

Princípios FAIR e melhores práticas do *Linked Data* na publicação de dados de pesquisa

Luciana Candida da Silva

Universidade Estadual Paulista Júlio Mesquita Filho – UNESP, E-mail: candida.luciana@gmail.com

José Eduardo Santarem Segundo

Universidade de São Paulo – USP, E-mail: santarem@usp.br

Marcel Ferrante Silva

Universidade Federal de Goiás – UFG, E-mail: marcelf@gmail.com

RESUMO

As tecnologias da Web Semântica, seguindo as melhores práticas do *Linked Data*, proporcionam benefícios para a publicação de dados na Web. Os princípios FAIR orientam que os dados de pesquisa sejam publicados para serem encontráveis, acessíveis, interoperáveis e reusáveis. Contudo, não sugerem tecnologias específicas para sua implementação. Dessa forma, este estudo tem como objetivo identificar a possibilidade de publicar dados de pesquisa seguindo os princípios FAIR: *Findable, Accessible, Interoperable, Re-usable*, a partir da adoção das tecnologias da Web Semântica, por meio do *Linked Data*, como solução efetiva de sua implementação no ambiente Web. Esta pesquisa é de natureza qualitativa, do tipo descritiva-exploratória. Para tanto, descreveu sobre os dados de pesquisa em seu contexto tipológico e, em seguida, sobre os princípios FAIR e sobre o *Linked Data*. O *Linked Data* foi detalhado no conjunto de melhores práticas para publicação de dados na Web e seus benefícios. No segundo momento, analisou as orientações dos princípios FAIR e buscou, nas melhores práticas do *Linked Data*, as diretrizes e tecnologias apropriadas para publicação de dados de pesquisa na Web, de forma a atender os princípios FAIR. Como resultado, obteve-se que das trinta e cinco melhores práticas recomendadas, a adoção das tecnologias indicadas nas categorias metadados, licença, proveniência, identificadores, formatos, vocabulários e acesso a dados atendem aproximadamente oitenta por cento (80%) das orientações de FAIR para publicação de dados de pesquisa na Web, de forma a permitir aos dados, novos significados em pesquisas colaborativas e interdisciplinares.

Palavras-chave: Dados de pesquisa. Princípios FAIR. Melhores práticas. *Linked Data*. Web Semântica.

1 INTRODUÇÃO

A pesquisa científica visa produzir conhecimentos contribuindo para o avanço da ciência e para o desenvolvimento social. Os resultados dessas pesquisas são divulgados geralmente em suas configurações finais, seja no formato de teses ou dissertações, publicações de livros, artigos de periódicos e comunicações de congresso. Neste cenário, os dados que levam aos resultados são descartados ou armazenados em mídias ou em servidores sem a devida gestão quando os projetos são concluídos (SAYÃO; SALES, 2015). Contudo, as fontes de dados de pesquisa são

fundamentais e não são divulgadas juntamente com as publicações gerais, ficando inacessíveis aos demais pesquisadores e à sociedade.

A comunicação científica encontra-se diante do surgimento de um modelo contemporâneo de publicação científica que visa disseminar, de forma aberta, o conhecimento produzido por instituições de pesquisa. O movimento denominado *Open Science*, cenário do quarto paradigma da Ciência Aberta se insere nas formas de produção colaborativa e interativa da informação e do conhecimento, em âmbito mundial. No Brasil é possível observar movimentos que incentivam projetos de dados abertos, tais como a Lei de Acesso Aberto, nº 12.527, de 18 de novembro de 2011, que destina a assegurar o direito fundamental de acesso à informação; a *Open Knowledge Brasil*, também conhecida como Rede de Conhecimento Livre, que atua com diversos projetos sobre dados abertos e foi uma iniciativa divulgada em 2014; o Manifesto de Acesso Aberto a Dados da Pesquisa para a Ciência Cidadã, lançado pelo IBICT em 2016, que busca promover o acesso aberto aos dados de pesquisa e apoiar movimentos para Ciência Aberta no Brasil.

Esses movimentos incentivam a divulgação de dados científicos em formato aberto, com propósito de validar os resultados das pesquisas e aumentar a transparência, principalmente, quando se tratam de pesquisas financiadas por instituições públicas. Além disso, reforçam a ideia de que os dados de pesquisa são parte integrante do registro acadêmico e devem estar disponíveis para reuso.

Sendo assim, para tornar possível o acesso aberto a dados de pesquisa de qualidade, faz-se necessária a publicação de tais dados seguindo princípios que orientam a adoção de uma estrutura semântica legível por humanos e máquinas. Nesse sentido, este estudo tem como objetivo identificar a possibilidade de publicar dados de pesquisa seguindo os princípios FAIR: *Findable, Accessible, Interoperable, Re-usable*, a partir da adoção das tecnologias da Web Semântica, por meio do *Linked Data*.

Este estudo se alicerça no fato de os princípios FAIR apresentarem características requeridas pelos ambientes modernos da publicação científica, mas não indicarem tecnologias ou soluções de implementação. Por outro lado, a *World Wide Web Consortium (W3C)*¹ recomenda um conjunto de melhores práticas para publicação de dados na Web com diretrizes e tecnologias da Web Semântica e *Linked Data*.

Acredita-se que a publicação de dados de pesquisa a partir das orientações de FAIR e aplicação das tecnologias e conceitos da Web Semântica, por meio do *Linked Data*, indicadas nas melhores práticas de publicação permitirá aos dados, novos significados em pesquisas colaborativas e interdisciplinares.

2 PROCEDIMENTOS METODOLÓGICOS

Esta pesquisa é de natureza qualitativa, do tipo descritiva-exploratória, e busca identificar a possibilidade de publicar dados de pesquisa seguindo os princípios FAIR: *Findable, Accessible, Interoperable, Re-usable*, a partir da adoção das tecnologias da Web Semântica, por meio do *Linked Data*, como solução para a sua efetiva implementação. Para tanto, a primeira etapa descreveu sobre os dados de pesquisa científica em seu contexto tipológico e, em seguida, descreveu sobre os princípios FAIR. Logo, foram identificados os elementos da Web Semântica, com foco na aplicação do *Linked Data*. O *Linked Data* foi detalhado em seu conjunto de melhores práticas (MP) para publicação de dados na Web.

No segundo momento deste estudo, foram analisadas as orientações dos princípios Encontrar, Acessar, Interoperar e Reutilizar, e buscou-se nas trinta e cinco (35) MPs recomendadas pelo Consórcio W3C, as orientações e tecnologias para implementação dos princípios FAIR.

