

Abordagens de reúso e a questão da reusabilidade dos dados científicos

Approaches for data reuse and the issue of scientific data reusability

Renata Curty*

RESUMO

As diretrizes governamentais e institucionais em torno do compartilhamento de dados de pesquisas financiadas com dinheiro público têm impulsionado a rápida expansão de repositórios digitais de dados afim de disponibilizar esses ativos científicos para reutilização, com propósitos nem sempre antecipados, pelos pesquisadores que os produziram/coletaram. De modo contraditório, embora o argumento em torno do compartilhamento de dados seja fortemente sustentado no potencial de reúso e em suas consequentes contribuições para o avanço científico, esse tema permanece acessório às discussões em torno da ciência de dados e da ciência aberta. O presente artigo de revisão narrativa tem por objetivo lançar um olhar mais atento ao reúso de dados e explorar mais diretamente esse conceito, ao passo que propõe uma classificação inicial de cinco abordagens distintas para o reúso de dados de pesquisa (reaproveitamento, agregação, integração, metanálise e reanálise), com base em situações hipotéticas acompanhadas de casos de reúso de dados publicados na literatura científica. Também explora questões determinantes para a condição de reúso, relacionando a reusabilidade à qualidade da documentação que acompanha os dados. Oferece discussão sobre os desafios da documentação de dados, bem como algumas iniciativas e recomendações para que essas dificuldades sejam contornadas. Espera-se que os

ABSTRACT

The availability of scientific assets through data repositories has been greatly increased as a result of government and institutional data sharing policies and mandates for publicly funded research, allowing data to be reused for purposes not always anticipated by primary researchers. Despite the fact that the argument favoring data sharing is strongly grounded in the possibilities of data reuse and its contributions to scientific advancement, this subject remains unobserved in discussions about data science and open science. This paper follows a narrative review method to take a closer look at data reuse in order to better conceptualize this term, while proposing an early classification of five distinct data reuse approaches (repurposing, aggregation, integration, meta-analysis and reanalysis) based on hypothetical cases and literature examples. It also explores the determinants of what constitutes reusable data, and the relationship between data reusability and documentation quality. It presents some challenges associated with data documentation and points out some initiatives and recommendations to overcome such problems. It expects to contribute not only for the conceptual advancement around the reusability and effective reuse of the data, but also to result in initiatives related to data documentation in order to increase the reuse potential of these scientific assets.

* Doutora em Information Science and Technology pela Syracuse University, NY, Estados Unidos. Professora Adjunta do Departamento de Ciência da Informação da Universidade Estadual de Londrina (UEL). Endereço: Campus Universitário. Rodovia Celso Garcia Cid (PR445), KM380. Caixa Postal: 6001 CEP 86051-990 Londrina-PR-Brasil. Telefone: (43) 3371-4348. E-mail: renatacurty@uel.br

argumentos apresentados contribuam não somente para o avanço conceitual em torno do reuso e da reusabilidade de dados, mas também reverberem em ações relacionadas à documentação dos dados de modo a incrementar o potencial de reuso desses ativos científicos.

Palavras-chave: Reuso de Dados; Reprodutibilidade Científica; Reusabilidade; Ciência Aberta; Dados de Pesquisa.

Keywords: Data Reuse; Scientific Reproducibility; Reusability; Open Science; Research Data

CONTEXTO E MOTIVAÇÃO

A ciência está alicerçada na possibilidade de avaliar criticamente a precisão e a validade das afirmações e conclusões produzidas pelos pesquisadores. As garantias de replicabilidade (replicability) e de reprodutibilidade (reproducibility) configuram-se como essenciais para a confirmação e a autocorreção de descobertas no contexto científico, sendo ambas diretamente dependentes da transparência dos dados e dos métodos empregados nas pesquisas. Além disso, as oportunidades de expansão do potencial de aplicabilidade de dados científicos preexistentes, de modo a otimizar custos e esforços associados à coleta de dados constituem-se como importantes argumentos em prol do amplo compartilhamento de dados de pesquisa.

As práticas de compartilhamento de dados científicos não são estranhas à ciência, pois sempre fizeram parte da essência das relações entre pesquisadores e da dinâmica da produção científica, mesmo que em maior ou menor grau, dependendo da área do conhecimento e da natureza dos dados. No entanto, a formalização e melhor sistematização dos processos para disponibilização dos dados, bem como o acesso amplo e aberto a coleções de dados por meio de repositórios na web, impulsionadas por órgãos governamentais, agências de fomento, entidades científicas e editoriais, têm refletido em uma maior legitimação do reuso de dados no contexto científico.

O caráter facultativo e ad hoc do compartilhamento de dados, sem compromisso com o futuro reuso destes dados, não mais interessa ao atual contexto de ciência aberta, que passa a adotar o sistema “carrots and sticks”, aliando abordagens de recompensa e punitivas para que os dados sejam, de fato, disponibilizados.

A expansão das políticas de compartilhamento de dados de pesquisas financiadas com recursos públicos, iniciada mais intensivamente há cerca de uma década na América do Norte e na Europa e, mais recentemente, no cenário nacional, tem resultado em um vertiginoso crescimento de repositórios de dados científicos de acesso aberto. Embora a disponibilidade dos dados seja condição sine qua non para que o reuso possa materializar-se, o reuso não pode ser simplesmente presumido como consequência certa do compartilhamento. O efetivo reuso requer não somente a possibilidade de identificar e acessar dados preexistentes, alinhados aos interesses dos pesquisadores; mas principalmente que os dados atendam a condições mínimas de reusabilidade.

A reusabilidade constitui o quarto pilar dos princípios FAIR, do acrônimo Findability, Accessibility, Interoperability e Reusability definidos pelo grupo multidisciplinar FORCE11, responsável por discussões sobre o futuro e a modernização da

comunicação científica (WILKINSON et al., 2016). A reusabilidade é determinada por características técnicas de conformidade, preconizadas pelo FAIR, e também representa o potencial de reuso aferido, em última instância, pelo pesquisador, considerando os propósitos de sua pesquisa e com base em seu melhor julgamento sobre os atributos e condições de qualidade dos dados.

Em outros termos, os dados precisam possuir certas características para serem considerados reutilizáveis, incluindo aspectos como relevância, completude, compreensibilidade e a confiabilidade. Tais atributos são melhor verificáveis e julgados quando os dados estão documentados, acompanhados de informações suplementares e descrições que informem sobre características como, por exemplo, a proveniência dos dados e o contexto em que eles foram obtidos, os métodos para coleta, registros, transcrição, codificação e/ou processamento dos dados, suas diferentes versões, entre outros.

