



Paulo Cezar Vieira Guanaes

Abertura e compartilhamento de dados de pesquisa subjacentes a artigos científicos: questões do direito autoral

**Tese de doutorado
Dezembro de 2020**



UFRJ



UNIVERSIDADE FEDERAL DO RIO DE JANEIRO - UFRJ
ESCOLA DE COMUNICAÇÃO - ECO
INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA - IBICT
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO - PPGCI

Paulo Cezar Vieira Guanaes

Abertura e compartilhamento de dados de pesquisa subjacentes a artigos científicos:
questões do direito autoral

Tese de Doutorado apresentada ao Programa de Pós-Graduação em Ciência da Informação, convênio entre o Instituto Brasileiro de Informação em Ciência e Tecnologia e a Universidade Federal do Rio de Janeiro/Escola de Comunicação, como requisito parcial à obtenção do título de Doutor em Ciência da Informação.

Orientadora: Profa. Dra. Sarita Albagli

RIO DE JANEIRO

2020

CIP - Catalogação na Publicação

GG913a GUANAES, PAULO CEZAR VIEIRA
Abertura e compartilhamento de dados de pesquisa subjacentes a artigos científicos: questões do direito autoral / PAULO CEZAR VIEIRA GUANAES. -- Rio de Janeiro, 2020.
233 f.

Orientador: Sarita Albagli.

Tese (doutorado) - Universidade Federal do Rio de Janeiro, Escola da Comunicação, Instituto Brasileiro de Informação em Ciência e Tecnologia, Programa de Pós-Graduação em Ciência da Informação, 2020.

1. Abertura e compartilhamento de dados de pesquisa. 2. Direito autoral. 3. Dados subjacentes. 4. Licenciamento. 5. Comunicação científica. I. Albagli, Sarita, orient. II. Título

Paulo Cezar Vieira Guanaes

Abertura e compartilhamento de dados de pesquisa subjacentes a artigos científicos: questões do direito autoral

Tese de Doutorado apresentada ao Programa de Pós-Graduação em Ciência da Informação, convênio entre o Instituto Brasileiro de Informação em Ciência e Tecnologia e a Universidade Federal do Rio de Janeiro/Escola de Comunicação, como requisito parcial à obtenção do título de Doutor em Ciência da Informação.

Orientadora: Prof. Dra. Sarita Albagli

Aprovada em

Profª. Dra. Sarita Albagli (Orientadora)
PPGCI-IBICT/UF RJ

Profª. Dra. Lena Vânia Ribeiro Pinheiro
PPGCI-IBICT/UF RJ

Profª. Dra. Luana Farias Sales Marques
PPGCI-IBICT/UF RJ

Profª. Dra. Vanessa de Arruda Joege
FIOCRUZ

Prof. Dr. Allan Rocha de Souza
PPED-IE-UF RJ / ITR-UFRRJ

Para Luciana e Rafael, meus amores.

AGRADECIMENTOS

À professora Sarita Albagli, minha orientadora. Nada melhor do que ser orientado por quem você admira intelectualmente. Competente, amiga, transmitiu-me com sua experiência, conhecimento e firmeza a segurança necessária para mergulhar em um tema ainda, de fato, pouco explorado na ciência da informação.

À minha esposa, a minha doce Luciana, amorosa companheira de todas as horas, um anjo da guarda atento, sempre com uma palavra de incentivo em momentos necessários, os quais só a cumplicidade de nosso longo relacionamento é capaz de detectar.

Ao meu filho Rafael, um constante incentivador, que muitas vezes aconteceu com ideias que vinham revestidas de carinho, mas também de muito senso prático, em favor do andamento do trabalho concreto.

Às professoras Lena Vânia Pinheiro Ribeiro e Luana F. Sales e ao professor Allan Rocha de Souza, pelas generosas e estruturantes contribuições feitas no exame de qualificação desse estudo.

Às professoras e professores do Programa de Pós-Graduação em Ciência da Informação do IBICT-UFRJ, pelos ricos ensinamentos, alargando nossas visões com aulas e temas multidisciplinares, desenvolvidos em ambiente de democráticas discussões. Às incansáveis Janete Desidério e Christine Alvarez, da secretaria acadêmica.

Ao Rodrigo Souto, amigo e guru da tecnologia da informação e comunicação, pelas impagáveis consultorias.

Aos editores científicos e gestores de repositórios, pelas entrevistas concedidas, que me proporcionaram um dos melhores momentos dessa jornada, pela troca, mas sobretudo pelo aporte de distintas visões à pesquisa.

Aos colegas da revista *Trabalho, Educação e Saúde*, da EPSJV-Fiocruz, pelo apoio e incentivo.

À professora Paulo Xavier dos Santos, coordenadora do Grupo de Trabalho em Ciência Aberta da Fiocruz, pelo convite para integrar esse GT, e aos demais integrantes pela troca enriquecedora, experiência que me trouxe até aqui.

Ao Grupo de Pesquisa em Ciência Aberta do Liinc, do IBICT/UFRJ, pelo apoio e interação.

[...] não temos mais pressa.

(Dino Buzzati, *Naquele exato momento*, 1986)

GUANAES, PAULO Cezar Vieira. **Abertura e compartilhamento de dados de pesquisa subjacentes a artigos científicos: questões do direito autoral**. 2020. Tese (Doutorado em Ciência da Informação) – PPGCI-IBICT/UFRJ, 2020.

RESUMO

O objetivo principal desta tese é analisar as implicações do direito autoral aplicável a dados de pesquisa, observando-se o processo de publicação e abertura dos dados de pesquisa subjacentes a artigos de revistas científicas, em particular as do campo da saúde. Inicialmente, incursionamos sobre o tema da ciência aberta, evidenciando seus vínculos com o movimento pelo acesso aberto à informação científica e o movimento de abertura de dados de pesquisa, ressaltando os distintos desdobramentos em países centrais e em países em desenvolvimento, especialmente os da América Latina. Para tanto, realizamos pesquisa bibliográfica, sobretudo no repositório PubMed Central, arquivo de acesso gratuito sobre literatura de periódicos biomédicos e de ciências biológicas da Biblioteca Nacional de Medicina do NIH. Também foram consultados livros jurídicos, documentos legais sobre proteção a dados pessoais e direitos autorais da União Europeia e dos Estados Unidos, as leis brasileiras que regulam o direito autoral e a proteção a dados pessoais e a Constituição. A consulta a essa compilação serviu de base a um estudo conceitual sobre direito autoral, não exaustivo, no qual tratamos, entre outros temas, do conteúdo e objeto do direito autoral, seu caráter internacional, as limitações a esses direitos, autoria e propriedade de dados, e proteção a bancos de dados, com especial atenção ao licenciamento de dados. Permitiu também considerações acerca da proteção à privacidade e confidencialidade de seres humanos envolvidos em pesquisa, particularmente os dados sensíveis da pessoa, mais propensos a usos discriminatórios ou lesivos. No intuito de mapear e analisar as políticas de publicação, abertura e compartilhamento de dados de pesquisa e o quadro normativo sobre direito autoral e licenciamento para reúso de dados de revistas científicas, repositórios de dados de pesquisa e agências de financiamento, desenvolvemos duas pesquisas, entre novembro de 2019 e setembro de 2020. A primeira consistiu em uma pesquisa observacional, de cunho qualitativo, nos *sites* de nove representantes do processo de comunicação da ciência de países centrais: revistas *Plos Medicine*, *theBMJ* e *Nature Medicine*; repositórios Dryad, Figshare e Zenodo; e agências National Institutes of Health, Wellcome Trust/Wellcome Open Research e Bill & Melinda Gates Foundation, em que examinamos as mencionadas políticas e as suas abordagens sobre direito autoral. A segunda tratou-se de uma pesquisa empírica, qualitativa, realizada em duas partes, com dois segmentos do sistema de comunicação científica: editores científicos da Fiocruz, do *hot site* SciELO Saúde Pública e da UFRJ, que representam, em sua maioria, uma parte do universo de revistas de saúde pública de origem nacional e latino-americana; e gestores de repositórios nacionais, da Fiocruz, IBICT, Metabuscador de dados da Fapesp, UFPR, UFSC, RNP, e do repositório internacional Dryad. O objetivo foi avaliar a implementação da abertura, compartilhamento e disponibilização em acesso aberto de dados de pesquisa subjacentes a artigos por revistas e repositórios de dados, ouvir e observar como vêm lidando com questões de direito autoral envolvidas. A primeira parte consistiu na autoaplicação de questionário eletrônico enviado por *e-mail*. A segunda foi efetivada com

entrevistas por videoconferência, visando ao aprofundamento dos pontos de discussão contidos no questionário eletrônico. O resultado indica que a temática da abertura de dados subjacentes a artigos e suas implicações com o direito autoral ainda não é tratada como prioridade, apesar de haver projetos de construção de estruturas e de políticas em andamento nessas instituições. Conclui-se que essa nova fronteira da ciência aberta apresenta um estágio avançado em termos de sua proposição como teoria, havendo o endosso de *players* de peso da comunicação científica, mas prescinde, na prática, de um padrão universal de requisitos para sua operacionalização em larga escala, de um sistema de recompensa que valorize os dados de pesquisa e de um quadro normativo que, no limite, exclua certas categorias de dados e informações científicos da proteção de direitos autorais ou proíba a imposição de restrições ao seu uso.

Palavras-chave: Abertura e compartilhamento de dados de pesquisa. Direito autoral. Dados subjacentes. Licenciamento. Comunicação científica. Ciência da Informação.

GUANAES, PAULO Cezar Vieira. **Abertura e compartilhamento de dados de pesquisa subjacentes a artigos científicos: questões do direito autoral.** 2020. Tese (Doutorado em Ciência da Informação) – PPGCI-IBICT/UFRJ, 2020.