A literatura consultada partiu de buscas nas bases de dados Brapci e Portal de Periódicos Capes, além de livros, artigos de periódicos e sites oficiais da W3C (Melhores práticas do *Linked Data*) e Force11 (Princípios FAIR). Os principais termos de busca foram “Dados de pesquisa”, “Dados científicos”, “Web Semântica”, “*Linked Data*”, “Princípios FAIR”, “Publicação de dados”. Foram realizadas buscas booleanas para “Boas práticas” e “Publicação de dados”, “Melhores práticas” e “Publicação de dados”, “*Linked Data*” e “Dados de pesquisa”. Os mesmos termos foram consultados no idioma Inglês, onde se obteve o maior índice de respostas. O recorte temporal utilizado possui ênfase nos períodos de 1999 a 2017 no que se refere a dados de pesquisa científica, 2014 a 2016 para os princípios FAIR e 2001 a 2017 para *Linked Data* e melhores práticas para publicação de dados.

3 DADOS DE PESQUISA CIENTÍFICA

Os dados que alicerçam as publicações científicas são comumente nomeados dados de pesquisa / dados primários / dados brutos, para se referirem a dados sem inferências, discussões e

interpretações. A *Organization for Economic Co-Operation and Development* (OECD) (2007, p. 13, tradução nossa) define dados de pesquisa como “registros factuais usados como fonte primária para a pesquisa científica e que são aceitos pelos pesquisadores como necessários para validar os resultados do trabalho científico”. Segundo o *National Research Council* (NRC) (1999, p.15, tradução nossa), “os dados podem ser constituídos de fatos, números, letras e símbolos que descrevem um objeto, ideia, condição, situação ou outros fatores”.

Para Borgman (2010), a noção de dados pode variar consideravelmente entre os colaboradores, e ainda mais entre as áreas do conhecimento. Para Borgman (2010, p. 3, tradução nossa), “alguns tipos de dados têm valor imediato e são duráveis, alguns adquirem valor ao longo do tempo, alguns têm valor transitório e outros são mais fáceis de serem recriados do que curados”. Muitas dessas distinções dependem da categoria dos dados. A *National Science Board* (NSB) (2005) classificou os dados segundo a sua origem em observacional, computacional, experimental e registros.

Os **dados observacionais** incluem medidas meteorológicas e levantamentos de altitude, que podem ser associados a lugares e horários específicos ou podem envolver múltiplos lugares e tempos. Os **dados computacionais** resultam da execução de um modelo de computador ou simulação, seja física ou realidade virtual cultural. Os **dados experimentais** incluem testes laboratoriais. Os **registros** do governo, negócios e vida pública e privada também fornecem dados úteis para pesquisa científica, social científica e humanística (NATIONAL SCIENCE BOARD, 2005, p. 19, tradução nossa).

Para Green, MacDonald e Rice (2009), os dados podem ser classificados, segundo a sua natureza, em: números, imagens, vídeos, software, algoritmos, equações, animações ou modelos e simulações. Segundo sua fase de pesquisa, em: dados brutos ou preliminares (*Raw data*), os quais vêm diretamente dos instrumentos científicos; dados derivados, que são resultados do processo ou combinação de dados brutos ou de outros dados; e dados canônicos ou referenciais, que são coleções de dados consolidados e arquivados geralmente em grandes centros de dados, por exemplo, sequência genética, estrutura química, etc.

Os dados de pesquisa, nas suas diversas tipologias, vêm recebendo reconhecimento pela comunidade científica como parte essencial das boas práticas de pesquisa. Do ponto de vista do pesquisador, o acesso aberto a dados de pesquisa proporciona maior visibilidade, pois podem ser citados por membros mais proeminentes de seu campo e o pesquisador pode receber amplo destaque na comunicação científica. Para as instituições de ensino superior, os dados abertos de

pesquisa contribuem para melhorar o monitoramento e avaliação das atividades científicas e ao mesmo tempo proporcionar prestígio e visibilidade, além de democratizar o acesso. Do ponto de vista das entidades financiadoras de pesquisa, o depósito em acesso aberto das investigações em projetos justifica o investimento, cria transparência e evita duplicação de financiamento. Para a pesquisa, portanto, pode avançar para beneficiar a comunidade científica global (TENOPIR et al., 2015).

Ainda segundo Tenopir et al. (2015), o acesso aberto a dados de pesquisa promoveu um bem público, uma medida de economia de custos, uma maneira de aumentar a transparência e permitir a verificação de resultados de pesquisas anteriores. Neste sentido, muitas instituições de fomento à pesquisa consideram necessário que os dados resultantes de projetos financiados sejam gerenciados e compartilhados de forma a garantir maior benefício para o avanço científico e tecnológico.

Segundo Simionato (2017), grandes agências de fomento internacionais como a *National Science Foundation*, *National Institutes of Health*, *National Endowment for the Humanities*, *Economic and Social Research Council*, *Wellcome Trust* e a Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) exigem a apresentação do Plano de Gestão dos Dados (PGD) como requisito para análise do financiamento da pesquisa. Segundo Sayão e Sales (2015, p.15), “um PGD constitui em um documento formal que estabelece um compromisso de como os dados serão tratados durante todo o desenvolvimento do projeto, e também após a sua conclusão”.

A elaboração do PGD envolve as ações relacionadas ao ciclo de vida dos dados que perpassa pelas etapas de planejamento, coleta e criação dos dados, assegura a qualidade e descreve o dado a partir do uso de metadados apropriados, preserva em arquivos e em ambientes digitais adequados, possibilita a descoberta a partir de metadados, integra com outros dados e analisa o ciclo (SAYÃO; SALES, 2015).

Para Wilkinson et al. (2016) o bom gerenciamento de dados é o principal canal para a descoberta e inovação de conhecimentos e à subsequente integração e reutilização de dados após o processo de sua publicação. Os autores apresentam os princípios FAIR, sigla de *Findable*, *Accessible*, *Interoperable*, *Re-usable*, para orientar na elaboração do plano de gestão e publicação de dados.