Em concordância com Niu (2009ab), Faniel e Jacobsen (2010) e Faniel e Zimmerman (2011), o presente artigo parte do pressuposto de que a reusabilidade dos dados depende, em grande parcela, da qualidade associada à documentação dos dados disponibilizados para reuso. Nesse sentido, pretende-se fomentar essa relação, ao passo que também busca-se definir mais explicitamente o conceito de reuso de dados e suas abordagens, a partir dos seguintes questionamentos: Como o reuso de dados pode ser entendido no contexto científico? Quais seriam as possíveis formas e estratégias para o reuso de dados de pesquisa? Quais aspectos são determinantes para a reusabilidade dos dados?

Com vistas a responder a essas questões, adota-se o método de revisão narrativa para análise crítica e interpretativa da literatura (ROTHER, 2007) sobre o reuso, aspectos de reusabilidade e documentação de dados científicos. Para tanto, em um primeiro momento, apresenta-se uma discussão em torno do conceito de reuso de dados no contexto da pesquisa científica, de modo a desvendar esse processo a partir da exemplificação de diferentes abordagens para o reuso de dados de pesquisa. Além da apresentação conceitual das abordagens, a título de ilustração de como cada uma delas pode ser materializada em pesquisas científicas, são indicados exemplos hipotéticos e sinalizados exemplos reais selecionados na literatura. Posteriormente, os atributos de reusabilidade serão apresentados, com ênfase para o papel da documentação de dados, discutindo as dificuldades associadas a esse processo, bem como as alternativas para ampliar o potencial de reuso de dados de pesquisa.

Espera-se que os resultados permitam avanços não apenas de ordem conceitual, mas também prático, na medida em que gestores de repositórios de dados, profissionais que atuam com a curadoria de dados científicos, além dos próprios pesquisadores que produzem, compartilham e reutilizam dados, atentem-se para questões capazes de incrementar o potencial de reuso dos dados científicos. Considerando o atual cenário de expansão da disponibilidade de dados científicos, é também expectativa deste artigo fomentar discussões em torno das modalidades de reuso de dados de pesquisa, traduzidas em ações, que visem intensificar e estimular o ensino de métodos e técnicas para a reutilização de dados, sendo uma alternativa para a geração de novas pesquisas e melhor escrutínio da qualidade das produções científicas.

REÚSO: DESMISTIFICANDO O TERMO E SUAS DERIVAÇÕES

O reúso de dados tem sido tema recorrente nos debates científicos. Diferentes modelos buscam representar o ciclo de vida de dados de pesquisa e preveem o reúso como uma ação desejada resultante dos esforços de curadoria e de gestão desses ativos científicos, assim como uma etapa que retroalimenta novas ideias e processos de pesquisa.

Embora frequentemente incorporado nesses diagramas e mencionado na literatura relacionada, menos frequentes são os que definem o termo reúso de dados explicitamente, ficando esse conceito à margem das discussões em torno do momentum da ciência de dados e da ciência aberta. Van de Sandt e colaboradores (2019) atestam esse fato a partir de um levantamento sistemático da literatura. Os autores constataram que dentre os 65 estudos identificados sobre reúso de dados científicos, menos de um terço deles conceituavam o termo, propondo uma nova definição ou mesmo tomando como base a definição de outros autores. O mesmo estudo denuncia a inexistência de conciliação sobre o significado do termo, uma vez que muitos autores apresentam definições consideradas contraditórias em uma mesma publicação. Conseqüentemente, a confusão terminológica e a falta de medidas práticas acerca do que compreende o reúso de dados requerem um esclarecimento conceitual que oriente diretrizes e políticas científicas (VAN DE SANDT et al., 2019).

Pasquetto, Randles e Borgman (2017) indicam que o problema central para compreensão do que constitui reúso de dados está na relação entre os dados e quem os reutiliza:

Na situação mais simples, os dados são coletados por um indivíduo, para um projeto de pesquisa específico, e o primeiro “uso” é feito por aquele indivíduo para responder a uma pergunta de pesquisa específica. Se esse mesmo indivíduo retornar ao mesmo conjunto de dados posteriormente, seja para o mesmo projeto ou para um projeto diferente, isso geralmente seria considerado um “uso”. Quando esse conjunto de dados é disponibilizado em um repositório, recuperado por outro pesquisador e implantado em outro projeto, normalmente, isso seria considerado “reúso”. No jargão comum das práticas de dados, o reúso geralmente implica o uso de um conjunto de dados por alguém que não seja responsável pela coleta e obtenção dos dados originais (PASQUETTO; RANGLES; BORGMAN, 2017, n.p., tradução nossa).

A distinção supracitada desconsidera que todo novo uso dos dados, incluindo os que buscam responder a novas questões de pesquisa, seja considerado reúso, mesmo no caso de quem reutiliza os dados tenha participado do processo de sua coleta/obtenção. A relação do sujeito com os dados originais primários é, portanto, o que predetermina a condição de reúso na visão dos autores.

Heaton (2004), Zimmerman (2003; 2008), Daniels (2014), Fear (2013) e Kim e Yoon (2017) defendem, de forma uníssona, que o reúso de dados corresponde a toda nova aplicação de dados coletados para um propósito, para o estudo de um novo problema de pesquisa, independentemente de quem reutiliza os dados preexistentes. Zimmerman (2003) e Kim e Yoon (2017) enfatizam que o reúso de dados corresponde ao uso de dados coletados com um propósito específico para o estudo de um novo problema. Similarmente, o tesouro eletrônico Data Thesaurus do

National Network of Libraries of Medicine (NNLM), ligado ao National Institute of Health (NIH), dos Estados Unidos, define o reúso como aquele que envolve o uso de dados de pesquisa para uma atividade de pesquisa ou finalidade diferente daquela para a qual foi originalmente planejada (NATIONAL..., 2018).

A lógica de que o reúso depende necessariamente da existência de um novo propósito de modo a responder a novas questões e problemas de pesquisa, com base em dados preexistentes, exclui situações em que pesquisadores reanalisam dados de pesquisas originais, utilizando não apenas os mesmos dados, mas também os mesmos métodos, com vistas à confirmação ou refutação parcial ou total dos resultados obtidos.

Dada as limitações dos conceitos mencionados, a definição mais ampla e inclusiva para o reúso de dados proposta por Van de Sandt e coautores (2019) parece mais plausível. Os autores defendem o reúso como todo novo uso de qualquer recurso de pesquisa/coleção de dados. Assim, independentemente de quando é utilizado, do propósito de uso, das características dos dados e da relação dos usuários com os dados originais deve ser considerado reúso.

