

DADOS DE PESQUISA: DISPONIBILIZAÇÃO OU PUBLICAÇÃO?

Emanuelle Torino
Sandra Milena Roa-Martínez
Silvana Aparecida Borsetti Gregorio Vidotti

1. INTRODUÇÃO

O desenvolvimento científico ocorre ao longo dos anos no processo de investigação e publicação dos resultados de pesquisa mediante prévia e criteriosa avaliação por pares, processo que retroalimenta a investigação e gera novos resultados e/ou valida os anteriores, o que traz a característica de movimento cíclico.

Vale destacar que os periódicos científicos se constituíram ao longo das décadas no principal canal formal de comunicação dos resultados de pesquisa, cujo objetivo, além da disponibilização, passa pelo reconhecimento e chancela da pesquisa, realizada pelo processo de avaliação por pares.

Para Garvey (1979), a comunicação científica compreende atividades relativas à produção, à disseminação e ao uso da informação, e envolve todo o processo de pesquisa, desde a concepção da ideia até a aceitação dos resultados pela comunidade científica. Meadows (1999, p. 7) afirma que “[...] a comunicação situa-se no próprio coração da ciência. É para ela tão vital quanto a própria pesquisa, pois a esta não cabe reivindicar com legitimidade este nome enquanto não houver sido analisada e aceita pelos pares”.

As discussões recentes acerca da Ciência Aberta definidas pela Organisation for Economic Co-Operation and Development (2015, p. 7, tradução nossa) como esforços de pesquisadores e comunidade científica, instituições, governos e agências de fomento para “[...] tornar os resultados primários de pesquisas financiadas com recursos públicos – publicações e dados da pesquisa – acessíveis ao público em formato digital sem restrição ou com restrição mínima, como forma de acelerar a pesquisa”, trazem à tona, entre outras questões, a abertura de dados de pesquisa (*open data research*), um objeto gerado no processo de pesquisa, mas que inicialmente tinha como responsáveis pela gestão e armazenamento os próprios pesquisadores, que o faziam em seus laboratórios, utilizando as tecnologias disponíveis.

Tal afirmação busca enfatizar que os dados de pesquisa são gerados ao longo dos séculos, tendo sido muitas vezes esquecidos ou relegados a uma relevância inferior a dos artigos científicos, suporte no qual os resultados de pesquisa tradicionalmente são comunicados. Dessa forma, os pesquisadores são reconhecidos e têm suas reputações científicas embasadas nos tradicionais periódicos científicos e seus artigos. Callaghan (2019, tradução nossa) compartilha da mesma opinião, ao argumentar o seguinte: “Quando os usuários estão acostumados a ter o documento como o principal (e único) objeto de autoridade, o conjunto de dados pode ser rebaixado como meramente ‘informações de suporte’ quando, na verdade, é o artigo que fornece as informações de suporte sobre os dados”.

Em pesquisa com especialistas que atuam como protagonistas no Acesso Aberto na América Latina, Appel e Albagli (2019) discutem desafios recentes advindos do movimento de Acesso Aberto e de Ciência Aberta, e apontam uma ampliação na variedade de objetos científicos em circulação. Muito embora os periódicos e os artigos tradicionais se mantenham como fontes centrais de disponibilização dos resultados de pesquisa, emergem novas plataformas de comunicação científica para a gestão de objetos digitais de todo o ciclo da pesquisa demandados pela Ciência Aberta. Dessa forma, a “[...] a interoperabilidade desses objetos digitais é considerada crucial para multiplicar os efeitos do Acesso Aberto e da Ciência Aberta. Igualmente, ressalta-se a relevância de revisão por pares desses objetos, seu registro e descrição em metadados [...]” (p. 198).

Nesse sentido, ampliam-se as alternativas propostas por editores científicos para que diferentes objetos digitais de pesquisa estejam acessíveis publicamente, dentre as quais destacam-se: a disponibilização em ambientes externos, a exemplo de repositórios de dados de pesquisa; repositórios digitais ou

institucionais; e ainda, a inclusão dos referidos objetos como material suplementar do próprio artigo. De maneira menos frequente, verifica-se a disponibilização de periódicos que possuem como foco a publicação de artigos de dados de pesquisa (*data paper*), artigos tecnológicos e, mais recentemente, de periódicos que indicam em seu foco “a publicação além das formas tradicionais”, destacando-as como recursos e artefatos acadêmicos não tradicionais.

No que tange especificamente aos dados de pesquisa, foco deste estudo, segundo a Organisation for Economic Co-Operation and Development (2007, p. 9, tradução nossa), “Os bancos de dados científicos estão rapidamente se tornando uma parte crucial da infraestrutura do sistema científico global”. Reforçam, ainda, a necessidade de acesso aos dados de pesquisa para otimizar o investimento público em pesquisa, criar fortes cadeias de valor de inovação e aprimorar a cooperação internacional.