ABSTRACT

The main objective of this thesis is to analyze the implications of copyright applicable to research data, observing the process of publication and openness of research data underlying scientific journal articles, particularly those in the health field. Initially, we explored the theme of open science, highlighting its links with the movement for open access to scientific information and the movement of opening research data, highlighting the different developments in central and developing countries, especially those in Latin America. To this end, we conducted a bibliographic search, especially in the PubMed Central repository, a free access archive on the literature of biomedical and biological sciences journals of the NIH National Library of Medicine. Legal books, legal documents on protection of personal data and copyrights of the European Union and the United States, Brazilian laws that regulate copyright and protection of personal data and the Constitution were also consulted. The consultation of this compilation served as the basis for a conceptual study on copyright, which is not exhaustive, in which we deal, among other topics, with the content and object of copyright, its international character, the limitations to these rights, authorship and data ownership, and protection of databases, with special attention to data licensing. It also allowed for considerations about the protection of privacy and confidentiality of human beings involved in research, particularly the sensitive data of the person, more prone to discriminatory or harmful uses. In order to map and analyze the policies for publishing, openness and sharing research data and the regulatory framework on copyright and licensing for reusing data from scientific journals, research data repositories and funding agencies, we developed two surveys, among November 2019 and September 2020. The first consisted of observational, qualitative research on the websites of nine representatives of the science communication process in central countries: the journals *Plos Medicine*, *theBMJ* and *Nature Medicine*; Dryad, Figshare and Zenodo repositories; and agencies National Institutes of Health, Wellcome Trust / Wellcome Open Research and Bill & Melinda Gates Foundation, in which we examine the aforementioned policies and their approaches to copyright. The second was an empirical, qualitative research, carried out in two parts, with two segments of the scientific communication system: scientific editors from Fiocruz, from the hot site SciELO Saúde Pública and UFRJ, which represent one part of the universe of public health journals of national and Latin American origin; and managers of national repositories, from Fiocruz, IBICT, Fapesp Data Meta-harvester, UFPR, UFSC, RNP and the international Dryad repository. The objective was to evaluate the implementation of the open, sharing and availability in open access of research data underlying articles by journals and data repositories, listening and observing how they have been dealing with copyright issues involved. The first part consisted of a self-administered electronic questionnaire sent by e-mail. The second was carried out with videoconference interviews, aiming to deepen the discussion points contained in the electronic questionnaire. The result indicates that the theme of openness data underlying articles and their implications for copyright is not yet treated as a priority, despite the fact that there are ongoing projects to build structures and policies in

these institutions. It is concluded that this new frontier of open science presents an advanced stage in terms of its proposition as a theory, with the endorsement of important players of scientific communication, but does not offer, in practice, a universal standard of requirements for its wide operation scale, a reward system that values research data and a regulatory framework that ultimately excludes certain categories of data and scientific information from copyright protection or prohibits the imposition of restrictions on its use.

Keywords: Openness and sharing research data. Copyright. Underlying data. Licensing. Scientific communication. Information Science.

LISTA DE FIGURAS

- Figura 1.** Mapa conceitual do artigo de dados, *53*
- Figura 2.** Ciclo de vida dos dados, *65*
- Figura 3.** Distribuição de licença de uso dos periódicos, *104*
- Figura 4.** Evolução de conjuntos de dados de pesquisa disponibilizados em repositórios de dados, 2009-2018, *127*
- Figura 5.** Disponibilização pública de dados de pesquisa, *172*
- Figura 6.** Adoção da abertura de dados de pesquisa, *175*
- Figura 7.** Indicação de repositórios, *177*
- Figura 8.** Repositório com embargo, *178*
- Figura 9.** Que repositório indicar?, *179*
- Figura 10.** Repositório e política de direitos autorais, *182*
- Figura 11.** Serviços de repositório: ordem de importância, *189*
- Figura 12.** O repositório na comunicação científica, *191*
- Figura 13.** Qual é a melhor licença para reuso de dados?, *194*
- Figura 14.** Domínio público para dados, *195*
- Figura 15.** Um direito novo para dados de pesquisa, *197*

LISTA DE QUADROS

- Quadro 1.** Meios práticos para proteção de privacidade e informações, *30*
- Quadro 2.** A gradação do aberto, *34*
- Quadro 3.** Tipos resumidos de políticas de dados dos periódicos Spring Nature, *59*
- Quadro 4.** Pesquisa sobre percepção de ciência aberta com pesquisadores da área de biologia britânicos. Visão geral da análise temática (2013-2014), *63*
- Quadro 5.** Os princípios orientadores do FAIR, *68*
- Quadro 6.** Lista não exaustiva de princípios que informam o direito autoral, *74*
- Quadro 7.** Licenças para permissão de reuso de dados de pesquisa em bases legais, *97*
- Quadro 8.** Requisitos das políticas de dados de pesquisa de cada revista, *125*
- Quadro 9.** Requisitos das políticas de dados de pesquisa de cada repositório, *142*
- Quadro 10.** Conjunto de periódicos científicos da Fiocruz, *164*
- Quadro 11.** Coleção temática de periódicos científicos da SciELO Saúde Pública, *165*

LISTA DE TABELAS

Tabela 1. Distribuição das licenças CC-BY no SciELO Brasil, em 2014, **103**

Tabela 2. Envio dos questionários eletrônicos aos editores científicos, **168**

Tabela 3. Envio dos questionários eletrônicos aos gestores de repositórios, **188**

LISTA DE ABREVIATURA E SIGLAS

AJOL - African Journals OnLine
ANDS - Australian National Data Service
APC - Article Processing Charges
Bireme - Centro Latino-Americano e do Caribe de Informação em Ciências da Saúde
BMC - BiomedCentral
BOAI - Budapeste Open Access Initiative
CC - Creative Commons
CENS - Center for Embedded Networked Sensing
CERN - Organisation Européenne pour la Recherche Nucléaire
CNPq - Conselho Nacional de Desenvolvimento Científico e Tecnológico
COAR - Confederation of Open Access Repositories
CODATA - Comitee on data
DOI - Digital object identifier
EMBL-EBI - European Bioinformatics Institute
EOSC - European Open Science Cloud
Fapesp - Fundação de Amparo à Pesquisa do Estado de São Paulo
Fiocruz - Fundação Oswaldo Cruz
GATT - General Agreement on Tariffs and Trade
IBICT - Instituto Brasileiro de Informação em Ciência e Tecnologia
ICMJE -International Committee of Medical Journal Editors
INASP - International Network for the Availability of Scientific Publications
NASA - National Aeronautics and Space Administration
NFS - National Foundation Science
NIH - National Institutes of Health
NSB - National Science Board
OAI - Open Archives Initiative
OASPA - Open Access Scholarly Publishers Association
OCDE - Organização para a Cooperação e Desenvolvimento Econômico

OKF - Open Knowledge Foundation

OMC - Organização Mundial do Comércio

OMPI - Organização Mundial da Propriedade Intelectual

OMS - Organização Mundial da Saúde

OPAS - Organização Pan-Americana da Saúde

ORCID - Open Researcher and Contributor ID

PLoS - Public Library of Science

RCKU - Research Councils of United Kingdom

RDA - Research Data Alliance

RNP - Rede Nacional de Pesquisa

SciELO - Scientific Eletronic Library Online

SPARC - the Scholarly Publishing and Academic Resources Coalition

TDM - Text and Data Mining

TIC - Tecnologias de informação e comunicação

TPM - Technological Protection Measure

TRIPS - Agreement on Trade-Related Aspects of Intellectual Property Rights

UCLA - Universidade da Califórnia

UnB - Universidade de Brasília

UFPR - Universidade Federal do Paraná

UFRGS - Universidade Federal do Rio Grande do Sul

UFRJ - Universidade Federal do Rio de Janeiro

UFSC - Universidade Federal de Santa Catarina

WCT - WIPO Copyright Treaty

WIPO - World Intellectual Property Organization

SUMÁRIO

- 1 INTRODUÇÃO, 20**
- 2 ABERTURA DE DADOS DE PESQUISA: NOVA FRONTEIRA DA CIÊNCIA ABERTA, 26**
 - 2.1 Privacidade, direito autoral e restrição a análise de conteúdo digital, 29**
 - 2.2 O acesso aberto à informação científica se amplia e agrega os dados de pesquisa, 35**
 - 2.3 O acesso aberto a artigos e dados de pesquisa ganha apoio mundial, 40**
 - 2.4 Dados de pesquisa: (in)definições, 42**
 - 2.5 O acesso aberto a dados de pesquisa, 48**
 - 2.6 O papel dos editores científicos na abertura: falta padrão?, 52**
 - 2.7 A percepção dos pesquisadores, 61**
- 3 IMPLICAÇÕES DO DIREITO AUTORAL NA DISPONIBILIZAÇÃO PÚBLICA DE DADOS DE PESQUISA SUBJACENTES, 71**
 - 3.1 Direito autoral: breve histórico, conceituação, conteúdo e objeto, princípios, 71**
 - 3.1.1 A face internacional do direito autoral, 75**
 - 3.1.2 Limitações e exceções ao direito autoral, 77**
 - 3.1.3 Autoria e propriedade de dados de pesquisa, 78**
 - 3.2 A aplicabilidade do direito autoral em dados de pesquisa, 80**
 - 3.2.1 Aplicação de direitos autorais em dados factuais, 88**
 - 3.3 A proteção legal a bancos de dados, 89**
 - 3.3.1 A proteção de bancos de dados sob a lei de direitos autorais, 90**
 - 3.3.2 A proteção de bancos de dados pelo direito *sui generis*, 91**
 - 3.4 Remover restrições: Licenciamento de dados de pesquisa e domínio público, 94**
 - 3.5 CC-BY é a licença mais aberta, mas não está livre de controvérsias, 99**
 - 3.5.1 SciELO adota a CC BY, 102**
- 4 ABERTURA DE DADOS DE PESQUISA E O DIREITO AUTORAL NAS POLÍTICAS DE REVISTAS DE SAÚDE, REPOSITÓRIOS E AGÊNCIAS, 106**
 - 4.1 O desafio das revistas de saúde, 108**
 - 4.1.1 *PLoS Medicine*, 111**
 - 4.1.2 Direitos autorais: incompatibilidades, 114**
 - 4.1.3 *theBMJ*, 115**
 - 4.1.4 Direito autoral: diretrizes para artigos e dados se confundem, 118**

