

Linked Data e Ciência da Informação: diretrizes para a publicação de *datasets* institucionais abertos

Januário Albino Nhacuongue

Vitor Rozsa

Moisés Lima Dutra

Universidade Federal de Santa Catarina – UFSC, Brasil

ORIGINAL

Resumo

Objetivo. O objetivo deste trabalho é propor diretrizes para a publicação de dados ligados, segundo os princípios do *Linked Data*, a partir de um olhar da Ciência da Informação.

Método. A metodologia usada foi a revisão bibliográfica e documental, seguida de um procedimento exploratório e aplicado, para avaliar os alcances dos princípios do *Linked Data* sobre a disseminação de recursos nas transformações sociais. Trabalhos-chave do criador e propagador das práticas de *Linked Data*, Tim Berners-Lee, foram utilizados como base da proposta. Além disso, escolheu-se analisar projetos e *datasets* da área educacional, por serem compostos, em sua maioria, por recursos abertos, acessíveis e reutilizáveis.

Resultados. Os resultados estão sintetizados nas principais diretrizes para a publicação de conjuntos de dados, reforçando a iniciativa de profissionais, pesquisadores e instituições para o reuso e interoperabilidade.

Conclusões. As principais conclusões desta discussão resumem-se em duas linhas. A primeira versa sobre as novas perspectivas que surgem a partir desta abordagem da Ciência da Informação sobre o *Linked Data*, no âmbito do seu objeto e suas manifestações. A segunda, sobre a necessidade de comprometimento dos profissionais da informação com questões relacionadas ao *Linked Data*, de modo a ampliar o universo informacional e o ambiente de satisfação das necessidades dos usuários.

Palavras-chave

Ciência da Informação; Datahub; Datasets; Linked Data

Linked Data and Information Science: guidelines for publishing open institutional datasets

Abstract

Objective. The purpose of this paper is to propose guidelines for the publication of data, according to the principles of Linked Data, from an Information Science perspective.

Method. The methodology used comprised a bibliographical and documentary review, followed by an exploratory and applied procedure, to evaluate the scope of Linked Data principles on the dissemination of resources in social transformations. Key works by the creator and propagator of Linked Data practices, Tim Berners-Lee, were used as the basis of the proposal. In addition, we chose to analyze projects and datasets from the educational area, since they are composed, mostly, of open, accessible and reusable resources.

Results. The results are summarized in the main guidelines for the publication of datasets, reinforcing the initiative of professionals, researchers and institutions for reuse and interoperability.

Conclusions. The main conclusions of this discussion are summarized in two lines. The first one deals with the new perspectives that arise from this approach of Information Science on Linked Data, within the scope of its object and its manifestations. The second relies on the need for commitment of information professionals with issues related to Linked Data, in order to broaden the informational universe and the environment of satisfaction of users' needs.

Keywords

Datahub; Datasets; Information Science; Linked Data

1 Introdução

A criação de repositórios institucionais para compartilhar tipos específicos de conteúdo ou atender a comunidades particulares enquadra-se nas questões relacionadas a *Open Educational Resources (OER)*, isto é, da descoberta de recursos em consonância com o paradigma da educação aberta (BARKER; CAMPBELL, 2016). Um paradigma que, segundo Blessinger e Bliss (2016), mesmo tendo começado no final dos anos 1960, com o estabelecimento de universidades abertas, só ganhou notoriedade no início do século XXI. Tal destaque deve-se ao desenvolvimento de recursos educacionais e tecnologias abertas, para garantir a descoberta, o uso e o reuso da informação no âmbito da democratização do acesso.

A criação de repositórios de dados se insere no conceito de abertura, cujas qualidades e características podem ser sintetizadas pelas seguintes dimensões: espacial, temporal e processual (BLESSINGER; BLISS, 2016). Na dimensão espacial, a publicação de dados permite o acesso de recursos por pessoas, grupos e instituições, independentemente da sua localização geográfica. A dimensão temporal garante ubiquidade, eliminando barreiras temporais. Na dimensão dos processos, o uso de vocabulários abertos, aliado à observância das premissas do *Linked Data*, proporciona a descoberta e o reuso dos dados, como a criação de *links* sob ferramentas adicionais dos usuários. *Linked Data* (dados vinculados ou dados ligados) é um método proposto por Tim Berners-Lee (2006) para estruturar e publicar dados conectados na Web por meio de *hyperlinks* e anotações, de maneira a permitir buscas semânticas. Essas anotações e ligações são fatores determinantes para o crescimento exponencial do conhecimento.

A iniciativa de se criar diretrizes para a publicação de conjuntos de dados está, portanto, atrelada a um movimento assistido nos últimos anos. Existem milhares de *datasets* (ou seja, conjuntos de dados) estocados em repositórios de dados (que são as infraestruturas de armazenamento dos *datasets*) do mundo todo, que procuram atender a distintas comunidades com os mais diferentes tipos de necessidade. Em consulta realizada no início de fevereiro de 2017, constavam 850 *datasets* na área de educação e comunicação do continente europeu¹, para citar apenas um caso. Um dos *datasets* desta área está voltado para o cálculo da taxa de emprego por nível de escolaridade, um conjunto de dados extremamente importante para a construção de ferramentas de análise e prospecção da economia e do grau de integração das políticas públicas.

Na Ciência da Informação (CI), nossa iniciativa está centrada no núcleo principal da área, através do respectivo objeto (informação). A CI é um dos campos emergentes que se consagrou na segunda metade do século XX (CAPURRO; HJORLAND, 2007), principalmente com o movimento acelerado das tecnologias de comunicação e informação. O seu desafio foi notabilizado por meio de abordagens epistemológicas, processuais, metodológicas e tecnológicas, na busca de soluções de caráter interdisciplinar para garantir a preservação, descoberta, o acesso e uso de recursos. Por isso, enquanto processo, o objeto da Ciência da Informação consiste no ato de informar, comunicar e mediar. Também permeia um conjunto de interações e ligações entre produtores, ambientes, usuários e recursos na busca de soluções para o conhecimento em ação. Como conhecimento, a informação é a exteriorização do que foi apreendido em objetos informativos. Como coisa, a informação é a representação tangível do conhecimento e dos processos (BUCKLAND, 1991).

A perspectiva de Buckland (1991) sobre o objeto da CI permite-nos afirmar que os fundamentos e diretrizes para a criação e publicação de *datasets* no estilo *Linked Data* aqui propostos preenchem alguns patamares do campo. Também, despertam um olhar crítico da área na sua contribuição em torno das transformações impostas pelos avanços científicos, tecnológicos e humanos.

O panorama traçado com base no olhar da CI também nos permite concluir que este trabalho possivelmente tem uma dimensão social (INGWERSEN, 1992; SARACEVIC, 1996) que enaltece os conceitos de abertura, disponibilização, acesso e uso de recursos dentro da democratização do acesso a informação. Os dados de repositórios abertos são livres de licenças ou contêm poucas restrições, são disponibilizados em formatos acessíveis por diferentes plataformas, e destinam-se ao reuso em diferentes níveis e contextos. Esta é uma das recomendações de Barker e Campbell (2016) que, mesmo enaltecendo os esforços para a abertura na prática educacional e a conscientização sobre licenças livres, ainda reconhecem a dificuldade para a descoberta de recursos na Web. Como solução, apontam algumas estratégias tecnológicas que incluem repositórios institucionais, sistemas de gerenciamento de conteúdo, agregadores e metadados descritivos. Como se pode depreender, nosso trabalho se insere nesta linha.