Visando alcançar seus benefícios, diversos países estão definindo diretrizes para que os dados de pesquisa estejam publicamente acessíveis – notadamente: a otimização de recursos para a coleta de dados; a possibilidade de comprovar os resultados e reproduzir a pesquisa, gerar novas abordagens, discussões e conclusões acerca de um conjunto de dados; ampliar as redes de colaboração nacional e internacional; e reduzir o tempo para que avanços sejam possíveis. Diante disso, a gestão dos dados de pesquisa e sua publicização pressupõe o reúso em benefício da própria comunidade científica.

Contudo, é importante esclarecer que o reúso necessita do cumprimento de diretrizes que tornem os dados de pesquisa passíveis de consumo por humanos e aplicações computacionais (GOFAIR, 2020; LÓSCIO; BURLE; CALEGARI, 2017; OPEN KNOWLEDGE FOUNDATION, 2020). Assim, é necessário que estejam adequadamente documentados, padronizados, tratados, descritos e armazenados, o que reforça a importância do planejamento da abertura dos dados de pesquisa e, de igual maneira, da forma de conectá-los aos demais objetos de pesquisa a eles relacionados, possibilitando o mapeamento do processo da pesquisa que primariamente gerou os dados e, ainda, das pesquisas secundárias geradas a partir deles.

Vale destacar que, no presente estudo, dados de pesquisa são entendidos como “[...] registros factuais (pontuações numéricas, registros textuais, imagens e sons) usados como fontes primárias de pesquisa científica e geralmente aceitos na comunidade científica como necessários para validar os resultados da pesquisa” (ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT,

2007, p. 13, tradução nossa). Além disso, é importante esclarecer que “Dados Abertos são dados que podem ser livremente usados, reutilizados e redistribuídos por qualquer pessoa – sujeitos, no máximo, à exigência de atribuição da fonte e compartilhamento pelas mesmas regras.” (OPEN KNOWLEDGE FOUNDATION, 2020).

Roa-Martínez, Vidotti e Sant’Ana (2017) destacam a relevância dos dados de pesquisa para o desenvolvimento científico e reconhecem que um dos objetivos principais do compartilhamento dos dados é reduzir os recursos e os esforços dos pesquisadores em troca de ampliar seu reconhecimento, mas alertam que “[...] para que estes dados sejam úteis, não devem ser apenas compartilhados, mas publicados” (p. 2, tradução nossa).

Diante disso, o presente estudo discute a disponibilização e a publicação de dados de pesquisa, visando esclarecer as divergências, similaridades e auxiliar pesquisadores, gestores de pesquisa e de repositórios – além de editores de periódicos – no tratamento desses objetos oriundos do processo de pesquisa científica.

2. DISPONIBILIZAÇÃO DE DADOS DE PESQUISA

Considerando que o termo disponibilizar é uma derivação de disponível, recorreremos a Cunha e Cavalcanti (2008) para entendê-lo no contexto do presente estudo, tendo como retorno o verbete disponibilidade (*availability*) que, ao se referir a documentos (*access, communication*), consiste na transmissão de uma informação, documento ou fonte, de um ponto a outro, tendo um ou mais destinatários, o que pode se fazer *in loco* ou para uso remoto, mediante uma requisição.

Na literatura estrangeira, o termo *data sharing* é empregado no sentido de disponibilizar ou compartilhar dados de pesquisa, definida por Roa-Martínez, Vidotti e Sant’Ana (2017, p. 3, tradução nossa) como “[...] a ação de disponibilizar dados de pesquisa de diversas maneiras, como uma troca privada ou direta a pedido do interessado ao autor ou a partir do armazenamento (um repositório público de dados, por exemplo) para que sejam utilizados com diferentes finalidades por outros pesquisadores”.

Nesse sentido, disponibilizar dados, por exemplo, “[...] pode ser uma planilha em um site, um conjunto de imagens em um arquivo institucional, um fluxo de leituras de uma estação meteorológica transmitida pela internet ou um artigo de dados revisado por pares descrevendo um conjunto de dados” (KRATZ; STRASSER, 2014, tradução nossa).

As definições esclarecem que a disponibilização de dados de pesquisa tem como principal característica a possibilidade de acesso, seja público ou privado, o que pode ser realizado independentemente de um processo formal. E podem, ainda, utilizar-se de diferentes canais ou tecnologias para fazê-lo, a exemplo de *website*, sistemas de informação interno, e-mail, repositório etc.

3. PUBLICAÇÃO DE DADOS DE PESQUISA

A publicação científica, como produto primário e original, considerando Cunha e Cavalcanti (2008), consiste em um documento de interesse da e distribuído à comunidade científica, que emprega metodologia científica, contribui para ampliar a compreensão de determinados problemas, é redigida conforme convenções usuais de cada área do conhecimento, e visa possibilitar a outros pesquisadores a replicação dos experimentos. Tais definições se aplicam aos dados de pesquisa, quando esses são considerados como objeto científico, primário e original.