3.1 PRINCÍPIOS FAIR

Os princípios FAIR originaram-se, em 2014, de um debate entre representantes de diversas áreas do conhecimento, entre eles pesquisadores, bibliotecários, arquivistas, editores e financiadores de pesquisas, membros da *The Future of Research Communications and e-Scholarship* (FORCE11), para melhorar o ecossistema dos dados de pesquisa e funcionar como diretrizes para aqueles que desejam aumentar a reutilização de seus dados, no contexto da *e-Science*. A partir dessa discussão, foi definido um conjunto de quatro princípios e práticas orientadoras para publicação de dados que fossem facilmente encontráveis, acessíveis, interoperáveis e reutilizáveis, por máquinas e humanos, frente a grande quantidade de informações geradas pela ciência contemporânea intensiva em dados (FORCE11, 2014).

A ideia central dos princípios FAIR está no aprimoramento da capacidade das máquinas de encontrar e usar automaticamente os dados, além de apoiar a sua reutilização por indivíduos. Sendo assim, tais princípios incorporam características que definem que os recursos, ferramentas, vocabulários e infraestrutura de dados contemporâneos deveriam ser exibidos para auxiliar na descoberta e reutilização de terceiros (FORCE11, 2014). Os princípios FAIR são:

Quadro 1 - Princípios FAIR

Para ser pesquisável /encontrável:

- F1. (meta) dados são atribuídos a um identificador globalmente exclusivo e persistente;
- F2. os dados precisam ser descritos com metadados ricos (definidos por R1 abaixo);
- F3. os metadados incluem clara e explicitamente o identificador dos dados que descreve;
- F4. (meta) dados são registrados ou indexados em um recurso pesquisável.

Para ser acessível:

- A1. (meta) dados são recuperáveis pelo seu identificador usando um protocolo de comunicação padronizado;
- A1.1 o protocolo é aberto, gratuito e universalmente implementável;
- A1.2 o protocolo permite um procedimento de autenticação e autorização, quando necessário;
- A2. os metadados são acessíveis, mesmo quando os dados não estão mais disponíveis;

Para ser interoperável:

- I1. (meta) dados usam uma linguagem formal, acessível, compartilhada e amplamente aplicável para a representação do conhecimento;
- I2. (meta) utilizam vocabulários e/ou ontologias que seguem os princípios FAIR;
- I3. (meta) incluem referências qualificadas para outros (meta) dados.

Para ser reutilizável:

- R1. os metadados são ricamente descritos com uma pluralidade de atributos precisos e relevantes;
- R1.1. (meta) dados são liberados com uma licença de uso de dados clara e acessível;
- R1.2. (meta) dados estão associados com a proveniência detalhada;
- R1.3. (meta) dados cumprem os padrões comunitários relevantes para o domínio.

* Destaca-se que a expressão “(meta) dados” é adotada nos casos em que o princípio deve ser aplicado a metadados e aos objetos de dados.

Fonte: Adaptado de Wilkinson et al. (2016).

Segundo Wilkinson et al. (2016), os elementos dos princípios FAIR estão relacionados, mas são independentes e separáveis e podem ser implementados em qualquer combinação, de forma incremental, à medida que os provedores de dados evoluem suas estruturas no sentido de atingir um grau maior dentro do propósito dos princípios FAIR. Os autores esclarecem que estes princípios precedem as escolhas de implementação e não sugerem tecnologias específicas para implementação. Sendo assim, este estudo adota as diretrizes do *Linked Data*, a partir de tecnologias da Web Semântica, para publicar dados semânticos de pesquisas com alta capacidade de encontrar, acessar, interoperabilizar e reutilizar dados.

4 LINKED DATA

O novo modelo de publicação científica tem proporcionado aos dados de pesquisa maior valor enquanto ativos publicáveis de forma autônoma e/ou ampliada. Para obter qualidade na recuperação dos dados de pesquisa é essencial publicar tais dados a partir de uma estrutura semântica legível por humanos e máquinas.

Segundo Berners-Lee, Lassila e Hendler (2001), a Web Semântica visa fornecer uma estrutura de significados semânticos aos dados da Web, criando um ambiente onde pessoas e softwares possam trabalhar em cooperação. Esse conceito se beneficia das potencialidades das tecnologias e amplia a ideia original da Web Semântica, que passou a consistir em fazer ligações entre conjuntos de dados, de modo que uma pessoa ou máquina possa explorar a Web de dados (BERNERS-LEE, 2006).

Como parte desse desenvolvimento surge o conceito de *Linked Data*, introduzido por Tim Berners-Lee, em 2006, para se referir a um conjunto de boas maneiras de como publicar e interligar dados semânticos na Web. Essas boas maneiras ficaram conhecidas como princípios de *Linked Data* para promover a ligação entre dados de diferentes fontes distribuídas na Web. Estes princípios são: 1 – usar *Uniform Resource Identifier* (URIs) como nomes para coisas; 2 – usar URI HTTP, para que as pessoas possam procurar esses nomes; 3 – quando alguém procurar um URI, fornecer informações úteis, usando padrões, como *Resource Description Framework* (RDF) e SPARQL; e 4 – incluir *links* para outros URIs, para que os itens relacionados possam ser descobertos (BERNERS-LEE, 2006).

Segundo Isotani e Bittercourt (2015), o *Linked Data* possui uma estrutura de dados baseada no modelo RDF que define a forma de representar o conhecimento por meio de triplas, estabelecendo relacionamentos entre os diversos recursos. As triplas são formadas pelos elementos sujeito, predicado e objeto. Assim, por exemplo, Waller, Andrew (sujeito) é autor (predicado) da obra *The librarians view* (objeto). O sujeito “Waller, Andrew” e o objeto “*The librarians view*” representam dois recursos que estão relacionados pelo predicado *dc:creator*, do vocabulário Dublin Core. O conjunto de estruturas em triplas forma um grafo RDF, sendo que um grafo pode possuir múltiplas triplas em um único documento RDF. Para que o computador possa interpretar os dados representados por estas triplas, é necessário que os elementos sejam descritos e referenciados por identificadores únicos de recursos. Os URIs proveem uma maneira única para identificar recursos e expressar relações entre eles. Assim, a relação RDF com URI é um conceito chave para referenciar e descrever de forma única e não ambígua os dados na Web (ISOTANI; BITTENCOURT, 2015).

Para Riley (2017), o *Linked Data* tornou-se o passo prático bem sucedido para a implantação da Web Semântica. Nos últimos anos, a W3C vem desenvolvendo tecnologias e aperfeiçoando boas práticas para o bom funcionamento da Web Semântica.

Neste contexto, a W3C recomenda publicar dados na Web seguindo um conjunto de Melhores Práticas (MP) para facilitar a interação entre publicadores e usuários de dados. A aplicação das melhores práticas garante os seguintes benefícios:

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

6. **Acesso** – facilita o acesso aos dados atualizados em diversas formas, tanto para os usuários como para os computadores.

