

XVIII ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2017

GT-5 – Política e Economia da Informação - Pôster

IMPLICAÇÕES DO ACESSO ABERTO AOS DADOS DE PESQUISA: QUESTÕES EM DEBATE

Luciana Monteiro Krebs (Universidade Federal do Rio Grande do Sul - UFRGS)

Sônia Elisa Caregnato (Universidade Federal do Rio Grande do Sul - UFRGS)

IMPLICATIONS OF OPEN ACCESS TO RESEARCH DATA

Modalidade da Apresentação: Pôster

Resumo: O Acesso Aberto a Dados de Pesquisa, que integra a agenda da Ciência Aberta, é um movimento de uso e compartilhamento de dados científicos em repositórios digitais, tendo em vista a cooperação e o avanço na ciência global, no qual se engajam órgãos científicos nacionais e sindicatos científicos internacionais de dezenas de países desenvolvidos. O presente estudo tem como objetivo discutir algumas das implicações no uso e compartilhamento de dados de pesquisa, tendo em vista que o Acesso Aberto a Dados de Pesquisa não é ainda uma prática muito disseminada no Brasil. É uma pesquisa de natureza básica, abordagem qualitativa e objetivo exploratório. Como ponto de partida da discussão, apresenta o histórico e conceitos do movimento de acesso aberto na Ciência, depois estrutura-se a partir de questionamentos que serviram de base para escolha intencional dos textos e desenvolvimento dos temas. A discussão engloba, neste trabalho, o acesso aos dados, a propriedade dos dados, a privacidade dos sujeitos da pesquisa e a qualidade da pesquisa a partir do reuso de dados. Conclui-se que o Acesso Aberto a Dados de Pesquisa tem implicações em todos estes aspectos, que devem ser levados em consideração no caso de uma política nacional de incentivo ao compartilhamento de dados de pesquisa.

Palavras-Chave: Acesso Aberto a Dados de Pesquisa; Ciência Aberta; e-Science.

Abstract: Open Access to Research Data, which is part of the Open Science agenda, is a movement for the use and sharing of scientific data in digital repositories, with a view to cooperation and advancement in global science, involving national and international scientific bodies. Trade unions from dozens of developed countries. The present study aims to discuss some of the implications in the use and sharing of research data, considering that Open Access to Research Data is not yet a widely disseminated practice in Brazil. It is a research of basic nature, qualitative approach and exploratory objective. As a starting point for the discussion, it presents the history and concepts of the open access movement in Science, then is structured based on questions that served as a basis for intentional choice of texts and development of themes. The discussion includes, in this work, access to data, data ownership, privacy of the individuals of the research and the quality of the research from the reuse of data. It is concluded that Open Access to Research Data has implications in all these aspects, which should be taken into account in the case of a national policy to encourage the sharing of research data.

Keywords: Open Access to Research Data; Open Science; e-Science.

1 INTRODUÇÃO

Os dados são intrínsecos à ciência, em especial nas áreas que utilizam métodos experimentais. Há diversos estudos tratando as questões éticas na pesquisa e seus desafios (PITHAN, 2013; GOLDIM, 2016), mas com foco nos dilemas relacionados à publicação acadêmica, lidando majoritariamente com a prevenção de más condutas no processo de publicação, tais como “[...] casos de plágio, fraude, falsificação de dados, duplicidade e segmentação de resultados, autoria indevida, conflito de interesses, etc.” (SHINKAI, 2011, p.2). Nesse trabalho, no entanto, os esforços estão voltados para o processo de uso e compartilhamento de dados abertos de pesquisa, que pode ocorrer antes (no caso do uso) e depois (no caso de compartilhamento) da publicação da pesquisa em periódicos acadêmicos, e por isso não envolve relatórios publicados e documentos baseados em dados analisados (como artigos, teses, dissertações e livros). Entendemos por dados abertos de pesquisa, a

[...] informação digital estruturada por metodologia formal com o objetivo de criar novas pesquisas ou bolsa de estudos. Esses dados podem estar em uma variedade de formatos adequados para comunicação, interpretação ou processamento, incluindo leituras de sensores, pesquisas e equipamentos de laboratório, modelos de simulação e bancos de dados compilados e arquivos de texto, entre outros. (PARHAM; BODNAR; FUCHS, 2012, p.10, tradução nossa).