Nessa perspectiva, os dados de pesquisa podem ser reutilizados integralmente, parcialmente ou de forma combinada, uma, duas ou múltiplas vezes, para propósitos iguais, semelhantes ou completamente diferentes daqueles para os quais os dados foram originalmente coletados. O reúso envolve todo e qualquer novo uso de dados preexistentes podendo ser realizado pelos pesquisadores que coletaram/produziram os dados primários, ou por outras partes não necessariamente envolvidas na pesquisa original.

A simplificação do conceito de reúso não exclui a necessidade de tipificar suas diferentes derivações, estas, sim, condizentes com as relações entre os propósitos do reúso. Nesse sentido, Heaton (2004) apresenta os conceitos de replicação e de síntese. Para a autora, a replicação de dados diz respeito ao tipo de reúso empregado com o objetivo de verificar, refutar ou refinar o estudo original. Já a síntese inclui diferentes técnicas, como agregação e integração de dados, que podem ser empregadas a fim de sintetizar o conhecimento decorrente de dados preexistentes.

O Evolutionary Informatics Working Group (EvoIO) (2011) apresenta uma distinção mais granular das técnicas que podem ser empregadas para o reúso de dados científicos. Além das abordagens tratadas por Heaton (2004), o grupo apresenta as possibilidades de reúso de dados por reaproveitamento e metanálise e, embora reconheça uma relação entre o reúso por síntese com a agregação e a integração, prefere distingui-los em três modalidades distintas.

O grupo EvoIO (2011) atribui à abordagem de síntese a capacidade de integrar e agregar dados, mas com maior implicação de novidade conceitual e criatividade, de modo a combinar dados de maneiras que não foram imaginadas. Entretanto, o presente artigo opta por não classificar a abordagem de síntese como uma modalidade de reúso de dados independente, pois acredita-se que a novidade conceitual perpetrada pelo pesquisador que reutiliza dados não é exclusiva da integração ou da agregação de dados, porque é um incremento natural do processo de pesquisa e pode permear qualquer uma das demais abordagens de reúso. Também prefere adotar o termo reanálise, ao invés de replicação, por motivos que serão detalhados mais adiante. Dessa forma, sugere-se a seguinte classificação composta por cinco abordagens para o reúso de dados de pesquisa: 1) reaproveitamento, 2) agregação, 3) integração, 4) metanálise e 5) reanálise.

O reúso por reaproveitamento é a abordagem mais comumente mencionada na literatura, pois, como visto, o conceito geral de reúso está centrado na possibilidade de um pesquisador fazer novo uso de dados preexistentes com propósitos diferentes daqueles do consumidor primário dos dados. Nessa modalidade, os dados de um único estudo são reutilizados integralmente, ou parcialmente selecionados e aproveitados, sem que sejam complementados ou integrados com dados provenientes de outras fontes. Esses dados são, então, submetidos a novas análises, de modo a responderem questões de pesquisas diferentes daquelas abordadas no estudo que originou os dados. Em outras palavras, o reutilizador reaproveita dados brutos e/ou processados de um único estudo com objetivos nem sempre antecipados ou aprofundados pelos pesquisadores que coletaram os dados, de modo a extrair novas inferências e relações.

O reúso por reaproveitamento é muito comum em estudos que reutilizam dados de levantamento com grandes amostras. Por exemplo, Chandra, Copen e Stephen (2014) utilizaram alguns dados do relatório da Pesquisa Nacional de Crescimento da Família dos Estados Unidos (National Survey of Family Growth - NSFG), que continham inúmeras variáveis, para avaliar e estimar o uso de serviços de infertilidade no país. Para tanto, selecionaram apenas algumas variáveis de interesse e utilizaram as observações coletadas, de modo a traçar correlações específicas.

Curty e coautores (2017) também reaproveitaram os dados abertos coletados por Tenopir e colaboradores (2015) sobre o perfil e as percepções de pesquisadores, provenientes de diferentes países e das mais diversas áreas do conhecimento, com relação ao compartilhamento e reúso de dados de pesquisa. Para isso, tiveram que reorganizar os dados, combinar e transformar algumas variáveis, a fim de que pudessem empreender técnicas de análises diferentes das adotadas pelos autores da pesquisa original, com vistas a explorar novas questões de pesquisa baseadas em hipóteses autênticas. Desse modo, o estudo não apenas tinha o reúso como tema e objeto de pesquisa, mas também adotou a técnica de reúso por aproveitamento para analisar o fenômeno de interesse para além do escopo do estudo original.

O reúso por agregação é aquele que reúne dados de um mesmo domínio provenientes de diferentes estudos/fontes para compor uma coleção de dados (dataset) mais completa. Uma situação hipotética que exemplificaria esse tipo de reúso seria o caso de uma pesquisadora da área de Climatologia que, em sua pesquisa, reúne dados de diferentes fontes internacionais acerca da série histórica das temperaturas dos cinco oceanos nos últimos 50 anos para analisar o impacto do aquecimento global. Embora alguns dados tenham de ser transformados dada a existência de diferentes escalas utilizadas nas múltiplas fontes (Celsius, Fahrenheit ou Kelvin), essa seria a única alternativa para estipular modelos preditivos com base em projeções estatisticamente mais robustas.

Com efeito, a partir dessa abordagem de reúso, o pesquisador busca compor uma coleção de dados, tomando como base variáveis ou construtos predefinidos, que permitam agregar casos, elementos (data points) ou indicadores que se encaixam ao estudo. Pesquisadores podem combinar dados de diferentes fontes preexistentes, inclusive com dados primários que tenham coletado. Seguindo o mesmo exemplo, a pesquisadora poderia ter produzido dados sobre a variação da temperatura dos oceanos nos últimos cinco anos, mas, para um modelo mais robusto, dependeria de uma série histórica mais longa, o que a faria recorrer a outras fontes para completar sua coleção de dados.

O estudo de Viccosi e colaboradores (2018), sobre a relação entre o vírus da Zika e a microcefalia, ilustra bem esse tipo de abordagem, considerando que os autores agregaram dados provenientes de quatro diferentes fontes que continham indicadores de contaminação pelo vírus e de nascimentos de crianças com microcefalia. A partir disso, reutilizaram esses dados em análises conjuntas com vistas a examinar associações geoespaciais entre o vírus e a incidência da doença em bebês recém-nascidos.

O reúso por integração pode ser entendido como aquele que combina dados de diferentes domínios e diferentes tipos de estudos. Esse tipo de integração depende de variáveis que interliguem estudos separados. A variável de integração pode ser um número de acesso, um nome de espécie ou uma localização geográfica, por exemplo. Nessa linha de raciocínio, um cientista interessado em investigar a relação entre o aumento da criminalidade e a questão da empregabilidade, em diferentes regiões de um município, poderia relacionar os indicadores utilizando dois conjuntos de dados (datasets) independentes (criminalidade e empregabilidade), mas que poderiam ser interligados pela variável “município”, considerando que ambos os estudos apresentavam esses dados.