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

8. **Processabilidade** – habilita os computadores a processar e a manipular automaticamente os dados contidos na Web (LÓSCIO; BURLE; CALEGARI, 2017).

Atualmente, a W3C, sob edição de Lóscio, Burle e Calegari (2017), recomenda a aplicação de trinta e cinco (35) melhores práticas para publicar dados na Web. Estas práticas estão distribuídas em categorias e para cada melhor prática obtém-se um conjunto de benefícios, como estão apresentados no quadro 2.

Quadro 2 - Melhores Práticas e Benefícios

Categoria	Melhor Prática	Benefícios
Metadados	Fornecer metadados para usuários humanos e máquinas	Reuso, compreensão, descoberta e processabilidade
	Fornecer metadados descritivos	Reuso, compreensão e descoberta
	Fornecer metadados estruturados	Reuso, compreensão e processabilidade
Licenças	Fornecer informações de licença de dados	Reuso e confiabilidade
Proveniência de dados	Fornecer informações completas sobre as origens dos dados e quaisquer alterações feitas	Reuso, compreensão e confiabilidade
Qualidade de dados	Disponibilizar informações sobre a qualidade de dados e adequações necessárias	Reuso e confiabilidade
Versão de dados	Atribuir versão para cada conjunto de dados	Reuso e confiabilidade
	Fornecer um histórico de versão	Reuso e confiabilidade
Identificadores de dados	Usar URIs persistentes de conjunto de dados	Reuso, interligação, descoberta e interoperabilidade
	Usar URIs persistentes como identificadores dentro de conjuntos de dados	Reuso, interligação, descoberta e interoperabilidade
	Atribuir URIs a versões de conjuntos de dados	Reuso, descoberta e confiança
Formatos de dados	Usar formatos de dados padronizados	Reuso e processabilidade
	Usar representações de dados locais neutras	Reuso e compreensão
	Fornecer dados em vários formatos	Reuso e processabilidade
Vocabulários de dados	Reutilizar vocabulários, de preferência padronizados	Reuso, processabilidade, compreensão, confiança e interoperabilidade
	Escolher o nível de formalização correto	Reuso, compreensão e interoperabilidade

Acesso a dados	Permitir o acesso completo (em massa)	Reuso e acesso
	Permitir o acesso parcial ao conjunto de dados	Reuso, acesso, interligação e processabilidade
	Disponibilizar dados em vários formatos	Reuso e acesso
	Permitir o acesso em tempo real	Reuso e acesso
	Fornecer dados atualizados	Reuso e acesso
	Explicar os motivos de quando os dados não estiverem mais disponíveis	Reuso e confiabilidade
	Disponibilizar dados através de uma API	Reuso, processabilidade, interoperabilidade e acesso
	Usar padrões da Web como base de APIs	Reuso, interligação, interoperabilidade, descoberta, acesso e processabilidade
	Fornecer documentação à medida que adicionar ou alterar uma API	Reuso e confiabilidade
	Evitar quebrar alterações na sua API	Reuso e interoperabilidade
Preservação de dados	Preservar o identificador e fornecer informações sobre o recurso arquivado	Reuso e confiabilidade
	Avaliar a cobertura de um conjunto de dados antes da sua preservação	Reuso e confiabilidade
Feedback	Coletar <i>feedback</i> dos consumidores	Reuso, compreensão e confiabilidade
	Disponibilizar publicamente o <i>feedback</i>	Reuso e confiabilidade
Enriquecimento de dados	Enriquecer dados gerando novos dados	Reuso, compreensão, confiabilidade e processabilidade
	Oferecer apresentações complementares	Reuso, compreensão, acesso e confiabilidade
Republicação	Fornecer <i>feedback</i> ao publicador original	Reuso, interoperabilidade e confiabilidade
	Seguir os termos de licença	Reuso e confiabilidade
	Citar a publicação original	Reuso, descoberta e confiabilidade

Fonte: Adaptado de Lóscio, Burle e Calegari (2017).

As melhores práticas se aplicam para a publicação de dados abertos e para publicação de dados não abertos. Segundo a *Open Knowledge Internacional* (2004), os dados são abertos quando qualquer pessoa pode livremente acessá-los, utilizá-los, modificá-los e compartilhá-los para qualquer finalidade, estando sujeito a, no máximo, as exigências que visem preservar sua proveniência e sua abertura. A publicação a partir da abordagem *Linked Open Data* é realizada sob uma licença aberta.

A publicação de dados abertos e conectados pode validar a pesquisa e colaborar para a geração do conhecimento científico. Para tanto, os dados precisam ser publicados seguindo orientações e tecnologias amplamente utilizadas, como as orientações FAIR e as tecnologias indicadas nas boas práticas, recomendadas pela W3C.

5 RESULTADOS E DISCUSSÕES: CORRELAÇÃO ENTRE OS PRINCÍPIOS FAIR E AS MELHORES PRÁTICAS DO LINKED DATA

Esta seção discute e correlaciona os princípios FAIR (encontrar, acessar, interoperar e reutilizar) com as categorias das melhores práticas do *Linked Data* recomendadas pela W3C para buscar atender as características requeridas pelos ambientes contemporâneos de publicação de dados científicos.

5.1 FINDABLE / ENCONTRÁVEL

As primeiras orientações dos princípios FAIR destinam-se a tornar os dados pesquisáveis e encontráveis a partir da descrição de metadados, da atribuição de identificadores persistentes e do registro em um recurso pesquisável. Nesse sentido, pode-se aplicar as melhores práticas relacionadas ao fornecimento de metadados (MP 1, 2 e 3), identificadores (MP 9, 10 e 11) e formatos padronizados legíveis por máquinas (MP 12), conforme apresentadas no quadro 3.

Quadro 3 - Faceta *Findable* e Melhores Práticas

FAIR		Melhores Práticas	
Encontrável (<i>Findable</i>)	F1. (meta) dados são atribuídos a um identificador globalmente exclusivo e persistente	Identificadores de dados	MP 9 - Usar URIs persistentes de conjunto de dados
			MP 10 - Usar URIs persistentes como identificadores dentro de conjuntos de dados
			MP 11 - Atribuir URIs a versões e séries de conjuntos de dados
	F2. os dados precisam ser descritos com metadados	Metadados	MP 1 – Fornecer metadados para humanos e máquinas
F3. os metadados incluem clara e explicitamente o identificador dos dados que descreve	MP 2 – Fornecer metadados descritivos		
		MP 3 – Fornecer metadados estruturados	
F4. (meta) dados são registrados ou indexados em um recurso pesquisável	Formato de dados	MP 12 – Usar formatos padronizados legíveis por máquina	

Fonte: Elaborado pelos autores (2018).

