and robustness of scientometrics indicators based on stochastic models: an evidence-based opinion paper. *Journal of Informetrics*, v. 4, n. 3, p. 313-319, 2010.

HEATH, T.; BIZER, C. *Linked data evolving the web into a global data space*. Amsterdam: Morgan & Claypool, 2011.

LEYDESDORFFA, L.; WAGNER, C. Macro-level indicators of the relations between research funding and research output. *Journal of Informetrics*, v. 3, n. 4, p. 353-362, 2009.

LINKED DATA. Linked Data: connect distributed data across the web. 2012a. Disponível em: <<http://linkeddata.org>>. Acesso em: 28 ago. 2012.

LINKED DATA. Linked Data: design issues. 2012b. Disponível em: <<http://www.w3.org/DesignIssues/LinkedData.html>>. Acesso em: 28 ago. 2013.

MENEGHINI, R. Citations to papers from Brazilian institutions: a more effective indicator to assess productivity and the impact of research in graduate programs. *Braz J Med Biol Res*, v. 44, n. 8, p. 738-747, 2011.

MEYER, M. *et al.* The scientometric world of Keith Pavitt: A tribute to his contributions to research policy and patent analysis. *Research Policy*, v. 33, n. 9, p. 1405-1417, 2004.

RAUTENBERG, S. *et al.* QUALISBRASIL: Disponibilizando dados via Linked Open Data para estudos cientométricos. In: CONGRESSO LINKED OPEN DATA BRASIL, 1., 2014, Florianópolis. *Anais...* Florianópolis: UFSC/EGC, 2014. p. 95-110.

RUSHTON, J. P. A scientometric appreciation of H. J. Eysenck's contributions to psychology. *Personality and Individual Differences*, v. 31, n. 1, p. 17-39, 2001.

SANTOS, R. N. M. dos; KOBASHI, N. Y. Bibliometria, cientometria, infometria: conceitos e aplicações. *Tendências da Pesquisa Brasileira em Ciência da Informação*, v. 2, n. 1, p. 155-172. 2009.

SCHMACHTENBERG, M.; BIZER, C.; PAULHEIM, H. State of the LOD Cloud. 2017. Disponível em: <http://lod-cloud.net/state/state_2014>. Acesso em: 13 abr. 2017.

SCIENTIFIC ELECTRONIC LIBRARY ONLINE (SCIELO). *SCIELO*. 2014. Disponível em: <http://www.scielo.br/scielo.php?script=sci_home&lng=en&nrm=iso>. Acesso: 16 jun. 2014.

SCIENCEDIRECT. *ScienceDirect.com*: search through over 11 million science, health, medical journal full text articles and books. 2014. Disponível em: <<http://www.sciencedirect.com/>>. Acesso: 16 jun. 2014.

SERRANO, S. J. V. *et al.* Spanish Medical Center collaboration on smoking research from 1999 through 2003 according to the Science Citation Index. *Arch Bronconeumol*, v. 43, n. 7, p. 378-385, 2007.

SILVA, J. A. da; BIANCHI, M. de L. P. Cientometria: a métrica da ciência. *Paidéia*, v. 11, n. 21, p. 5-10, 2001.

STOCK, W. G.; STOCK, M. *Handbook of Information Science*. Berlin:Walter de Gruyter GmbH, 2015.

van RAAN, A. F. Scientometrics: state-of-the-art. *Scientometrics*, v. 38, n. 1, p. 205-218, 1997.

WORLD WIDE WEB CONSORTIUM (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.

ZHOU, P.; LEYDESDORFF, L. Fractional counting of citations in research evaluation: a cross- and interdisciplinary assessment of the Tsinghua University in Beijing. *Journal of Informetrics*, v. 5, n. 3, p. 360-368, 2011.