

## **Dados abertos conectados e gestão do conhecimento: estudos de caso cientométricos em uma universidade brasileira**

**Sandro Rautenberg**

***Doutor em Engenharia e Gestão do Conhecimento  
(UFSC) Professor do Programa de Pós-Graduação  
em Administração da Universidade Estadual do  
Centro-Oeste (UNICENTRO)***

**Sandro Kaue Motyl**

***Graduando em Ciência da Computação,  
Universidade Estadual do Centro-Oeste  
(UNICENTRO)***

**Alessandra Cassiana Burda**

***Graduando em Ciência da Computação,  
Universidade Estadual do Centro-Oeste  
(UNICENTRO)***

**Anderson Silvério**

***Mestre em Engenharia de Produção, Universidade  
Tecnológica Federal do Paraná (UFPR)***

**Fabrício Marom de Moura**

***Mestre em Matemática, Universidade Estadual de  
Ponta Grossa (UEPG)***

**<http://dx.doi.org/10.1590/1981-5344/2885>**

*Este artigo se baseia na relação de três conceitos: Gestão do Conhecimento, Cientometria e Dados Abertos Conectados. No contexto das universidades, objetiva-se discutir a promoção da Gestão do Conhecimento institucional amparada por Dados Abertos Conectados advindos do domínio da Cientometria. Para tanto, baseando-se nos preceitos da Web Semântica, pontualmente, no processo metodológico Linked Data Lifecycle, duas bases de dados são abertamente mantidas na Web de Dados: a) o histórico do Índice Qualis (2005-2015); e b) um conjunto de registros de publicações em periódicos extraído da Plataforma Lattes. Como resultado,*

*três estudos de caso reais são desenvolvidos, os quais consomem, relacionam e exploram os dados dessas fontes. Mediante os estudos realizados, conclui-se que as práticas relativas aos Dados Abertos Conectados contribuem na formação de bases de conhecimento institucionais, podendo suportar diversos estudos bibliométricos e cientométricos ou as atividades aderentes à Gestão do Conhecimento universidades brasileiras.*

**Palavras-chave:** *Gestão do Conhecimento; Dados Abertos Conectados; Web Semântica; Cientometria; Produtividade Institucional.*

## **Linked open data and knowledge management: scientometric study cases in a brazilian university**

*This paper is based on the relation of three concepts: Knowledge Management, Scientometrics and Linked Open Data. In the university context, we discuss the promotion of the institutional Knowledge Management with Linked Open Data, considering the Scientometrics field. Based on the Semantic Web best practices, mainly, on the Linked Data Lifecycle process, two datasets are maintained on the Web of Data: a) the historical data of the Qualis Index (2005-2015); and b) a set of periodical publications records extracted from the Lattes Platform. As a result, three case studies are developed, which consume and explore data from these sources. Through these studies, it is concluded that the practices related to Linked Open Data contribute for developing institutional knowledge bases, which can support bibliometric and scientometric studies or some of the Knowledge Management activities in Brazilian universities.*

**Keywords:** *Knowledge Management; Linked Open Data; Scientometrics; Semantic Web; Organization Productivity.*

Recebido em 03 08 2016 Aceito em 19 09 2017

## 1 Introdução

As universidades são organizações consumidoras, produtoras e disseminadoras de conhecimento. Considerando as áreas do conhecimento<sup>1</sup> difundidas pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPQ, 2016), organizar, formalizar e compartilhar indicadores sobre o conhecimento produzido e disseminado é uma tarefa desafiadora a estas instituições.

Em uma universidade, uma das formas de gerir alguns dos indicadores internos é qualificar a produção de artigos científicos de seus professores pesquisadores. Em geral, estes pesquisadores são instigados a produzir conhecimento, disseminar sua produção por meio da publicação de artigos e registrar as disseminações em seus curriculum.

Neste enredo, duas bases de dados abertos são promovidas pelos órgãos de pesquisa, desempenhando papel fundamental: a Plataforma Lattes (CNPQ, 2016) e a Plataforma Sucupira (SUCUPIRA, 2016) - anteriormente Sistema WebQualis (CAPES, 2013). No contexto das instituições de pesquisa, relacionar os dados dessas plataformas fomenta uma imensa base de conhecimento que pode ser explorada no contexto da Gestão do Conhecimento.

Embora, os dados das plataformas Lattes e Sucupira são abertamente disponíveis, integrá-los requer um esforço computacional considerável para compatibilizar seus formatos. Para contornar esse desafio, atualmente, são difundidos os Dados Abertos Conectados (Linked Open Data). Metodologicamente, estes se baseiam em práticas para publicar e reutilizar dados abertos na Web de Dados.

Diante dessa potencialidade, o objetivo deste trabalho é investigar o alinhamento das práticas de Dados Abertos Conectados para o tratamento, a organização, o cruzamento e a exploração de dados aderentes aos estudos da Cientometria e da Gestão do Conhecimento em universidades brasileiras. Neste sentido, são apresentados três estudos de caso, nos quais os Dados Abertos Conectados formalizam e relacionam dados abertos das plataformas Lattes e Sucupira, produzindo conhecimento contextualizado a uma universidade pública.

Para tanto, além desta seção introdutória, este artigo aborda: i) a fundamentação teórica, estabelecendo o entendimento dos conceitos Gestão do Conhecimento, Cientometria e Dados Abertos Conectados; ii) os materiais e métodos utilizados, principalmente, apontando o processo metodológico na publicação de Dados Abertos Conectados, os conjuntos de dados abertos considerados e os vocabulários usados para representar os dados abertos na web; iii) os esforços na publicação dos Dados Abertos Conectados conforme o procedimento metodológico adotado; iv) os estudos de caso no consumo de dados abertos cientométricos no contexto

---

<sup>1</sup> Atualmente são 49 áreas de conhecimento difundidas pelo referido conselho. Dentre os exemplos de áreas de conhecimento, têm-se: (i) Administração Pública e de Empresas, Ciências Contábeis e Turismo; (ii) Comunicação e Informação; ou (iii) Interdisciplinar.

da Gestão do Conhecimento; v) a síntese dos resultados, discutindo a interdisciplinaridade dos conceitos Gestão do Conhecimento, Cientometria e Dados Abertos Conectados; e vi) as considerações finais e os trabalhos futuros.

## 2 Fundamentação teórica

No âmbito deste trabalho, três conceitos se relacionam: Gestão do Conhecimento, Cientometria e Dados Abertos Conectados. Estes conceitos são brevemente discutidos, considerando as universidades brasileiras como o ambiente de aplicação.

### 2.1 Gestão do conhecimento

Como domínio, constata-se que a Gestão do Conhecimento é recente, sendo suas raízes definidas a pouco menos de três décadas. Trata-se de um domínio multidisciplinar, com os aportes originados e não limitados às disciplinas das Ciência da Computação, Ciência da Informação, Ciências Organizacionais e Ciências Cognitivas (WIIG, 2002; DALKIR, 2011). Diante dessa concepção, existe uma variedade de visões para conceituar a Gestão do Conhecimento (STEFANO *et al.*, 2014; GONZALEZ; MARTINS, 2015). Pontualmente, esta seção destaca a dimensão geográfica da Gestão do Conhecimento, apresentando resumidamente duas visões de mundo para o termo, a japonesa e a americana.

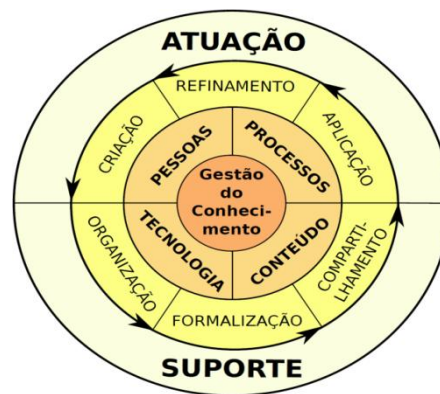
A visão japonesa da Gestão do Conhecimento tem suas raízes baseadas nas Ciências Organizacionais e Ciências Cognitivas. Sua visão fundamenta-se na premissa do indivíduo como o indutor principal da transformação do conhecimento tácito em conhecimento explícito, ao comunicar-se com um coletivo, considerando um fluxo contínuo de socialização, externalização, combinação e internalização de elementos de conhecimento (NONAKA, 1994). Neste sentido, Melhores Práticas, Mentoria, Mapas de Conhecimento, Comunidades de Prática estão entre os Instrumentos de Gestão do Conhecimento (MAIER, 2007) utilizados para facilitar a comunicação organizacional nas atividades intensivas em conhecimento.