Os metadados permitem a descrição detalhada de itens, a identificação e o acesso aos dados com qualidade (MP 1, 2 e 3), atendendo as subfacetas F1, F2 e F3 dos princípios FAIR.

Neste contexto, os editores de dados têm a importante contribuição de facilitar a acessibilidade aos seus conjuntos de dados, adicionando metadados descritivos com informações que colaboram para a compreensão do significado e auxiliam na localização, no uso e reuso dos dados, contemplando a subfaceta reuso (R1) dos princípios FAIR.

Para a publicação de dados de pesquisa, no contexto semântico, é necessário oferecer metadados interpretáveis pelos usuários humanos e máquinas. O uso de metadados é indicado para encontrar e acessar objetos de dados que estejam descritos suficientemente para pessoas e máquinas distinguirem um objeto de outro. Para implementar metadados para leitura humana, Lóscio, Burle e Calegari (2017) recomendam fornecer metadados como parte de uma página da Web HTML e como um arquivo de texto separado. Enquanto que para a interpretação de máquinas, os metadados podem ser fornecidos em um formato de serialização Turtle e JSON, ou podem ser incorporados na página HTML (HTML-RDFA ou JSON-LD), e reutilizar padrões existentes e vocabulários populares, como por exemplo, o padrão de metadados Dublin Core. Segundo Simionato (2017), o padrão Dublin Core possui grande influência na descrição de recursos digitais, tornando-se um dos principais padrões para a representação de dados científicos. O padrão Dublin Core apresenta elementos na versão simplificada e na versão qualificada que refinam a semântica dos elementos. De acordo com as melhores práticas, os metadados possibilitam os benefícios do reuso, compreensão, descoberta e processabilidade, enquanto que os princípios FAIR destacam a importância e uso dos metadados em todas as suas facetas.

A subfaceta F1, identificador globalmente exclusivo e persistente, é contemplada pelas MPs 9, 10 e 11. Para publicar dados de pesquisa na Web, segundo os conceitos do *Linked Data*, o identificador único de recurso é um elemento determinante, visto que os URIs podem estabelecer relações entre os recursos. Para os recursos serem encontráveis, os URIs indicados devem ser persistentes para garantir que os dados fiquem disponíveis ao longo do tempo. A MP 9 recomenda usar serviços de redirecionamento, como o *purl.org*, para encaminhar a um novo endereço, caso a localização antiga deixe de funcionar. A MP 10 indica que o reuso de URIs como identificadores persistentes dentro dos conjuntos de dados são aconselháveis, quando possível. Da mesma maneira, a MP 11 orienta a atribuição de indicadores para versões individuais de conjuntos de dados, bem como à série geral (LÓSCIO; BURLE; CALEGARI,

2017). Destaca-se que o uso de URIs permite realizar conexões entre vários *datasets* agregadores e atribuir valores omissos aos dados brutos existentes. Além disso, com o uso de URIs os dados de pesquisa podem se conectar a outras publicações científicas relacionadas a elas, como artigos de revistas e publicações de eventos, livros e capítulos de livros, teses e dissertações. Ademais, os dados destas publicações ampliadas podem ser enriquecidos através da conexão com vocabulários e ontologias como *Library of Congress Subject Headings* (LCSH), *Friend of a Friend* (FOAF), *Virtual International Authority File* (VIAF), GeoNames dentre outros.

Para os dados serem encontráveis, a subfaceta F4 orienta que sejam registrados ou indexados em um recurso pesquisável. Nesse sentido, a MP 12 recomenda o uso de formatos padronizados e legíveis por máquinas ao publicar dados na Web, pois na medida em que os dados se tornam mais onipresentes e os conjuntos de dados se tornam maiores e mais complexos, o processamento por computadores torna-se indispensável. Espera-se que as máquinas possam ler e processar os dados publicados na Web e os humanos possam usar ferramentas disponíveis no domínio relevante para trabalhar com eles. Sendo assim, entre os formatos indicados incluem, mas não limita, a sintaxe de serialização CSV, XML, HDF5, JDON e RDF como RDF/XML, JSON ou Turtle (LÓSCIO; BURLE; CALEGARI, 2017).

5.2 ACCESSIBLE / ACESSÍVEL

As melhores práticas do *Linked Data* se propõem a fornecer fácil acesso a dados na Web, permitindo que humanos e máquinas aproveitem os benefícios do compartilhamento usando infraestrutura da Web (LÓSCIO; BURLE; CALEGARI, 2017). As melhores práticas, 17 a 26, apresentam abordagens que podem colaborar com os princípios FAIR para tornar os dados acessíveis na Web. Porém, não foi possível correlacionar com exatidão todos os itens da faceta *accessible*/acessível, conforme exposto no quadro 4.

Quadro 4 - Faceta *Accessible* e Melhores Práticas

FAIR		Melhores Práticas	
Acessível (<i>Accessible</i>)	A1. (meta) dados são recuperáveis pelo seu identificador usando um protocolo de comunicação padronizado	Acesso a dados	MP 23 – Disponibilizar dados por meio de uma API
	A1.1 o protocolo é aberto, gratuito e universalmente implementável		MP 24 – Usar padrões da Web como base de API
	A1.2 o protocolo permite um procedimento de autenticação e autorização, quando necessário		*Não foram identificadas diretrizes específicas que atendem este requisito
	A2. os metadados são acessíveis, mesmo quando os dados não estão mais disponíveis		MP 22 – Fornecer uma explicação para dados que não estão disponíveis (<i>MP Adaptada para a faceta A2</i>)

Fonte: Elaborado pelos autores (2018).

No contexto da comunicação científica, as tecnologias da Web Semântica por meio do *Linked Data* possibilitam o acesso ao conjunto completo de dados de uma determinada pesquisa. Esse conjunto pode ser constituído por dados primários de pesquisa e também resultado de uma publicação ampliada composta por dados de pesquisa, tese ou dissertação, artigos de revistas e de comunicações de eventos. Para o fácil acesso a estes conjuntos de dados, a MP 17 recomenda que a infraestrutura da Web deva ser implantada de modo a permitir o acesso em massa de um conjunto de dados completo com apenas um pedido, evitando inconsistência no acesso individual de dados ao longo de muitas recuperações, bem como permitir o fornecimento de subconjuntos de dados (MP18), caso os consumidores não precisem do conjunto completo. Para a implementação de um conjunto de dados que contenham vários arquivos é necessário tornar os dados acessíveis para *download* por meio de um URI (LÓSCIO; BURLE; CALEGARI, 2017).