4.1.5 *Nature Medicine*, 120

4.1.6 Direitos autorais: poucas orientações, 123

4.2 Os repositórios se reinventam, 126

4.2.1 Repositórios de dados generalistas e o direito autoral, 128

4.2.2 Dryad, 129

4.2.3 Direitos autorais na perspectiva do Dryad, 130

4.2.4 Por que o CC0?, 132

4.2.5 Figshare, 134

4.2.6 Direitos autorais: além das licenças CC, 136

4.2.7 Zenodo, 138

4.2.8 Direitos autorais: *download* não transfere direitos, 140

4.3 As agências de financiamento batem pé: acesso aberto a dados de pesquisa já!, 142

4.3.1 National Institutes of Health (NIH), 144

4.3.2 Direitos autorais: sem licenças Creative Commons, 147

4.3.3 Wellcome Trust/Welcome Open Research, 149

4.3.4 Direito autoral: negociação justa (*fair dealing*) facilita a pesquisa, 152

4.3.5 Fundação Bill & Melinda Gates, 154

4.3.6 Direitos autorais como estratégia para garantir acesso, 157

4.4 Conflito aberto, 158

5 ABERTURA DE DADOS DE PESQUISA: UMA VISÃO DE EDITORES DO CAMPO DA SAÚDE E GESTORES DE REPOSITÓRIOS, 161

5.1 Editores científicos: Questionário e Entrevista, 163

5.1.2 Um exame observacional dos *sites* das revistas selecionadas, 169

5.2 Abertura de dados e direito autoral: uma visão de editores do campo da saúde, 172

5.3 Gestores de repositórios: Questionário e Entrevista, 183

5.4 Abertura de dados de pesquisa e direito autoral: a perspectiva dos gestores de repositórios, 189

6 CONCLUSÃO, 199

REFERÊNCIAS, 205

APÊNDICE A - Questionário eletrônico de autoaplicação para editores científicos, 226

APÊNDICE B - Questionário eletrônico de autoaplicação para gestores de repositórios institucionais e de dados, **229**

APÊNDICE C – Roteiro de perguntas para entrevista com editores científicos, **232**

APÊNDICE D – Roteiro de perguntas para entrevista com gestores de repositórios digitais e de dados, **232**

APÊNDICE E – Termo de Consentimento Livre e Esclarecido (enviado aos editores científicos), **233**

APÊNDICE F – Termo de Consentimento Livre e Esclarecido (enviado aos gestores de repositórios), **233**

APÊNDICE G – Roteiro das perguntas de uma entrevista realizada com umx Product Manager & Make Data Count Principal Investigator, do repositório internacional Dryad, **234**

1 INTRODUÇÃO

Esta tese tem como objetivo principal analisar questões do direito autoral aplicável a dados de pesquisa no processo de publicação e abertura dos dados de pesquisa subjacentes a artigos de revistas científicas, em particular as do campo da saúde. A abordagem do tema tem a ciência aberta em perspectiva, considerando-se a abertura de dados de pesquisa como um desdobramento do movimento de acesso aberto à informação científica.

O movimento de acesso aberto à informação científica ampliou-se em razão de um recrudescimento do movimento pela ciência aberta – que o influencia e por ela é influenciado –, o que enseja uma cascata de práticas abertas, turbinadas pelas tecnologias de informação e comunicação (TIC). Com potencial para gerar uma mudança paradigmática no atual modo de fazer ciência, tais práticas consistem, entre outras, na revisão por pares aberta, cadernos científicos abertos, ciência cidadã, recursos educacionais abertos e, talvez, a principal narrativa atual – a abertura dos dados de pesquisa (THE ROYAL SOCIETY, 2012; BORGMAN, 2012; 2015; ALBAGLI; CLINIO; RAYCHTOCK, 2014). A ciência aberta e essa nova agenda são abordadas no capítulo 2.

Ante a alegação, basicamente, de aprimoramento da reprodutibilidade, da transparência e da reutilização de dados de pesquisa financiada por recursos públicos, a abertura de dados de pesquisa passou a ser uma exigência de agências de fomento à pesquisa¹ e de editores de revistas científicas² dos países centrais³, que vêm elaborando variadas políticas de disponibilização pública dos chamados dados de pesquisa subjacentes aos artigos (HRYNASZKIEWICZ et al., 2020). São os dados que fundamentam a pesquisa desenvolvida ou, mais precisamente, “os dados necessários para validar os resultados de pesquisa apresentados em publicações científicas” (COMISSÃO EUROPEIA, 2017, p. 9).

Na esteira da abertura, os repositórios específicos para depósito de dados de pesquisa (subjacentes a artigos ou independentes) “se tornaram uma parte vital da infraestrutura de

¹ Ver em: National Institutes of Health (NIH) <https://grants.nih.gov/grants/policy/data_sharing/>; e em: Wellcome Trust <<https://wellcome.ac.uk/funding/guidance/data-software-materials-management-and-sharing-policy>>.

² Ver em: PLoS <https://grants.nih.gov/grants/policy/data_sharing/>; e em: Nature <<https://www.nature.com/news/announcement-where-are-the-data-1.20541>>.

³ Neste estudo, utiliza-se o termo “países centrais” (da Europa e EUA) em contraposição a “países periféricos” (da Índia, Ásia, América Latina), conforme o conceito de Rosa Luxemburgo de que o sistema capitalista mundial, no seu processo de constituição histórica, gerava o subdesenvolvimento na periferia como um aspecto complementar do desenvolvimento nos países centrais (LOUREIRO, 2015).

pesquisa” (OCDE, 2017, p. 8). Além de oferecerem depósito, curadoria, preservação digital, identificadores como o *digital object identifier* (DOI) para cada conjunto de dados, garantindo a devida citação, gestão administrativa e legal do compartilhamento e reuso, fornecem fontes de dados globais a campos da pesquisa dependentes dessa disponibilidade, o que tem feito surgir uma nova geração desses repositórios (COAR, 2017) e estudos sobre o seu papel como componente fundamental da infraestrutura pública de pesquisa (UHLIR, 2006).

A operação de abertura e compartilhamento de dados de pesquisa abrange todo o ciclo de pesquisa – plano, coleta, garantia, descrição, preservação, descoberta, integração, análise, de acordo com o DataONE (Figura 2) – e ainda o depósito, curadoria, acesso e reuso dos dados resultantes. No entanto, são práticas que podem esbarrar em diversos tipos de proteção jurídica a conjuntos de dados: direitos autorais, patentes, segredos industriais e outros, no universo da propriedade intelectual; direitos à privacidade e confidencialidade de dados pessoais, que são “informações relacionadas a pessoa natural identificada ou identificável” (BRASIL, 2018, p. 25), no âmbito das garantias individuais e coletivas; direitos relacionados à segurança nacional; e direitos protegidos por contratos (CARROL, 2015; GUIBAULT; MARGONI, 2015; KOBAYASHI; KANE; PATON, 2018; GUANAES et al., 2018; DONEDA, 2019).

Para fins deste trabalho, no entanto, embora reconheçamos que conjuntos de dados de pesquisa possam ser abrigados por vários tipos de proteção jurídica, como os já citados a título de exemplos, vamos nos fixar na análise do direito de autor, ou direito autoral, relacionado à abertura de dados de pesquisa subjacentes aos artigos submetidos para publicação em revistas científicas. Também não se cogitou desenvolver estudos sobre as implicações da ética nesse tema.

Parte importante deste cenário é o licenciamento de dados de pesquisa. Liberar dados sem deixar claro os termos de uso pode ser contraproducente para quem deseja ver seus dados reutilizados. No capítulo 3, ao lado de um breve estudo conceitual sobre direito autoral e a perspectiva de sua eventual incidência na abertura de dados de pesquisa, o tema do licenciamento é examinado por meio de uma análise das licenças públicas, com ênfase nas da Creative Commons (CC), lançadas em 2001. Com o mote “Alguns direitos reservados”, que permitem cópia e compartilhamento de obras literárias, educativas e científicas com menos restrições, contrapõem-se ao tradicional regime de “Todos os direitos reservados”. As licenças têm sido encampadas por agências de financiamento à pesquisa, periódicos científicos e

repositórios, mas não escapam de polêmicas sobre qual licença é a mais adequada para trabalhos científicos, por exemplo, com uma delas retratada no item 3.5 do capítulo 3.

Esses três atores da comunicação científica – agências de financiamento à pesquisa, periódicos e repositórios – e suas implicações com a proteção do direito autoral a dados de pesquisa subjacentes a artigos são centrais nesta tese, que apresenta, no capítulo 4, o resultado de uma pesquisa sobre eles, objetivando identificar o que suas políticas de abertura de dados dizem sobre direito autoral. Os três segmentos têm sido instados a buscar um conjunto comum de requisitos de políticas de abertura de dados, respeitadas as diferenças disciplinares, visando à adoção de um padrão para publicação e abertura desses dados em conexão com os artigos científicos. É um novo cenário que sugere práticas de mudança social que pressionam mudanças institucionais que, por sua vez, podem ditar novos comportamentos, mormente na comunicação científica, mas também suscitar questões legais.