A publicação de dados, do inglês *data publication* ou *data publishing*, utiliza-se da disponibilização, mas possui rigor na realização, visando atender às seguintes características: disponibilização pública por tempo indeterminado; documentação que possibilite a reprodução ou reuso (artigo de dados); citabilidade, a exemplo do que se faz com artigos e livros para reconhecer o trabalho do pesquisador que gerou os dados; e validação por pares, sendo essa uma característica não consensual na comunidade científica (KRATZ; STRASSER, 2014).

Dentre as estratégias de publicação dos dados de pesquisa, Pampel e Dallmeier-Tiessen (2014) destacam quatro: como objeto científico independente em um repositório de dados de pesquisa; como documentação textual estruturada; como artigo de dados; e como enriquecimento de um artigo ou publicação enriquecida.

Quando se trata de publicação de dados de pesquisa, a escolha do ambiente informacional para a disponibilização precisa ser feita de forma criteriosa, considerando a necessidade de uma infraestrutura que permita, entre outras coisas, que os dados sejam adequadamente armazenados e representados, utilizando padrões de metadados internacionalmente aceitos e que possibilitem, por exemplo, a interoperabilidade e a recuperação (LÓSCIO; BURLE; CALEGARI, 2017). Esse é um aspecto relevante para que os dados sejam FAIR (*Findable, Accessible, Interoperable, Reusable*) (GOFAIR, 2020). Além disso, devem ser privilegiados os repositórios confiáveis e que possibilitem a preservação digital, para que o pesquisador e a comunidade científica tenham assegurado o armazenamento e o acesso aos dados em longo prazo.

Vale lembrar que a disponibilização dos dados em um repositório pode significar o acesso público imediato ou ser realizada atendendo à necessidade de restrição de acesso por prazo determinado. A maioria dos sistemas utilizados para a disponibilização de repositórios possibilita gerenciar o acesso ao(s) arquivo(s) individualmente, assim, os dados podem ficar embargados e a representação publicamente acessível, por meio dos metadados; ou, ainda, nos casos em que os dados são constituídos por mais de um objeto ou acompanhados da documentação, é possível estabelecer as condições de acesso de forma individual.

A disponibilização pública de um conjunto de dados nem sempre é suficiente para o reuso, uma vez que isso implica na compreensão dos parâmetros utilizados na coleta, o que justifica a relevância da documentação dos dados.

Esse procedimento pode ser realizado por meio de artigo de dados (*data paper*), um objeto científico que descreve minuciosamente todos os elementos necessários à compreensão do conjunto de dados, incluindo a justificativa e os métodos de coleta, e sua publicação pode ser realizada em periódico convencional ou em periódico de dados (*data journal*). Não há consenso entre os periódicos quanto aos elementos mínimos dessa documentação, mas Roa-Martínez, Vidotti e Sant'Ana (2017) propõem uma estrutura que auxilia aos pesquisadores nesta atividade.

Além do artigo de dados, Kratz e Strasser (2014) apresentam outras duas formas de documentar os dados de pesquisa: disponibilizar a documentação juntamente com os dados em um repositório confiável, o que possibilitará, de igual maneira, a compreensão do conjunto. Considerando que esse documento não possui estrutura descrita, para beneficiar a compreensão e reuso dos dados, recomenda-se que ao fazê-lo se utilize a estrutura de artigo de dados.

Ademais, há a disponibilização dos dados como material suplementar de um artigo tradicional, prática essa não recomendável por questões como: um conjunto de dados pode ser utilizado por diversos artigos e um artigo pode utilizar diversos conjuntos de dados; ausência de representação dos dados que permita a recuperação e reúso; e ausência de identificador persistente que permita a citação dos dados de pesquisa.

Parte importante da documentação dos dados é a sua representação, que deve ser completa e exaustiva, considerando a granularidade dos dados. Para isso, deve-se utilizar metadados adequados ao domínio, a fim de que sejam recuperados por humanos e aplicações computacionais. Os tipos de metadados e suas aplicações são descritos em Gilliland (2008), Méndez Rodríguez (2002) e Riley (2017). Destaca-se que a infraestrutura adequada de metadados não é suficiente para que os dados de pesquisa sejam recuperados, por isso é imprescindível que a descrição seja realizada adequadamente, com vistas a servir de ponto de acesso aos dados descritos. A relevância dos metadados é pautada nos princípios FAIR (GOFAIR, 2020).