Já a visão americana da Gestão do Conhecimento, originalmente, baseia-se na codificação dos elementos de conhecimento mediante o uso de tecnologias (WALTZ, 2003). Constitutivamente, a referida escola tem os aportes advindos das Ciência da Informação e Ciência da Computação para definir os processos e as tecnologias, permitindo ao indivíduo o acesso aos conteúdos de conhecimento de forma customizada (conteúdo para determinada pessoa, no local certo e em tempo oportuno). Nessa perspectiva, tecnologicamente, encoraja-se o uso de Instrumentos da Gestão do Conhecimento (MAIER, 2007) como Portais Corporativos,

Sistemas de Gestão de Conteúdo, Sistemas Baseados em Conhecimento, Sistemas de Apoio à Decisão, entre outros.

Aderente à visão americana, este trabalho considera que a Gestão do Conhecimento envolve um processo que, como uma cadeia de valor, parte-se da matéria-prima (dados primários) em direção a produtos (conhecimento). O processo envolve aquisição (de dados), triagem, filtragem, indexação e organização (informação), raciocínio (análise e síntese) para criação e disseminação do conhecimento. Neste contexto, conforme é representado na Figura 1, para efetivar a Gestão do Conhecimento (parte central da figura), quatro dimensões devem ser harmonizadas, considerando os processos de atuação e de suporte no ciclo de vida do conhecimento. As referidas dimensões são:

Figura 1 – Arcabouço conceitual da Gestão do Conhecimento – dimensões perante o ciclo de vida do conhecimento



Fonte: Os autores.

a) pessoas: a Gestão do Conhecimento deve considerar as culturas e as estruturas que permeiam a criação de conhecimento em uma organização (INAZAWA, 2009). Em um ambiente profícuo, as pessoas cooperam mediante suas capacidades de integração do conteúdo informacional e de ação (FUKUNAGA *et al.*, 2016; MACEDO *et al.*, 2017). Neste sentido, o indivíduo é o agente principal da Gestão do Conhecimento, cabendo a ele a criação, a aplicação e o refinamento do conhecimento (NISSEN, 2006), também utilizando seu conhecimento tácito envolto de aprendizagem, discernimento e experiência;

b) tecnologia: se as pessoas desempenham o papel principal em tarefas intensivas em conhecimento (FILIPPIM; LIMA, 2014; STEFANO *et al.*, 2014), à tecnologia reserva-se o papel coadjuvante de suporte às pessoas, expandindo as possibilidades de criação, aplicação e refinamento do conhecimento (SILVA; SOUZA, 2015). Para tanto, a tecnologia

é incorporada em Instrumentos de Gestão do Conhecimento (MAIER, 2007), promovendo a cooperação entre as pessoas. Ou seja, nas perspectivas de Nissen (2006) e de Maier (2007), um Instrumento de Gestão do Conhecimento é utilizado para organizar, formalizar e compartilhar o conteúdo, propriamente, os elementos de conhecimento.

c) conteúdo: na Gestão do Conhecimento, o elemento estrutural básico é denominado elemento de conhecimento (MAIER, 2007). Resumidamente, um elemento de conhecimento consiste em um pacote atômico de conteúdo que pode ser mantido via tecnologias e disseminado às pessoas com o propósito da aplicação do conhecimento. Exemplos de elementos de conhecimento são: um documento descrevendo uma lição aprendida ou uma boa prática, uma mensagem eletrônica entre os indivíduos, um arquivo de áudio ou vídeo, uma planilha eletrônica, um relatório, uma patente, uma entrada em um fórum ou sistema de informação, dentre outros;

d) processos: organizacionalmente, na Gestão do Conhecimento adota-se um ciclo de vida do conhecimento, em que seus processos subsidiam as pessoas nas atividades intensivas em conhecimento. É oportuno destacar que existem diversos ciclos de vida do conhecimento sugeridos na literatura. Neste sentido, Supyuenyong e Islam (2006) e Hädrich (2008) revisam alguns desses ciclos, evidenciando como os pesquisadores renomados da Gestão do Conhecimento rotulam os diferentes processos para a criação, a organização, a disseminação e a utilização do conhecimento. Pontualmente, neste artigo adota-se o ciclo de vida do conhecimento proposto por Nissen (2006). Em sua obra, o referido autor enumera seis processos de conhecimento, caracterizados como de atuação (processos reservados às pessoas – vide a parte superior da Figura 1) e de suporte (com uso intensivo das Tecnologias de Informação e Comunicação – parte inferior da Figura 1). Os seis processos são:

criação: é a fase inicial do ciclo de vida do conhecimento, na qual o conhecimento é gerado por uma pessoa,

organização: diz respeito ao uso de Tecnologias da Informação e Comunicação para mapear os elementos de conhecimento, empregando taxonomias ou ontologias, por exemplo,

formalização: é o registro de novos elementos de conhecimento na base de conhecimento organizacional,

privilegiando as formas de codificação e de armazenamento computacional,

compartilhamento: utiliza os meios de acesso, distribuição ou transferência de elementos de conhecimento, geralmente, utilizando como meio as redes de computadores (intranets ou *web*),

aplicação: é a utilização do conhecimento pelas pessoas para a resolução de problemas ou tomada de decisão em seu cotidiano,

refinamento: é a evolução do conhecimento realizada pelas pessoas, refletindo os aspectos da aprendizagem organizacional ao longo do tempo.

Diante a apresentação da visão, das dimensões e de seus processos, entende-se a Gestão do Conhecimento como a gestão de elementos do conhecimento, apoiando as pessoas com os Instrumentos da Gestão do Conhecimento (MAIER, 2007) nos processos de criação, organização, formalização, compartilhamento, aplicação e refinamento do conhecimento (NISSEN, 2006).

Nesse entendimento, considerando as universidades e seu tripé de atuação (ensino, pesquisa e extensão), admite-se que parte da Gestão do Conhecimento nas referidas instituições perpassa pela organização, formalização e uso de dados cientométricos, mediante a utilização de novos paradigmas tecnológicos. Para basilar este entendimento, a seguir discorre-se sobre a Cientometria como uma disciplina da Ciência da Informação, interdisciplinar à Gestão do Conhecimento, que suporta a formalização de alguns elementos de conhecimento inerentes às universidades.

## 2.2 Cientometria

No escopo deste trabalho, outro construto pertinente é a Cientometria. Cientificamente, no domínio da Ciência da Informação, a Cientometria surge em 1969, ano em que os autores russos Nalimov e Mul'chenko publicaram a obra "*Scientometrics. The Study of Science as an Information Process*" (GARFIELD, 2009). Desde sua origem, o entendimento do referido conceito reserva-se à pesquisa quantitativa da ciência e da tecnologia (van RAAN, 1997). De forma didática, Santos e Kobashi (2009, p. 159) pontuam que:

[...] A Cientometria preocupa-se com a dinâmica da ciência, como atividade social, tendo como objetos de análise a produção, a circulação e o consumo da produção científica [...].

Metodologicamente, os estudos cientométricos utilizam os registros de (STOCK; STOCK, 2015): documentos científicos (resumos, artigos ou livros), autores de comunicações científicas, periódicos, congressos, dados

socioeconômicos, instituições científicas, dentre outros, como seus dados primários em busca de respostas às questões que mensuram a informação científica/tecnológica. Neste sentido, particularmente, o ato de mensurar enseja a utilização dos ferramentais da Matemática e da Estatística (VANTI, 2011) para realizar inferências a partir dos conjuntos de dados científicos disponíveis.

Ao se considerar o presente, é notório que se presencia a massificação da produção de dados e informação e, por conseguinte, se experimenta maior apropriação da comunicação científica. Muito em decorrência do uso da Internet como plataforma global de compartilhamento das comunicações científicas. Por isso, os métodos cientométricos são úteis para entender as dinâmicas do relacionamento da ciência e da tecnologia (SILVA; BIANCHI, 2001) no contexto da Sociedade da Informação e do Conhecimento.