De acordo com a Organização Mundial de Propriedade Intelectual – OMPI (World Intellectual Property Organization – WIPO), *direito de autor* “é um termo legal usado para descrever os direitos que os criadores têm sobre suas obras literárias e artísticas” (OMPI, 2020, *on line*). Bittar (2008) cita os dois tipos de conteúdo nos direitos autorais vinculados ao titular da criação intelectual: os *direitos morais*, que se referem ao direito de reivindicar a paternidade e a integridade da obra; e os *direitos patrimoniais*, que correspondem à utilização econômica da obra.

A Convenção de Berna, um pacto internacional adotado em 1886, com última atualização em 1979, referendado pelo Brasil, trata da proteção das obras e dos direitos de seus autores. A proteção de obras intelectuais originadas no campo científico está estabelecida no artigo 2, n. 1, no qual a Convenção diz que a expressão “obras literárias e artísticas” inclui “toda produção no domínio literário, científico e artístico, qualquer que seja o modo ou a forma de sua expressão, como livros, panfletos e outros escritos” (OMPI, 2020, *on line*, grifos nossos), o que abrange o artigo científico e pode abarcar o conjunto de dados de pesquisa. Não protege, porém, fatos e ideias. Conforme Carroll (2015, *on line*) ensina, os direitos autorais são aplicáveis ao compartilhamento de dados de pesquisa se “o criador, o proprietário ou o repositório que abriga os dados quiserem limitar direitos como a cópia, a distribuição ou outros reúsos de dados”.

Pelo fato da tese dar ênfase ao campo da saúde, trazemos também considerações acerca da proteção à privacidade e confidencialidade dos dados de sujeitos de pesquisa em

humanos, particularmente os dados sensíveis da pessoa, mais propensos a usos discriminatórios ou lesivos.

Neste cenário em reconfiguração por força das agendas da ciência aberta, ainda há o desafio de gerenciar a abundância de dados gerados pela tecnológica e-Science⁴ em pesquisas colaborativas internacionais, obrigando à observância de diversas leis nacionais sobre direitos autorais.

Diante do quadro sumário apresentado, ganhou relevância a realização deste estudo, que vem revestida da questão norteadora – “Como as revistas científicas, particularmente as da área da saúde, estão lidando com a questão da abertura de dados de pesquisa subjacentes aos artigos que publicam em face do direito autoral?” – e também de perguntas subsidiárias, tais como “As eventuais mudanças das práticas de comunicação científica repercutem no regime de direito autoral dos dados de pesquisa?” ou “Em que medida o processo de publicação e abertura de dados de pesquisa indicaria a emergência de um novo regime de direito autoral no campo da ciência?”

Cabe reiterar a ressalva de que estamos cientes de que o debate sobre proteção de direitos sobre os dados e conjuntos de dados de pesquisa é amplo, mas o foco deste trabalho é a proteção que os direitos autorais podem exercer sobre dados de pesquisa.

Uma das razões de concentrar a pesquisa proposta no campo da saúde é a minha atuação há 15 anos na Fundação Oswaldo Cruz (Fiocruz) como tecnologista em saúde, exercendo a função de editor executivo da revista *Trabalho, Educação e Saúde*, publicada pela Escola Politécnica de Saúde Joaquim Venâncio, uma de suas unidades científicas. E também pela participação, até 2018, no Grupo de Trabalho em Ciência Aberta da Fiocruz como coordenador do Grupo de Trabalho em Ciência Aberta – Marcos Legais, cuja missão era fornecer subsídios e diretrizes jurídicos para o desenvolvimento da Política de Abertura de Dados para Pesquisa da instituição, ainda hoje em fase de elaboração. Nesta passagem pelo GT Ciência Aberta – Marcos Legais, contribuí na elaboração de dois trabalhos de pesquisa de autoria coletiva: *Livro Verde – Ciência aberta e dados abertos: mapeamento e análises de*

⁴ Termo criado por John Taylor, diretor geral do Escritório de Ciência e Tecnologia do Reino Unido, em 1999, é a aplicação da tecnologia da computação para a realização de pesquisa científica moderna, com grande volume de dados, via internet, incluindo a preparação, experimentação, coleta de dados, disseminação de resultados, armazenamento a longo prazo e acessibilidade de todos os materiais gerados pelo processo científico, incluindo modelagem e análise de dados, cadernos de laboratório eletrônicos/digitalizados, conjuntos de dados brutos e integrados, versões de rascunhos, pré-impressões e publicações impressas e/ou eletrônicas (BOHLE, 2013). Uma discussão sobre ciência aberta e e-Science pode ser encontrada em Appel, Maciel e Albagli (2016).

políticas, infraestruturas e estratégias em perspectiva nacional e internacional (SANTOS; ALMEIDA; HENNING, 2017), do qual fui um dos autores; e *Marcos legais nacionais em face da abertura de dados para pesquisa em saúde: dados pessoais, sensíveis ou sigilosos e propriedade intelectual* (GUANAES et al., 2018), do qual fui o organizador e um dos autores, juntamente com Allan Rocha de Souza, Danilo Doneda e Francisco José T. do Nascimento, ambos publicados pela Fiocruz. Trata-se do relatório de uma pesquisa documental efetuada no ordenamento jurídico brasileiro para identificar as normas legais que incidem sobre questões relacionadas à abertura de dados para pesquisa em saúde, entre elas a utilização e proteção de dados pessoais, sensíveis ou sigilosos, segurança da informação, reuso e compartilhamento de dados e direito autoral.

A proposta metodológica desta tese é de uma pesquisa exploratório-descritiva, qualitativa quanto à sua natureza, a fim de compreender como as revistas científicas, em particular as do campo da saúde, estão lidando com a questão da abertura de dados de pesquisa subjacentes aos artigos que publicam no que diz respeito ao direito autoral.

Primeiro, no intuito de efetivar os objetivos declarados, a proposta metodológica contemplou uma pesquisa bibliográfica e documental na literatura especializada nacional e internacional, na qual reuniram-se estudos e documentos sobre a publicação e a abertura de dados de pesquisa subjacentes a artigos, com foco na inter-relação dos seguintes atores-chave desse ecossistema: pesquisadores, revistas científicas, preferencialmente da área da saúde, repositórios de dados e agências de financiamento à pesquisa, visando a obter uma compreensão de suas práticas. Identificaram-se estudos prévios sobre o direito autoral como instrumento de proteção a direitos de obras científicas, em especial de conjuntos de dados de pesquisa, o que permitiu uma análise do quadro normativo tanto no que concerne à abertura de dados de pesquisa quanto aos direitos autorais.

Em segundo lugar, foi feita uma pesquisa de campo, que é relatada no capítulo 5. A pesquisa desenvolveu-se em dois momentos, com dois segmentos do sistema de comunicação científica: editores científicos da Fiocruz, da coleção temática internacional SciELO Saúde Pública e da Universidade Federal do Rio de Janeiro (UFRJ); e gestores de repositórios institucionais e de dados de pesquisa da Rede de Repositórios de Dados Científicos do Estado de São Paulo, gerenciada pela Fundação de Amparo à Pesquisa do Estado de São Paulo (Fapesp), Universidade Federal do Paraná (UFPR), Universidade Federal de Santa Catarina (UFSC), Universidade Federal do Rio Grande do Sul (UFRGS), Universidade de Brasília

(UnB), da Rede Nacional de Pesquisa (RNP), do (Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT) e repositório generalista internacional Dryad.

Com o emprego de dois instrumentos de coleta – questionário e entrevista –, a primeira parte consistiu de envio de dois questionários eletrônicos específicos a editores e gestores (Apêndices A e B), contendo perguntas fechadas, abertas e de múltipla escolha (VIEIRA, 2009), produzidos no Google Formulários. A segunda parte foi executada mediante a realização de entrevistas com editores científicos e gestores de repositórios, apoiadas por dois roteiros com perguntas semiestruturadas (Apêndices C e D), também específicas para cada um dos segmentos, realizadas e gravadas na ferramenta Zoom, devido à necessidade de aprofundar os pontos de discussão dos questionários eletrônicos. Também foi realizada uma entrevista com uma gestora do repositório internacional Dryad, que respondeu a perguntas por *e-mail*, que podem ser acessadas no Apêndice G.

No capítulo conclusivo, derivamos algumas conclusões do estudo efetuado em torno da questão que o norteou: “Como as revistas científicas, particularmente as da área da saúde, estão lidando com a questão da abertura de dados de pesquisa subjacentes aos artigos que publica em face do direito autoral?”. Reitera-se que o debate sobre a proteção legal de dados e conjuntos de dados é amplo, mas concentramos a pesquisa em eventuais questões potenciais de direitos autorais. Destacamos ainda que, em razão do persistente descompasso regulatório entre direito autoral e ciência aberta, são bem-vindas pesquisas futuras que explorem temas como o impacto do direito autoral em plataformas digitais de publicação de resultados de pesquisa; e as restrições impostas por regulamentos internacionais às exceções e limitações ao direito autoral, mediante medidas tecnológicas de proteção.

2 ABERTURA DE DADOS DE PESQUISA: NOVA FRONTEIRA DA CIÊNCIA ABERTA

O termo “ciência aberta” não surgiu neste e tampouco no século passado. É o que nos revela Paul David num extenso ensaio, em que examina a economia do mecenato na produção de conhecimento e sua influência na formação histórica de elementos-chave no *ethos* e na estrutura organizacional da ciência aberta financiada publicamente. David assinala que o surgimento da ideia e prática da “ciência aberta” ocorreu no final do século XVI e início do XVII, constituindo-se em um aspecto organizacional distintivo e vital da Revolução Científica. O autor acrescenta que isso representou uma ruptura com o *ethos* de segredo anteriormente dominante na busca dos “segredos da natureza” e uma passagem a um novo conjunto de normas, incentivos e estruturas organizacionais que reforçaram os compromissos dos pesquisadores científicos com a rápida divulgação de novos conhecimentos (DAVID, 2008).