Algumas das estratégias citadas podem ser vislumbradas a partir da capacidade de prospecção de ações de investigação no contexto da CI, de modo a transformar paradigmas, nas relações sociais, em processos de gestão e organização (RIBEIRO JUNIOR, 2013). Em outras palavras, embora esta proposta possa resultar na solução de problemas, comporta um olhar sobre outras perspectivas do campo. Tais perspectivas não cabem apenas no núcleo tecnológico, pelo contrário, suscitam uma abordagem epistemológica e fenomenológica para a compreensão do problema como um todo. A nossa posição vai ao encontro da percepção de Sant'Ana (2016, p.117), sobre a necessidade de uma "estrutura básica" do campo que permita maiores reflexões sobre o atual cenário informacional.

D'Aquin *et al.* (2013), por sua vez, reconhecem os esforços atuais das instituições no domínio da educação, por meio de tecnologias de *Linked Data*, para publicar informações. Essas informações incluem ofertas de cursos, recursos educacionais abertos e facilidades educacionais de uma forma acessível e reutilizável. Contudo, mesmo enfatizando as soluções oferecidas pelo *Linked Data* para transformar a educação, os autores admitem que estas ainda não são amplamente adotadas no campo. Desta forma, este trabalho se propõe também a contribuir com a difusão do conhecimento que poderá gerar novas ações em prol dos princípios do *Linked Data* e das iniciativas sobre a educação aberta no todo.

Esta é uma pesquisa aplicada, do ponto de vista de sua natureza, pois ambiciona gerar conhecimentos voltados à aplicação prática, e exploratória, quanto aos objetivos. Emprega como metodologia uma revisão bibliográfica e documental, para avaliar os alcances dos princípios do *Linked Data* sobre a disseminação de recursos nas transformações sociais.

Utilizou-se como base para esta proposta dois artigos-chave do criador e disseminador das práticas de *Linked Data*, Tim Berners-Lee (BERNERS-LEE, 2006; BIZER; HEATH; BERNERS-LEE, 2009), complementados por projetos de *Linked Data* pesquisados no portal *Eu Open Data* ou referenciados por Barker e Campbell, na obra que publicaram em 2016, que discute estratégias tecnológicas para a disseminação de recursos educacionais na Web.

Além disso, os autores desta pesquisa compactuam com a visão de D'Aquin *et al.* (2013) a respeito dos recursos abertos educacionais, por isso, *datasets* compostos por dados abertos e acessíveis, provenientes do contexto educacional, foram escolhidos para serem analisados aqui. Como limitação deste trabalho, é preciso destacar que o corpus analisado não se configura como uma lista exhaustiva de iniciativas, mas tão somente como um conjunto que os autores consideram empiricamente representativo de todo o espectro de projetos atualmente existentes relacionados à temática dos dados ligados abertos e sua publicação.

Os resultados deste trabalho permitiram gerar um conjunto de diretrizes para a publicação de *datasets Linked Data*, que, espera-se, possa servir como ferramenta útil de auxílio a profissionais, pesquisadores e instituições durante o processo de uso, reuso e interoperabilidade (transferência entre diferentes sistemas de informação) de dados e informações.

Finalmente, os autores entendem que a distância atualmente existente entre as tecnologias de *Linked Data* e sua discussão na CI é uma lacuna que precisa começar a ser preenchida.

2 Trabalhos relacionados

O compartilhamento de *datasets* na Web é desenvolvido em torno dos princípios de *Linked Data*. Trata-se de parte dos fundamentos sobre a Web Semântica, que visa tornar a informação consistente, acessível, ubíqua, interoperável e reutilizável em diferentes contextos. Sendo assim, esta não é uma ideia recente, muitas iniciativas têm movido pesquisas e instituições no mundo inteiro. Esta seção apresenta uma lista não-exaustiva de iniciativas de compartilhamento de recursos informacionais sobre educação na Web.

O trabalho de D'Aquin *et al.* (2013), por exemplo, cita a iniciativa do projeto *LinkedUp*² na exploração e adoção de dados públicos e abertos disponíveis na Web, em particular no contexto da Educação. O projeto *LinkedUp* tem como objetivo principal facilitar o desenvolvimento de aplicativos abertos, inovadores e robustos, que exploram dados em larga escala na Web em cenários educacionais. Alguns dos seus resultados são sintetizados no gerenciamento de informações da Web, na transferência de tecnologia³, no cruzamento de

dados do setor público e privado, e na coleção de dados de qualidade para a educação (OPEN KNOWLEDGE INTERNATIONAL, 2018).

No livro “*Open Education: International Perspectives in Higher Education*”, Barker e Campbell (2016) descrevem uma variedade de abordagens profissionais e institucionais para desenvolver e utilizar repositórios. Essas tecnologias são voltadas para a disseminação, descoberta, uso e reuso de recursos no âmbito da educação aberta.

Quadro 1. Estratégias tecnológicas para a disseminação de recursos sobre a educação aberta.

Instituição	Tecnologia	Descrição
Massachusetts Institute of Technology (MIT)	MIT OpenCourseWare (MIT OCW)	Recursos derivados de cursos do MIT (currículos de cursos, palestras, etc.), na sua forma original. Alguns são hospedados em plataformas externas (<i>Youtube</i>). Os metadados são exportados para agregadores como <i>OER Commons</i> e <i>RSS feeds</i> podem ser coletados para outras fontes.
University of Nottingham	Repositório U-Now	Recursos internos (currículos de cursos, textos, vídeos) ou <i>links</i> a recursos externos, na sua maioria sob licenças <i>COERtive Commons</i> ou livre de licenças. Disponibilidade de ferramentas para apoiar a criação de recursos (<i>Xerte Community</i>), sua descoberta (<i>Xpert</i>), etc. <i>RSS feeds</i> para agregadores como <i>iTunes U</i> .
University of Oxford	OpenSpires, Podcasts, Great Writers Inspire, etc.	Várias iniciativas para a educação aberta (ex: arquivos digitais de conteúdo de bibliotecas e museus). Uso de <i>podcasts</i> (sindicação de registros e metadados por <i>feeds RSS</i> , coletados por agregadores como <i>iTunes U</i> , que lhes dá maior circulação e visibilidade.
The Open University	OpenLearn	Reúne recursos de maior escala sobre <i>OER</i> no Reino Unido (ex: séries de TV da BBC) que podem ser utilizados na educação a distância de vários cursos. Também fornece ferramentas para a criação e remixagem de conteúdo através da plataforma <i>OpenLearn Works</i> .
Várias	Massive Open Online Course (MOOC)	Coleções amplamente disseminadas de recursos de aprendizagem, mas não são repositórios propriamente ditos. Alguns conteúdos <i>MOOC</i> são disponíveis em plataformas externas como o <i>Youtube</i> , todavia, às vezes, o acesso dos recursos em algumas instituições é restrito.

Fonte: Elaboração própria, com base em Barker e Campbell (2016).