O estudo de Brunton e colaboradores (2019) pode ser considerado um exemplo de reúso por integração, uma vez que para investigar o consumo de energéticos e seus efeitos em adolescentes do Reino Unido os autores utilizaram coleções de dados de cinco diferentes estudos sobre sintomas e efeitos comportamentais associados a indicadores de consumo de bebidas energéticas na região. Os estudos tinham focos específicos sobre o efeito de bebidas energéticas e alguns haviam sido realizados com pessoas de diferentes idades, mas foram associados e integrados considerando os indicadores específicos obtidos para a faixa etária de interesse do estudo.

O reúso por metanálise caracteriza-se como aquele que combina análises de dados provenientes de múltiplos estudos independentes com perguntas de pesquisa e hipóteses idênticas ou muito semelhantes de modo a abordar questões além do escopo ou poder de uma única análise. Em suma, essa abordagem de reúso permite empreender uma análise de análises e extrair conclusões capazes de traduzir o estado da arte em determinado tema de pesquisa. Ao invés de reutilizar os dados primários de cada estudo conjuntamente, a metanálise toma por base os dados processados, inferências e resultados dos estudos de interesse. Portanto, essa modalidade de reúso requer cautela pelo fato de ser altamente sensível a erros e limitações metodológicas derivadas de múltiplos estudos.

A título de ilustração, essa abordagem de reúso de dados poderia ser conduzida por uma pesquisadora da área de nutrição desportiva e metabologia ao realizar a compilação dos resultados de centenas de estudos clínicos independentes, publicados sobre o efeito de suplementação antes e após o treino na hipertrofia, identificados por meio de revisão sistemática da literatura. Os resultados obtidos referentes às variáveis de interesse são extraídos de cada um dos estudos independentes e compilados em um novo conjunto de dados para que uma análise global permita conclusões mais contundentes sobre a existência de algum efeito diferenciado na administração de suplementos.

Camilo e coautores (2018) desempenharam um estudo com base na abordagem de reúso por metanálise reutilizando resultados de oito estudos que envolveram um total de mais de 20 mil indivíduos sobre o comportamento sedentário e o estado nutricional de idosos. Os autores analisaram, de forma conjunta, os resultados dos diferentes estudos, reconhecendo algumas diferenças metodológicas entre os

diferentes estudos, mas que, ainda assim, permitiram inferir relações entre hábitos sedentários e o estado nutricional. Além disso, demonstraram que o panorama de pesquisas sobre o tema indica que o comportamento sedentário não está necessariamente associado a maiores chances de sobrepeso e/ou obesidade.

O reuso por reanálise é aquele voltado para verificação dos resultados obtidos pelo estudo que gerou os dados, por meio de nova análise dos dados, utilizando os mesmos métodos e técnicas empregados no estudo original. Por conseguinte, essa abordagem de reuso está diretamente atrelada à questão da reprodutibilidade em que dados são submetidos à verificação, com vistas à confirmação ou refutação das conclusões anteriores.

Apesar de o grupo Evolo (2011) preferir o termo replicação (replication), entende-se que replicação busca realizar um novo estudo de modo a verificar se os resultados obtidos no estudo original são atingidos em condições semelhantes. Dessa maneira, a replicação não subentende o reuso dos dados propriamente, mas tão somente tomá-los como base comparativa para reafirmar ou confrontar os resultados obtidos em novas pesquisas. Estudos por replicação buscam determinar se as descobertas do estudo original podem ser aplicadas a outros participantes e circunstâncias, e, portanto, não são considerados reuso por essência.

Mas se o reuso por reanálise é uma repetição exata da análise dos mesmos dados, seguindo a mesma abordagem analítica empregada pelo estudo original, qual seria a sua relevância para a ciência? Essa abordagem é essencial para a autocorreção da ciência a partir do escrutínio de resultados de pesquisa.

A reanálise ocorre principalmente em ocasiões de suspeitas de fraudes por fabricação ou distorção de dados, que, caso confirmadas, podem incorrer em retratação pública e na remoção da publicação, dentre outras penalidades.

Após criteriosa avaliação de denúncias de má conduta em pesquisa, fabricação e deturpação de dados, 17 artigos de um mesmo pesquisador da área de marketing e de comportamento consumidor, filiado à Cornell University, foram retratados, seis deles publicados pela família de periódicos do The Journal of the American Medical Association (JAMA) (YEAGER, 2018). Vários outros periódicos, como a exemplo Nature, BioMed Central e Cell, tiveram artigos retratados nos últimos anos e adotam a política de informar quando o artigo está embargado por motivo de retratação ou foi removido. O observatório de retratações científicas Retraction Watch permite busca de artigos retratados em diferentes áreas do conhecimento e em centenas de periódicos científicos, inclusive especificando uma ou mais dentre os quase 100 tipos de violações de dados possíveis capazes de gerar contestações e retratações.

Embora seja considerada uma abordagem de reuso mais controversa e menos atrativa aos cientistas, o reuso por reanálise não acontece apenas nos bastidores dos periódicos ou dos comitês que avaliam denúncias de possíveis fraudes. Um exemplo com base na abordagem de reanálise, é o artigo de autoria de Borsboom e colaboradores (2017) publicado no Journal of Abnormal Psychology. Os autores realizaram uma reanálise utilizando os mesmos dados e os mesmos métodos adotados por Forbes e coautores (2017), que haviam publicado no mesmo periódico um estudo criticando a replicabilidade de estudos que utilizam a metodologia de análise de redes para o estudo de sintomas e patologias psíquicas. Com o título inusitado “False alarm? A comprehensive reanalysis of “Evidence that psychopathology symptom networks have limited replicability” by Forbes, Wright, Markon, and Krueger (2017)”, os autores contestaram os resultados obtidos por Forbes e outros (2017), apontando diferentes falhas cometidas pelos pesquisadores

capazes de explicar as conclusões equivocadas acerca da validade do método em questão.

Fato é que a intensificação das iniciativas em prol da transparência dos dados científicos amplia a disponibilidade de dados passíveis de serem reutilizados. Independentemente da abordagem de reuso adotada, seja para fins de investigação de novos problemas a partir de uma mesma coleção de dados, de reunir diferentes fontes de dados de forma a compor uma coleção de dados mais completa, de interligar e relacionar dados provenientes de múltiplos estudos, extrair inferências e identificar padrões em múltiplas coleções de dados acerca de um mesmo fenômeno de pesquisa, seja para verificar os dados a partir de reanálise buscando confirmar ou refutar resultados, é necessário que os dados apresentem condições mínimas de reusabilidade, conforme será tratado a seguir.