Neste sentido, as pesquisas cientométricas são desenvolvidas para: (i) propor o uso de 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. Dentre as aplicações citadas, o presente trabalho visa mapear o conhecimento de uma universidade pública brasileira, organizando os dados primários em consonância às boas práticas dos Dados Abertos Conectados (conforme é apresentado na seção 5). Contudo, manipular os dados primários, subsidiando os estudos cientométricos, pode ser uma tarefa complexa (SANTOS; KOBASHI, 2009). A referida complexidade envolve conhecimentos diversos, em especial de computação, visto que existem desafios na coleta, na organização e no relacionamento de dados pertinentes. Tais desafios são potencializados porque os dados podem estar distribuídos em várias fontes, apresentados em formatos incompatíveis ou proprietários, dificultando sua manipulação.

No contexto das universidades, por exemplo, duas fontes de dados abertos são amplamente difundidas e disponibilizadas na Internet: a Plataforma Lattes (CNPQ, 2016) e a Plataforma Sucupira (SUCUPIRA, 2016). Na Plataforma Lattes, armazena-se os dados sobre pesquisadores e suas respectivas comunicações científicas. Já na Sucupira, os índices de qualidade de algumas dessas comunicações estão estratificados, formando o Índice Qualis. Interligando essas fontes, é possível quantificar e qualificar algumas informações cientométricas alinhadas à Gestão do Conhecimento. Dentre os potenciais insumos a se explorar deste relacionamento estão:

- a) considerando todos os colaboradores, em quais áreas do conhecimento demonstra-se maior competência?
- b) quais dos pesquisadores cooperam, dada uma área de conhecimento?



c) quais pesquisadores poderiam colaborar para incrementar os indicadores científicos da instituição?

Numa visão geral, tais questões podem nortear as ações estratégicas ou o entendimento institucional da evolução das pesquisas, considerando os pesquisadores individuais, os grupos de pesquisa, as áreas do conhecimento, os programas de pós-graduação, entre outras dimensões.

Corroborando a complexidade apontada por Santos e Kobashi (2009), aferir informações a partir das Plataformas Lattes e Sucupira é uma tarefa custosa. Nessas plataformas, os dados estão disponíveis em formatos distintos. Originalmente, os currículos Lattes estão em páginas de Internet e o Índice Qualis, historicamente, é acessado nos formatos PDF (*Portable Document Format* - Formato Portátil de Documento) ou XLS (*eXcel Spreadsheet* - formato de planilha eletrônica da Microsoft). Isso dificulta a extração, o cruzamento dos dados e a exploração de um relacionamento intrínseco.

Para contornar tais dificuldades, a *Web Semântica* oferece o suporte dos Dados Abertos Conectados - *Linked Open Data* (AKSW, 2016). Metodologicamente, os Dados Abertos Conectados se baseiam em um conjunto de melhores práticas para organizar, publicar, conectar e compartilhar dados na *web* (LINKED DATA, 2016) de forma aberta e transparente.

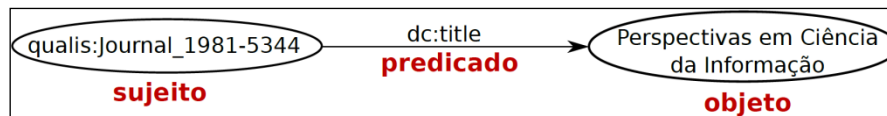
## 2.3 Dados Abertos Conectados

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. Constitutivamente, esta percepção é vinculada a dois entendimentos: a) o que são dados abertos; e b) como os dados são conectados.

Dados são considerados abertos quando “podem ser livremente usados, reutilizados e redistribuídos por qualquer pessoa - sujeitos, no máximo, à exigência de atribuição da fonte e compartilhamento pelas mesmas regras” (OPEN KNOWLEDGE INTERNATIONAL, 2016).

Na *web*, os dados abertos podem estar conectados a outros dados, constituindo os Dados Abertos Conectados. Porém, isto somente é possível com a utilização de um modelo padrão para relacionar os dados de diversas origens. Sob este prisma, tem-se o *Resource Description Framework* (RDF), uma linguagem que usa um modelo padrão para conectar dados na *web*. Em suma, o modelo se baseia em triplas, as quais realizam a descrição de um recurso em três partes (sujeito → predicado → objeto), relacionando um sujeito a um objeto através de um predicado.

Figura 2 – Representação de uma tripla RDF



Fonte: Os autores.

Como exemplo, a

Figura 2 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 acerca de um assunto gera uma grande estrutura de ligações, formando um grafo RDF. Neste sentido, o exemplo expresso na

Figura 2 faz parte de um grafo, o qual é discutido na subseção "3.2.1 QualisBrasil".

Ressalta-se que os dados abertos são classificados de acordo com seu nível de abertura e 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 em que estão publicados. Assim, estes dados podem ser lidos, impressos, armazenados, modificados, compartilhados ou usados como dados de entrada em outros sistemas.

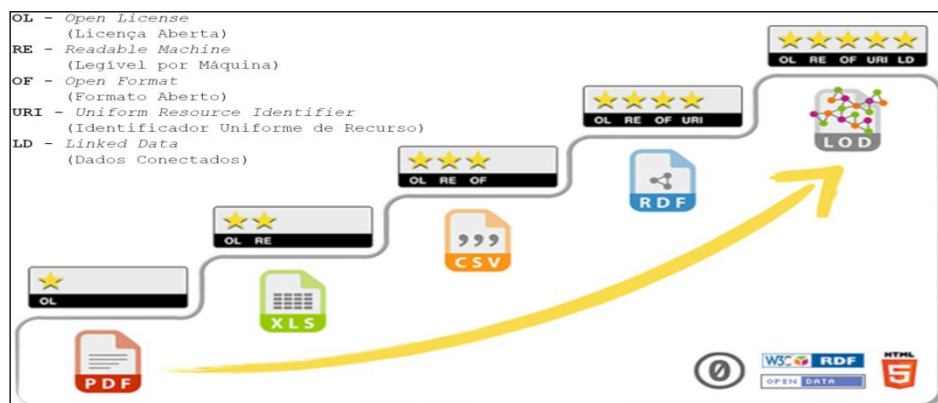
2ª Estrela - é conferida à publicação de dados estruturados legíveis por máquinas (*Readable Machine* - RE). Os dados podem ser diretamente processados por *softwares* proprietários e podem ser convertidos para outros formatos.

3ª Estrela - é concedida aos dados que são publicados em formato aberto não proprietário (*Open Format* - OF). Neste patamar, a manipulação dos dados é realizada sem a necessidade do uso de um *software* proprietário.

4ª Estrela - é designada à utilização de Identificadores Uniforme de Recursos (*Uniform Resource Identifier* - URI) para nomear os dados, permitindo que outros usuários criem ligações e façam reuso dos dados disponibilizados.

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.

Figura 3: Representação da Classificação 5-Estrelas



Fonte: Adaptado de 5-STAR (2016).

Mediante a classificação anteriormente citada, entende-se que o estabelecimento dos Dados Abertos Conectados é alcançado no patamar da 5ª Estrela. Esse estabelecimento fomenta um imenso grafo RDF globalmente disponível, a *Web* de Dados. Neste sentido, a publicação de Dados Abertos Conectados tem como objetivo usar a *web* para promover dados estruturados em uma escala global. Dessa forma, incentiva-se o (re)uso de um conjunto de dados universal por diferentes pessoas e aplicações ao redor do mundo.

### 3 Materiais e métodos

Para subsidiar a realização de estudos de caso no âmbito da Gestão do Conhecimento de universidades brasileiras com dados cientométricos, esta seção aborda: i) o processo metodológico utilizado para publicação dos Dados Abertos Conectados; ii) os conjuntos de dados abertos considerados no domínio da Cientometria; iii) as ontologias e os vocabulários utilizados para representar os dados abertos na *web*; iv) as ferramentas tecnológicas utilizadas para publicar os Dados Abertos Conectados; e v) a consulta basilar para consumir os Dados Abertos Conectados nos estudos de caso.