No Quadro 1 estão sintetizadas algumas das iniciativas para a disseminação e utilização de recursos, apontadas por Barker e Campbell (2016). Os autores também destacam a importância do uso de agregadores para repositórios de recursos de domínio específico, descrição e vocabulários. Por exemplo, *HumBox* foi construído sobre a plataforma *EdShare* da Universidade de Southampton, baseado no repositório *open source ePrints*, com maior incidência sobre o domínio da educação em humanidades. Com o uso de agregadores de conteúdo, o repositório reúne recursos de cerca de seis instituições de ensino superior no Reino Unido, organizados em slides, textos, imagens, registros de áudio e vídeo, entre outros.

Embora versem sobre *Linked Data*, os trabalhos de D’Aquin *et al.* (2013) e de Barker e Campbell (2016) são voltados para a educação. Não obstante o fato da nossa discussão congregar alguns aspectos comuns, também carrega uma vertente sobre os olhares e fazeres da Ciência da Informação. Por exemplo, realçamos a

importância da descrição dos recursos criando interseções práticas com os processos de mediação dos profissionais da informação.

O trabalho de Sant'Ana (2016) contribui para a CI por meio da proposta de novas abordagens sobre o ciclo de vida de dados que ampliem a aproximação entre os usuários e os recursos. A partir do conceito e das potencialidades do *Big Data* (modelo que versa sobre volume, variedade de formatos, velocidade de acesso e processamento de dados, veracidade, variabilidade, valor e visualização de dados), o autor constrói novas perspectivas centradas nos dados ou recursos. O seu modelo do ciclo de vida de dados é orientado pelas necessidades e competências impostas pelos próprios dados e envolve as atividades de coleta, armazenamento, recuperação e descarte. Estas atividades concentram-se sobre aspectos que, por sua vez, suscitam novas reflexões na área. Tais aspectos consistem na privacidade, na integração, qualidade, disseminação, preservação e nos direitos autorais.

A discussão de Sant'Ana (2016) é indispensável à Ciência da Informação e possui estreita ligação com esta proposta. Um exemplo elucidativo pode ser destacado a partir da proposta da construção de repositórios institucionais, com base no modelo do ciclo de vida. Em última análise, o ciclo de vida de dados pode facilitar o relacionamento de recursos no estilo *Linked Data*. Por isso, o atual cenário informacional suscita maior comprometimento dos profissionais da área. Embora esse comprometimento possa transparecer novas abordagens, as práticas já consagradas ainda se mostram necessárias para incorporar as novas tendências.

3 Princípios do *Linked Data*

O termo *Linked Data* refere-se a um conjunto de princípios para publicar e conectar dados estruturados na Web. Bizer, Heath e Berners-Lee (2009) afirmam que a sua gênese está relacionada à insuficiência da estrutura semântica para aferir sobre a relação dos objetos informacionais na linguagem de marcação do hipertexto - *HyperText Markup Language* (HTML) e em outros formatos tradicionais na Web. Além deste fator, Keßler, D'Aquin e Dietze (2013), por um lado, apontam a incompatibilidade dos esquemas de descrição⁴ e exposição de recursos. Por outro lado, a limitação da interoperabilidade devido à heterogeneidade das tecnologias de armazenamento e interfaces usadas ao longo dos anos. Por isso, *Linked Data* é uma extensão da Web para a conexão de dados globais sobre pessoas, organizações, livros, publicações científicas, filmes, músicas, programas de televisão, ensaios clínicos, etc. Ou seja, *Linked Data* é um conjunto de dados publicados na Web, em formatos que sejam legíveis por máquinas, com estruturas semânticas bem definidas, que estejam ligados a outros dados externos. É uma maneira de compartilhar, reutilizar e conectar recursos e usuários.

Os princípios do *Linked Data* desdobram-se em três principais linhas: a navegação, a pesquisa e o suporte de aplicações. Em relação à navegação, a busca por informações em determinadas fontes de dados (*datasets*) pode ser estendida para outras fontes com base em *links* de relacionamento. Quanto à pesquisa, os mecanismos de busca podem rastrear a Web através de *links* para fornecer consultas expressivas. Finalmente, aplicações inovadoras podem sugerir novos modelos de inferência a partir de dados de diferentes fontes (BIZER, HEATH; BERNERS-LEE, 2009).

Berners-Lee (2006) estabeleceu um conjunto de regras que se tornaram princípios básicos para publicar dados na Web de acordo com o *Linked Data*: a) usar *Uniform Resource Identifiers* (URIs) como nomes para as coisas; b) usar URIs *Hypertext Transfer Protocol* (HTTP) para a verificação dos nomes; c) quando alguém procurar em uma URI, fornecer informações usando *Resource Description Framework* (RDF) e *SPARQL Protocol and RDF Query Language* (SPARQL); d) incluir *links* para outras URIs, para possibilitar a descoberta de novo conhecimento.

No contexto do *Linked Data*, URIs, RDF e SPARQL são tecnologias relacionadas que viabilizam a publicação de dados ligados na Web. As URIs servem como identificadores únicos dos recursos e podem ser utilizadas para acessar informações sobre o recurso (quando são HTTP URIs). RDF é o *framework* utilizado para descrever recursos na Web. Por fim, o SPARQL é o meio utilizado para consultar recursos disponíveis em *datasets*.

Para que os recursos no *Linked Data* possam ser referenciados, de modo a descrevê-los ou recuperá-los, é necessário um método de identificação não ambíguo. Esse método é fornecido pelas URIs, que permitem que recursos virtuais (ex.: páginas Web, organizações, vocabulários, etc.) e físicos (ex.: pessoas e lugares) possuam

um identificador único na Web. Basicamente, as URIs seguem o seguinte padrão (WOOD *et al.*, 2014) “*http://{authority}/{container}/{item_key}*” no qual

- a) *{authority}* – é utilizado para identificar um servidor na Web;
- b) *{container}* – é utilizado para separar diferentes contextos de dados. Por exemplo, podemos criar o contexto “funcionários” para agrupar identificadores de funcionários em uma organização; e o contexto “clientes” para agrupar a identificação de clientes daquela organização; e
- c) *{item_key}* – é utilizado para identificar unicamente um recurso específico.

Uma vez que é possível identificar de forma exata os recursos na Web, o RDF permite realizar a descrição destes recursos por meio de triplas que são compostas por três elementos: sujeito, predicado e objeto. O sujeito é um recurso do qual se está falando. O predicado se refere ao que se está falando do sujeito e pode ser entendido como uma propriedade do sujeito. Já o objeto é o que se está falando do sujeito e pode ser entendido como o valor da propriedade, sendo que o objeto pode ser um valor literal (número, texto, data, etc.) ou a referência para outro recurso.

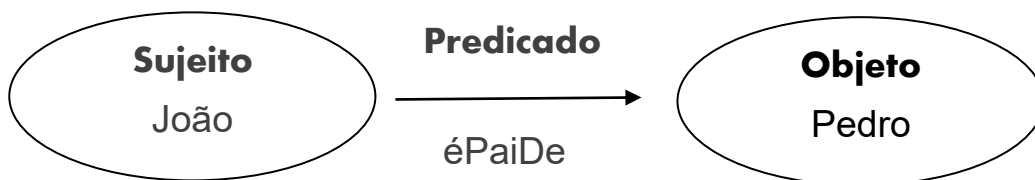


Figura 1. Representação em triplas – elaboração própria

Os recursos descritos pelas triplas podem ser, por exemplo, documentos, pessoas, objetos físicos, conceitos, ideias, dados provenientes de sensores, etc., e, como mencionado, são identificáveis por meio de URIs. Ainda, os predicados podem ser utilizados para relacionar os recursos entre si, formando assim uma base de dados em forma de grafo. Na Figura 1, está representada a relação de paternidade entre as pessoas (recursos) “João” e “Pedro”, utilizando o predicado “éPaiDe”. Ou seja, a tripla contém a informação de que “João é pai de Pedro”.