Wittenburg, Hellström e Zwölf (2017) destacam a necessidade de os dados de pesquisa receberem no momento da disponibilização um identificador persistente (*Persistent Identifier* - PID) que possibilite a sua recuperação e citação. E, para beneficiá-las, recomenda que o PID seja atribuído considerando a granularidade dos dados de pesquisa à menor unidade. Tal recomendação também é obedecida nos princípios FAIR, situando o PID como o mais importante, pois sem ele não há garantias de acesso unívoco aos dados (GOFAIR, 2020).

Os PIDs, apesar de fornecerem persistência no acesso aos dados beneficiando a característica de ser citável, não indicam a forma de fazê-lo. Assim, a *Research Data Alliance* disponibiliza diretrizes para que os dados utilizados em pesquisas possam ser adequadamente citados (RAUBER; ASMI; UYTVANCK; PROELL, 2015), resultando em reconhecimento aos seus autores.

No que tange à validação dos dados, considera-se um aspecto por esquadrihar durante o processo de publicação de dados de pesquisa, pois embora sejam objetos científicos, suas características podem inviabilizar a validação análoga à dos artigos ou requer mecanismos de avaliação próprios pela especificidade e natureza de cada conjunto de dados. Apesar disso, não resta dúvidas de que há a necessidade de validação. Consequentemente, ao se referir a um conjunto de dados que passou adequadamente pelo processo de publicação, Callaghan (2019, tradução nossa) afirma que “como esse conjunto de dados é completo,

bem documentado e com garantia de qualidade, pode ser considerado um artefato científico de primeira classe, com capacidade de ser citado”. Nesse sentido, já é apontada a possibilidade de alteração no processo de validação, o que pode certamente se aplicar aos dados. “No caso da publicação científica, a figura do *peer review*, a quem cabe operacionalizar o filtro de qualidade e de certificação da produção científica, pode vir a ceder terreno para a figura da curadoria, que remete ao estar junto, à atenção e ao cuidado, à coprodução” (ALBAGLI, 2018, p. 9).

4. DADOS DE PESQUISA: ALGUMAS DISCUSSÕES

As recomendações para a disponibilização de dados na web (GOFAIR, 2020; LÓSCIO; BURLE; CALEGARI, 2017; OPEN KNOWLEDGE FOUNDATION, 2020) estão alicerçadas na infraestrutura que suporta a abertura dos dados e, por isso, objetivam favorecer o processamento de agentes computacionais em benefício da comunidade.

Por outro lado, tais diretrizes não possuem como foco o pesquisador e, assim sendo, utilizam indistintamente os termos abertura, disponibilização, compartilhamento e publicação, ao se referirem ao processo de tornar os dados de pesquisa públicos e passíveis de recuperação, acesso e processamento computacional, visando seu reúso.

Nesse sentido, o presente estudo discute a disponibilização e a publicação de dados de pesquisa, com o objetivo de esclarecer as divergências e similaridades. Tal discussão pretende auxiliar os pesquisadores na tomada de decisões assertivas quanto à forma de tratamento dos dados de pesquisa gerados, escolhendo quando disponibilizá-los e quando publicá-los.

Cumprе ressaltar que cada um dos termos utilizados para se referir ao que se chama na Ciência Aberta de *open data* possui implicações que devem ser reconhecidas por pesquisadores, instituições, agências de fomento, editores científicos, gestores de repositórios e demais envolvidos no processo de pesquisa científica.

A abertura de dados de pesquisa, segundo a Open Knowledge Foundation (2020), requer que a disponibilização seja realizada com a utilização de uma licença de direitos autorais que solicite o reconhecimento de autoria e o

compartilhamento pela mesma licença. Esse fato possibilitará o reuso e o *re-mix* dos dados por terceiros.

Destaca-se que a legislação de direitos autorais brasileira (BRASIL, 1998), embora tenha natureza dúplce e proteja o autor e as suas criações expressas ou fixadas em um suporte tangível ou intangível – ou seja, a sua obra –, é omissa quanto à titularidade dos direitos de dados de pesquisa, uma vez que esses não são “criações do espírito” do pesquisador, mas sim coletados, gerados ou sistematizados por ele. A referida legislação, por outro lado, entende no art. 7º, inciso XIII, que as bases de dados são protegidas e impõe, no capítulo VII, as formas legais de sua utilização. Verifica-se assim que os dados não são protegidos por direitos autorais, o que por um lado fragiliza o pesquisador, visto que o reuso dos dados de pesquisa independe da sua autorização; por outro lado, sendo as bases de dados protegidas pela legislação, a publicação dos dados de pesquisa como dados abertos de pesquisa ou seja, a disponibilização em uma base de dados utilizando uma licença, torna-os protegidos, ao mesmo tempo que possibilita o estabelecimento dos limites de uso, o que favorece, além do pesquisador, a própria abertura dos dados.