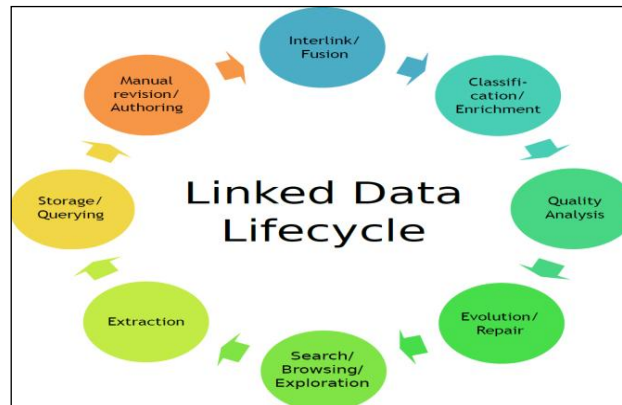
#### 3.1 *Linked Data Lifecycle*: o ciclo de vida para dados conectados

Ilustrado na Figura 4, o *Linked Data Lifecycle* é um processo metodológico que contempla um conjunto atividades para publicar dados conectados na *web*. É um processo difundido pelo Instituto de Pesquisa *Agile Knowledge and Semantic Web*<sup>2</sup>, o qual também contribui no

<sup>2</sup> Disponível em: <<http://aksw.org>>. Acesso em: 9 set. 2017.

desenvolvimento de ferramentas computacionais de suporte à *Web* de Dados (AUER, 2014). Seu processo compreende oito atividades:

Figura 4: Processo metodológico Linked Data Lifecycle



Fonte: AUER, 2014.

- a) *Extraction* (Extração) – é o mapeamento dos dados não-estruturados ou estruturados e em diferentes formatos para o modelo RDF.
- b) *Storage/Querying* (Armazenamento/Consulta) - é a disponibilização dos dados RDF com o uso de sistemas gerenciadores (*Triple Stores*) para potencializar as tarefas de publicação e de consumo de dados.
- c) *Manual Revision/Authoring* (Revisão Manual/Autoria) – engloba as tarefas de criação ou editoração de novos dados em um grafo RDF.
- d) *Interlink/Fusion* (Interligação/Fusão) – estabelece o relacionamento de dados de um grafo RDF para com os dados de outra fonte para explorar novas informações.
- e) *Classification/Enrichment* (Classificação/Enriquecimento) - viabiliza a expressividade e a riqueza semântica de um conjunto de dados em relação a um contexto, representando os dados mediante o uso de ontologias ou vocabulários.
- f) *Quality Analysis* (Análise de Qualidade) – pontualmente, trata dos aspectos de integridade, precisão, consistência e validade de dados. E de forma geral, verifica os requisitos de compreensibilidade, disponibilidade e proveniência do modelo de dados.
- g) *Evolution/Repair* (Evolução/Reparação) - uma vez encontradas inconsistências nos dados ou no modelo de

representação, ações corretivas devem ser tomadas a fim de corrigir as não-conformidades.

- h) *Search/Browsing/Exploration* (Busca/Navegação/Exploração) - usa técnicas de busca, navegação, exploração para visualizar os dados RDF.

Na publicação de Dados Abertos Conectados, tais atividades são abordadas de forma incremental, sendo combinadas conforme os desafios encontrados no ambiente de aplicação. Tais desafios incluem:

- a) o mapeamento e a publicação de dados de sistemas legados com vocabulários amplamente utilizados na *Web de Dados*;
- b) o reparo dados com a finalidade de garantir a qualidade na publicação dos dados; ou
- c) o compartilhamento de dados locais, conectando-os a demais dados na *web*.

## 3.2 Conjuntos de dados abertos utilizados

Neste trabalho são considerados dois conjuntos de dados abertos: a) o histórico do Índice Qualis (grafo QualisBrasil); e b) os registros de artigos publicados em periódicos de uma universidade (grafo LattesProduction). Tais conjuntos de dados são apresentados a seguir.

### 3.2.1 QualisBrasil

O índice Qualis foi coletado ao longo dos últimos dez anos, a partir do Sistema WebQualis (CAPES, 2013) e a Plataforma Sucupira (SUCUPIRA, 2016). A Tabela 1 resume a coleta e o pré-processamento dos dados, associando: um período de referência para construção de histogramas, a fonte de dados, o formato e as tuplas validadas. Cabe ressaltar que a publicação do índice Qualis como Dados Abertos Conectados é um esforço constantemente realizado pelos autores, iniciado em 2014 (RAUTENBERG *et al.*, 2014; RAUTENBERG; BURDA, 2016).

Tabela 1 – Captura e pré-processamento do índice Qualis nos últimos dez anos

Ano de Coleta	Período referência	Origem	Formato	Tuplas validadas
2007	2005-2007	WebQualis	XLS	35.020
2009	2008-2010	WebQualis	PDF	54.233
2013	2011-2013	WebQualis	PDF	107.429
2015	2014	Internet <sup>3</sup>	XLS	108.622
2016	2015	Sucupira	XLS	44.463

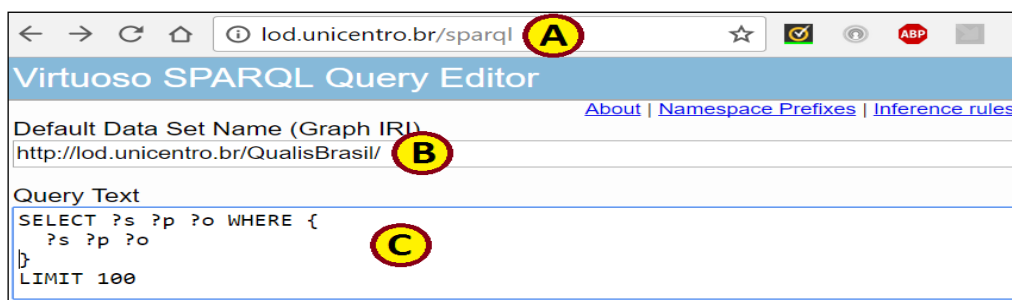
Fonte: RAUTENBERG; BURDA (2016).

3 NIEVINSKI, F. G. [ciência aberta] Planilha Qualis (em anexo). [mensagem eletrônica]. Disponível em: <<https://lists.okfn.org/pipermail/cienciaaberta/2014-October/000559.html>>. Acesso em: 2 fev. 2015.

Considerando a Classificação 5-Estrelas, originalmente, o Índice Qualis encontrava-se nas 1ª ou 2ª Estrelas (nos formatos PDF e XLS, respectivamente). Isso incute algumas limitações. Por exemplo, os índices coletados em 2009 e 2013 estavam disponíveis em formato proprietário, ou seja, na 1ª Estrela. Tal fato exige um esforço considerável para extrair, converter e utilizar seus dados a cada nova aplicação/estudo. Outro fato a considerar é a dimensão tempo. Neste quesito, a série histórica do Índice Qualis não está disponível em sua totalidade. O Sistema WebQualis não é mais acessível e somente parte do histórico é recuperado a partir da Plataforma Sucupira. Isso dificulta a realização de estudos mais fidedignos, considerando a temporalidade na classificação das publicações em periódicos.

Neste sentido, como contribuição adicional, constantemente, eleva-se os dados do referido índice à 5ª Estrela, mantendo os dados como Dados Abertos Conectados. Isso permite o reuso facilitado do Índice Qualis atualizado por outras pessoas em outros estudos.

Figura 5 – Acessando os Dados Abertos Conectados do Índice Qualis



Fonte: Os autores.

Conforme a Figura 5, os Dados Abertos Conectados referentes ao Índice Qualis podem ser acessados: (A) a partir do *endpoint* `http://lod.unicentro.br/sparql`; (B) no grafo `http://lod.unicentro.br/QualisBrasil/`; e (C), na interface disponibilizada, de acordo com uma consulta escrita em linguagem SPARQL<sup>4</sup> (acrônimo de *SPARQL Protocol And RDF Query Language*), os dados podem ser retornados em diversos formatos não-proprietários, minimizando os esforços de extração e manipulação.

### 3.2.2 LattesProduction

A Plataforma Lattes é um sistema de informação integrado mantido pelo Ministério da Ciência, Tecnologia e Inovação do Brasil (CNPQ, 2016). Neste sistema, o *Curriculum Lattes* (CVLattes) é o documento principal de processamento, tornando públicas as informações de pesquisadores.

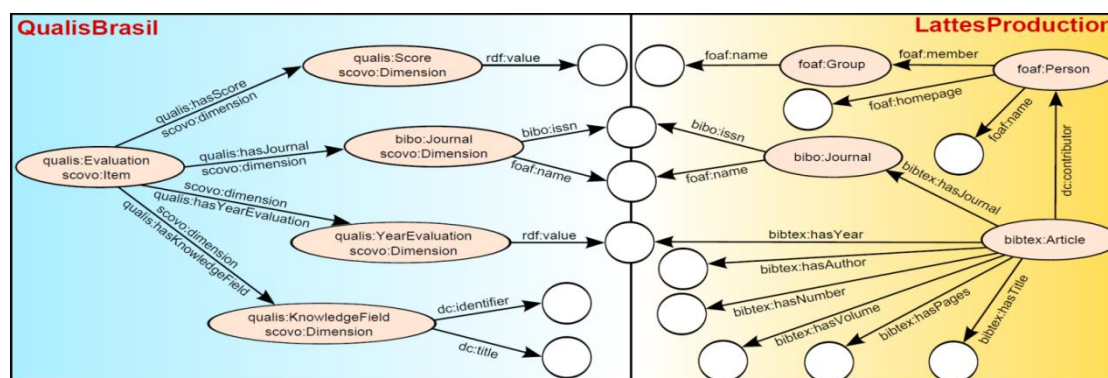
4 Disponível em: <<https://www.w3.org/TR/rdf-sparql-query>>. Acesso em: 9 set. 2017.