Hoje, em geral, a ciência aberta envolve um feixe de movimentos e iniciativas que visam tornar a produção de pesquisa financiada com recursos públicos o mais amplamente acessível em formato digital à comunidade científica e à sociedade em geral. Entre as iniciativas de ciência aberta, podemos citar os recursos educacionais abertos, software e hardware abertos, cadernos de pesquisa abertos, revisão por pares aberta e, provavelmente, a mais notável iniciativa, que é o movimento pelo acesso aberto à literatura acadêmica e científica que se desdobrou na abertura de dados de pesquisa, tema deste estudo ao lado das implicações do direito autoral que daí decorrem (ALBAGLI; CLINIO; RAYCHTOCK, 2014).

Segundo a Organização para a Cooperação e Desenvolvimento Econômico (OCDE), não existe uma definição formal de ciência aberta. No entanto, em seu relatório *Making Open Science a Reality* (OCDE, 2015a), que analisa o desempenho de seus países-membros sobre apresentação de resultados de pesquisa – artigos e dados de pesquisa – de forma acessível ao pesquisador e inovador, essa organização intergovernamental advoga que

O termo [ciência aberta] refere-se a esforços de pesquisadores, governos, agências de fomento à pesquisa ou da própria comunidade científica para

tornar o volume de produção primária dos resultados da pesquisa financiada publicamente – artigos e dados da pesquisa – acessível ao público em formato digital com nenhuma ou com a mínima restrição possível, como um meio para acelerar a pesquisa. Esses esforços refletem o interesse em aumentar a transparência e a colaboração e promover a inovação (OCDE, 2015a, p. 7).

Ainda segundo a OCDE (2015a), a ciência aberta também representa o encontro entre a antiga tradição de abertura proporcionada pela ciência e as modernas ferramentas das TIC, que reconfiguraram o empreendimento científico.

A internet e as plataformas *on line* têm permitido a criação de novas formas de organização e publicação de artigos científicos e grandes conjuntos de dados, a fim de que sejam imediatamente disponibilizados aos demais pesquisadores ou outros usuários da internet. E as TIC possibilitam a coleta de grandes quantidades de dados e informações, por meio da mineração de texto e dados, que podem se transformar em base de experimentos e pesquisas científicas e contribuir para tornar a ciência cada vez mais orientada a dados. Repositórios e arquivos *on line* oferecem a possibilidade de armazenar, acessar, usar e reutilizar pesquisas e seus resultados científicos (artigos e conjuntos de dados) e agilizar a transferência de conhecimento entre pesquisadores e entre campos científicos, abrindo novas formas de colaboração e novos métodos de pesquisa (OCDE, 2015a).

O termo “ciência aberta” tem, portanto, uma ampla gama de significados, desde a publicação dos resultados, a partir do início do ciclo da pesquisa, até o compartilhamento gratuito de todos os elementos da pesquisa, incluindo também protocolos e planos de análise. Os principais elementos que a caracterizam são transparência, acessibilidade, compartilhamento e colaboração (FELL, 2019). Albagli, Clinio e Raychtock (2014, p. 435) afirmam tratar-se de um “termo guarda-chuva, que engloba diferentes significados, tipos de práticas e iniciativas, bem como envolve distintas perspectivas, pressupostos e implicações”. Segundo Abdo (2014, p. 462), inclui ainda “a participação direta de não cientistas e não especialistas no fazer ciência”, no que se convencionou chamar de “ciência cidadã” ou ainda “ciência amadora, [...] *hacking e making*, ciência *open-source*, ciência distribuída”.

De acordo com o relatório *Science as an open enterprise*, produzido pela The Royal Society (2012), ciência aberta são “dados abertos (dados disponíveis, inteligíveis, avaliáveis e utilizáveis) combinados com acesso aberto a publicações científicas e comunicação eficaz de

seus conteúdos” (THE ROYAL SOCIETY, 2012, p. 16, tradução nossa). Tais definições espelham o foco contemporâneo na abertura da ciência, que evoluiu no contexto da internet pública e nas oportunidades de comunicação que ela proporcionou, incluindo a ampliação do empreendimento científico com o aparecimento de novas instituições em todo o mundo e outras motivações relacionadas, como o direito do contribuinte aos resultados de pesquisas públicas e a capacidade de qualquer membro da sociedade de examinar, avaliar, desafiar e reproduzir alegações científicas (THE ROYAL SOCIETY, 2012).

Esse cenário representa um avanço contínuo desde o início dos anos 2000, a partir da reunião fundamental Iniciativa de Acesso Aberto de Budapeste (BUDAPESTE, 2002) e das seguintes, em Berlim (BERLIN DECLARATION, 2003), e em Bethesda (BETHESDA STATEMENT..., 2003). Foram encontros que lançaram as bases do movimento pelo acesso aberto à informação científica (artigos e dados), culminando com movimentos substanciais de abertura da ciência, como o incentivo ao arquivamento público *on line* gratuito de artigos científicos, como o PubMed Central e o arXiv.org, o lançamento da Public Library of Science (PLOS), em 2003, que viria a se tornar um exemplo de revistas sob o modelo de acesso aberto “autor paga”, em oposição ao velho modelo de assinatura, e o dilúvio de dados moderno (THE ROYAL SOCIETY, 2012; NATIONAL ACADEMIES OF SCIENCE, 2018).

Em uma análise sobre o estado da ciência aberta, a National Academy of Science (2018) destaca algumas abordagens internacionais que refletem em alguma medida um movimento evolutivo: o extenso (e influente) relatório sobre questões relacionadas à ciência aberta elaborado pela The Royal Society (2012); a adoção da ciência aberta e da reutilização dos dados de pesquisa, em 2015, pelo Conselho Europeu e pelo Grupo dos Sete (G7); o endosso dos princípios FAIR pela Science Europe, com apoio do G20 na cúpula de Hangzhou, em 2016 (MONS et al., 2017); o compromisso do G7⁵, em 2017, de incentivar a realização de atividades científicas abertas e fornecer infraestruturas globais de pesquisa; a criação em 2013 da Research Data Alliance (RDA)⁶, organização global voltada para a comunidade, com o apoio europeu, norte-americano e australiano, para acelerar o compartilhamento de dados e a inovação orientada a dados; a ação do Comitê on data (CODATA)⁷, para melhorar a disponibilidade, usabilidade e interoperabilidade dos dados de

⁵ Ver em: <http://www.g7italy.it/sites/default/files/documents/ANNEX%204_WG%20Open%20Science/index.pdf>.

⁶ Ver em: <<https://www.rd-alliance.org/node/51727>>.

⁷ Ver em: <<https://codata.org/about-codata/our-mission/>>

pesquisa, incluindo a interoperabilidade legal; o lançamento da European Open Science Cloud (EOSC)⁸ em 2016, ambiente global onde pesquisadores, empresas e cidadãos podem publicar, encontrar e reutilizar dados e ferramentas uns dos outros com fins de pesquisa, educação e inovação, incluindo atividades do projeto Horizonte 2020.

Contudo, ao traçar um paralelo entre aspiração e realidade na ciência aberta, a principal sociedade científica do Reino Unido afirma no citado relatório (THE ROYAL SOCIETY, 2012) que o estágio de uma grande parte da prática científica de hoje fica aquém dos ideais de abertura. Apesar de reconhecer o poder da abertura dos dados de pesquisa, o estudo assevera que muita ciência é ininteligível quando vai além de sua própria disciplina especializada e os dados que fundamentam a pesquisa não são tornados acessíveis de maneira consistente para outros cientistas. Algo parecido se passa com a literatura científica primária aberta, incluindo acesso total e imediato a todos os trabalhos de pesquisa publicados. Segundo Swan, as políticas de acesso aberto conseguiram avanços bem-sucedidos nessa área, conteúdos em acesso aberto têm se acumulado em repositórios e sido publicados em periódicos de acesso aberto, mas esse acervo atinge uma parte acanhada do volume total da literatura publicada. Além disso, segundo SPARC, apesar de cientistas explorarem no dia a dia enormes volumes de dados e da capacidade de computação da era digital, sua abordagem ainda é associada à era do papel (THE ROYAL SOCIETY, 2012; SWAN, 2016; NATIONAL ACADEMIES OF SCIENCE, 2018; SPARC, 2019).

2.1 Privacidade, direito autoral e restrição a análise de conteúdo digital

Ademais, ainda persistem fatores em torno dos dados que limitam a maior desenvoltura da ciência aberta no meio científico, tais como a proteção da privacidade de indivíduos e o direito autoral. No tocante ao fator privacidade de indivíduos, dados coletados no decorrer de uma pesquisa médica em geral contêm informações pessoais, como registros médicos. Quando eles são abertos, os direitos dos titulares desses dados devem ser respeitados. Não significa, no entanto que não possam ser abertos, mas para isso é necessária a implementação de procedimentos de proteção para impedir ou aumentar a dificuldade de extrair informações incorporadas nos dados por meio da análise de dados, que revela o contexto no qual os dados estão inseridos e sua organização e estrutura (OCDE, 2015a). No

⁸ Ver em: <https://ec.europa.eu/research/openscience/pdf/realising_the_european_open_science_cloud_2016.pdf>.

Quadro 1, seguem exemplos de meios práticos para dificultar a análise de dados, o que pode também afetar a utilidade deles.

Quadro 1. Meios práticos para proteção de privacidade e informações

“Redução ou minimização de dados – É talvez o meio mais forte para impedir a extração de informações, porque nenhum dado é coletado, nenhuma informação pode ser extraída. Os titulares dos dados podem reter ou recusar o fornecimento de dados.

Criptografia – É um meio tecnológico que incorpora princípios, meios e métodos para a transformação de dados, a fim de ocultar seu conteúdo de informações, estabelecer sua autenticidade, impedir sua modificação não detectada, rejeição e/ou uso não autorizado.