Conforme discussão realizada anteriormente, a disponibilização de dados de pesquisa consiste em torná-los acessíveis publicamente, desconsiderando a relevância de aspectos que são prioritários no processo de publicação dos dados de pesquisa. O objetivo principal da disponibilização é que os dados possam ser acessíveis e reutilizáveis pela comunidade científica.

Por outro lado, a publicação de dados de pesquisa se faz de modo análogo à publicação de artigos, levando em consideração aspectos normativos e de padronização, cuidadosa explicitação metodológica que permitirá a interpretação e o reuso dos dados por terceiros, gerando com isso citações. Contudo, para que os dados de pesquisa (ainda que publicados) sejam recuperados, há de se atentar para o ambiente de disponibilização, que deve oferecer infraestrutura para que tais dados sejam FAIR, o que os tornará recuperáveis, passíveis de reuso e de citação.

Outro aspecto importante é que tais dados possam ser validados por meio de dois aspectos: o científico, que consiste na revisão dos processos utilizados na coleta, sistematização e documentação dos dados, o que deve ser feito antes de torná-los publicamente acessíveis; e o técnico, mais relacionado ao processo de representação dos dados por metadados, etapa a ser realizada quando da disponibilização em um ambiente informacional digital como um repositório de dados de pesquisa.

Embora todo o processo de publicação tenha etapas já conhecidas, a que mais possui controvérsias é a validação, pois o processo adotado tradicionalmente pelos periódicos, no caso da revisão de artigos, pode não ser aplicável aos dados de pesquisa.

A disponibilização ou publicação dos dados de pesquisa os torna publicamente acessíveis, o que possibilita o compartilhamento, entendido como o (re)uso e a (re)distribuição de um objeto.

Já há consenso de que a disponibilização dos dados para a reutilização apropriada reduz recursos empregados na coleta. Kratz e Strasser (2014) afirmam que documentar, publicar e armazenar dados requer esforços, mas geralmente menos significativos que repetir a coleta. Consideram, ainda, que em casos efêmeros a coleta não pode ser refeita, como por exemplo em registros climáticos, de ventos ou astronômicos.

A importância crescente do reconhecimento de outros objetos oriundos do processo de pesquisa, além dos tradicionais artigos científicos, é dirigida aos dados de pesquisa por meio da *Joint Declaration of Data Citation Principles* (DATA CITATION SYNTHESIS GROUP, 2014, tradução nossa), que afirma que “[...] os dados devem ser considerados produtos legítimos e citáveis da pesquisa. A citação de dados, como a citação de outras evidências e fontes, é uma boa prática de pesquisa e faz parte do ecossistema acadêmico que apoia o reúso de dados”. A Declaração estabelece oito princípios orientadores para dados: importância, crédito e atribuição, evidência, identificação única, acesso, persistência, especificidade e verificabilidade, interoperabilidade e flexibilidade. E afirma que as citações de dados devem ser compreensíveis por humanos e aplicações computacionais.

Embora as questões de publicação de dados de pesquisa tenham sido discutidas, deve-se notar que dentre os principais benefícios, destacam-se:

- uso e otimização de recursos (tempo, financeiro, entre outros) em projetos financiados;
- redução de pesquisas duplicadas com coleta de dados e resultados semelhantes;
- maior transparência nos resultados, permitindo a replicação de experimentos e verificação de hipóteses;

- nova análise de evidências no mesmo conjunto de dados de pesquisa;
- geração de novas perguntas e/ou hipóteses a partir de dados de pesquisa existentes que permitem o progresso em pesquisa e inovação;
- reconhecimento científico de pesquisadores;
- garantia contra a fabricação e falsificação de dados de pesquisa;
- maximização do potencial dos dados da pesquisa.

Por outro lado, como oportunidades, na espera de propostas de soluções a serem abordadas pela Ciência da Informação, se encontram aspectos relacionados com:

- incerteza sobre o local de armazenamento dos dados de pesquisa;
- garantias de uso adequado dos dados de pesquisa disponibilizados;
- falta de citação e reconhecimento aos criadores do conjunto de dados de pesquisa;
- falta de tempo, incentivos e recursos financeiros que levam à disponibilidade dos dados de pesquisa;
- alto esforço e falta de conhecimento técnico e especializado para o processamento e disponibilidade de dados de pesquisa coletados pelos pesquisadores para que possam ser acessados por terceiros;
- falta de metadados capazes de garantir o acesso aos dados de pesquisa durante seu ciclo de vida e de garantir sua preservação;
- falta de padronização de modelos, metodologias, tecnologias e ferramentas utilizadas pelos pesquisadores no gerenciamento de dados de pesquisa.

5. RECOMENDAÇÕES

Visando contribuir com as discussões acerca da temática abordada, são apontadas recomendações a pesquisadores, gestores de pesquisa, gestores de repositórios e editores de periódicos.