3.1 Consulta SPARQL

Para acessar os dados armazenados em *datasets*, é necessário o emprego de uma ferramenta para consulta. Sendo assim, o SPARQL é uma linguagem de consulta e manipulação de dados armazenados em bases de dados no formato RDF. Basicamente, uma consulta em SPARQL é composta por um ou mais padrões de triplas, com sujeito, predicado e objeto, sendo que pelo menos um desses elementos é uma variável (YANG *et al.*, 2013). Esse conjunto representa um subgrafo do grafo RDF que, se existir, será retornado como resultado da consulta.

Além do subgrafo a ser encontrado, uma consulta SPARQL possui os seguintes elementos (SANTAREM SEGUNDO, 2014):

- a) **Prefixos** - utilizados para que não seja necessário escrever URIs por extenso nas consultas (facilita a legibilidade e escrita da consulta);
- b) **Conjunto de dados** - define quais conjuntos de dados RDF estão sendo consultados;

- c) **Cláusula de resultado** - identifica os valores alvo da consulta (a consulta pode, pode exemplo, especificar diferentes variáveis, mas estar interessada em apenas uma delas).
- d) **Padrão de triplas** - que descreve o conjunto de triplas que se espera obter a partir do conjunto de dados RDF;
- e) **Modificadores** - utilizados para modificar o resultado da pesquisa, como limitar o número de resultados, ordená-los, recuperar apenas um subconjunto, entre outros.

O exemplo a seguir (Figura 2) ilustra uma consulta em SPARQL que pode ser lida como “Quem é o pai de Pedro?”. Considerando a tripla descrita no exemplo anterior (Figura 1), o valor “João” será atribuído à variável “?pai” como resultado da consulta. Ou seja, “João é pai de Pedro”.

```
PREFIX pessoa: <http://pessoas.com/res#>
PREFIX relacao: <http://relacoes.com/res#>

SELECT ?pai
WHERE
{
    ?pai relacao:ePaiDe pessoa:Pedro .
}
```

Figura 2. Exemplo de consulta SPARQL – elaboração própria

Uma das limitações da versão inicial do SPARQL é fornecer apenas operações de leitura sobre o conjunto de dados. Em 2013, o World Wide Web Consortium publicou uma nova versão do SPARQL, o SPARQL 1.1 Update, que fornece a possibilidade de atualizar o conjunto de dados em RDF (W3C, 2013). O SPARQL 1.1 Update expande a funcionalidade do SPARQL fornecendo operações para atualizar, criar e remover grafos RDF em um *dataset*.

Os princípios para a publicação de dados na Web também implicam o uso de vocabulários abertos, amplamente estruturados e conhecidos. Por outro lado, a interoperabilidade pode ser afetada pela heterogeneidade de representação dos vocabulários e linguagens usadas (KEßLER; D’AQUIN; DIETZE, 2013). O W3C (2015) recomenda a reutilização dos vocabulários já existentes para evitar a duplicação e viabilizar a interoperabilidade de dados entre diferentes *datasets*.

O uso de vocabulários abertos é fundamental para a descrição de conceitos e suas relações. Um vocabulário reúne as definições de um conjunto de classes e propriedades para descrever tipos específicos de conceitos em um determinado domínio. Os termos de vocabulários também fornecem os *links* em dados ligados. As definições de termos fornecidos pelos vocabulários contêm uma semântica que permite que os dados se tornem significativos. Embora esta semântica seja construída com base em RDF e na instanciação de ontologias em *Web Ontology Language* (OWL), também pode ser reforçada a partir das lexicalizações dos termos nas linguagens naturais e no vocabulário controlado. O uso de termos de vocabulários extraídos do léxico e do vocabulário controlado, com a respectiva semântica, permite uma maior familiaridade do usuário com o vocabulário, facilitando a decisão sobre o seu uso na descrição de recursos em repositórios de dados.

Em fevereiro de 2015, o W3C atualizou o conjunto de princípios para operações HTTP em recursos da Web, algumas baseadas em RDF, para fornecer uma arquitetura do *Linked Data* (W3C, 2015). No geral, os princípios são divididos em quatro partes. A primeira discute as técnicas dos padrões HTTP e RDF usadas para construir clientes e servidores⁵ que criam, registram e descrevem os recursos na plataforma do *Linked Data*. Também contêm as melhores práticas a serem adotadas e especificações a serem evitadas. A segunda parte destina-se ao tipo específico de recursos (*containers*), que envolvem a coleta de várias fontes, por exemplo: autores, obras

publicadas, locais de publicação, áreas temáticas, etc. A terceira parte destina-se às especificações adicionais dos dados ligados. A última parte é voltada para grandes volumes de dados e suas representações (W3C, 2015).

4 Publicação de *datasets* abertos

Por meio da publicação de dados na Web, de acordo com os princípios do *Linked Data*, os repositórios tornam visíveis seus dados para serem encontrados e utilizados de formas variadas - e muitas vezes não esperadas - por diferentes aplicações (BIZER; HEATH; BERNERS-LEE, 2009). Além da observância dos princípios do *Linked Data* (descritos anteriormente) na criação dos *datasets*, outras diretrizes são consideradas relevantes neste processo. Nesta seção descrevemos um conjunto de diretrizes para a publicação de *datasets* abertos na Web, baseadas principalmente nos trabalhos de Heath e Bizer (2011) e Wood *et al.* (2014) sobre *Linked Data*. Estas diretrizes estão voltadas para a adequação do *dataset* para publicação, de modo que o mesmo possa ser identificado, compreendido e utilizado corretamente.

Inicialmente, é justificável ressaltar a importância do papel das URIs na publicação de dados ligados, uma vez que fornecem a identificação e localização (quando se usa URIs HTTP) dos recursos disponibilizados nos *datasets*. URIs, de modo geral, permitem identificar unicamente recursos nos *datasets*. URIs HTTP permitem, além disso, que a identificação do recurso seja utilizada para obter informações complementares sobre o mesmo. Heath e Bizer (2011) fornecem diretrizes de como criar URIs adequadas aos propósitos do *Linked Data*:

- a) **Utilize apenas URIs de domínios que você controla** - esta diretiva se refere ao caso em que outro repositório disponibiliza determinado recurso (ex.: um filme) com uma URI específica sob seu controle. Assim, sugere-se não utilizar esta URI, uma vez que está sob autoridade de terceiro, que controla o conteúdo recuperado quando a URI é acessada e decide a respeito da integridade desta ao longo do tempo. Sendo assim, sugere-se gerar uma nova URI no *dataset* local e declarar que ela é equivalente àquele recurso mantido por terceiro;
- b) **Oculte detalhes de implementação** - esta diretiva sugere evitar adicionar detalhes de implementação na URI, tais como a identificação do servidor de hospedagem e a linguagem de implementação da página Web. Por exemplo, o *link* "http://main_server1.ufsc.com.br/alunos.php?id=joao_silva" é um exemplo de URI que apresenta detalhes de implementação. Além de aumentar a complexidade da URI, caso estas tecnologias sejam alteradas, a URI tornar-se-á inválida. Uma URI adequada teria, por exemplo, o seguinte formato: "http://ufsc.com.br/alunos/joao_silva.rdf";
- c) **Use chaves naturais nas URIs** - esta diretiva sugere o uso de termos que possuam significado dentro do contexto de modelagem dos dados e sejam únicos. Por exemplo, em um *dataset* que contém dados sobre alunos, professores e servidores, nome e sobrenome combinados poderiam ser utilizados como identificadores das pessoas. Além de garantirem a identificação única de cada indivíduo, também permitiriam a criação de URIs mais legíveis e mnemônicas.