Desidentificação - Abrange um conjunto de práticas que variam de anonimização a pseudonimização. Impedem a extração de atributos de identificação (ou seja, a re-identificação). A **anonimização** é um processo no qual as informações de identificação são excluídas ou mascaradas para que não possa ser, ou se torne muito dispendiosa para ser reconstruída. Algumas pesquisas sugerem que, quando vinculadas a outros dados, a maioria dos dados anonimizados pode ser desanonimizada; isto é, as informações de identificação podem ser reconstruídas. A **pseudonimização** é uma prática por meio da qual os atributos mais identificáveis (os identificadores) em um registro de dados são substituídos por identificadores artificiais exclusivos (ou seja, pseudônimos).

Desvinculação e separação funcional - Resulta de processos para garantir que os processadores de dados não possam distinguir se os itens de interesse estão relacionados ou não. A desidentificação é um meio de permitir a desvinculação, mas não garante a impossibilidade de conexão. Outros meios técnicos incluem separação e distribuição funcional (descentralização).

Adição de ruído e desinformação - Pode permitir que a análise baseada no conjunto de dados completo permaneça significativa enquanto oculta os atributos de dados confidenciais. Desinformação é informação falsa ou imprecisa, espalhada intencionalmente para enganar. As técnicas de adição de ruído ajudam a proteger a privacidade e a confidencialidade nos bancos de dados, mantendo todos os conjuntos de dados estatisticamente próximos dos originais.”

Na União Europeia, entrou em vigor, em maio de 2018, o Regulamento Geral de Proteção de Dados⁹. O Brasil editou, em agosto do mesmo ano, a Lei Geral de Proteção de Dados (BRASIL, 2018). Ambos abarcam, inclusive, o tratamento de dados pessoais em saúde pública.

O uso de conjuntos de dados com informações pessoais é vital para pesquisas nas Ciências Médicas (sobretudo em casos de emergência em saúde, como na pandemia da Covid-19) e Sociais. Tal uso apresenta desafios consideráveis para a governança das informações, devido ao potencial de comprometer a privacidade individual. Há risco sobretudo quanto aos dados pessoais sensíveis, relativos à condição racial, sexual, social ou política do indivíduo (BRASIL, 2018), que podem ser usados para explorar, estigmatizar, discriminar ou violar a autonomia pessoal de indivíduos (HRYNASZKIEWICZ; ALTMAN, 2009; OCDE, 2015a; TUCKER et al., 2016; KOBAYASHI; KANE; PATON, 2018; MALIN et al., 2018; DONEDA, 2019). Há ainda questões éticas que o pesquisador responsável por uma pesquisa envolvendo seres humanos deve observar, tais como “utilizar o material e os dados obtidos na pesquisa exclusivamente para a finalidade prevista no seu protocolo, ou conforme o consentimento do participante”, norma prevista na seção III, item q, da Resolução nº 466, de 12 de dezembro de 2012, do Conselho Nacional de Saúde¹⁰.

Quanto ao fator direito autoral, os avanços da tecnologia deram origem a tensões crescentes nos últimos anos. De acordo com o relatório da The Royal Society (2012), o direito autoral confere aos criadores de obras literárias, científicas e artísticas originais – incluindo artigos científicos e conjuntos dados de pesquisa – o direito de impedir que outras pessoas copiem a expressão de suas ideias em uma obra. Não protege as ideias, nem fatos. Boa parte dos periódicos científicos da era eletrônica ainda exigem transferência de direitos autorais em troca de avaliação por pares e distribuição. O prazo para o criador ou detentor do direito autoral exercê-lo é o seu tempo de vida mais 70 anos após a morte do autor na União Europeia e no Brasil (1998), e deve ser pelo menos de 50 anos após a morte do autor para todos os membros da Organização Mundial do Comércio (OMC). Entretanto, em vista do estágio em que as tecnologias de informação e comunicação se encontram, de acordo com o documento da The Royal Society, “o controle sobre a cópia em um ambiente digital é quase

⁹ Ver em: <<https://eur-lex.europa.eu/legal-content/PT/TXT/HTML/?uri=CELEX:32016R0679&from=PT>>.

¹⁰ Ver em: <<https://conselho.saude.gov.br/resolucoes/2012/Reso466.pdf>>

sem sentido” (THE ROYAL SOCIETY, 2012, p. 45) e pode apenas limitar a reutilização de resultados de pesquisa – porque impede análises adicionais dos dados, como, por exemplo, a mineração de texto e dados (ELLIOT, 2005).

O caso da mineração de texto e dados na ciência aberta, abrangendo pesquisa em grande escala de artigos e dados de pesquisa, tem sido exemplar da urgência de uma alteração nas legislações de proteção a direitos autorais (ou direitos de patentes ou de bancos de dados), pois

[...] limitam o poder das técnicas de análise de conteúdo digital, como mineração de texto e dados (para texto e dados) ou mineração de conteúdo (para análise computadorizada de conteúdo em todos os formatos), fatores que também criam desigualdades no acesso à descoberta de conhecimento na era digital (THE HAGUE DECLARATION, 2015, *on line*),

para enfrentamento de grandes desafios como mudanças climáticas, epidemias globais aceleração de desenvolvimento econômico e social.

A defesa do uso da mineração de texto e dados na pesquisa acadêmica e científica na Europa, que inclui a Declaração de Haia (THE HAGUE DECLARATION, 2015) sobre a descoberta do conhecimento na era digital, o incentivo ao uso de licenças abertas e ferramentas de dedicação ao domínio público por editores e agências de fomento à pesquisa (THE ROYAL SOCIETY, 2012; COMISSÃO EUROPEIA, 2016; BORGMAN, 2018; NATURE, 2019), levou a Comissão Europeia a incorporar uma exceção ao direito autoral expressa na Recomendação UE 2018/790, de abril de 2018. Ela trata da definição e aplicação de políticas claras para o acesso à informação científica e a sua preservação, desenvolvimento de infraestruturas para a ciência aberta, gestão de dados de pesquisa, acesso aberto a publicações científicas, incentivos e recompensas etc. (COMISSÃO EUROPEIA, 2018).

De acordo com o documento *Open Innovation, Open Science, Open to the World – a vision for Europe*, “A ‘Mineração de Texto e Dados’ (Text and Data Mining - TDM) é um conjunto de métodos que usam computadores para pesquisar, filtrar e interpretar automaticamente grandes quantidades de conteúdo digital e *on line*” (COMISSÃO EUROPEIA, 2016, p. 51, tradução nossa). A Comissão Europeia estima que os desenvolvimentos tecnológicos que sustentam a TDM criarão, até 2030,







[...] vastas quantidades de informações novas e dados digitais heterogêneos, decorrentes de iniciativas como políticas de acesso a dados gratuitos e abertos, expansão contínua das mídias sociais, *feeds* de dados sensoriais em tempo real à medida que a Internet das coisas evolui, bem como resultado de novas infraestruturas e plataformas sociais que permitem especialistas e cidadãos produzirem grandes volumes de dados (COMISSÃO EUROPEIA, 2016, p. 51, tradução nossa).

Duas disposições contidas na seção “Acesso aberto às publicações científicas”, item 1, da citada Recomendação UE 2018/790, lidas conjugadamente, revelam uma facilitação para a área da pesquisa acadêmica e, por extensão, à ciência aberta quanto ao uso de ferramenta para mineração de texto e dados. A primeira disposição estabelece que, em virtude da evolução da tecnologia, as licenças não devem limitar “a pesquisa [mineração] de textos e dados de publicações resultantes de investigação financiada por fundos públicos”, sem prejuízo da lei aplicável de direitos autorais. A segunda refere-se à relação editores científicos-pesquisadores. Quando celebrarem contratos, os pesquisadores devem conservar os direitos de propriedade intelectual (para fins deste estudo, o direito autoral) “necessários [...] para cumprir os requisitos da política em matéria de acesso aberto”, no que se refere especialmente ao autoarquivamento e à reutilização, notadamente através de mineração de dados (COMISSÃO EUROPEIA, 2018, *on line*).

Há questões que causam tensão e ainda precisam ser enfrentadas no que concerne aos limites da abertura na ciência. Esses limites coincidem com a linha divisória entre a ciência financiada pelo setor público e pelo setor privado? Até onde devem ir os interesses comerciais na exploração de dados, informações e conhecimentos científicos? A abertura pode ser economicamente benéfica ou socialmente desejável em alguns setores? Como a privacidade e a confidencialidade podem ser mantidas adequadamente? Em que medida os direitos autorais devem (ou não) intervir na paisagem da ciência aberta e regular conflitos de propriedade de dados de pesquisa? (THE ROYAL SOCIETY, 2012)

O Quadro 2, que reproduz o guia “How Open is it?”, resultado de uma colaboração entre SPARC, PLOS e OASPA, tem o objetivo de auxiliar na compreensão dos componentes de uma política editorial que irão definir o quanto um periódico é mais ou menos de acesso aberto. Esses componentes recebem uma gradação que vai do “acesso aberto” ao “acesso fechado”, ajudando na tomada de decisão sobre onde publicar.