Sem a pretensão de esgotar o tema, o presente estudo apresenta algumas reflexões que visam contribuir para o debate acerca das implicações éticas do Acesso Aberto a Dados de Pesquisa (AADP) no país. A pesquisa é qualitativa e tem caráter exploratório, por tratar de um tema ainda pouco desenvolvido na literatura especializada no país. Para dar início à construção do trabalho, elaboraram-se questões éticas de partida, que se originam do acesso aberto aos dados de pesquisa ou são afetadas por ele. A discussão engloba os impactos do AADP no acesso à informação, propriedade dos dados, privacidade dos participantes e qualidade da pesquisa. As questões serviram então como base para a realização de um levantamento bibliográfico seletivo, ou seja, a partir da consulta a fontes de informação nacionais e internacionais, realizou-se a triagem e seleção de alguns trabalhos de interesse para a discussão do tema.

2 DESENVOLVIMENTO

Com a formação do Centro Mundial de Dados (*World Data Centre - WDC*), estabeleceu-se o conceito de acesso aberto a dados de pesquisa. O objetivo da criação do

centro foi minimizar o risco de perda de dados e maximizar a acessibilidade aos mesmos. Assim, o Conselho Internacional para a Ciência¹ estabeleceu vários centros de dados mundiais e recomendou que os dados coletados de pesquisas observacionais fossem disponibilizados, ainda 1955, em formato legível por máquina. (NATIONAL..., [201-]).

Em 2007, a Organização para a Cooperação e Desenvolvimento Econômico (OCDE)² publicou os Princípios e Diretrizes da OCDE para o acesso aos dados da investigação financiada por fundos públicos, com base em fundamentos comumente acordados entre os Ministros da Ciência de todas as nações da organização a partir de 2004. Com o objetivo de melhorar o acesso e o compartilhamento internacional de dados de pesquisa geradas a partir de financiamento público, o documento é uma recomendação jurídica flexível. (OECD, 2004; OECD, 2007).

No Brasil, Sayão e Sales (2013) vêm discutindo diversos aspectos do AADP nos últimos anos. Em setembro de 2016 o IBICT lançou o Manifesto Aberto de Acesso Aberto a Dados da Pesquisa Brasileira para Ciência Cidadã (IBICT, 2016) e em 2017 teve início o projeto Acesso Aberto a Dados de Pesquisa da Rede Nacional de Pesquisa (RNP)³, que remete à:

[...] disponibilização de dados de pesquisa em plataformas digitais de livre acesso, respeitando os direitos autorais, períodos de embargo, questões de confidencialidade, privacidade, especificidades das áreas científicas, entre outros aspectos. (RNP, 2017).

Embora a prática já seja comum fora do país, no Brasil ainda não está consolidada, e por esta razão, nas subseções deste trabalho, discorreremos sobre alguns questionamentos éticos envolvendo o acesso aberto a dados de pesquisa.

2.1 Dados abertos de pesquisa e o acesso à informação

Se assumirmos que dados de pesquisa são oriundos de projetos planejados e executados por pesquisadores, seriam eles os únicos a terem direito a acessar esses dados? O que diz a lei a respeito do acesso a informações científicas? A Lei de Acesso à Informação

¹ O *International Council for Science (ICS)* dedica-se à cooperação no avanço da ciência e tem como membros 120 órgãos e sindicatos científicos que representam 140 países. Site: <https://www.icsu.org/about-us>. Acesso em: 28 jul. 2017.

² São membros da OCDE: Austrália, Áustria, Bélgica, Canadá, República Checa, Dinamarca, Finlândia, França, Alemanha, Grécia, Hungria, Islândia, Irlanda, Itália, Japão, Coreia, Luxemburgo, México, Países Baixos, Nova Zelândia, Noruega, Polônia, Portugal, República Eslovaca, Espanha, Suécia, Suíça, Turquia, Reino Unido e Estados Unidos, além da Comissão das Comunidades Europeias. (OECD, 2007).

³ Para mais informações, ver http://www.mcti.gov.br/noticia/-/asset_publisher/epbV0pr6elS0/content/rnp-e-ibict-lancam-chamada-para-financiar-projeto-sobre-acesso-a-dados-de-pesquisa.

XVIII ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2017
23 a 27 de outubro de 2017 – Marília – SP

(Lei nº 12.527, de novembro de 2011) menciona a pesquisa científica nos capítulos II (Do acesso a informações e sua divulgação) e IV (Das restrições de acesso à informação). No capítulo II, está vetado ao cidadão o acesso a informações referentes a projetos de pesquisa e desenvolvimento científicos ou tecnológicos cujo sigilo seja imprescindível à segurança da sociedade e do Estado, como demonstra o grifo no capítulo II, artigo 7º, inciso VII, parágrafo 1º:

§ 1o O acesso à informação previsto no *caput* não compreende as informações referentes a **projetos de pesquisa e desenvolvimento científicos ou tecnológicos cujo sigilo seja imprescindível à segurança da sociedade e do Estado**. (BRASIL, 2011, documento eletrônico, grifo nosso).