Além da criação de URIs, existe, também relacionada aos princípios fundamentais do *Linked Data*, a diretriz de se gerar ligações entre diferentes *datasets*. Este processo consiste em se "apontar" para recursos fora do *dataset* original. A geração de ligações é um critério a ser seguido que permite unir *datasets* em um espaço global e interconectado de dados para que aplicações possam descobrir informações adicionais a partir de um primeiro ponto de partida (HEATH; BIZER, 2011). Estas ligações são construídas, basicamente, criando-se triplas RDF nas quais o sujeito é um recurso no *dataset* local e o objeto se refere a um recurso em um *dataset* externo, unidos por algum predicado conveniente para a ligação. Uma vez acessado o recurso externo, podem-se descobrir ligações para outros recursos e assim por diante.

Os principais tipos de ligações são de relacionamento, identidade e vocabulário (HEATH; BIZER, 2011). As ligações de relacionamento permitem indicar informações relacionadas a um recurso. Por exemplo, informações

associadas a uma pessoa (o recurso) poderiam ser a localidade em que ela habita e seus livros de interesse. Neste caso, seriam criados *links* entre o *dataset* que descreve esta pessoa e *datasets* que descrevem lugares e livros.

As ligações de identidade (HEATH; BIZER, 2011) visam associar um mesmo recurso presente em diferentes *datasets*. Por exemplo, considerando-se a existência de um *dataset* que mantenha informações gerais sobre pessoas (ex.: idade, gênero, local de moradia, etc.), outro *dataset* que contenha informações acadêmicas e um *dataset* que possua informações profissionais, poderiam ser criadas relações de identidade que indicassem que a mesma pessoa está presente nos três conjuntos de dados. Esta ação permitiria que aplicações obtivessem informações combinadas a partir destes três *datasets* sobre uma mesma pessoa para, por exemplo, elaborar automaticamente o seu currículo. Para expressar ligações de identidade, é possível utilizar a propriedade da linguagem OWL “*owl:sameAs*”⁶.

Por fim, as ligações de vocabulário (HEATH; BIZER, 2011) permitem identificar conceitos e propriedades com o mesmo significado, mas distribuídos por diferentes vocabulários. Desta forma, é possível que aplicações compreendam e integrem dados a partir destes vocabulários distintos. No exemplo anterior, sobre geração automática de currículo a partir de diversos *datasets*, o *dataset* sobre informações gerais poderia utilizar um vocabulário que descrevesse o nome da pessoa por meio da utilização do termo “nome”. O *dataset* sobre informações acadêmicas poderia utilizar o termo “estudante”. E o *dataset* com as informações profissionais poderia utilizar o termo “profissional”. Apesar de serem termos diferentes, todos serviriam ao propósito de identificar uma mesma pessoa. Para expressar a relação de igualdade entre conceitos e propriedades de diferentes vocabulários pode-se utilizar, respectivamente, as propriedades do OWL “*owl:equivalentClass*”⁷ e “*owl:equivalentProperty*”⁸.

Uma forma de evidenciar a presença do *dataset* na Web é por meio da geração de ligações a partir de outros *datasets* para o *dataset* em questão. Wood *et al.* (2014) sugerem a conexão com o *dataset DBPedia*⁹. O *DBPedia* é uma iniciativa da comunidade que visa extrair informações estruturadas do Wikipédia e disponibilizar estas informações abertamente na Web. Uma vez que o *dataset DBPedia* é amplamente conhecido e utilizado na comunidade do *Linked Data*, a geração de ligações a partir dele contribui para a identificação e uso de *datasets* ligados. O próprio *DBPedia* fornece orientações de como este processo de geração de ligações pode ser realizado, no seguinte endereço: <https://github.com/DBPedia/links>.

4.1 Vocabulários

O reuso de vocabulários é um aspecto importante a ser observado durante a construção de *datasets* para publicação. A compreensão do contexto, ou seja, quais são os termos e relações a serem modeladas, revela as necessidades de representação e direciona a busca por vocabulários a serem empregados na descrição semântica dos dados. O uso de vocabulários compartilhados é um segundo critério que - assim como a geração de ligações entre diferentes *datasets* - implica na disponibilização de vocabulários comuns entre diferentes repositórios, estratégia que, por sua vez, viabiliza a criação de aplicações genéricas capazes de consumir dados de diferentes *datasets* como se estivessem operando sobre o mesmo espaço global de dados (HEATH; BIZER, 2011). Heath e Bizer (2011) descrevem alguns pontos a serem considerados durante a seleção de vocabulários:

- a) **Uso** - o vocabulário é amplamente aceito e utilizado pela comunidade;
- b) **Manutenção e governança** - o vocabulário é mantido por meio de um processo claro de governança;
- c) **Cobertura** - o vocabulário atende a uma boa parte das necessidades de descrição dos dados;
- d) **Expressividade** - o vocabulário possui expressividade adequada ao *dataset* e ao cenário de aplicação.

Devido à alta heterogeneidade de conceitos presentes na maioria dos contextos, pode ser necessária a utilização de diferentes vocabulários para a descrição semântica dos dados. Por exemplo, uma instituição de

ensino pode apresentar a necessidade de representar pessoas (ex.: professores, alunos e servidores), recursos físicos (ex.: secretaria, salas de aula e laboratórios) e material acadêmico e científico (ex.: cursos, apostilas e artigos). Schaible, Gottron e Scherp (2014) identificaram em pesquisa que os repositórios de *Linked Data* buscam manter um equilíbrio entre vocabulários populares e específicos de determinado domínio, o que permite fornecer uma estrutura dos dados de modo a facilitar seu consumo. Sendo assim, podem-se empregar diferentes vocabulários, caso as necessidades de representação sejam variadas.

Uma vez que os dados possam ser compreendidos por terceiros devido ao emprego de vocabulários compartilhados, torna-se necessário prover informações sobre conteúdo e uso do *dataset*, bem como de suas ligações com outros *datasets*. Estas informações facilitam que os conteúdos dos *datasets* sejam encontrados, vinculados, agregados e reutilizados (WOOD *et al.*, 2014). Ao invés de tratar de apenas um recurso, passamos a nos referir a um conjunto de recursos. Em vista disso, uma das principais formas para a descrição de *datasets* e suas ligações é a publicação de um arquivo baseado no *Vocabulary of Interlinked Datasets* (VoID)¹⁰ em conjunto com o *dataset*.