Apesar dessa publicidade, extrair dados desses documentos é uma atividade custosa. Atualmente, a Plataforma Lattes utiliza CAPTCHA (*Completely Automated Public Turing test to tell Computers and Humans Apart*). Isso dificulta a tarefa de empregar extratores automáticos, como por exemplo, o ScriptLattes (MENA-CHALCO; CESAR-JR, 2009), para capturar os dados abertos. Entretanto, com uma requisição formal, universidades brasileiras podem acessar uma visão dos CVLattes de seus pesquisadores para atualizar seus sistemas legados. Neste contexto, o Sistema de Avaliação e Acompanhamento de Programas Institucionais (SAAPI) é um desenvolvimento tecnológico interno de uma universidade brasileira. Resumidamente, O SAAPI acessa os dados provenientes da Plataforma Lattes, sincronizando-os para com outros sistemas legados. Isto constitui uma forma automatizada e alternativa para acessar um conjunto de dados abertos provenientes de diversos CVLattes. Ao utilizar o SAAPI, o subconjunto de registros de publicações em periódicos de colaboradores foi capturado para, posteriormente, ser publicado como Dados Abertos Conectados conforme a 5ª Estrela.

### 3.3 Ontologias e vocabulários usados na representação dos dados abertos

Para elevar os dados abertos do Índice Qualis e da Plataforma Lattes à 5ª Estrela, alguns vocabulários disponíveis na *web* são utilizados. A Figura 6 ilustra como os dados são representados no modelo RDF. Neste sentido, o modelo RDF é baseado nos seguintes vocabulários ou ontologias:

Figura 6 – Representação dos grafos QualisBrasil e LattesProduction



Fonte: Os autores.

- a) SCOVO<sup>5</sup> (*The Statistical CORE VOcabulary*) - é um vocabulário simples para representar dados estatísticos na

<sup>5</sup> Disponível em: <<http://vocab.der.i.e/scovo>>. Acesso em: 9 set. 2017.



*web*. Neste trabalho, é usado para organizar o Índice Qualis na forma multidimensional;

- b) DC<sup>6</sup> (Dublin Core) – é um vocabulário amplamente utilizado para descrever recursos. É utilizado para: a) melhor representar as áreas de conhecimento no grafo QualisBrasil (elementos *dc:identifier* e *dc:title*); e b) relacionar um indivíduo a um artigo científico como um coautor (*dc:contributor*);
- c) BIBO<sup>7</sup> (*Bibliographic Ontology Specification*) – é uma ontologia que modela os conceitos e as propriedades de referências bibliográficas. Seus elementos são usados para representar os periódicos (*Journals*) nos grafos;
- d) FOAF<sup>8</sup> (*Friend-of-a-Friend*) – é um vocabulário utilizado para relacionar entidades a informações na *web*. No grafo LattesProduction, por exemplo, mapeia grupos (cursos, departamentos ou centros) a seus membros;
- e) BIBTEX<sup>9</sup> (*Transformation of bibTeX into an OWL ontology*) – é uma ontologia que define os elementos de referências bibliográficas. É usada para mapear as referências capturadas da Plataforma Lattes.

### 3.4 Ferramentas utilizadas na publicação dos dados abertos

Para automatizar a publicação dos dados abertos na *Web* de Dados, são utilizadas as seguintes ferramentas:

- a) Sparqlify<sup>10</sup> – a característica desta ferramenta é, com base em um arquivo de mapeamento, converter dados primários para triplas RDF. Ressalta-se que no estudo, os dados primários são capturados de bases legadas, no formato CSV. Ou seja, a Sparqlify efetua a passagem dos dados da 3ª à 4ª Estrela;
- b) LIMES<sup>11</sup> – é uma ferramenta que relaciona dados entre um grafo de origem e outro de destino. No estudo, a LIMES é usada para interligar os identificadores de periódicos do grafo QualisBrasil (*bibo:ISSN*) com o grafo DBpedia<sup>12</sup>, potencializando a navegação e a descoberta de informações

<sup>6</sup> Disponível em: <<http://dublincore.org/documents/dcmi-terms/>>. Acesso em: 9 set. 2017.

<sup>7</sup> Disponível em: <<https://github.com/structuredynamics/Bibliographic-Ontology-BIBO/blob/master/bibo.owl>>. Acesso em: 9 set. 2017.

<sup>8</sup> Disponível em: <<http://xmlns.com/foaf/spec/>>. Acesso em: 9 set. 2017.

<sup>9</sup> Disponível em: <<http://zeitkunst.org/bibtex/0.1/>>. Acesso em: 9 set. 2017.

<sup>10</sup> Disponível em: <<http://aksw.org/Projects/Sparqlify.html>>. Acesso em: 9 set. 2017.

<sup>11</sup> Disponível em: <<http://aksw.org/Projects/LIMES.html>>. Acesso em: 9 set. 2017.

<sup>12</sup> Um esforço conjunto para extrair informação estruturada da Wikipedia. Disponível em: <<http://dbpedia.org>>. Acesso em: 9 set. 2017.

relacionadas. Com os dados do QualisBrasil interligados na *Web de Dados*, alcança-se a 5ª Estrela;

- c) Open Link Virtuoso<sup>13</sup> - é considerado um sistema universal para acesso, integração e gerenciamento de dados relacionais e/ou grafos baseados no modelo RDF. O Virtuoso é usado para armazenar e disponibilizar os grafos QualisBrasil e LattesProduction (conforme a Figura 5).

### 3.5 A consulta basilar para consultar os dados abertos

Para consultar os dados abertos dos grafos QualisBrasil e LattesProduction, uma consulta em linguagem SPARQL é codificada, conforme a Listagem 1.

Listagem 1 – Consulta basilar a ser reformulada nos estudos de caso

```
01 PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
02 PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
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 PREFIX bibtex: <http://purl.org/net/nknouf/ns/bibtex/>
07 PREFIX prod: <http://lod.unicentro.br/LattesProduction/>
08 PREFIX qualis: <http://lod.unicentro.br/QualisBrasil/>
09
10 SELECT
11   ?qualisYearEvaluationValue AS ?ano
12   ?qualisKnowledgeFieldTitle AS ?areaConhecimento
13   ?qualisScoreValue
14   COUNT(*) AS ?qtdeArtigos
15 WHERE {
16   ?evaluation rdf:type qualis:Evaluation .
17   ?evaluation qualis:hasJournal ?qualisJournal .
18   ?evaluation qualis:hasYearEvaluation ?qualisYearEvaluation .
19   ?evaluation qualis:hasKnowledgeField ?qualisKnowledgeField .
20   ?evaluation qualis:hasScore ?qualisScore .
21
22   ?qualisJournal bibo:issn ?qualisJournalId .
23   ?qualisYearEvaluation rdf:value ?qualisYearEvaluationValue .
24   ?qualisScore rdf:value ?qualisScoreValue .
25   ?qualisKnowledgeField dc:title ?qualisKnowledgeFieldTitle .
26
27   ?paper rdf:type prod:PeriodicalPaper .
28   ?paper bibtex:hasJournal ?paperJournal .
29   ?paper bibtex:hasYear ?qualisYearEvaluationValue .
30   ?paperJournal bibo:issn ?qualisJournalId .
31 }
32 GROUP BY ?qualisYearEvaluationValue ?qualisKnowledgeFieldTitle
33 ?qualisScoreValue
```

Fonte: Os autores.

Como característica, a referida consulta percorre os grafos QualisBrasil e LattesProduction da seguinte forma: a) são recuperados os recursos de uma avaliação Qualis (linhas 16-20); b) os dados referentes aos ISSN, ano, área de conhecimento, *score* Qualis dessas avaliações são adquiridos nas linhas 22-25; c) os recursos de artigos publicados são capturados nas linhas 27-30; e d) o relacionamento entre os anos e ISSN de uma avaliação Qualis para com um artigo é codificada nas linhas 29-

<sup>13</sup> Disponível em: <<http://virtuoso.openlinksw.com/>>. Acesso em: 9 set. 2017.