Porém, não fica claro quem julga quais projetos de pesquisa estão ou não dentro do espectro de sigilo previsto, ou de que forma essa decisão é tomada. Mesmo a Seção II (Da Classificação da Informação quanto ao Grau e Prazos de Sigilo) do Capítulo IV é vaga. No Artigo 23 dessa seção, Inciso VI, está dito que as informações passíveis de serem classificadas como “sigilosas” são aquelas cuja divulgação ou acesso irrestrito possam:

[...] VI - prejudicar ou causar risco a **projetos de pesquisa e desenvolvimento científico ou tecnológico, assim como a sistemas, bens, instalações ou áreas de interesse estratégico nacional**; [...] (BRASIL, 2011, documento eletrônico, grifo nosso)

Em outras palavras, não é transparente para o cidadão quais projetos de pesquisa trariam ameaça à segurança da sociedade e do Estado caso seus dados fossem compartilhados. Outra preocupação que se coloca nesse âmbito é de que, com acesso aberto e irrestrito aos dados, não está devidamente esclarecido quais entidades seriam responsáveis por verificar casos de conflito de interesse em projetos avaliados. Isso, se de fato a análise for feita caso a caso e por especialistas do campo de conhecimento em questão. Assim, havendo futuramente a recomendação por parte da Rede Nacional de Pesquisa (RNP) para que os grupos de pesquisa procedam com compartilhamento de seus dados digitais, será interessante elaborar políticas específicas para estas práticas.

2.2 A propriedade dos dados abertos de pesquisa

Davis e Vickery (2007) afirmam que os dados primários tornaram-se a principal moeda da ciência. Sendo assim, a quem pertencem os dados de pesquisas realizadas com recursos públicos? Seria sensato afirmar que esses dados primários são de propriedade de quem financiou a pesquisa, e, portanto, devem ser abertos para acesso do público em geral?

Um levantamento feito pela organização sem fins lucrativos Batelle, que mede os investimentos mundiais em pesquisa, indica que o Brasil investe 1,3% do seu PIB em

pesquisa e desenvolvimento. Em 2013 o valor foi de US\$ 31 bilhões, o que o coloca na 36ª posição no ranking mundial (considerando a relação do valor investido e sua representatividade no PIB do país). Em termos de investimento absoluto, o país fica em 10º lugar no ranking. (CALEIRO, 2016).

Mas o que isso significa? Para alguns especialistas, é um mau investimento. A revista *Nature* publicou em 2014 um *ranking* que indica que o Brasil possui uma das piores relações custo-benefício no mundo para o investimento em ciência. Figurando em 50º lugar entre 53 países avaliados, o país só é melhor que o Egito, Turquia e Malásia. Para chegar ao *ranking*, dividiu-se o número de artigos publicados em 68 revistas científicas internacionais de alto prestígio pelo valor total investido em pesquisa no país. O Brasil, com investimento de cerca de US\$ 30 bilhões ao ano em ciência e desenvolvimento, publicou 670 artigos nessas revistas em 2013. A Arábia Saudita lidera o *ranking*, com investimento de US\$ 500 milhões ao ano – incluindo público e privado –, e tendo publicado 288 artigos. (ALVES, 2014).

Há quem defenda, porém, que a pesquisa brasileira é menosprezada. Para Bonorino (2015), no país não se conhece nem valoriza o potencial da pesquisa inclusive em trazer retorno financeiro (como é o caso do Projeto Genoma Humano, que possui um retorno de US\$ 141,00 por cada dólar investido). A pesquisadora relata que os cortes de custos refletem uma dura realidade: que projetos começam e acabam ao sabor das orientações partidárias então no poder. E assim, os projetos de P&D ganham e perdem relevância segundo uma lógica estritamente política e nunca conforme os interesses reais da população, quem, no final do dia, é que paga por todo o processo. (BONORINO, 2015).

Resta saber se o AADP traria maior transparência (com efeitos na melhoria da eficiência na pesquisa científica do país) ou se apenas oneraria o pesquisador, que além das funções de docente e pesquisador também precisaria agregar a função de gestor de dados.

2.3 A privacidade dos participantes da pesquisa no AADP

Uma questão muito relevante é o respeito à privacidade dos indivíduos participantes. De que formas pode-se garantir o sigilo quanto à identidade dos sujeitos das pesquisas em plataformas de AADP?