O VoID é um vocabulário que fornece termos e padrões para a descrição de metadados sobre *datasets* RDF. Podemos entender como principal objetivo do VoID tornar acessível o conteúdo de *datasets* aos usuários. A utilização do VoID permite que (ALEXANDER *et al.*, 2009): a) o *dataset* seja encontrado e agregado por ferramentas de buscas; b) sejam disponibilizadas informações sobre licenciamento dos dados, informando aos consumidores como os dados podem ser utilizados e como devem ser creditados; e c) consumidores possam obter informações sobre interfaces de acesso aos dados, como *Application Programming Interfaces* (API)¹¹ e *SPARQL EndPoints*¹². Além disso, o VoID também permite a descrição do conteúdo do *dataset*, sua localização, vocabulários utilizados e ligações com outros *datasets* (WOOD *et al.*, 2014).

Finalmente, para possibilitar o acesso ao *dataset* por terceiros, é necessária a publicação do mesmo na Web. Para que isso ocorra, pode-se criar e manter um repositório próprio de dados, disponibilizando um *SPARQL EndPoint* ou APIs de acesso a estes dados; pode-se submeter o *dataset* em algum servidor centralizador, ou ambos. Neste trabalho, focamos na submissão de *datasets* em servidores centralizadores como, por exemplo, o *Datahub*¹³.

O *Datahub* é uma ferramenta centralizadora de dados que funciona como um catálogo no qual é possível acessar ou registrar *datasets*. Esta foi a ferramenta utilizada pela iniciativa *LinkedUp* para a criação de um catálogo de *datasets* de teor acadêmico, o *LinkedEducationCatalog* (D'AQUIN, 2013). No total, o catálogo conta com o registro de 36 *datasets* que podem ser utilizados na construção de aplicações voltadas para propósitos acadêmicos¹⁴. Convém notar que o *Datahub* também permite a criação de entradas apenas para registrar metadados sobre *datasets*, como: documentação de APIs e endereço de *SPARQL Endpoints*, exemplos de triplas, arquivos de licença e arquivos de VoID. Nesse contexto, mesmo que seja adotada a abordagem de manter um repositório próprio de *Linked Data*, é relevante utilizar meios para divulgar informações sobre os dados que foram abertos. Esta divulgação pode ocorrer, por exemplo, por meio da catalogação do *dataset* no *Datahub*.

O *Datahub* fornece indicações de como um usuário pode cadastrar-se na ferramenta e registrar *datasets* em seu catálogo. Estes passos são sumarizados a seguir:

- 1) Criar um novo cadastro de usuário a partir do endereço <https://datahub.io/user/register> para obter acesso à ferramenta;
- 2) Selecionar a opção “+ AddDataset” na página principal de visualização de *datasets*. Nesse ponto devemos observar que para criar um novo *dataset* o *Datahub* exige que o usuário faça parte de uma “organização”, que pode ser entendida como o grupo responsável pelo *dataset*;
- 3) Uma vez selecionada a opção de criação, serão solicitadas informações sobre o novo *dataset*. Entre estas informações estão: título, descrição, marcadores para busca, *link* de origem dos dados e versão do *dataset*. Nesta página também é possível

especificar o tipo de licença de uso do *dataset* e informações sobre os autores e mantenedores do *dataset*;

- 4) O próximo passo permite que seja submetido um arquivo contendo os dados a serem disponibilizados no *Datahub*, além de adicionar informações adicionais sobre o arquivo submetido (nome, descrição e formato).

4.3 Diretrizes

Uma vez publicado, o *dataset* pode ser encontrado e utilizado em aplicações diversas. Assim, aplicações capazes baseadas em *Linked Data* estarão habilitadas a localizar os *datasets* por meio do catálogo ou de ligações provenientes de outros *datasets*. Serão capazes de compreender automaticamente os dados por meio dos vocabulários compartilhados e fornecer funcionalidades a partir destes dados. No Quadro 2 estão sintetizadas as principais diretrizes a serem seguidas para viabilizar a descoberta e uso de *Linked Data* na Web:

Quadro 2. Diretrizes para a Publicação de *Datasets* Abertos

Diretriz	Descrição
Utilização de URIs adequadas	As URIs são os identificadores únicos dos recursos nos <i>datasets</i> . Então, além dos princípios básicos do <i>Linked Data</i> , considere: <ul style="list-style-type: none"> - Utilizar apenas URIs de domínios que você controle; - Ocultar detalhes de implementação; - Utilizar chaves naturais nas URIs.
Vínculo com outros <i>datasets</i>	De modo a conectar-se a outros <i>datasets</i> e aumentar a visibilidade dos dados (tanto próprios como de terceiros) crie: <ul style="list-style-type: none"> - Ligações de relacionamento; - Ligações de identidade; e - Ligações de vocabulários. Também, quando possível, gere ligações a partir de outros <i>datasets</i> para o seu (ex.: a partir do <i>DBPedia</i>).
Reuso de Vocabulários	Buscar reutilizar vocabulários populares para que os dados possam ser compreendidos por aplicações que consomem dados de diferentes repositórios. Manter equilíbrio entre vocabulários populares e específicos de domínio.
Disponibilização de Metadados	Gerar metadados sobre o <i>dataset</i> por meio da utilização do vocabulário VoID, para facilitar seu reconhecimento por potenciais usuários.
Publicação do <i>dataset</i>	Submeta o <i>dataset</i> ou crie uma referência para ele em uma ferramenta centralizadora (ou catálogo) <i>online</i> , como o <i>Datahub</i> .

Fonte: Elaboração Própria

Por fim, convém salientar que as diretrizes aqui expostas são baseadas na literatura de *Linked Data*, buscando refletir as principais práticas utilizadas no processo de publicação de *datasets* abertos. Aliado a estas diretrizes, sugere-se também a consulta a grupos de referência na Web Semântica, como o W3C, e participação em listas de discussão sobre *Linked Open Data*, como o “*Public LOD*”¹⁵.

Os serviços do repositório de dados podem se concentrar na criação de recursos, na sua utilização ou em ambas. O mais importante é a iniciativa dos pesquisadores, dos profissionais, e das instituições mantenedoras, no sentido de tornarem os recursos abertos e interoperáveis, garantindo não só o seu reaproveitamento como também de seus metadados descritivos.

Para garantir uma maior disseminação da coleção de dados, Barker e Campbell (2016) recomendam que repositórios institucionais de suporte sejam interligados àqueles que incidem sobre conteúdos específicos como *Youtube* (vídeos) ou *Flickr* (imagens). Para a Ciência da Informação, o fundamento dos autores ganha maior repercussão pelos processos e ambientes de disseminação e apropriação da informação, mediados pelos profissionais de informação. Muitos usuários não são especialistas em tipos específicos de conteúdo. Logo, para satisfazer às suas necessidades informacionais, recorrem aos ambientes que se destacam na disseminação de tipos específicos de recursos que motivam o seu comportamento de busca, como, por exemplo, o *Youtube*. O uso dos princípios do *Linked Data* para disseminar os recursos nessas plataformas por meio de APIs e outras tecnologias alarga o escopo de uso e aplicação dos dados. Por exemplo, em um conjunto de bibliotecas digitais, esta pode ser uma alternativa para representar e vincular todas as manifestações de uma mesma obra, ainda que catalogadas em *datasets* distintos, para que sejam recuperadas em conjunto quando necessárias.