Para pesquisadores:

- elaborar plano de gestão de dados de pesquisa;
- definir um repositório digital confiável para a publicização dos dados de pesquisa;
- optar pela publicação dos dados de pesquisa, sempre que possível, visando privilegiar o reúso;
- documentar os dados de pesquisa para que possam ser reutilizados por meio da publicação de *data paper* ou descrição similar, disponível juntamente com os dados de pesquisa no repositório;
- considerar questões éticas e legais no momento da coleta e da publicização dos dados de pesquisa;
- anonimizar dados sensíveis;
- disponibilizar dados de pesquisa com cuidadosa representação, considerando que os metadados são vitais para sejam recuperados e citados;
- utilizar licenças para estabelecer os limites de uso dos dados de pesquisa;
- indicar, em todas as submissões para a publicação de objetos de pesquisa, o PID dos dados e demais objetos relacionados, utilizando metadados;
- buscar dados em repositórios de dados de pesquisa antes de iniciar a coleta, visando seu reúso;
- utilizar dados de pesquisa publicados sempre que possível;
- utilizar dados de pesquisa de terceiros atentando-se aos limites estabelecidos na licença;

- citar os dados de pesquisa coletados por terceiros sempre que utilizados.

Para gestores de pesquisa:

- Requerer plano de gestão de dados de pesquisa;
- tornar mandatária a publicação de dados de pesquisa, ainda que atendendo a períodos de embargo;
- possibilitar o processo de curadoria dos dados de pesquisa;
- fornecer infraestruturas robustas e adequadas para a gestão dos objetos de pesquisa;
- reconhecer os dados de pesquisa e demais objetos oriundos do processo de pesquisa (em métricas, na necessidade da destinação de horas para coleta, tratamento e documentação, entre outros);
- disponibilizar ambiente de informação que permita gerenciar o ciclo de vida da pesquisa e o relacionamento de diferentes objetos que a compõem.

Para gestores de repositórios:

- Considerar as especificidades disciplinares e a granularidade dos dados de pesquisa no perfil de aplicação de metadados, para que eles sejam adequadamente representados;
- fornecer campos de metadados para que outros objetos de pesquisa relacionados aos dados publicizados possam ser indicados em metadados;
- fornecer revisão técnica da representação dos dados de pesquisa, considerando a relevância para a recuperação;
- solicitar a documentação dos dados de pesquisa, quer seja indicando, por meio de metadados, a URI do artigo de dados (*data paper*) publicado, ou mediante a disponibilização de descrição similar juntamente com os dados;
- indicar nos metadados a licença dos dados de pesquisa;
- recomendar que os dados de pesquisa sejam submetidos à curadoria;

- fornecer identificador persistente para dados de pesquisa e assegurar seu funcionamento por meio da atualização das URIs;
- possibilitar o versionamento dos dados de pesquisa;
- assegurar a preservação digital dos dados de pesquisa;
- possibilitar o relacionamento de diferentes objetos que compõem o processo de pesquisa.

Para editores de periódicos:

- impossibilitar a submissão de dados de pesquisa como material suplementar de artigo científico tradicional;
- requerer que os dados de pesquisa estejam publicados em um repositório de dados de pesquisa;
- disponibilizar, na submissão do artigo, metadado específico para relacionar a publicação aos dados de pesquisa, por meio da URI dos dados;
- possibilitar que outros objetos de pesquisa relacionados ao artigo submetido possam ser indicados em metadados;
- disponibilizar periódico de dados (*data journal*) para a publicação de artigo de dados ou, quando não for possível, incluir esse tipo de artigo no escopo do periódico científico, publicando-o em sessão específica;
- estabelecer estrutura mínima necessária à publicação de artigo de dados.

6. CONSIDERAÇÕES FINAIS

A presente discussão, embora não tenha o objetivo de se esgotar, pretende contribuir com as reflexões dos atores do processo de pesquisa e auxiliá-los no planejamento de políticas e ações que envolvam a gestão e a publicização dos dados de pesquisa e demais objetos oriundos do processo de pesquisa científica.

É relevante destacar que a adoção de um termo em uma política precisa ser pensada para que o objetivo seja alcançado. Salienta-se que o emprego do termo disponibilização se refere ao simples ato de tornar dados de pesquisa acessíveis, utilizando-se de qualquer meio. Enquanto a publicação refere-se a um conjunto mais estruturado de ações, que envolve a documentação, a escolha de um repositório digital confiável, o uso de PID e a validação dos dados para ser disponibilizados. Para que sejam considerados abertos, os dados de pesquisa necessitam de uma licença que estabeleça claramente as formas de uso. E, a partir da disponibilização, os dados de pesquisa podem ser compartilhados e reutilizados.