Uma técnica de proteção da privacidade dos sujeitos da pesquisa é a anonimização. À medida que os dados vão se tornando acessíveis para fora do núcleo de pesquisa é preciso aumentar o cuidado com a anonimização dos indivíduos de forma eficiente. Além disso,

deve-se ter em mente que apenas ocultar nomes ou IPs pode não ser mais suficiente, devido à triangulação que se torna possível a partir do uso secundário e terciário (cruzamento dos dados de pesquisa com outras bases para gerar novas informações).

Mayer-Schonberger e Cukier (2013) alertam para o fato de que a anonimização funciona em um pequeno universo de dados, e há aí uma intersecção entre dados abertos de pesquisa e o fenômeno *big data*. Segundo os autores, com o aumento de quantidade e variedade de informações, fica mais fácil reidentificar os participantes. As preocupações com a privacidade tendem a influenciar a elaboração de legislações específicas, como a “*US Do Not Track Online Act*” em 2011 nos Estados Unidos (FANG *et al*, 2015), permitindo que um usuário proíba a coleta ou uso de suas informações pessoais ou por exemplo com o “direito a ser esquecido” (*Right to be Forgotten*) na União Européia (BUNN, 2015).

Com o AADP parece que a necessidade de Acordos de Confidencialidade claros e precisos é ainda maior, mencionando a possibilidade de compartilhamento dos dados em plataformas de acesso aberto e todos os procedimentos de anonimização a serem realizados antes da publicação dos dados. Também deve-se buscar a supressão de todos os dados dos participantes que não sejam de extrema relevância para a interpretação e análise dos mesmos, evitando a possibilidade de triangulação no uso secundário e terciário dos dados.

2.4 Impacto na qualidade da pesquisa com uso e reuso de dados abertos de pesquisa

Em que medida o reuso de dados pode comprometer os resultados das investigações? Quando dados são usados para responder diferentes questões de pesquisa das previstas inicialmente, ou quando são combinados com diferentes conjuntos de dados provenientes de outras fontes ou utilizando diferentes métodos, os resultados obtidos podem apresentar vieses ou distorções. Isto pode ocorrer simplesmente porque há o risco de o pesquisador perder de vista todos os critérios e métodos de coleta, limpeza e tratamento dos dados disponíveis, pois estes processos foram executados por terceiros, muitas vezes alheios ao reuso posterior dos dados.

Segundo Curty (2016), os dados depositados são fruto de pesquisas específicas, criados ou coletados em circunstâncias particulares, seguindo procedimentos e técnicas específicas de coleta de dados. A autora estudou a percepção dos pesquisadores quanto a benefícios e riscos do AADP e identificou, entre outros, receios relacionados a este aspecto: o medo de infringir códigos de ética (em caso de confidencialidade, por exemplo, ou uso

secundário não previsto em acordos prévios); o medo de más interpretações, por incorreções ou uso equivocado não intencional de dados alheios; e a vulnerabilidade a erros escondidos, ou seja, o receio de que dados coletados por outros pesquisadores contenham erros que não facilmente identificáveis pelos usuários de dados abertos de pesquisa.

Tendo em vista que a descontextualização dos dados pode prejudicar as análises se não levada em consideração apropriadamente, constata-se que a transparência e o rigor metodológico precisam ser redobrados em caso de compartilhamento de dados de pesquisa em plataformas de acesso aberto.

3 CONSIDERAÇÕES FINAIS

Neste trabalho abordaram-se alguns aspectos éticos imbricados na prática de Acesso Aberto a Dados de Pesquisa. Em relação ao acesso à informação, evidencia-se a falta de clareza sobre quais projetos de pesquisa podem ter seus dados abertos sem ameaçar a segurança da sociedade e do Estado. Quanto à propriedade dos dados, é preciso compreender se a relação custo-benefício de investimento em pesquisa no Brasil poderia ser positivamente afetada com o AADP. A proteção da privacidade dos participantes das pesquisas também deve ser uma preocupação em especial no acesso aberto. E por fim discutiu-se as possíveis implicações da qualidade da pesquisa com o reuso de dados, pois a mudança metodológica pode gerar deslizos neste início de *modus operandi*. As questões elencadas neste artigo e muitas outras serviriam para iniciar um diálogo consciente sobre o tema entre governo, pesquisadores e sociedade, considerando não apenas os benefícios do AADP, mas também seus riscos e implicações.

REFERÊNCIAS

ALVES, Gabriel. Gasto brasileiro com ciência é muito pouco eficiente, diz 'Nature'. **Folha de São Paulo**, São Paulo, Coluna Ciência, 16 nov. 2014.