30. Como resultado, são contabilizadas as quantidades dos artigos publicados, considerando o ano, a área de conhecimento e o *score* Qualis, conforme é parcialmente visto na Tabela 2.

Tabela 2 – Sumarização e classificação das publicações em revistas de acordo com o ano, a área de conhecimento e score qualis

Ano	Área de Conhecimento	Score	# artigos
[...]	[...]	[...]	[...]
2015	INTERDISCIPLINAR	B1	75
2015	INTERDISCIPLINAR	B2	61
2015	CIÊNCIAS AGRÁRIAS I	B1	56
2015	ENGENHARIAS III	B5	50
2015	CIÊNCIAS AMBIENTAIS	B1	46
[...]	[...]	[...]	[...]

Fonte: Os autores.

Salienta-se que a consulta SPARQL apresentada é utilizada como base à customização nos estudos de caso apresentados na Seção “5 Estudos de caso: Gestão do Conhecimento em uma universidade brasileira com Dados Abertos Conectados”.

## 4 Publicando dados na *Web de Dados*

Para publicar os grafos QualisBrasil e LattesProduction, dois *workflows* são estabelecidos e executados como descritos a seguir:

- a) QualisBrasil: os dados primários são extraídos de uma base de dados legada e convertidos ao formato CSV (3ª Estrela). Em seguida, utiliza-se a ferramenta Sparqlify, convertendo a massa de dados para o formato RDF (4ª Estrela). Então, as triplas RDF são armazenadas no *Triple Store* Open Link Virtuoso. Para alcançar a 5ª Estrela, os recursos que representam os jornais/revistas são relacionados aos recursos PeriodicalPapers da DBpedia, utilizando a ferramenta LIMES. Como último passo deste *workflow*, os recursos de relacionamento aferidos no passo anterior são armazenados junto ao grafo QualisBrasil;
- b) LattesProduction: os dados primários sobre publicações em revistas/jornais são extraídos dos CVLattes de colaboradores de uma universidade pelo SAAPI, no formato CSV (3ª Estrela). Com a ferramenta Sparqlify, os dados em CSV são convertidos para triplas RDF (4ª Estrela). Por fim, com as triplas RDF é criado o grafo LattesProduction. A 5ª Estrela é alcançada intrinsecamente, visto que para cada registro de publicação é possível haver uma correspondência (dada pelos ISSN da revista/jornal e ano de publicação) com uma avaliação Qualis no grafo QualisBrasil.

## 5 Estudos de caso: Gestão do Conhecimento em uma universidade brasileira com Dados Abertos Conectados

Com os grafos QualisBrasil e LattesProduction publicados na *Web de Dados*, alguns estudos de caso cientométricos podem ser realizados. Sob o prisma da Gestão do Conhecimento, a publicação dos referidos grafos subsidia as atividades de organização, formalização e compartilhamento de elementos para criar, refinar ou aplicar conhecimento sobre fenômenos internos de uma universidade brasileira. Nesta seção, três estudos de caso reais são explorados: a) internamente, localizar os professores pesquisadores com determinado perfil de publicação para atuar em um programa de pós-graduação; b) dado um grupo de professores, aferir o perfil de publicação de artigos científicos de acordo com as áreas de conhecimento; e c) institucionalmente, entender a evolução das pesquisas de um grupo de professores de um departamento em face do investimento realizado em capital humano.

### 5.1 Localizando pesquisadores em uma universidade

Um modo de utilizar o relacionamento dos grafos QualisBrasil e LattesProduction é na forma de uma base de conhecimento em um Instrumento de Gestão do Conhecimento denominado *Expertise Locator* (MAIER, 2007). Considerando um Mestrado Profissional institucionalizado em 2014, com a base de conhecimento resultante da união dos grafos, pretende-se localizar os potenciais professores pesquisadores a se engajarem no corpo docente no programa de pós-graduação. Neste estudo de caso, o perfil de colaboradores se restringe àqueles que continuamente publicam nos periódicos bem ranqueados na Área do Conhecimento "Administração Pública e de Empresas, Ciências Contábeis e Turismo"<sup>14</sup>. Atualmente, a universidade em questão dispõe de cerca de 649 professores que publicaram cerca de 5.600 artigos científicos em periódicos nos anos 2005-2015. Diante desse número, a localização de especialistas se torna uma tarefa desafiadora, uma vez que, é necessário extrair os dados sobre publicações dos respectivos CVLattes e classificar as publicações de acordo com o Índice Qualis da referida área.

Tabela 3 – Ponderação dos artigos em periódicos na área Administração Pública e de Empresas, Ciências Contábeis e Turismo

	QUALIS							
	A1	A2	B1	B2	B3	B4	B5	C
Ponderação	100	80	60	40	30	20	10	0

Fonte: BRITO; LUCA; TEIXEIRA (2016).

<sup>14</sup> Originalmente, a área era denominada "Administração, Ciências Contábeis e Turismo" conforme a base de dados Qualis disponibilizada na Plataforma Sucupira no momento do desenvolvimento dos estudos de caso (SUCUPIRA, 2016).

Por isso, tomam-se os grafos QualisBrasil e LattesProduction como fontes de conhecimento. Em suma, são consideradas as publicações de 2012 em diante, representando a janela de um quadriênio de análise. Este conjunto de publicações é classificado de acordo com o Índice Qualis mais atual (ano referência 2015) na área de "Administração Pública e de Empresas, Ciências Contábeis e Turismo". As classificações são agrupadas por professor pesquisador, ponderando-as conforme a Tabela 3. Salienta-se que a referida tabela é advinda de um relatório da área, sendo continuamente utilizada nas avaliações trienais e quadriennais dos cursos de pós-graduação.

Tabela 4 – Seleção de possíveis professores para um curso de pós-graduação

<b>Professor/Pesquisador</b>	<b>Score</b>
Nome_01	1030
Nome_02	800
Nome_03	710
Nome_04	690
Nome_05	590
[...]	[...]

Fonte: Os autores.

Readequando a consulta SPARQL da Listagem 1, foi possível extrair da base de conhecimento uma lista de professores pesquisadores conforme a Tabela 4. Na tabela, o nome dos professores pesquisadores é omitido. Entretanto, os nomes foram apresentados ao coordenador do programa de pós-graduação, objetivando um convite de colaboração junto aos possíveis novos professores.

## 5.2 Avaliando as publicações um grupo *versus* áreas de conhecimento

Os grafos QualisBrasil e LattesProduction também podem ser utilizados para prospecção de novos cursos de pós-graduação, conforme neste estudo de caso. Neste sentido, um grupo de 12 professores almeja desenvolver uma proposta de abertura de curso a nível de mestrado.

Tabela 5 – Classificação dos artigos de um grupo de professores nas principais áreas

<b>Área de Conhecimento</b>	<b>A1</b>	<b>A2</b>	<b>B1</b>	<b>B2</b>	<b>B3</b>	<b>B4</b>	<b>B5</b>	<b>C</b>
INTERDISCIPLINAR	3	2	4	8	10	7	3	0
EDUCAÇÃO	2	0	1	4	1	0	7	4
ENGENHARIAS I	2	0	1	0	3	5	4	2
ENSINO	2	0	0	1	0	2	2	0

Fonte: Os autores.

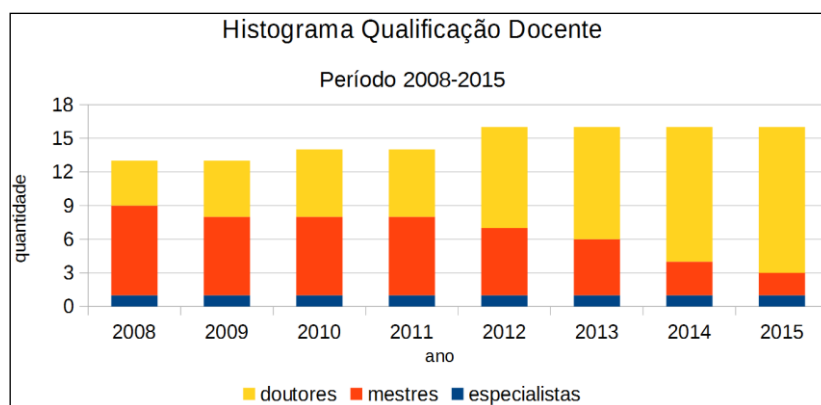
Com uma nova consulta SPARQL derivada da Listagem 1 foi possível qualificar as 45 publicações do referido grupo de pesquisadores em todas

as áreas do conhecimento. A Tabela 5 relaciona o retorno da consulta desenvolvida, a qual subsidiou as conversas iniciais na prospecção de um novo programa de pós-graduação.