REUSABILIDADE E SUAS RELAÇÕES COM A DOCUMENTAÇÃO DOS DADOS

Diferentemente do percurso convencional da pesquisa, em que o pesquisador participa de todo o ciclo de planejamento, coleta, organização e transcrição e/ou decodificação dos dados, nas situações de reuso os pesquisadores não possuem controle algum sobre a qualidade dos dados.

A reutilização de dados de pesquisa é processo complexo que requer não apenas a capacidade de descobrir e acessar dados inteligíveis, confiáveis e relevantes (THESEN et al., 2011), mas principalmente exige dos pesquisadores uma capacidade interpretativa, de tradução, e de recontextualização dos dados primários coletados por terceiros, a fim de aplicá-los aos seus próprios fins, sem que incorram em interpretações equivocadas. Desse modo, antes de decidir por reutilizar dados coletados por terceiros, ou empreender nova coleta, os pesquisadores devem avaliar as condições de reusabilidade.

No entanto, assim como o termo reuso, mesmo que frequentemente enaltecida como uma qualidade importante dos dados de pesquisa, a reusabilidade ainda permanece relegada conceitualmente. Neste artigo, a reusabilidade representa o nível de conformidade dos dados de pesquisa com características determinantes para que eles sejam considerados reutilizáveis.

A reusabilidade constitui um dos quatro pilares do FAIR. Os dados de pesquisa considerados reutilizáveis devem estar em conformidade com os seguintes critérios: 1) apresentar uma pluralidade de atributos precisos e relevantes; 2) estar disponíveis com uma licença de uso de dados clara e acessível; 3) estar associados à sua proveniência; 4) atender aos padrões da comunidade relevantes ao domínio (disciplina/área de especialidade); 5) estar alinhados aos três demais princípios do FAIR; 6) contemplar descrição completa e rica de metadados, de modo a permitir uma ligação em integração com outras fontes de dados; e 7) prover metadados e informações de proveniência suficientes para permitir citação apropriada dos dados (WILKINSON et al., 2016).

Além das condições explicitadas, de acordo com David (1991), ao considerar o reuso como uma opção, os pesquisadores devem avaliar um conjunto de atributos na coleção de dados de interesse, incluindo: a adequação à pesquisa em desenvolvimento, a autoridade dos dados (quem produziu os dados?), a confiabilidade (o método adotado era rigoroso/adequado?), e a integridade e consistência dos dados (os dados estão completos e sem erros?).

Em todo caso, o reúso de dados requer prévio conhecimento detalhado do conjunto de dados e documentação suplementar (FANIEL; JACOBSEN, 2010). A documentação de dados desempenha um papel preponderante na reusabilidade dos dados disponíveis na web, especialmente porque a interação com pesquisadores do estudo original para o esclarecimento e melhor compreensão de possíveis questões relacionadas aos dados não pode ser garantida (FANIEL; ZIMMERMAN, 2011).

Ao reutilizarem dados, os pesquisadores dependem diretamente de descrições e documentos suplementares que detalhem os dados e o estudo original para auxiliar na capacidade interpretativa do reutilizador sobre os pontos fortes, implicações e limitações dos dados. Embora a experiência em pesquisa e em análise de dados auxiliem a capacidade avaliativa dos cientistas, a documentação de dados é a fonte imediata e principal de conhecimento sobre os dados (NIU, 2009ab).

A baixa qualidade ou a insuficiência de documentação de dados aumenta a suscetibilidade dos reutilizadores, mesmo os mais experientes, a possíveis desvios analíticos e semânticos (PIGGOT et al., 2001), além do inadvertido mau uso dos dados. Nesse sentido, a despeito dos desafios inerentes ao processo de documentação de dados, o reúso efetivo é diretamente dependente da qualidade das descrições e informações suplementares que acompanham os dados.

Embora a documentação de dados possa variar muito em termos de extensão, número de documentos comprovativos e nível de detalhamento, dependendo do estudo original e dos dados que reporta, alguns requisitos mínimos para permitir a reutilização dos dados são esperados. A preparação da documentação de dados para o compartilhamento é considerada uma atividade complexa e que exige tempo, esforço, recursos financeiros apoio e técnico (NIU, 2009ab).

DOCUMENTAÇÃO DE DADOS ABERTOS DE PESQUISA: DESAFIOS E NOVAS POSSIBILIDADES

A documentação de dados para reúso futuro não é tarefa trivial. Trata-se de um processo ainda mais complexo e laborioso quando executado para terceiros. Markus (2001) articula as diferenças entre documentar para si mesmo, documentar para outros semelhantes e não semelhantes com relação aos níveis de detalhamento e informações contextuais incluídas. No contexto científico, semelhantes podem ser entendidos como os pares envolvidos de forma direta ou indireta na pesquisa que originou os dados. Em suma, o autor estabelece que o nível de complexidade do processo de documentação aumenta com o grau de separação entre produtores de conhecimento e reutilizadores. Isso se agrava quando tratamos de dados que estão em um nível de abstração mais elevado do que fontes de conhecimento explícito como as que são exemplificadas pelo autor.

Markus (2001) explica que, quando documentamos informações para reúso próprio, a documentação gerada é, muitas vezes, um “subproduto do próprio trabalho” (p.14) e se reduz ao fornecer justificativa suficiente para várias decisões tomadas ao longo do processo. A documentação para outros semelhantes, por outro lado, deve ser fundamentada em alguns padrões para permitir a reutilização. No entanto, como é esperado que outras pessoas semelhantes compartilhem práticas, visões e entendimentos comuns, os dados, muitas vezes, são suficientemente inteligíveis com base em informações contextuais. Já a documentação para outros dissimilares é uma tarefa mais complexa, pois os produtores de dados precisam documentá-los de

forma que consigam capturar o contexto e todos os procedimentos para a geração dos dados em detalhes.

Seja para reuso futuro próprio ou para semelhantes ou não-semelhantes, no atual contexto da ciência aberta, em que os dados estão distribuídos na web, exige-se que os dados sejam anonimizados de modo a ocultar informações confidenciais e sigilosas antes de torná-los públicos. Desse modo, e considerando que os dados abertos de pesquisa devem ser documentados para um público diverso e também levar em conta possíveis colaborações científicas interdisciplinares, as comunidades científicas e entidades, dedicadas a promover a transparência e a reusabilidade em ciência, têm se preocupado em definir estratégias e parâmetros para a documentação de dados.

Uma documentação de dados adequada deve garantir que os mesmos sejam rastreáveis, inteligíveis, compreendidos e reutilizados de forma não ambígua por usuários futuros. Essa documentação necessita incluir descrições sobre o processo para a obtenção dos dados, a descrição dos dados propriamente, assim como informações sobre a proveniência, indicando as possíveis mudanças na coleção de dados e todos os históricos de edições. Sob esse aspecto, o United Kingdom Data Archive (UKDA) sugere que os dados sejam acompanhados de três níveis de documentação: 1) do estudo; 2) dos dados; e 3) metadados de catalogação (CORTI et al., 2014).