Quadro 2. A gradação do aberto

Acesso	Direitos dos leitores	Direitos de reutilização	Direitos autorais	Direitos de depósito dos autores	Depósito automático	Acessibilidade por máquina	Acesso
 OPEN ACCESS 	Acesso Aberto a todos os artigos imediatamente após a publicação	Direitos generosos de reutilização e adaptação (p. ex., licença CC BY)	O autor retém os direitos autorais sem restrições	O autor pode depositar qualquer versão num repositório ou <i>site</i> na <i>web</i>	As revistas disponibilizam cópias dos artigos de forma automática em repositórios confiáveis de terceiros (p. ex. PubMed Central) imediatamente após a publicação	O texto integral do artigo, metadados, citações e dados, incluindo dados suplementares, são fornecidos em formatos padrão legíveis por máquinas através de uma API* ou protocolo	 OPEN ACCESS 
	Acesso Aberto a todos os artigos depois de um período de embargo não superior a 6 meses	Reutilização e adaptação sujeitas a algumas restrições ou condições (p. ex. licenças CC BY-NC e CC BY-SA)	O autor retém os direitos autorais da versão publicada com algumas restrições de utilização da versão publicada	O autor pode depositar a versão final com revisão por pares <i>postprint</i> em qualquer repositório ou <i>site</i> na <i>web</i>	As revistas disponibilizam cópias dos artigos de uma forma automática em repositórios confiáveis de terceiros (p. ex. PubMed Central) no prazo de 6 meses	O texto integral do artigo, metadados, citações e outros dados podem ser extraídos através de uma API ou protocolo padrão	
	Acesso Aberto a todos os artigos depois de um período de embargo superior a 6 meses	Reutilização (sem permissão de adaptações) sob algumas restrições ou condições (p. ex. licenças CC BY-ND)	O editor retém os direitos autorais da versão publicada com algumas permissões de utilização pelos autores e leitores	O autor pode depositar a versão final com revisão por pares <i>postprint</i> em determinados repositórios ou <i>site</i> na <i>web</i>	As revistas disponibilizam cópias dos artigos de uma forma automática em repositórios confiáveis de terceiros (p. ex. PubMed Central) no prazo de 12 meses	O texto integral do artigo, metadados e citações podem ser extraídos ou acessíveis sem uma permissão especial ou registro	
	Acesso Aberto e imediato a alguns, mas não a todos os artigos (incluindo modelos híbridos)	_____	O editor retém os direitos autorais da versão publicada com algumas permissões de utilização pelos autores	O autor pode depositar a versão submetida/ou o rascunho da versão final <i>preprint</i> em determinados repositórios ou <i>sites</i> na <i>web</i>	_____	O texto integral do artigo, metadados e citações podem ser extraídos ou acessíveis com permissão	
	Assinatura, adesão, pagamento por artigo ou outras taxas para acessar os artigos	Sem direitos de reutilização além do uso justo ou limitações e exceções de direitos autorais (todos os direitos reservados)	O editor retém os direitos autorais da versão publicada sem permissão de utilização pelo autor para além do uso justo	O autor não pode depositar nenhuma versão em repositórios ou <i>sites</i> na <i>web</i>	Não é permitido o depósito automático em repositórios	O texto integral do artigo e metadados não estão disponíveis em formato legível por máquinas	
CLOSED ACCESS 						CLOSED ACCESS 	

Fonte: SPARC, PLOS e OASPA. 2014. Disponível em: <https://sparcopen.org/wp-content/uploads/2016/01/PLOS-OAS_HOII-Portuguese0603-FINAL.pdf>. Nota: *API - é o acrônimo de Application Programming Interface ou, em português, Interface de Programação de Aplicativos <<https://blog.cedrotech.com/o-que-e-api-significado/>>.

2.2 O acesso aberto à informação científica se amplia e agrega os dados de pesquisa

É comum considerar a conferência Iniciativa de Acesso Aberto de Budapeste (Open Access Initiative - BOAI), realizada pelo Open Society Institute, em 2001, naquela cidade, como o marco do movimento pelo acesso aberto à literatura científica (BUDAPESTE, 2002). Não há dúvida de que tem importância central. Foi lá que se usou o termo “acesso aberto” pela primeira vez e que os participantes traçaram as estratégias originais de disseminação do movimento: a via Verde, que se refere “ao autoarquivamento pelo autor, no qual os artigos revisados por pares e *preprints* revisados por pares ou não são publicados *on line* em um repositório institucional e/ou temático ou em um *site* pessoal” e a via Dourada, em que “artigos de pesquisa deviam estar acessíveis gratuitamente no momento da publicação” (TENNANT et al. 2016, *on line*) em versões digitais de periódicos (BUDAPESTE, 2002; GUÉDON, 2004; HARNAD et al., 2004). Além disso, a BOAI articulou a definição de acesso aberto:

Por ‘acesso aberto’ a esta literatura [artigos revisados por pares ou *preprints*], queremos dizer sua disponibilidade gratuita na internet pública, permitindo que qualquer usuário leia, faça *download*, copie, distribua, imprima, pesquise ou crie *links* para os textos completos desses artigos, rastreie-os para indexação, transmita-os como dados para software ou use-os para qualquer outra finalidade legal, sem barreiras financeiras, legais ou técnicas que não sejam aquelas inseparáveis de obter acesso à própria internet. A única restrição à reprodução e distribuição e o único papel dos direitos autorais nesse domínio devem ser o de dar aos autores controle sobre a integridade de seu trabalho e o direito de serem devidamente reconhecidos e citados (BUDAPESTE, 2002, *on line*).

Também as já citadas Declaração de Bethesda¹¹ sobre Publicação em Acesso Aberto (junho/2003) e a Declaração de Berlim¹² sobre Acesso Aberto ao Conhecimento em Ciências e Humanidades (outubro/2003) foram definidoras dos conceitos desse movimento.

¹¹ Ver em <https://dash.harvard.edu/bitstream/handle/1/4725199/suber_bethesda.htm?sequence=1> a íntegra da Declaração de Bethesda.

¹² Ver em <<https://openaccess.mpg.de/>> a íntegra da Declaração de Berlim.

Um evento menos mencionado, revelador das origens do acesso aberto, é a Convenção de Santa Fé, realizada em Santa Fé, nos EUA, em outubro de 1999, que levou ao estabelecimento da Iniciativa de Arquivos Abertos (Open Archives Initiative - OAI), que por sua vez desenvolveu o Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH), um mecanismo para a interoperabilidade de repositórios, utilizado para coleta e distribuição de metadados (VAN DE SOMPEL; LAGOZE, 2000; LAGOZE; VAN DE SOMPEL, 2015).

Há, contudo, um consenso de que o movimento de acesso aberto à literatura científica, desenvolvido na internet, foi “uma reação ao domínio das editoras no processo editorial [...] e a consequências como o alto custo das assinaturas” de periódicos adquiridos por bibliotecas, conforme afirma Pinheiro (2014, p. 156), no que corrobora Guédon (2004, p. 315), ao assegurar que, “Historicamente, o Acesso Aberto surgiu em grande parte como uma reação ao rápido aumento dos preços de periódicos acadêmicos e científicos. [...] [e] os altos preços dos periódicos obviamente limitavam o acesso por meios econômicos”.

Já para a Coalizão de Publicações Acadêmicas e Recursos Acadêmicos (The Scholarly Publishing and Academic Resources Coalition - SPARC), instituição que incentiva o compartilhamento aberto de resultados de pesquisa e materiais educacionais, o sistema atual de comunicação de pesquisa, apesar dos avanços obtidos, ainda

[...] é prejudicado por um modelo secular que não foi atualizado para aproveitar a tecnologia do século XXI: [...] Nosso sistema atual de comunicação de pesquisa utiliza um modelo baseado em impressão na era digital. [...] Embora a pesquisa seja amplamente produzida com dinheiro público por pesquisadores que a compartilham livremente, os resultados estão ocultos por barreiras técnicas, legais e financeiras. Essas barreiras artificiais são mantidas por editoras e restringem o acesso a uma pequena fração de usuários, bloqueando a maior parte da população mundial e impedindo o uso de novas técnicas de pesquisa (SPARC, 2019, *on line*, tradução nossa).

Swan (2016, p. 24) confirma esse entendimento ao afirmar que “o acesso ainda é uma grande preocupação para pesquisadores”. SPARC prossegue e assevera que

[...] essa incompatibilidade fundamental entre o que é possível com a tecnologia digital – um sistema aberto para a comunicação de resultados de pesquisas no qual qualquer pessoa, em qualquer lugar, pode contribuir – e nosso sistema de publicação desatualizado levou ao movimento de acesso aberto (SPARC, 2019, *on line*, tradução nossa).

Entretanto, há quem assegure que o movimento pelo acesso aberto já dera um grande salto em 1991, com o lançamento do arXiv, repositório fundado por Paul Ginsparg, originalmente conhecido por seu endereço xxx.lanl.gov, anterior à World Wide Web (BORGMAN, 2015). O arXiv mudou-se dos Laboratórios Nacionais de Los Alamos para a Universidade de Cornell, onde está hospedado até hoje. Em 2019, o repositório recebeu 155.866 novos envios, contabilizando um aumento de 11% em relação a 2018, e cerca de 260 milhões de downloads de todo o mundo (ARXIV, 2020).

Segundo Borgman (2015, p. 39, tradução nossa), algumas lições do lançamento do arXiv são importantes para uma abordagem atual do acesso aberto a dados de pesquisa:

- [...] o sistema era consequência da cultura de troca ativa de *preprints* na física de alta energia;
- [...] interrompeu uma infraestrutura de conhecimento existente que suportava o fluxo de informações nas redes de colegas próximos, alterando as relações entre autores, editores, bibliotecas e leitores como partes interessadas na comunicação acadêmica da física, de modo que pesquisadores e estudantes, em países ricos e pobres, obtiveram acesso a artigos muito antes da data oficial de publicação;
- e, [...] apesar do sucesso, não foi transferido rapidamente a outros campos.

O acesso aberto refere-se, portanto, à remoção dos principais obstáculos ao acesso, compartilhamento e reutilização dos resultados de pesquisas acadêmicas. A lógica é facilitar o processo de pesquisa e garantir um acesso rápido e amplo aos resultados, de modo que todas as comunidades tenham a oportunidade de reusá-los, além de suprir o *gap* da divisão global da produção de pesquisa Norte-Sul. Para tanto, atualmente existem mais de 700 políticas e

mandatos de acesso aberto em todo o mundo registrados por institutos de pesquisa e órgãos de financiamento, de acordo com o Registry of Open Access Repository Mandates and Policies (ROARMAP)¹³ (LAAKSO; BJÖRK, 2012; TENNANT et al., 2016).