### 5.3 Avaliando a evolução científica de um grupo

Nas universidades, a realização de pesquisa científica está intimamente ligada ao investimento realizado na formação de seu corpo docente. Como reflexo, almeja-se melhores condições de ensino, pesquisa e extensão. Neste estudo de caso, os grafos QualisBrasil e LattesProduction subsidiam a observação da capacidade científica do corpo docente de um departamento universitário ao longo do tempo.

Figura 7 – Qualificação docente ao longo dos anos, dados extraídos de



Fonte: CNPQ (2016).

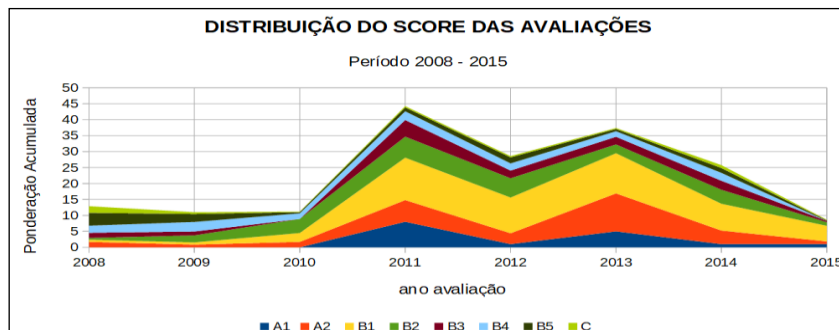
Conforme a Figura 7, admite-se que um investimento importante foi realizado na capacitação do corpo docente em questão. Atualmente, o departamento conta com 16 professores efetivos. Ao observar o histograma, em 2008 quatro de 13 professores figuravam como doutores (~31%). Em 2015, são 13 de 16 professores com a referida titulação (~81%).

Tabela 6 – Ponderação dos artigos em periódicos segundo Regulamento de Pesquisa

	QUALIS							
	A1	A2	B1	B2	B3	B4	B5	C
Ponderação	1.0	0.85	0.70	0.55	0.40	0.25	0.10	0.05

Fonte: UNIVERSIDADE ESTADUAL DO CENTRO-OESTE (UNICENTRO, 2011).

Figura 8 – Avaliação dos artigos publicados por pesquisadores de um departamento, utilizando os grafos QualisBrasil e LattesProduction



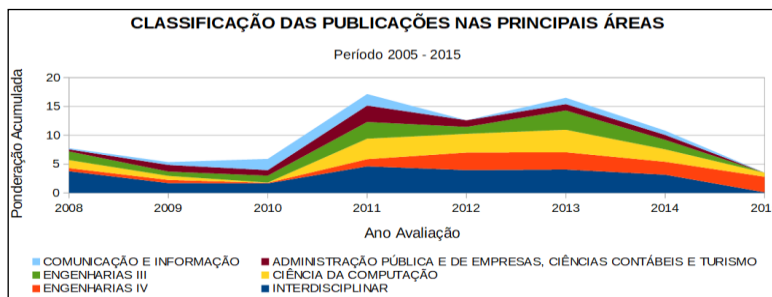
Fonte: Os autores.

Ao traçar um entendimento do investimento realizado frente à consultas dos grafos LattesProduction e QualisBrasil, nota-se que o aumento da capacitação docente se relaciona positivamente ao período de maior produção qualificada de artigos em periódicos, como pode ser percebido na

Figura 8. Na referida figura está representada a classificação do conjunto de artigos publicados pelo grupo de docentes com o índice Qualis (independentemente de áreas de conhecimento), ponderando as classificações conforme as regras do Regulamento Interno de Pesquisa da universidade em investigação (

Tabela 6). Historicamente, independente das áreas de conhecimento, observa-se que a evolução na quantidade de publicações nos estratos Qualis mais importantes (A1, A2 e B1) está ligada ao aumento da quantidade de doutores. Conjectura-se que os novos doutores obtiveram êxito na disseminação do conhecimento produzido durante o processo de doutoramento nos periódicos de maior impacto científico. A diante, o conjunto de artigos publicados é reclassificado conforme as áreas de conhecimento. Essa classificação é sumarizada e representada na Figura 9.

Figura 9 – Avaliação da qualidade dos artigos publicados de professores de um departamento, de acordo com o Regulamento Interno de Pesquisa



Fonte: Os autores.

Na

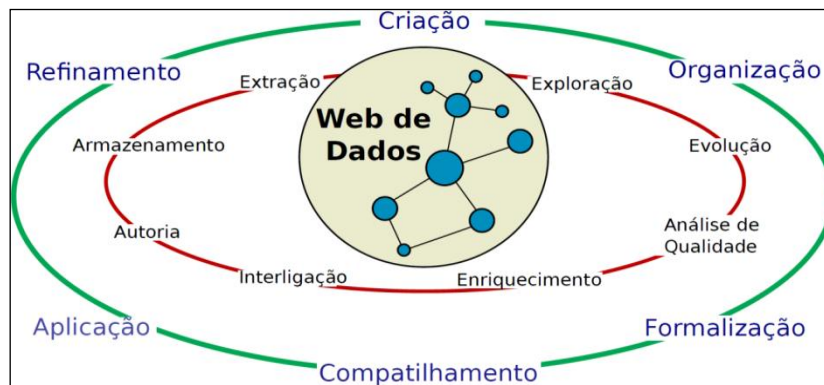
Figura 9, percebe-se que as contribuições científicas decorrentes da formação daquele corpo docente se concentram nas áreas tecnológicas. Destaca-se que o referido grupo de pesquisadores atua em um curso de Bacharelado em Ciência da Computação. No gráfico apresentado, a área Ciência da Computação é demarcada na cor amarela. E considerando o investimento concretizado *versus* publicações na referida área, constata-se a adequação das pesquisas realizadas frente a atuação do corpo docente. Ademais, essa compreensão vem ao encontro de outro fato histórico. O curso de atuação deste corpo docente obteve o Conceito Preliminar de Curso (CPC) igual a três em 2011, passando a cinco em 2014. Por isso, pondera-se que o investimento realizado foi importante para que o referido curso figurasse entre os cinco melhores cursos de Bacharelado em Ciência da Computação do Brasil, conforme Conceito Preliminar de Curso (INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA - INEP, 2016).

## 6 Síntese do trabalho

Diante os estudos de caso apresentados, a presente investigação é caracterizada conforme a Figura 10. Resumidamente, os dados abertos oriundos das Plataformas Lattes e Sucupira são elevados ao patamar da 5ª Estrela e publicados na *Web* de Dados. Para tanto, foi utilizado o procedimento metodológico *Linked Data Lifecycle* (AUER, 2014) - círculo interno da figura -, perfazendo as atividades de: i) extração de dados abertos de sistemas legados; ii) conversão dos dados para o modelo RDF; iii) armazenamento dos grafos RDF; iv) interligação dos dados com recursos RDF da DBpedia; e v) exploração dos dados abertos nos estudos de caso propostos. Nos referidos estudos, de acordo com os preceitos da Gestão do Conhecimento propostos por Nissen (2006) – círculo externo da figura-, foi possível organizar, formalizar e compartilhar elementos de conhecimento úteis para criar, refinar ou aplicar conhecimento novo não explicitado isoladamente nas Plataformas Lattes e Sucupira.

Figura 10 – Representando os pressupostos do trabalho





Fonte: Os autores.

Baseando-se nos três estudos de caso desenvolvidos, confirma-se o potencial da *Web de Dados* como uma plataforma global em que dados abertos são disponibilizados para o reuso em diversos contextos. Neste sentido, como contribuição adicional, o histórico do Índice Qualis está publicado abertamente nesta plataforma e disponível a outros pesquisadores para reutilização em outros estudos bibliométricos ou cientométricos (vide a Figura 5 da seção “3.2 Conjunto de dados abertos utilizados”).