As manifestações, frequentemente designadas como versões, edições ou instanciações de uma obra ou recurso, são artefatos de pontos individuais no momento em que uma determinada obra é estabelecida, ou seja, registrada para a preservação ou disseminação. A obra é o registro de um conhecimento cujas instanciações classificam-se em: derivadas, quando não há alteração da semântica; ou do conteúdo ideacional e mutação, quando há alteração da semântica ou do conteúdo ideacional, ou mesmo de ambos (SMIRAGLIA, 2005). Por exemplo, se todas as edições de um livro (derivações) forem vinculadas pelos princípios do *Linked Data* e o filme inspirado nele (mutação) estiver disponível no *Youtube*, o recurso será recuperado com um todo. Consequentemente, os respectivos metadados descritivos não serão produzidos com base nas várias manifestações ocorridas, acarretando conflitos. Em outras palavras, serão coletados e reutilizados a partir da primeira derivação.

5 Conclusões e perspectivas

Este trabalho propõe um conjunto de diretrizes, para a publicação de *datasets* abertos na Web, de maneira que os mesmos possam ser identificados, compreendidos e utilizados corretamente. O conjunto de diretrizes aqui proposto toma por base os trabalhos de Tim Berners-Lee (2006); Bizer, Heath e Berners-Lee (2009); Heath e Bizer (2011) e Wood *et al.* (2014) sobre *Linked Data*. Procurou-se traçar um cenário cujas perspectivas remodelam os fundamentos e práticas incipientes na gênese da Ciência da Informação. Tais perspectivas podem ser destacadas sobre a capacidade de resiliência do campo, que a maioria das pesquisas aponta, face às mudanças impostas pela atual complexidade informacional. Por exemplo, ao dar a primazia sobre os dados, através dos revolucionários processos de descrição e compartilhamento, abre-se o espaço para a ecologia da informação e conhecimento em ação na CI.

Ao lado das pesquisas, existe uma infinidade de dados coletados e processados sob diversas tecnologias e perspectivas de análise. A sua disseminação não só evita a replicação da coleta, como garante novas inferências inseridas no contexto do uso. Assim, também se garante o crescimento exponencial do conhecimento, independentemente das limitações geográficas, temporais, tecnológicas, entre outras. Por estas razões, ao longo da nossa discussão reforçamos a necessidade de publicação dos dados em formatos abertos.

As constatações feitas em prol da disseminação, do acesso, e do reuso dos recursos informacionais nos permite concluir sobre a necessidade do comprometimento dos profissionais de informação, por meio de iniciativas de caráter diferente. Por exemplo, os tradicionais serviços de disseminação de informação como Bibliotecas, Arquivos e Museus podem usar os princípios do *Linked Data*, a partir dos estudos de usuários, para agregar APIs de outras plataformas em seus catálogos. Deste modo, o usuário pode acessar recursos hospedados fora destes ambientes, enriquecendo a satisfação da sua necessidade informacional. Estas tecnologias, se por um lado podem atrair mais usuários para estas unidades de informação, por outro lado podem expandir o âmbito de aplicação da maioria dos seus serviços.

5.1 Trabalhos futuros

Ao abordar o tema desta pesquisa em um nível conceitual, deparamo-nos com alguns aspectos que podem suscitar questionamentos em relação à aplicabilidade no contexto da CI. Assim, futuramente, pretendemos criar e publicar um conjunto de dados do nosso próprio grupo de pesquisa, seguindo os princípios do *Linked Data*.

Outra agenda de pesquisa consiste em aprofundar o conceito de *Linked Data* na Ciência da Informação, explorando outras vertentes da sua aplicação. Uma dessas vertentes envolve a área dos sistemas de recuperação da informação e da busca semântica.

Finalmente, para suprir as limitações desta pesquisa, trabalhos futuros deverão analisar também *datasets* de outras áreas, de maneira a alargar o espectro de projetos existentes relacionados à temática dos dados ligados abertos, no que tange a publicação, uso, reuso e interoperabilidade.

Referências

- ALEXANDER, Keith et al. Describing Linked Datasets: On the Design and Usage of void, the "Vocabulary of InterLinked Datasets". In: LINKED DATA ON THE WEB WORKSHOP, 2., 2009, Madri, Espanha. **LINKED DATA ON THE WEB**. Madri: LDOW2009, 2009. p. 1 - 10.
- BARKER, Phil; CAMPBELL, Lorna M. Technology strategies for open educational resource dissemination. In: BLESSINGER, Patrick; BLISS, Tj (Ed.). **Open education: international perspectives in high education**. Cambridge, Reino Unido: Openbook Publishers, 2016. p. 51-71.
- BERNERS-LEE, Tim. **Linked Data**. Disponível em: <<https://www.w3.org/DesignIssues/LinkedData.html>>. Acesso em: 21 maio 2018.
- BIZER, Christian; HEATH, Tom; BERNERS-LEE, Tim. **Linked Data: The Story So Far**. 2009. Disponível em: <<http://tomheath.com/papers/bizer-heath-berners-lee-ijswis-linked-data.pdf>>. Acesso em: 21 maio 2018.
- BLESSINGER, P.; BLISS, TJ. Introduction to open education: towards a human rights theory. In: Blessinger, P. e Bliss, TJ (ed.) **Open education: international perspectives in high education**. Cambridge, Reino Unido, Openbook Publishers, 2016. p. 11-30.
- BUCKLAND, Michael K. Information as thing. **Journal Of The American Society For Information Science**, Berkeley (CA), EUA, v. 5, n. 45, p.351-360, jun. 1991.
- CAPURRO, Rafael; HJORLAND, Birger. O conceito de informação. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 12, n. 1, p.148-207, jan. 2007. Disponível em: <<http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/54/47>>. Acesso em: 21 maio 2018.
- D'AQUIN, Mathieu et al. **LinkedUp Veni Competition: Linked and Open Data for Education**. 2013. Disponível em: <http://ceur-ws.org/Vol-1124/linkedup_veni2013_introduction.pdf>. Acesso em: 21 maio 2018.
- DIETZE, Stefan et al. Interlinking educational Resources and the Web of Data: a Survey of Challenges and Approaches. **Data Technologies And Applications**, [s.l.], v. 47, n. 1, p.60-91, jan. 2013. Disponível em: <<https://stefandietze.files.wordpress.com/2009/01/dietze-et-al-linkededucation-survey-final.pdf>>. Acesso em: 21 maio 2018.
- HEATH, Tom; BIZER, Christian. **Linked Data: Evolving the Web into a Global Data Space**. San Rafael (CA), EUA: Morgan & Claypool Publishers, 2011.
- INGWERSEN, P. Conceptions of information science. In: VAKKARI, P.; CRONIN, B. (Ed.). **Conceptions of library and information science: historical, empirical and theoretical perspectives**. Los Angeles: Taylor Graham, 1992. p.299-312.
- KEßLER, Carsten; D'AQUIN, Mathieu; DIETZE, Stefan. **Linked Data for Science and Education**. 2009. Disponível em: <<https://pdfs.semanticscholar.org/9c89/005b0e7e417b52c596c4eb5d3cc8fd1f43cf.pdf>>. Acesso em: 21 maio 2018.
- OPEN KNOWLEDGE INTERNATIONAL. **LinkedUp**. Disponível em: <<https://okfn.org/projects/linkdup/>>. Acesso em: 21 maio 2018.
- RIBEIRO JUNIOR, Divino Ignácio. Repositórios de dados para E-Science: Open Data, Linked Data e suas tecnologias. **Ciência da Informação**, Brasília, v. 42, n. 2, p.274-284, maio 2013. Disponível em: <<http://revista.ibict.br/ciinf/article/view/1386/1564>>. Acesso em: 21 maio 2018.
- SANT'ANA, Ricardo César Gonçalves. Ciclo de vida dos dados: uma perspectiva a partir da ciência da informação. **Informação & Informação**, Londrina, v. 21, n. 2, p.116-142, maio 2016. Disponível em: <<http://www.uel.br/revistas/uel/index.php/informacao/article/view/27940/20124>>. Acesso em: 21 maio 2018.
- SARACEVIC, Tefko. CIÊNCIA DA INFORMAÇÃO: ORIGEM, EVOLUÇÃO E RELAÇÕES. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 1, n. 1, p.41-62, jan. 1996. Disponível em: <<http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/235/22>>. Acesso em: 21 maio 2018.
- SCHAIBLE, Johann; GOTTRON, Thomas; SCHERP, Ansgar. Survey on Common Strategies of Vocabulary Reuse in Linked Open Data Modeling. In: EUROPEAN SEMANTIC WEB CONFERENCE (ESWC), 11., 2014, Creta, Grécia. **The Semantic Web: Trends and Challenges**. Creta, Grécia: ESWC2014, 2014. p. 457 - 472. Disponível em: <<https://pdfs.semanticscholar.org/9dbf/2c4352960c45f7b3183952c6f287b8cd7310.pdf>>. Acesso em: 21 maio 2018.