Destaca-se, assim, que disponibilizar não é o suficiente para que os dados de pesquisa possam ser reutilizados e que, além disso, o processo de publicação é que dará a eles características de qualidade, confiabilidade e veracidade. Ademais, o processo de publicação auxilia para que os dados sejam FAIR, o que permitirá que sejam encontrados, acessados, interoperáveis e reutilizados.

Nesse contexto, é necessária no Brasil a criação de periódico de dados (*data journal*) para que a documentação dos dados, por meio de artigo de dados, possa ser publicada e esteja acessível à comunidade científica, recebendo a mesma visibilidade e possibilidade de citação que um artigo científico tradicional.

No que tange aos ambientes para a gestão dos dados de pesquisa e sua documentação, é imprescindível que sejam confiáveis, possibilitem preservação digital e forneçam infraestrutura adequada de metadados. Salienta-se que os metadados fornecem a estrutura adequada para que os dados de pesquisa possam ser recuperados por humanos e aplicações computacionais, além de serem utilizados para enriquecimentos semânticos no contexto do *linked data*.

É importante ainda reconhecer que os diferentes objetos que compõem o processo de pesquisa científica se relacionam no que tange a dados, conteúdos e processos. Assim, é fundamental atentar-se ao ciclo de vida e aos atores envolvidos neste processo (TORINO, VIDOTTI, SANT'ANA, 2019), para que os ambientes de gestão dos objetos de pesquisa possam interoperar nos níveis sintático e semântico, favorecendo a comunicação científica ampliada e o mapeamento do ciclo de vida da própria pesquisa.

7. REFERÊNCIAS

ALBAGLI, S. Ciência Aberta em questão. *In*: ALBAGLI, S.; MACIEL, M. L.; ABDO, A. H. (org.). **Ciência Aberta, questões abertas**. Brasília: IBICT; Rio de Janeiro: UNIRIO, 2015. p. 9-26.

APPEL, A. L.; ALBAGLI, S. Acesso aberto em questão: novas agendas e desafios. **Inf. & Soc.:Est.**, João Pessoa, v.29, n.4, p. 187-208, out./dez. 2019. Disponível em: <https://periodicos.ufpb.br/ojs2/index.php/ies/article/view/50113/29300>. Acesso em: 14 jan. 2020.

BRASIL. Lei nº 9.610, de 19 de fevereiro de 1998. **Diário Oficial da União**, Brasília, DF, 20 fev. 1998. Seção 1. Disponível em: http://www.planalto.gov.br/ccivil_03/leis/L9610.htm. Acesso em: 15 nov. 2018.

CALLAGHAN, S. Research Data Publication: moving beyond the metaphor. **Data Science Journal**, v. 18, n. 1, 2019. Disponível em: <http://doi.org/10.5334/dsj-2019-039>. Acesso em: 19 jan. 2020.

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

DATA CITATION SYNTHESIS GROUP. **Joint Declaration of Data Citation Principles**. San Diego CA: FORCE11, 2014. Disponível em: <https://doi.org/10.25490/a97f-egyk>. Acesso em: 16 nov. 2018.

GARVEY, W. D. **Communication**: the essence of science. Oxford: Pergamon Press, 1979.

GILLILAND, A. J. Setting the stage. *In*: BACA, M. (ed.). **Introduction to metadata**. Los Angeles: Getty, c2008. p. 1-19. Disponível em: <http://d2aohiyo3d3idm.cloudfront.net/publications/virtuallibrary/0892368969.pdf>. Acesso em: 15 jul. 2018.

GOFAIR. **FAIR principles**. Disponível em: <https://www.go-fair.org/fair-principles/>. Acesso em: 14 jan. 2020.

KRATZ, J; STRASSER, C. Data publication consensus and controversies [versão3]. **F1000Research**, n. 94, 2014. Disponível em: <https://doi.org/10.12688/f1000research.3979.1>. Acesso em: 14 jan. 2020.

LÓSCIO, B. F.; BURLE, C.; CALEGARI, N. (ed.). **Data on the web best practices**. 2017. Disponível em: <https://www.w3.org/TR/dwbp/>. Acesso em: 14 jan. 2020.

MEADOWS, A. J. **A comunicação científica**. Brasília: Briquet de Lemos, 1999.

MÉNDEZ RODRÍGUEZ, E. **Metadatos y recuperación de información**: estándares, problemas y aplicabilidad en bibliotecas digitales. Gijón: Trea, 2002.

OPEN KNOWLEDGE FOUNDATION. **The open data handbook**. Disponível em: <http://opendatahandbook.org/>. Acesso em: 14 jan. 2020.

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT. Making open science a reality. **OECD Science, Technology and Industry Policy Papers**, n. 25. Paris: OECD Publishing, 2015. Disponível em: [10.1787/5jrs2f963zs1-en](https://doi.org/10.1787/5jrs2f963zs1-en). Acesso em: 14 jan. 2020.