BONORINO, Cristina. Pesquisa científica no Brasil é menosprezada: Cristina Bonorino, professora e pesquisadora 1C do CNPq, escreve sobre a importância de se investir no conhecimento. **Zero Hora**, Caderno Frutos da Ciência. 15 ago. 2016.

BRASIL. Lei nº 12.527, de 18 de novembro de 2011. Regula o acesso a informações previsto no inciso XXXIII do art. 5º, no inciso II do § 3º do art. 37 e no § 2º do art. 216 da Constituição Federal; altera a Lei no 8.112, de 11 de dezembro de 1990; revoga a Lei no 11.111, de 5 de maio de 2005, e dispositivos da Lei no 8.159, de 8 de janeiro de 1991; e dá outras providências. **Diário Oficial [da] República Federativa do Brasil**: seção 1, Brasília, DF, Edição Extra, p. 1 (Publicação Original), 18 nov. 2011.

XVIII ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2017
23 a 27 de outubro de 2017 – Marília – SP

BUNN, A. The curious case of the right to be forgotten. **Computer Law & Security Review**, v. 31, n. 3, jun. 2015.

CALEIRO, João Pedro. 15 países que mais investem em pesquisa (e o Brasil em 36º). **Exame**, São Paulo, Economia, 13 set. 2016.

CURTY, R. G. Factors Influencing Research Data Reuse in the Social Sciences: An Exploratory Study. **International Journal of Digital Curation (IJDC)**, v. 11, n. 1, p.96-117, 2016. Disponível em < <http://www.ijdc.net/index.php/ijdc/article/view/11.1.96> >. Acesso em 12 jul. 2017.

DAVIS, H.M.; VICKERY, J.N.. Datasets, a shift in the currency of scholarly communication: implications for library collections and acquisitions. **Serials Review**, v. 33, n. 1. p.26-32, 2007. Disponível em: < <http://www.tandfonline.com/doi/abs/10.1080/00987913.2007.10765089> >. Acesso em 12 jul. 2017.

FANG, H. et al. A survey of big data research. **IEEE network**, v. 29, n. 5, p.6-9, set-out./2015. Disponível em: < <https://doi.org/10.1109/MNET.2015.7293298> >. Acesso em ago. 2017.

IBICT. **Ibict lança Manifesto de Acesso Aberto a Dados de Pesquisa Brasileira para Ciência Cidadã**. 03 out. 2016.

GOLDIM, J.R. Integridade na pesquisa: um desafio sempre atual. IN: PITHAN, L. H. BARCELLOS, M. L. L.. **Integridade na pesquisa e propriedade intelectual na Universidade**. Porto Alegre: EDIPUCRS, 2016.

NATIONAL CENTERS FOR ENVIRONMENTAL INFORMATION. **World Data Centers**. [201-] Disponível em: < <https://www.ncdc.noaa.gov/customer-support/world-data-centers> >. Acesso em: 12 jul. 2017.

OECD. **OECD Principles and Guidelines for Access to Research Data from Public Funding**. 2007. 23 p. Disponível em: < <http://www.oecd.org/science/sci-tech/oecdprinciplesandguidelinesforaccesstoresearchdatafrompublicfunding.htm> >. Acesso em: 01 jul. 2017.

OECD. **Science, Technology and Innovation for the 21st Century**: Meeting of the OECD Committee for Scientific and Technological Policy at Ministerial Level, 29-30 January 2004 - Final Communique. [S.l.]: OECD, Science and Technology policy. 30 jan. 2004.

PARHAM, S. W.; BODNAR, J.; FUCHS, S.. Supporting tomorrow's research: assessing faculty data curation needs at Georgia Tech. **College & Research Libraries News**, p. 10-13, jan. 2012.

PITHAN, L. H.; VIDAL, T. R. A.. O plágio acadêmico como um problema ético, jurídico e pedagógico. **Direito & Justiça**, v. 39, n. 1, 2013.

RNP. **Acesso Aberto a Dados de Pesquisa**. [Pôster apresentado no wRNP 2017]. 2017. Disponível em: <

XVIII ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2017
23 a 27 de outubro de 2017 – Marília – SP

http://wrnp.rnp.br/sites/wrnp2017/files/11_wrnp2017_cartaz_aadp_design.pdf >. Acesso em: 12 jul. 2017.

SAYÃO, L. F.; SALES, L. F.. Dados de pesquisa: contribuição para o estabelecimento de um modelo de curadoria digital para o país. **Tendências da Pesquisa Brasileira em Ciência da Informação**, v. 6, n. 1, 2013.

SHINKAI, R. S.. Integridade na pesquisa e ética na publicação. **Scientia Medica**, Porto Alegre, v. 21, n. 1, p.2-3, 2011.