Ainda segundo Lóscio, Burle e Calegari (2017), por padrão, a Web oferece acesso usando métodos de protocolo de transferência de hipertexto (HTTP) para *download* simples, em massa, de um arquivo. Ainda que os dados estejam distribuídos em vários URIs, estes podem ser organizados em um modelo de contêiner, através do protocolo de transferência de arquivos, para facilitar o acesso em massa aos dados. Os dados distribuídos em vários arquivos podem ser recuperados através de uma interface de programação de aplicativos (API), método de recuperação mais sofisticado. Uma API geralmente é a abordagem mais flexível para servir subconjuntos de dados, pois permite a personalização de quais dados são transferidos, tornando os subconjuntos disponíveis muito mais prováveis de fornecer os dados necessários (MP18). De

acordo com a MP 23, uma API geralmente é a abordagem mais flexível e oferece capacidade de processamento para os consumidores de dados. A MP 20 indica o uso de API para o fornecimento de dados em tempo real e a MP 23 indica a API para ativar o uso de dados em tempo real. No caso de dados de pesquisa, o tempo dependerá do tipo de pesquisa que está sendo realizada, enquanto que para o acesso a dados de uma publicação ampliada poderá não ocorrer em tempo real em função dos trâmites das defesas públicas, avaliações por pares e publicações em repositórios.

A MP 24 reforça que as APIs criadas com base nos padrões da Web aproveitam pontos fortes, como por exemplo, usar verbos HTTP como métodos e URIs que mapeiam diretamente para recursos individuais e ajudam a evitar um acoplamento rígido entre solicitações e respostas (LÓSCIO; BURLE; CALEGARI, 2017).

A subfaceta A2 orienta que os metadados devam ficar acessíveis, mesmo quando os dados não estão disponíveis, enquanto que a W3C orienta, por meio da MP 22, fornecer explicações para dados que não estão disponíveis, informando sobre como podem ser acessados e quem pode acessá-los. Nesse caso, as metodologias FAIR e MP podem se complementar disponibilizando metadados, mesmo quando os dados não estiverem disponíveis, e ainda assim oferecer mensagens explicativas. Uma possível abordagem de sua implementação é a publicação de um documento HTML que forneça tais explicações legíveis para a indisponibilidade de dados. As mensagens podem ser apresentadas a partir de códigos de status HTTP, como por exemplo, as mensagens 303 (ver outro), 410 (permanentemente removido), 503 (serviços indisponíveis), dentre outros.

Para a subfaceta A1.2, não foram identificadas nas MPs diretrizes específicas que atendam com exatidão o procedimento de autenticação e autorização de dados.

5.3 INTEROPERABLE / INTEROPERÁVEL

Os dados podem ser interoperáveis, quando são processáveis por máquina e permitem a troca efetiva de conteúdo entre sistemas, por meio de metadados que descrevem esse conteúdo, para que os sistemas possam efetivamente perfilar material entrante e combiná-los com suas estruturas internas (RILEY, 2017). Para os dados se tornarem interoperáveis, são necessários

identificadores (MP 9 e 10) e metadados (MP 1 a 3), discutidos no item 5.1, além de vocabulários (MP 15 e 16), conforme apresentados no quadro 5.

Quadro 5 - Faceta *Interoperable* e Melhores Práticas

	FAIR		Melhores Práticas
Interoperável <i>(Interoperable)</i>	I1. (meta) dados usam uma linguagem formal, acessível, compartilhada e amplamente aplicável para a representação do conhecimento	Vocabulários de dados	MP 15 - Reutilizar vocabulários, preferencialmente padronizados para codificar dados e metadados.
	I2. (meta) utilizam vocabulários e/ou ontologias que seguem os princípios FAIR		MP 16 – Escolher o nível de formalização correto
	I3. (meta) incluem referências qualificadas para outros (meta) dados	Acesso a dados	MP 22 – Fornecer uma explicação para dados que não estão disponíveis (<i>MP adaptada</i>)

Fonte: Elaborado pelos autores (2018).

Para promover a interoperabilidade entre conjuntos de dados é importante adotar vocabulário de dados. No cenário da Web Semântica, os vocabulários (MPs 15 e 16) exercem papel fundamental na integração dos dados, pois definem os conceitos e relacionamentos usados para descrever e representar uma área de interesse. A recomendação de Lóscio, Burle e Calegari (2017) é a de fazer uso de vocabulários amplamente utilizados e padronizados para codificar dados e metadados. Entre os vocabulários que atendem a essa recomendação, encontram-se o padrão de metadados Dublin Core, FOAF e o *Simple Knowledge Organization System* (SKOS). Para Lóscio, Burle e Calegari (2017), os vocabulários oferecem, além da interoperabilidade, os benefícios de processabilidade, compreensão, confiabilidade e reuso dos dados.

Para atender a subfaceta I3, acerca de referências qualificadas entre metadados, recomendam-se as diretrizes da MP 22 para o uso de uma explicação contextual sobre a situação dos dados e de seus metadados. A implementação pode ocorrer por meio da publicação de um documento HTML, com explicações legíveis para a situação dos dados.

5.4 RE-USABLE / REUSÁVEL

Para os dados serem reutilizáveis precisam ser ricamente descritos, compatíveis com os princípios anteriores (encontrável, acessível e interoperável) e integrados com outras fontes de dados, para permitir a citação adequada (WILKINSON et al. 2016). Assim também, para atingir o benefício de reuso dos dados, é recomendável a adoção das atuais trinta e cinco (35) melhores

práticas para publicação de dados na Web. Contudo, este estudo restringiu a descrição das MPs que correlacionam com as facetas apresentadas no quadro 6.