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT. **OECD principles and guidelines for access to research data from public funding**. 2007. Paris: OECD Publishing. Disponível em: <https://doi.org/10.1787/9789264034020-en-fr>. Acesso em: 14 jan. 2020.

PAMPEL, H.; DALLMEIER-TIESEN, S. open research data: from vision to practice. In: BARTLING, S., FRIESIKE, S. (ed.). **Opening science**. Cham: Springer, 2014. p. 213-224. Disponível em: https://link.springer.com/chapter/10.1007/978-3-319-00026-8_14. Acesso em: 19 jan. 2020.

SUBSTITUIR POR: RAUBER, A.; ASMI, A.; UYTVANCK, D; PROELL, S.. **Data citation of evolving data**: recommendations of the Working Group on Data Citation (WGDC). 2015. Disponível em: <https://dx.doi.org/10.15497/RDA00016>. Acesso em: 17 jan. 2020.

RILEY, J. **Understanding metadata**: what is metadata, and what is it for?. Baltimore: National Information Standards Organization (NISO), c2017. Disponível em: https://groups.niso.org/apps/group_public/download.php/17446/Understanding%20Metadata.pdf. Acesso em: 15 jul. 2018.

ROA-MARTÍNEZ, S. M.; VIDOTTI, S. A. B. G.; SANT'ANA, R. C. G. Estructura propuesta del artículo de datos como publicación científica. **Revista Española de Documentación Científica**, v. 40, n. 1, jan./mar. 2017. Disponível em: <http://redc.revistas.csic.es/index.php/redc/article/view/971>. Acesso em: 14 jan. 2020.

TORINO, E.; VIDOTTI, S. A. B. G.; SANT'ANA, R. C. G. Ciclo de vida de dados no processo de publicação e acesso à produção científica. *In*: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 20., 2019, Florianópolis. **Anais** [...] Florianópolis: Universidade Federal de Santa Catarina, 2019. Disponível em: <https://conferencias.ufsc.br/index.php/enancib/2019/paper/viewFile/576/612>. Acesso em: 15 out. 2019.

WITTENBURG, P.; HELLSTRÖM, M.; ZWÖLF, C.-M. (ed.). **Persistent identifiers: consolidated assertions**. 2017. Disponível em: <https://dx.doi.org/10.15497/RDA00027>. Acesso em: 17 jan. 2020.

8. SOBRE OS AUTORES



Emanuelle Torino

Lattes: <http://lattes.cnpq.br/5042296869081637>

Orcid: <https://orcid.org/0000-0002-3791-9884>

Doutoranda em Ciência da Informação pela Universidade Estadual Paulista (Unesp), Mestre em Gestão da Informação pela Universidade Estadual de Londrina (UEL), Bacharel em Biblioteconomia pela Universidade Estadual de Londrina. Membro do Grupo de Pesquisa Novas Tecnologias em Informação (GPnti). Bibliotecária na Universidade Tecnológica Federal do Paraná (UTFPR) e Coordenadora dos ambientes de Gestão da Informação em Acesso Aberto da UTFPR.



Sandra Milena Roa Martínez

Lattes: <http://lattes.cnpq.br/7188015719101807>

Orcid: <https://orcid.org/0000-0002-2271-6101>

Doutora em Ciência da Informação da Universidade Estadual Paulista (Unesp), Mestre em Engenharia com ênfases em Eletrônica da Universidad del Valle (Colômbia), Graduação em Engenharia de Sistemas da Universidad Industrial de Santander (Colômbia). Professora Titular e membro do Grupo de Investigación en Inteligencia Computacional do Departamento de Sistemas da Faculdade de Engenharia Eletrônica e Telecomunicações da Universidad del Cauca (Colômbia).



Silvana Aparecida Borsetti Gregorio Vidotti

Lattes: <http://lattes.cnpq.br/7390573927636069>

Orcid: <https://orcid.org/0000-0002-4216-0374>

Doutora em Educação pela Universidade Estadual Paulista (Unesp), Mestre em Ciências pela Universidade de São Paulo (USP). Licenciada em Matemática pela Universidade Estadual Paulista (Unesp). Professora do Departamento e do Programa de Pós-Graduação em Ciência da Informação da Unesp. Líder do Grupo de Pesquisa Novas Tecnologias em Informação (GPnti). Assessora da Pró-Reitoria de Graduação e Presidente do Comitê Gestor de Acesso Aberto da Unesp.

9. COMO CITAR

TORINO, E.; ROA-MARTÍNEZ, S. M.; VIDOTTI, S. A. B. G. Dados de pesquisa: disponibilização ou publicação?. *In*: SHINTAKU, M.; SALES, L. F.; COSTA, M. (org). **Tópicos sobre dados abertos para editores científicos**. Botucatu, SP: ABEC, 2020. p. 183-201. DOI: 10.21452/ 978-85-93910-04-3.cap15.