SEGUNDO, José Eduardo Santarem. Web Semântica: Introdução a recuperação de dados usando SPARQL. In: ENCONTRO NACIONAL DE PESQUISAS EM CIÊNCIA DA INFORMAÇÃO (ENANCIB), 14., 2014, Belo Horizonte. **Além das nuvens: expandindo as fronteiras da ciência da informação**. Belo Horizonte: Ufmg, 2014. p. 3242 - 3261. Disponível em: <https://www.researchgate.net/publication/304216019_WEB_SEMANTICA_INTRODUCAO_A_RECUPERACAO_DE_DADOS_USANDO_SPARQL>. Acesso em: 21 maio 2018.

SMIRAGLIA, Richard P. Content Metadata: An Analysis of Etruscan Artifacts in a Museum of Archeology. **Cataloging & Classification Quarterly**, Filadélfia (PA), EUA, v. 40, n. 3, p.135-151, ago. 2005. Disponível em: <<https://pdfs.semanticscholar.org/8e67/44aa77ac12476510d44a8ffb3370938f7654.pdf>>. Acesso em: 21 maio 2018.

WOOD, D. et al. **Linked Data**. Shelter Island (NY), EUA: Manning Publications, 2014.

W3C. **SPARQL 1.1 Update**: W3C Recommendation 21 March 2013. 2013. Disponível em: <<https://www.w3.org/TR/sparql11-update/>>. Acesso em: 21 maio 2018.

_____. **Linked Data Platform 1.0**: W3C Recommendation 26 February 2015. 2015. Disponível em: <<https://www.w3.org/TR/2015/REC-ldp-20150226/>>. Acesso em: 21 maio 2018.

YANG, Tao et al. Efficient SPARQL Query Evaluation via Automatic Data Partitioning. In: INTERNATIONAL CONFERENCE ON DATABASE SYSTEMS FOR ADVANCED APPLICATIONS (DASFAA), 18., 2013, Wuhan, China. **Database Systems for Advanced Applications**. Wuhan, China: DASFAA, 2013. p. 244 - 258. Disponível em: <<http://iir.ruc.edu.cn/~jchchen/rdfDasfaa13.pdf>>. Acesso em: 21 maio 2018.

Dados dos autores

Januário Albino Nhacuongue

Doutor em Ciência da Informação pela Universidade Estadual Paulista Júlio de Mesquita Filho, em estágio pós-doutoral no Programa de Pós-graduação em Ciência da Informação da Universidade Federal de Santa Catarina (PGCIN/UFSC).

januanita@hotmail.com

Vitor Rozsa

Mestrando do Programa de Pós-graduação em Ciência da Informação da Universidade Federal de Santa Catarina (PGCIN/UFSC).

vitor.rozsa@hotmail.com

Moisés Lima Dutra

Docente do Programa de Pós-graduação em Ciência da Informação da Universidade Federal de Santa Catarina (PGCIN/UFSC).

moises.dutra@ufsc.br

Recebido - Received: 2017-02-18

Aceitado - Accepted: 2018-12-11

¹ <http://data.europa.eu/euodp/en/data>

² O projeto teve início em novembro de 2012, vinculado à Universidade de Hanover Gottfried Wilhelm Leibniz, Alemanha (OPEN KNOWLEDGE INTERNATIONAL, 2018).

³ Transferência de conhecimentos técnicos, científicos ou *know-how* para a produção de bens e serviços. Esses conhecimentos podem resultar de pesquisas científicas de investigadores ou inovações tecnológicas de empresas. A transferência pode consistir em contratos (Instituto Nacional de Propriedade Industrial – INPI, no caso do Brasil) de caráter econômico ou não, permitindo que os usuários criem novos serviços, produtos, processos, materiais, etc.

⁴ Alguns dos esquemas citados são Dublin Core e IEEE Learning Object Metadata (LOM), este último voltado especificamente para a educação (KEßLER, D'AQUIN e DIETZE, 2013).

⁵ Cliente é um programa que estabelece conexões com a finalidade de enviar uma ou mais solicitações HTTP. Servidor é um programa que aceita conexões para atender a solicitações HTTP enviando respostas HTTP.

⁶ Mais informações em <https://www.w3.org/TR/2004/REC-owl-ref-20040210/#sameAs-def>

⁷ Mais informações em <https://www.w3.org/TR/2004/REC-owl-ref-20040210/#equivalentClass-def>

⁸ Mais informações em <https://www.w3.org/TR/2004/REC-owl-ref-20040210/#equivalentProperty-def>

⁹ Mais informações em <http://wiki.DBpedia.org/>

¹⁰ Mais informações em <https://www.w3.org/TR/void/>

¹¹ No contexto deste trabalho, APIs podem ser entendidas como ferramentas de *software* que fornecem recursos de acesso aos *datasets* e que são utilizadas na confecção de aplicações que consomem os dados.

¹² Um SPARQL EndPoint pode ser entendido como um servidor Web que recebe requisições (consultas) em SPARQL sobre o *dataset*.

¹³ Disponível em <https://Datahub.io/>

¹⁴ O LinkedEducationCatalog está disponível em <https://Datahub.io/organization/linkedin-education>

¹⁵ Disponível em <https://lists.w3.org/Archives/Public/public-lod/>



This work is licensed under a Creative Commons Attribution 4.0 United States License.



This journal is published by the [University Library System](#) of the [University of Pittsburgh](#) as part of its [D-Scribe Digital Publishing Program](#) and is cosponsored by the [University of Pittsburgh Press](#).