Quadro 6 - Faceta *Re-Usable* e Melhores Práticas

FAIR		Melhores Práticas	
Reutilizável (Re-usable)	R1. (meta) dados são ricamente descritos com uma pluralidade de atributos precisos e relevantes	Metadados	MP 1 – Fornecer metadados
			MP 2 – Metadados descritivos
			MP 3 – Metadados estruturais
	R1.1. (meta) dados são liberados com uma licença de uso de dados clara e acessível	Licença de dados	MP4 – Fornecer informações de licença de dados
R1.2. (meta) dados estão associados com a proveniência detalhada	Proveniência de dados	MP5 – Fornecer informações completas sobre as origens dos dados e quaisquer alterações feitas.	
R1.3. (meta) dados cumprem os padrões comunitários relevantes para o domínio		*Não identificadas diretrizes específicas para este requisito	

Fonte: Elaborado pelos autores (2018).

Para os dados serem reutilizáveis, os princípios FAIR orientam metadados ricamente descritos com uma pluralidade de atributos precisos e relevantes (R1), os quais estão contemplados nas MPs 1 a 3 referentes ao fornecimento de metadados descritivos e estruturais, conforme apresentados no item 5.1.

Considerando a abordagem da publicação e compartilhamento de dados em formato aberto, a subfaceta R1.1 orienta o fornecimento de licença dos dados, ou seja, disponibilizar as condições que o publicador dos dados dá ao usuário para utilização dos mesmos. Da mesma maneira, a W3C recomenda, por meio da MP 4, o fornecimento de termos da licença, seja o *link* ou cópia do contrato da licença que controla o uso dos dados. O termo pode ser explicitado pelo elemento *dcterms:license* do padrão DCTERMS, *cc:license* do *Creative Commons*, *schema:license* do Schema.org ou *xhtml:license* do XHTML. De acordo com Pires (2015), sem a informação da licença de uso dos dados, não há garantias de que a utilização esteja resguardada pelos princípios de abertura de dados.

A subfaceta R1.2, metadados associados com a proveniência detalhada, é inerente ao fornecimento de informações completas sobre as origens dos dados e quaisquer alterações feitas. A MP 5 recomenda a disponibilidade de metadados que incluam o criador, o editor e data de

emissão, por meio dos metadados *dc:creator*, *dc:publisher* e *dc:issued*, como forma de apresentar informações sobre a proveniência e proporcionar confiança e credibilidade dos dados que estão sendo compartilhados.

Além das melhores práticas correlacionadas no quadro 6, Lóscio, Burle e Calegari (2017) recomendam as MPs 12 a 14 referentes aos formatos de dados para obter os benefícios da processabilidade e do reuso de dados. A MP 12 aconselha o uso de formatos padronizados e legíveis por máquinas para permitir a interoperabilidade e reusos futuros, conforme apresentados no item 5.1.

A MP 13 orienta disponibilizar informações sobre parâmetro de localidade, para evitar dificuldades de compreensão dos dados que se modificam de um idioma para outro, como por exemplo, datas e moedas; o recomendável é usar estruturas e valores de dados localmente neutros ou fornecer metadados sobre a localidade usada pelos valores de dados, como por exemplo *dc:language* do padrão Dublin Core. O valor pode ser conectado ao catálogo de códigos para representação de nomes de idiomas da *International Organization for Standardization* (ISO) 639, publicada pela *Library of Congress* como `<dc:language xsi:type="dcterms:ISO639-2">en</dc:language>`.

A MP 14 recomenda sempre que possível a disponibilidade de dados em múltiplos formatos, quando mais de um se adequa ao seu uso pretendido. São exemplos de formatos indicados pela W3C, o RDF/XML e o Turtle.

Para a subfaceta R1.3, não foi possível identificar nas melhores práticas diretrizes específicas que cumprem os padrões comunitários relevantes para o domínio.

Os autores Lóscio, Burle e Calegari (2017), recomendam ainda as categorias qualidade, indicadores de versão, preservação, *feedback* e republicação de dados para melhorar o reuso ao publicar dados na Web.

6 CONSIDERAÇÕES FINAIS

Este estudo apresentou possibilidades de publicar dados de pesquisa seguindo os princípios FAIR: *Findable*, *Accessible*, *Interoperable*, *Re-usable*, a partir da adoção das tecnologias da Web Semântica, indicadas nas melhores práticas do *Linked Data*.

Os princípios FAIR trazem orientações para publicação de dados de pesquisa, mas não indicam tecnologias específicas para a sua implementação. Assim, analisamos nas melhores práticas do *Linked Data* as diretrizes e as tecnologias consideradas apropriadas para publicar dados de pesquisa na Web, segundo os princípios FAIR. Essa análise permitiu identificar que as principais categorias das MPs que atendem aos princípios FAIR são metadados, licença, proveniência, identificadores persistentes, formatos de dados, vocabulários e acesso a dados. A categoria enriquecimento de dados, apesar de não estar explícita nas facetas de FAIR, é altamente recomendada, pois permite a agregação de valores omissos aos dados existentes e melhora a qualidade da recuperação dos dados. Além disso, as tecnologias necessárias a sua implementação estão presentes nas categorias indicadas neste estudo. Outras categorias que podem colaborar na publicação dos dados são a qualidade, indicação da versão dos dados, preservação, *feedback* e republicação de dados. Contudo, fazem-se necessários estudos mais aprofundados acerca de diretrizes e tecnologias que permitam um procedimento de autenticação e autorização, quando necessário (A1.2), referências qualificadas para outros metadados (I3) e padrões comunitários relevantes para o domínio (R1.3). Os itens A1.2, I3 e R1.3 dos princípios de FAIR não foram identificados ou foram adaptados nas MPs do *Linked Data*. Outrossim, foi possível concluir que aproximadamente oitenta por cento (80%) das orientações de FAIR estão contempladas nas melhores práticas do *Linked Data* para publicação de dados na Web.

Logo, pode-se aferir que as categorias das melhores práticas recomendadas pela W3C oferecem diretrizes e tecnologias que vêm ao encontro das propostas dos princípios de FAIR de tornar os dados encontráveis, acessíveis, interoperáveis e reutilizáveis, permitindo a eles, novos significados em pesquisas colaborativas e interdisciplinares.

Vale salientar que outras tecnologias de publicação de dados podem ser testadas para publicar dados de pesquisa científica seguindo os princípios FAIR.

Além da correlação entre as metodologias de FAIR e MPs, este estudo deixou evidente os desafios trazidos pelo novo modelo de publicação científica que requer dos pesquisadores e dos profissionais da informação o conhecimento de novos instrumentos, formatos e tecnologias para descrição e disponibilização dos dados, de forma a tornar fácil encontrá-los e compreendê-los.