A documentação do estudo compreende informações contextuais capazes de fornecer informações de alto nível sobre o escopo da pesquisa, o tipo e abordagem metodológica adotados, os procedimentos de coleta de dados utilizados, preparações e manipulações a que os dados foram submetidos. Incluem ainda detalhes da fonte de dados como a descrição geral de sujeitos/participantes, informações técnicas sobre o software usado para criar e/ou manipular os dados e outras informações adicionais com relação à conversão de dados e portabilidade. Além de documentos referentes do comitê de ética (NIU, 2009ab).

A documentação dos dados propriamente fornece informações no nível de variáveis em um banco de dados ou em objetos individuais, como transcrições de entrevistas ou imagens. Além disso, inclui os materiais de codificação e classificação aplicados no estudo (dicionário de dados) e informações sobre confidencialidade e anonimização.

Os metadados de catalogação contemplam informações pesquisáveis estruturadas que ajudam os usuários a encontrar recursos de dados existentes e também citá-los em caso de reuso efetivo. Ainda incluem dados de identificação como, por exemplo, autoria, título da coleção de dados, localização geográfica, características temporais, condições de acesso e licenças de uso, identificadores persistentes, entre outros. A maior parte dos repositórios que oferece catálogos de dados na web ou portais de descoberta segue padrões ou esquemas internacionais de metadados. Em caso de repositórios disciplinares, é comum o uso de vocabulários e tesouros controlados que ajudam a manter a consistência e a qualidade da documentação dos dados.

Um exemplo de iniciativa que define padrões para documentação de dados é a Data Documentation Initiative (DDI), promovida pela organização Data Alliance. Dedicada aos dados estatísticos provenientes de pesquisas de levantamento (survey) e questionários das Ciências Sociais, a DDI definiu um padrão de metadados homônimo, aberto e não proprietário que teve sua primeira versão publicada em meados dos anos 1990.

Esse padrão vem sofrendo constantes atualizações ao longo dos anos. Em sua versão mais atual oferece, além de um padrão para a elaboração de dicionários de dados interativos (DDI Codebook), um modelo modular e extensível para o gerenciamento

de todo o ciclo de vida dos dados (DDI Lifecycle) baseado em XML , essencial para coleções de dados longitudinais. Com uma longa lista de repositórios e publicações ampliadas, que adotam o DDI para o registro e documentação de dados, a Data Alliance também tem desenvolvido algumas ferramentas que promovem a descoberta de dados e procuram ampliar as possibilidades de reuso dos dados a partir da documentação padronizada.

Um exemplo é a ferramenta de banco de perguntas (question banks) que permitem aos pesquisadores realizarem buscas federadas e mais granulares nos diferentes estudos arquivados, de modo a identificar perguntas utilizadas em questionários nos temas de interesse. Esses bancos de questões possibilitam que pesquisadores encontrem perguntas e conjuntos de respostas idênticos e semelhantes ao tema de interesse para fins de reuso, comparem dados, harmonizem ou desenvolvam novos questionários, partindo das questões identificadas, pois essa ferramenta também conta com recursos de avaliação de perguntas que ajudam o pesquisador a melhorar o design das questões, caso opte por criar as suas próprias.

Outro caso de banco de dados de questões é o Q-bank, coordenado pelo National Center for Health Statistics (NCHS), que tem como missão a melhoria de pesquisas do tipo survey por meio de colaboração. A plataforma permite buscas por temas e subtemas predefinidos (vocabulário controlado) e/ou termos livres, além de alguns outros filtros de busca. A Figura 1 exemplifica uma busca sobre questões relacionadas à deficiência e mobilidade, que tenham especificamente tratado da questão da mobilidade ao caminhar.

Figura 1: Recursos de busca do Q-Bank

The screenshot shows the Q-Bank search interface. On the left, there are search filters for Survey Type (Population, Establishment), Surveys, Question Topic (Disability/ Impairment, Mobility), Sub Topic, Agency, and Keywords (walking). A 'Search Questions' button is at the bottom of the filter panel. The main area displays 'There are 41 questions matching your criteria' and a table of results. The table has columns for Survey, Question, and Response Option(s). Below the table, there is a summary table with columns: Main Topic, Sub Topic, Disability/ Impairment, Keywords, Mode, and Year Tested.

Survey	Question	Response Option(s)
UNICEF Multiple Indicators Survey	Compared with children of the same age, does [he/she] have difficulty walking a short distance? for e.g. across a large room.	1) No difficulty 2) Some difficulty 3) A lot of difficulty 4) Cannot do at all
UNICEF Multiple Indicators Survey	Compared with children of the same age, does [he/she] have difficulty walking a longer distance? For e.g. to a store?	1) No difficulty 2) Some difficulty 3) A lot of difficulty 4) cannot do at all
Not Applicable	Do you have difficulty walking or climbing steps?	a. No -no difficulty b. Yes - some difficulty c. Yes - a lot of difficulty d. Cannot do at all
Module on Child Functioning and Disability	Compared with children of the same age, does [he/she] have difficulty walking? Would you say...	1) No difficulty 2) Some difficulty 3) A lot of difficulty 4) Cannot do at all

Main Topic	Sub Topic	Disability/ Impairment	Keywords	Mode	Year Tested
		Disability/ Impairment	Mobility	IA	2010

Fonte: <https://wwwn.cdc.gov/qbank/Search>

A partir do acesso à questão de interesse, o pesquisador pode verificar o relatório do estudo em detalhes, incluindo informações contextuais e metodológicas da pesquisa, ter acesso às outras questões do instrumento, salvar e compartilhar as questões desejadas e indicar citação. Apesar das importantes funções da plataforma, os dados primários obtidos pelas pesquisas não estão interligados aos registros das questões ou ao estudo de modo geral a que estão relacionadas, o que requer buscas

adicionais em outras plataformas, caso o pesquisador queira reutilizar os dados da pesquisa.

Outro recurso que tem sido adotado com vistas à otimização do reuso de dados é o vínculo de estudos que reutilizaram uma mesma coleção de dados à sua documentação. Para Niu (2009ab), a indicação de estudos relacionados a uma mesma coleção de dados enriquece a documentação dos dados visto que torna mais evidentes diferentes demonstrações práticas do processo de reuso dos dados.