## 7 Considerações finais

Este artigo apresenta um estudo interdisciplinar, envolvendo alguns elementos da Cientometria e dos Dados Abertos Conectados para promover a Gestão do Conhecimento em uma universidade pública brasileira.

No domínio das instituições voltadas à pesquisa científica, ressalta-se que o caráter inovador do trabalho delinea a estruturação de um processo para utilizar os Dados Abertos Conectados nas atividades de organização, formalização, compartilhamento, relacionamento e exploração de dados bibliométricos/cientométricos.

Neste sentido, considera-se os esforços despendidos como um exemplo profícuo da utilização de Dados Abertos Conectados no subsídio da Gestão do Conhecimento de uma organização. Pontualmente, o trabalho corrobora um projeto maior que é o desenvolvimento de um “Modelo Tecnológico ao Compartilhamento de Dados para Estudos Cientométricos baseado em *Linked Open Data*”. Por isso, admite-se que o trabalho pode ser ampliado à medida que novas fontes abertas de dados primários sejam incorporadas (dados a respeito de resumos e artigos publicados em congressos, projetos de pesquisa, projetos de extensão, entre outros). Este avanço suportaria outros estudos cientométricos ou as atividades aderentes à Gestão do Conhecimento. Neste sentido, com a experiência adquirida, como trabalhos futuros são traçados:

- a) a manutenção do compartilhamento ao nível da 5ª Estrela do histórico do Índice Qualis ao longo do tempo, disponibilizando dados primários a outros estudos no campo da Ciência da Informação;
- b) o início do compartilhamento de demais índices cientométricos, como por exemplo, o SCImago *Journal Rank* (SJR) e *Source Normalized Impact per Paper* (SNIP); e
- c) a prospecção de novos estudos de casos envolvendo a Cientometria e a Gestão do Conhecimento no âmbito das universidades brasileiras.

## 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*. 2016. Disponível em: <<http://5stardata.info/en>>. Acesso em: 8 jun. 2016.
- AKSW. Agile knowledge and semantic web. 2014. Disponível em: <<http://aksw.org/About.html>>. Acesso em: 8 jun. 2016.
- AUER, S. Introduction to LOD2. In: AUER, S.; BRYL, V.; TRAMP, C (Ed.). *Linked Open Data: creating knowledge out of interlinked data*. Heidelberg: Springer-Verlag, 2014.
- BRITO, E. P. Z.; LUCA, M. M. M. de; TEIXEIRA, A. J. C. *Relatório de Avaliação 2010-2012 Trienal 2013: Administração, Ciências Contábeis e Turismo*. 2016. Disponível em: <<https://docs.google.com/viewer?a=v&pid=sites&srcid=Y2FwZXMuZ292LmJyfHRyaWVuYWwtMjAxM3xneDo2YzNkMjU2ZTdmY2M0ODAx>>. Acesso em: 11 jun. 2016.
- CONSELHO NACIONAL DE DESENVOLVIMENTO CIENTÍFICO E TECNOLÓGICO (CNPq). *Dados Abertos: Portal CNPq*. 2016. Disponível em: <[http://cnpq.br/dados\\_abertos](http://cnpq.br/dados_abertos)>. Acesso em: 8 jun. 2016.
- CONSELHO NACIONAL DE DESENVOLVIMENTO CIENTÍFICO E TECNOLÓGICO (CNPq). *Plataforma Lattes*. 2016. Disponível em: <<http://lattes.cnpq.br/>>. Acesso em: 8 jun. 2016.
- DALKIR, K. *Knowledge management in theory and practice*. Cambridge: The MIT Press, 2011.

FILIPPIM, E. S.; LIMA, A. A. de. Conhecimento tácito em organizações: estudo de caso no setor metalúrgico. *Perspectivas em Gestão & Conhecimento*, v. 4, n. 1, p. 80-97, 2014.

FUKUNAGA, F. *et al.* Indicadores bibliométricos da produção acadêmica mundial sobre o conceito do trabalhador do conhecimento. *Perspectivas em Gestão & Conhecimento*, v. 5, Número Especial, p. 42-56, 2015.

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.

GONZALEZ, R. V.; MARTINS, M. F. Gestão do conhecimento: uma análise baseada em fatores contextuais da organização. *Production*, v. 25, n. 4, p. 834-850, 2015.

HÄDRICH, T. *Situation-oriented provision of knowledge service*. 2008. 415f. Tese (Doutorado em Economia) – der Juristischen und Wirtschaftswissenschaftlichen Fakultät, der Martin-Luther- Universität, Halle-Wittenberg, 2008.

INAZAWA, F. K. O papel da cultura organizacional e da aprendizagem para o sucesso da gestão do conhecimento. *Perspectivas em Ciência da Informação*, v. 14, n. 3, p. 206-220, 2009.

INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA (INEP). *CPC-INEP*. 2016. Disponível em: <<http://portal.inep.gov.br/educacao-superior/indicadores/cpc>>. Acesso em: 20 jun. 2016.

LINKED DATA. *Linked Data: connect distributed data across the web*. 2016. Disponível em: <<http://linkeddata.org>>. Acesso em: 14 jun. 2016.

MACEDO, V. *et al.* Tipologia do trabalhador do conhecimento: papéis e processos. *Perspectivas em Gestão & Conhecimento*, v. 7, Número Especial, p. 94-108, mar. 2017.

MAIER, R. *Knowledge management systems information and communication technologies for knowledge management*. 3. ed. Heidelberg: Springer, 2007.

MENA-CHALCO, J. P.; CESAR-JR, R. M. scriptLattes: an open-source knowledge extraction system from the Lattes platform. *Journal of the Brazilian Computer Society*, v. 15, n. 4, p. 31-39, 2009.

NISSEN, M. E. *Harnessing knowledge dynamics*. Hershey: Idea Group Inc., 2006.

NONAKA, I. A Dynamic theory of organizational knowledge creation. *Organization Science*, v. 5, n. 1, p. 14-37, 1994.

OPEN KNOWLEDGE INTERNATIONAL. *O que são dados abertos?* 2016. Disponível em: <[http://opendatahandbook.org/guide/pt\\_BR/what-is-open-data/](http://opendatahandbook.org/guide/pt_BR/what-is-open-data/)>. Acesso em: 14 jun. 2016.

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.

RAUTENBERG, S.; BURDA, A. 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.

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.*

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

SILVA, N. B. X.; SOUSA, M. R. F. de. a dimensão tecnológica da gestão do conhecimento e a contribuição da arquitetura da informação: uma análise da Plataforma Podio. *Perspectivas em Gestão & Conhecimento, v. 5, n. 2, p. 186-200, 2015.*

STEFANO, N. M. *et al.* Gestão de ativos intangíveis: implicações e relações da Gestão do Conhecimento e Capital Intelectual. *Perspectivas em Gestão & Conhecimento, v. 4, n. 1, p. 22-37, 2014.*

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

SUCUPIRA. *Plataforma Sucupira.* 2016. Disponível em: <<https://sucupira.capes.gov.br/sucupira/public/consultas/coleta/veiculoPublicacaoQualis/listaConsultaGeralPeriodicos.jsf>>. Acesso em: 3 abr. 2016.

SUPYUENYONG, V; ISLAM, N. Knowledge management architecture: building blocks and their relationships. *In: PORTLAND INTERNATIONAL CENTER FOR MANAGEMENT OF ENGINEERING AND TECHNOLOGY, 2006, Istambul, Turquia, Julho. Proceedings...* PICMET: Istambul, 2006. p. 1210-1219.

UNIVERSIDADE ESTADUAL DO CENTRO-OESTE (UNICENTRO). Resolução nº 5-CEPE/UNICENTRO. Aprova o Regulamento de Pesquisa da UNICENTRO.

2011. Disponível em: <<http://www2.unicentro.br/pesquisa/files/2012/10/Resolucao-5-2011.pdf>>. Acesso em: 11 jun. 2016.

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

VANTI, N. A Cientometria revisitada à luz da expansão da ciência, da tecnologia e da inovação. *PontodeAcesso, v. 5, n. 3 p. 5-31, 2011.*

WALTZ, E. *Knowledge management in the intelligence enterprise*. Boston: Artech House, 2003.

CAPES. *Sistema WebQualis*: Portal Capes. 2013. Disponível em: <<http://qualis.capes.gov.br/webqualis/principal.seam>>. Acesso em: 25 ago. 2013.

WIIG, K. M. Knowledge management in public administration. *Journal of Knowledge Management*, v. 6, n. 3, p. 224-239, 2002