Entende-se que, para haver uma boa recuperação e compreensão, os dados devem ser ricamente descritos. Nesse sentido, sugerimos para trabalhos futuros, estudos que abarcam padrões

para representação da informação, tais como metadados, controle de autoridades, vocabulários de assunto, tesouros, dentre outros que favorecem a qualidade dos dados ao serem publicados em um sistema de gerenciamento de dados. Sugerimos também a publicação de relatos de experiência de pesquisadores e profissionais que estejam reutilizando e publicando dados de pesquisa no Brasil.

Fair Principles and best practices of the Linked Data in the publication of research data

ABSTRACT

Linked Data, as Semantic Web technologies, following the best practices of Linked Data, provide benefits for the publication of data on the Web. FAIR principles guide research data to be published, to be findable, accessible, interoperable and reusable, yet does not suggest technologies implementation. Thus, this study aims to identify the possibility of publishing research data following FAIR principles: Findable, Accessible, Interoperable, Re-usable, from the adoption of Semantic Web technologies, through Linked Data, as an effective solution implementation in the Web environment. This is a qualitative research, a kind of descriptive-exploratory one. To do so, It described the research data in its typological context and then described it on the principles of FAIR and Linked Data. Linked Data was detailed in the set of best practices for publishing data on the Web and its benefits. In the second stage, it analyzed the guidelines of the FAIR principles and researched in Linked Data's best practices for the appropriate guidelines and technologies for publishing research data on the Web in order to comply with FAIR principles. As a result, the adoption of the technologies indicated in the categories metadata, license, provenance, identifiers, formats, vocabularies and data access met approximately eighty percent (80%) of FAIR guidelines for the thirty-five best practices recommended for publication of research data on the Web in order to allow data to have new meanings in collaborative and interdisciplinary research.

Keywords: Search data. FAIR principles. Best practices. Linked Data. Semantic Web.

REFERÊNCIAS

BERNERS-LEE, T.; HENDLER, J.; LASSILA, O. The Semantic Web: a new form of Web content that is meaningful to computers will unleash a revolution of new possibilities. **Scientific American**, New York, 17 may 2001.

BERNERS-LEE, T. **Linked Data**. 2006. Disponível em: <https://www.w3.org/DesignIssues/LinkedData.html>. Acesso em: 13 jun. 2018.

BORGMAN, C. Research data: who will share what, with whom, when, and why? In: CHINA--NORTH AMERICAN LIBRARY CONFERENCE, 5., 2010, Beijing. Disponível em: http://www.ratswd.de/download/RatSWD_WP_2010/RatSWD_WP_161.pdf>. Acesso em: 13 jun. 2018.

FORCE11. The Future of Research Communications and e-Scholarship. **Guiding principles for findable, accessible, interoperable and reusable data publishing version B1.0**. 2014. Texto digital. Disponível em: <https://www.force11.org/fairprinciples>. Acesso em: 13 jun. 2018.

GREEN, A.; MACDONALD, S.; RICE, R. **Policy-making for research data in repositories: a guide**. May 2009 Version 1.2. Disponível em: <https://www.coar-repositories.org/files/guide.pdf>. Acesso em: 13 jun. 2018.

IBICT. INSTITUTO BRASILEIRO DE INFORMAÇÃO E TECNOLOGIA. **Manifesto de acesso aberto a dados de pesquisa brasileira para Ciência Cidadã**. 2016. Texto digital. Disponível em: http://www.ibict.br/Sala-de-Imprensa/noticias/2016/ibict-lanca-manifesto-de-acesso-aberto-a-dados-da-pesquisa-brasileira-para-ciencia-cidada/#_ftn1. Acesso em: 7 dez. 2017.

ISOTANI, S.; BITTENCOURT, Ig I. **Dados abertos conectados**. São Paulo: Novatec, 2015. Disponível em: <http://ceweb.br/livros/dados-abertos-conectados/>. Acesso em: 10 nov. 2017.

LÓSCIO, B. F.; BURLE, C.; CALEGARI, N. **Data on the Web best practices**. W3C, 2017. Texto digital. Disponível em: <https://www.w3.org/TR/dwbp/>. Acesso em: 02 jun. 2018.

NATIONAL RESEARCH COUNCIL. **A question of balance: private rights and the public interest in scientific and technical databases**. Washington: NRC, 1999. Disponível em: <https://www.nap.edu/read/9692/chapter/1#ii>. Acesso em: 02 jun. 2018.

NATIONAL SCIENCE BOARD. **Long-lived digital data collections: enabling research and education in the 21st century**. Alexandria, USA: National Science Foundation, 2005. Disponível em: <https://www.nsf.gov/pubs/2005/nsb0540/nsb0540.pdf>. Acesso em: 2 jul. 2018.

PIRES, M. T. **Guia de dados abertos**. São Paulo: NIC.br, 2015. Disponível em: <http://ceweb.br/guias/dados-abertos/>. Acesso em: 08 ago. 2018.

OPEN KNOWLEDGE INTERNATIONAL. **The open data handbook**. 2004. Disponível em: <http://opendatahandbook.org/guide/en/what-is-open-data/>. Acesso em: 24 set. 2017.

ORGANIZATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT (OECD). **Principles and Guidelines for Access to Research Data from Public Funding**. OECD, 2007. Disponível em: <http://www.oecd.org/sti/sci-tech/38500813.pdf>. Acesso em: 10 nov. 2017.

RILEY, J. **Understanding metadata: what is metadata, and what is it for?** Baltimore, MD: NISO Primer, 2017. Disponível em: http://www.niso.org/apps/group_public/download.php/17446/Understanding%20Metadata.pdf. Acesso em: 02 jun. 2018.

SAYÃO, L. F.; SALES, L. F. **Guia de gestão de dados de pesquisa para bibliotecários e pesquisadores**. Rio de Janeiro: CNEN, 2015.

SIMIONATO, A. C. Mapeamento dos metadados para dados científicos. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 18. 2017. Marília, SP: UNESP, 2017. Disponível em: <http://enancib.marilia.unesp.br/index.php/xviiienancib/ENANCIB/paper/viewFile/563/874>. Acesso em: 6 jun. 2018.

TENOPIR, C. et al. Changes in data sharing and data reuse practices and perceptions among scientists worldwide. **PLoS One**, v. 10, n.8, 2015. DOI 10.1371/journal.pone.0134826

WILKINSON, M. et al. The FAIR guiding principles for scientific data management and stewardship. **Sci Data**, n. 3, 2016. DOI: 10.1038/sdata.2016.18