O Inter-university Consortium for Political and Social Research (ICPSR) permite que usuários consultem uma base de dados de publicações relacionadas (artigos de periódicos, livros, capítulos de livros, relatórios governamentais, dissertações e teses, artigos de conferências, artigos de jornais, documentos audiovisuais, entre outros) que citam os dados arquivados em seu repositório. Além disso, as publicações relacionadas a um conjunto de dados são indicadas de forma dinâmica e atualizada como parte integrante do registro e da documentação dos dados, conforme pode ser visualizado nas Figuras 2 e 3.

Figura 2: Exemplo de coleção de dados no ICPSR

The screenshot displays the ICPSR website interface for the 'Youth Development Study, 1988-2011 [St. Paul, Minnesota] (ICPSR 24881)'. The page features a navigation bar with options like 'FIND DATA', 'SEARCH/COMPARE VARIABLES', and 'DATA-RELATED PUBLICATIONS'. Below the navigation bar, the study title and version information are shown. A table lists data files with columns for Name, Size, Preview, and Download. The table includes files such as 'D50 Study-Level Files' (640 KB) and 'D51 G1, Wave 1 (combined file)' (20 MB). To the right of the table, there are statistics: '190,275 Downloads * Usage Report * past three years' and '66 Data-related Publications'. A 'Notes' section contains information about data availability and version control. The ICPSR logo and a brief description of the organization's mission are also visible at the bottom of the page.

Name	Size	Preview	Download
D50 Study-Level Files	640 KB		
D51 G1, Wave 1 (combined file)	20 MB		
D52 G1, Wave 4 (combined file)	22 MB		
D53 G2, Wave 1 (combined file)	39 MB		
D54 G2, Wave 2 (combined file)	24 MB		
D55 G2, Wave 3 (combined file)	20 MB		
D56 G2, Wave 4 (combined file)	28 MB		

Fonte: <https://www.icpsr.umich.edu/icpsrweb/ICPSR/studies/24881/datadocumentation>

Figura 3: Detalhamento das Publicações Relacionadas aos Dados

The screenshot shows the ICPSR website interface. At the top, there is a navigation bar with 'Find & Analyze Data' and 'Log In/Create Account'. Below this, a purple header contains navigation links: 'FIND DATA', 'SEARCH/COMPARE VARIABLES', 'DATA-RELATED PUBLICATIONS', 'RESOURCES FOR STUDENTS', and 'HELP'. The main content area is titled 'Youth Development Study, 1988-2011 [St. Paul, Minnesota] (ICPSR 24881)'. It includes the version date (Dec 18, 2015), the principal investigator (Jeylan T. Mortimer), and the DOI. There are buttons for 'Download' and 'Analyze Online (0)'. A tabbed interface shows 'Data-related Publications' selected. A message box says 'More published analyses of the data in this study/series may exist. If you know of any not listed here, please suggest a new citation.' Below this, there is a search and sort section with 'Showing 1 to 50 of 66 entries' and 'Sort by: Pub Date (newest)'. A table lists publications with columns for Type, Year, and Citation. The first four entries are:

Type	Year	Citation
	2016	Vuolo, Mike, Mortimer, Jeylan T., Staff, Jeremy The value of educational degrees in turbulent economic times: Evidence from the Youth Development Study. <i>Social Science Research</i> , 57, 233-252. Full Text Options: DOI WorldCat Google Scholar Export Options: RIS EndNote related studies/series
	2015	Elason, Scott R., Mortimer, Jeylan T., Vuolo, Mike The transition to adulthood: Life course structures and subjective perceptions. <i>Social Psychology Quarterly</i> , 78, (3), 205-227. Full Text Options: DOI WorldCat Google Scholar PubMed Central Export Options: RIS EndNote related studies/series
	2015	Hitlin, Steven, Kirkpatrick, Johnson, Monica Reconceptualizing agency within the life course: The power of looking ahead. <i>American Journal of Sociology</i> , 120, (5), 1429-1472. Full Text Options: DOI WorldCat Google Scholar Export Options: RIS EndNote related studies/series
	2015	Johnson, Monica Kirkpatrick, Mortimer, Jeylan T. Reinforcement or compensation? The effects of parents' work and financial conditions on adolescents' work values during the Great Recession. <i>Journal of Vocational Behavior</i> , 87, 89-100. related studies/series

On the right side, there are statistics: '190,275 Downloads * Usage Report * past three years' and '66 Data-related Publications'. Below that is a 'Notes' section with two bullet points. At the bottom right, there is an 'ICPSR' logo and a short description of the organization's mission.

Fonte: <https://www.icpsr.umich.edu/icpsrweb/ICPSR/studies/24881/publications>

A partir do exemplo ilustrado na Figura 3, é possível constatar que 66 publicações foram baseadas na mesma coleção de dados, investigando temas diversos relacionados aos dados. Há também grande variedade dos periódicos e tipos de publicações geradas, assim como de autoria, indicando que o reuso dos dados tem sido feito por pesquisadores de diferentes filiações e não necessariamente envolvidos no estudo original.

A relação entre as produções geradas a partir de coleções de dados depende da correta citação e atribuição dos dados utilizados nessas produções, o que é constantemente incentivado pelo ICPSR por meio de ferramentas automáticas para citação e do uso de identificadores persistentes. Além do monitoramento automático de publicações baseadas nos dados custodiados pelo repositório, o ICPSR encoraja que os reutilizadores comuniquem e submetam suas produções científicas para complementar essa base relacional.

A bibliografia relacionada associada à documentação dos dados permite que pesquisadores identifiquem diferentes pesquisas, realizadas com base em uma mesma coleção de dados, seguindo diferentes abordagens de reuso. Possibilita ainda identificar quais questões acerca dos dados já foram investigadas e esgotadas, bem como oportunidades para novos estudos. Proporciona também que pesquisadores identifiquem implicações e relações interdisciplinares dos dados e, ainda, se aprofundem em questões metodológicas por vezes melhor discutidas e debatidas nas publicações científicas.

ALGUMAS REFLEXÕES FINAIS

No atual contexto científico, o compartilhamento de dados de pesquisa de modo sistemático e comprometido com o potencial de reuso futuro deixa de ser opcional e passa a ser imperativo. No entanto, o reuso de dados não pode ser assumido como uma mera consequência do compartilhamento, mas, sim, como uma prática que também requer estratégias e incentivos para que se materialize de forma comprometida com os princípios da ciência. Consequentemente, questões acerca do reuso e da reusabilidade de dados, tanto em nível conceitual quanto prático, devem ser amplamente debatidas.

O artigo ofereceu uma discussão conceitual acerca do reuso de dados e da reusabilidade no contexto científico e, a partir de uma revisão narrativa, oferece uma classificação inédita que considera as abordagens de reaproveitamento, agregação, integração, metanálise e reanálise como possibilidades de reuso de dados científicos de modo exemplificado. Também buscou explorar o efeito da documentação na reusabilidade de dados e apresentar iniciativas que viabilizam uma melhor verificação do potencial de reuso dos dados científicos.

Embora as tipologias aqui apresentadas possam não contemplar todas as abordagens de reuso possíveis e tenham se baseado em apenas alguns exemplos hipotéticos e casos da literatura, acredita-se que este é um primeiro passo para que o processo de reuso de dados seja abordado menos periféricamente e outras possíveis abordagens sejam identificadas de forma mais palpável e concreta para permitir que as diferentes comunidades científicas identifiquem oportunidades de reuso, tanto em suas disciplinas como em projetos interdisciplinares. Mesmo caso se aplica às questões concernentes à reusabilidade e ao incremento das ações voltadas para padrões e boas práticas de documentação de dados, que devem ser debatidas de modo a identificar soluções e alternativas para elevar o potencial de reuso dos dados que começam a se acumular em repositórios.

Artigo recebido em 08/07/2019 e aprovado em 25/11/2019.

REFERÊNCIAS

BORSBOOM, D. et al. False alarm?: a comprehensive reanalysis of "Evidence that psychopathology symptom networks have limited replicability" by Forbes, Wright, Markon, and Krueger (2017). *Journal of Abnormal Psychology*, v. 126, n. 7, p. 989-999, 2017.

BRUNTON, G.; KNEALE, D.; SOWDEN, A. Caffeinated energy drinks and effects in UK young people: a secondary analysis of population-level datasets. London: Department of Health & Social Care Reviews Facility, 2019.

CAMILO, B. de F. et al. Sedentary behavior and nutritional status among older adults: a meta-analysis. *Revista Brasileira de Medicina do Esporte*, v. 24, n. 4, p. 310-315, 2018.

CHANDRA, A.; COPEN, C. E.; STEPHEN, E. H. Infertility service use in the United States: data from the national survey of family growth, (1982-2010). *National Health Statistics Reports*, Hyattsville, n. 73, p.1-21, 2014.

CORTI, L. et al. Managing and sharing research data: a guide to good practice. London: SAGE, 2014.

- CURTY, R. G. et al. Attitudes and norms affecting scientists' data reuse. *PLoS One*, v. 12, n. 12, p. e0189288, 2017.
- DANIELS, M. G. Data reuse in museum contexts: experiences of archaeologists and botanists. 2014. Tese (Doutorado em Filosofia da Informação) - University of Michigan, Ann Arbor, 2014.
- DAVID, M. The science of data sharing: documentation. In: SIEBE, J. E. (ed.). *Sharing social science data: advantages and challenges*. Thousand Oaks: Sage Publications, 1991. p. 91-115.
- EVOIO WORKING GROUP. Reuse cases. 2011. Disponível em: http://www.evoio.org/wiki/Reuse_Cases Acesso em: 09 jun. 2019.
- FANIEL, I. M.; JACOBSEN, T. E. Reusing scientific data: how earthquake engineering researchers assess the reusability of colleagues' data. *Computer Supported Cooperative Work*, v. 19, n. 3, 2010, p. 355-375, 2010.
- FANIEL, I. M.; ZIMMERMAN, A. Beyond the data deluge: a research agenda for large-scale data sharing and reuse. *International Journal of Digital Curation*, v. 6, n. 1, p. 58-69, 2011.
- FEAR, K. M. Measuring and anticipating the impact of data reuse. Tese (Doutorado em Filosofia da Informação) - University of Michigan, Ann Arbor, 2013.
- FORBES, M. K. et al. Evidence that psychopathology symptom networks have limited replicability. *Journal of Abnormal Psychology*, v. 126, n. 7, p. 969-988, 2017.
- HEATON, J. *Reworking Qualitative Data*. Thousand Oaks: Sage Publications, 2004.
- KIM, Y.; YOON, A. Scientists' data reuse behaviors: a multilevel analysis. *Journal of the Association for Information Science and Technology*, v. 68, n.12, p.2709-2719, 2017
- MARKUS, L. M. Toward a theory of knowledge reuse: types of knowledge reuse situations and factors in reuse success. *Journal of Management Information Systems*, v. 18, n. 1, p. 57-93, 2001.
- NATIONAL NETWORK OF LIBRARIES OF MEDICINE. Data reuse. Disponível em: <https://nnlm.gov/data/thesaurus/data-reuse> Acesso em: 9 jun. 2019.
- NIU, J. Overcoming inadequate documentation. *Proceedings of the American Society for Information Science and Technology*, v. 46, n. 1, p. 1-14, 2009a.
- NIU, J. Perceived documentation quality of social science data. 2009. Tese (Doutorado em Filosofia da Informação) - University of Michigan, Ann Arbor, 2009b. Disponível em: http://deepblue.lib.umich.edu/bitstream/handle/2027.42/63871/niuif_1.pdf?sequence=1 Acesso em: 10 jun. 2019.
- PASQUETTO, I. V.; RANGLES, B. M.; BORGMAN, C. L. On the reuse of scientific data. *Data Science Journal*, v. 16, n. 8, 2017.
- PIGOTT, D. et al. An approach to managing repurposing of digitized knowledge assets. *Australasian Journal of Information Systems*, Malden, v. 9, n. 1, p. 92-103, 2001.
- TENOPIR, C. et al. Changes in data sharing and data reuse practices and perceptions among scientists worldwide. *PLoS One*, v. 10, n. 8, p. e0134826, 2015.
- THESSSEN, A. E. et al. Data issues in the life sciences. *ZooKeys*, v. 15, n.150, p. 15-51, 2011.
- VAN DE SANDT, T. et al. The definition of reuse. *Data Science Journal*, v. 18, n. 1, 2019.

VISSOCI, J. R. N. et al. Zika virus infection and microcephaly: evidence regarding geospatial associations. *PLoS Neglected Tropical Diseases*, v. 12, n. 4, p. e0006392, 2018.

WILKINSON, M. D. et al. The FAIR guiding principles for scientific data management and stewardship. *Scientific Data*, v. 3, 2016.

YEAGER, A. JAMA Journals retract six papers by Cornell researcher. *The Scientist*, 19 Sept. 2018. Disponível em: <https://www.the-scientist.com/news-opinion/jama-journals-retract-six-papers-by-cornell-food-scientist-64828> Acesso em: 09 jun. 2019.

ZIMMERMAN, A. S. Data sharing and secondary use of scientific data: experiences of ecologists. 2003. Tese (Doutorado em Filosofia da Informação) - University of Michigan, Michigan, 2003.

ZIMMERMAN, A. S. new knowledge from old data: the role of standards in the sharing and re-use of ecological data. *Science, Technology & Human Values*, v. 33, n. 5, p. 631-652, 2008.