Como já dito preliminarmente, o movimento de acesso aberto dissemina a literatura científica por meio de duas estratégias: as vias Dourada e Verde. A via Dourada, que se caracteriza por periódicos em acesso aberto, registra pelo menos três modelos de negócio: no primeiro, o periódico oferece acesso aberto aos artigos e cobra taxas de processamento de artigo ao autor, não havendo cobrança para o leitor ou a biblioteca; no segundo, oferece o chamado modelo “híbrido”, que combina o modelo clássico de assinaturas com o de pagamento de taxas de processamento de artigos para autores que queiram ter seu artigo em acesso livre; e o terceiro é a chamada “via Diamante”, em que o periódico oferece acesso livre aos artigos sem cobrar taxas de processamento do autor, do leitor ou de bibliotecas. Os periódicos são mantidos por voluntários ou associações, sociedades ou organizações sem fins lucrativos (ACESSO ABERTO USP, 2020).

A via Verde, estratégia de autoarquivamento do autor de artigos revisados por pares ou *preprints* com ou sem revisão por pares em um repositório institucional, temático e/ou em um *site* pessoal, passou a depender, geralmente, de políticas de periódicos ou de editores sobre autoarquivamento. Há editores que exigem um período de embargo antes que o depósito em repositórios públicos seja permitido. Os embargos são aplicados para evitar reduções de receita do modelo de assinaturas devido a esse autoarquivamento, embora a prática dos editores seja discutível. Essa via também é ativada pela retenção de direitos de autor, não exclusivos, que os concede em decorrência de contratos de trabalho com instituições empregadoras antes mesmo de publicar qualquer trabalho (HARNAD, 2013; WEITZEL, 2014; TENNANT et al., 2016; BJÖRK; KORKEAMÄK, 2020). Por essas duas vias, “quase 25% de todos os documentos acadêmicos arquivados na *web* agora são obtidos por acesso aberto em algum lugar da internet”, afirma Tennant et al. (2016, *on line*, tradução nossa).

Contudo, uma questão central permanece: o acesso universal a aproximadamente 75% dos artigos não é diretamente possível. Ou só é permitido àqueles que, em uma posição privilegiada, em uma instituição de pesquisa ou universidade, tenha acesso por assinatura a uma parcela desses artigos ou possua recursos suficientes para pagar por artigo em uma base. As assinaturas de todos os periódicos revisados por pares não são acessíveis para nenhum

¹³ Ver em: <<http://roarmap.eprints.org/>>.

indivíduo, instituto de pesquisa ou universidade (ODLYZKO, 2006; SUBER, 2012; TENNANT et al., 2016). Desta forma, segundo Tennant et al. (2016), o impacto potencial dos artigos científicos nunca é totalmente realizado, o que impede o progresso científico pela falta de utilização e de modo negativo tanto o reconhecimento de pesquisadores individuais quanto de financiadores de pesquisa.

Ocorre algo parecido com os dados de pesquisa, que, apesar de várias iniciativas para torná-los acessíveis e reutilizáveis, entre elas os Panton Principles (MURRAY-RUST et al., 2010) e os princípios FAIR (WILKINSON et al., 2016), nem sempre estão prontamente disponíveis, pois o compartilhamento é comum apenas em alguns campos. Em 2012, Borgman anunciava que “O dilúvio de dados chegou”. No artigo “The conundrum of sharing research data” (O enigma do compartilhamento de dados de pesquisa), confirmava a intensificação do movimento pelo compartilhamento de dados científicos. Mas por que é urgente compartilhar dados?, perguntava-se, admirada com a repercussão do tema nas mídias tradicional e acadêmica, afirmando que compartilhamento de dados não é um tópico novo de discussão em círculos acadêmicos e políticos. E citava relatórios sobre como melhorar o compartilhamento e a curadoria de dados elaborados a partir da década de 1980 até o início do século XXI (BORGMAN, 2012).

Um exemplo da sinergia dos movimentos de acesso às publicações científicas e dados de pesquisa foi o lançamento do documento Guidelines on Implementation of Open Access to Scientific Publications and Research Data (COMISSÃO EUROPEIA, 2017). Esse documento encoraja beneficiários de financiamento do programa europeu de pesquisa Horizonte 2020 (COMISSÃO EUROPEIA, 2014) a garantirem acesso aberto e gratuito para qualquer usuário a todas as publicações científicas revisadas por pares relacionadas a seus resultados e, da mesma forma, facilitar o acesso, reutilização e preservação dos dados gerados durante o trabalho de pesquisa (COMISSÃO EUROPEIA, 2017).

As agências de fomento à pesquisa vêm se constituindo em grandes propulsoras da abertura e do compartilhamento de dados, pois ao concederem financiamento exigem a liberação imediata de dados de pesquisa subjacentes, tão logo o artigo seja publicado, em repositórios de acesso aberto. Também os periódicos participam dessa cruzada e muitos também exigem o depósito de dados subjacentes a artigos em repositórios, adaptando-se às exigências das instituições financiadoras, conforme se observa em pesquisa desenvolvida e apresentada no capítulo 4 desta tese.

2.3 O acesso aberto a artigos e dados de pesquisa ganha apoio mundial

Desde meados da década de 2000, um número cada vez maior de governos, organismos internacionais, universidades, instituições de pesquisa e editores tem apoiado o movimento de abertura de artigos e dados de pesquisa e adotado políticas de acesso aberto.

Nos EUA, além do apoio ao movimento de universidades, tais como a de Harvard, a da Califórnia, o Instituto de Tecnologia de Massachusetts, o Instituto de Tecnologia da Califórnia, em 2013, o governo Obama determinou que artigos e dados de pesquisas financiadas pelo governo fossem disponibilizados gratuitamente ao público dentro de um ano após a publicação (BORGMAN, 2015; OFFICE OF SCIENCE..., 2013).

Em 2012, o órgão Conselhos de Pesquisa do Reino Unido (Research Councils of United Kindom - RCUK) anunciou que todos os artigos de periódicos e de conferências, revisados por pares, resultantes total ou parcialmente de seu financiamento, deveriam ser submetidos a periódicos de acesso aberto, a partir de abril de 2013 (BORGMAN, 2015). Em 2017, a Comissão Europeia estendeu o projeto piloto criado para incentivar o acesso aberto a artigos científicos, originados de pesquisas financiadas pela Comissão, ao acesso aberto a dados de pesquisa. A partir de então, o acesso aberto se tornava padrão para dados de pesquisa gerados no programa Horizon 2020, assim como os artigos por ele financiados. Ambos os resultados deverão ser publicados em acesso aberto, sem embargo, a partir de 2021 (COMISSÃO EUROPEIA, 2014; 2017; COAlition S, 2018; 2019).

Na África, em 1998, a Rede Internacional para o Avanço da Ciência e da Política (International Network for the Availability of Scientific Publication (INASP), organização sem fins lucrativos sediada em Oxford, na Inglaterra, lança os Jornais Africanos OnLine¹⁴ (African Journals OnLine - AJOL), plataforma de periódicos acadêmicos africanos publicados naquele continente. A AJOL, baseada na África do Sul, é uma organização sem fins lucrativos que hospeda 526 revistas, com 264 em acesso aberto.

Na América Latina, excluído o Brasil, que trataremos à parte, destacam-se vários documentos de apoio e iniciativas, entre eles: a Redalyc, base de dados bibliográficos e biblioteca digital de revistas de acesso aberto, patrocinada pela Universidade Autônoma do

¹⁴ Ver em: <<https://www.ajol.info/index.php/ajol>>

Estado do México; a Declaração de Cuba a favor do acesso aberto, em 2007; a criação pelo governo argentino do Sistema Nacional de Repositório Digital (SNRD), por meio de uma resolução ministerial, em 2011; a criação da Red Mexicana de Repositórios Institucionais (ReMeRI), em 2011; o lançamento em 2012 da Federação Latino-Americana de Repositórios Institucionais de Documentação Científica na América Latina (LA Referencia), com o apoio da RedCLARA, o patrocínio do Fundo Regional de Bens Públicos do Banco Interamericano de Desenvolvimento (BID) e o objetivo de promover a criação de uma estratégia regional para repositórios institucionais; o Repositório Nacional Digital de Ciência, Tecnologia e Inovação em Acesso Aberto, no Peru, em 2013, caracterizado como um *'site central'* onde são armazenadas informações digitais relacionadas à produção de ciência, tecnologia e inovação no país (COSTA; LEITE, 2016); a Declaração do Panamá sobre Ciência Aberta¹⁵, que estabeleceu 11 estratégias de políticas públicas para implementação de ciência aberta, em 2018; e a criação da AmeliCA¹⁶, infraestrutura de comunicação para publicação acadêmica e ciência aberta, sob a forma de cooperativa sustentada, focada no modelo de publicação sem fins lucrativos, em 2019.

No Brasil, não há uma política de governo que oriente o acesso aberto a resultados de pesquisa. Um projeto de lei de 2011¹⁷, que dispunha sobre o processo de registro e disseminação da produção técnico-científica pelas instituições de educação superior, bem como pelas unidades de pesquisa no Brasil, foi arquivado em 2018. Mesmo assim, universidades, instituições de pesquisa e agências de apoio à pesquisa avançaram em apoio ao movimento. Dentre outras iniciativas, destacamos: a SciELO¹⁸, biblioteca eletrônica de coleção selecionada de periódicos científicos brasileiros, apoiada pela Fundação de Amparo à Pesquisa do Estado de São Paulo (Fapesp), em parceria com o Centro Latino-Americano e do Caribe de Informação em Ciências da Saúde (Bireme) e o Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), lançada em 1997; o Manifesto Brasileiro de Apoio ao Acesso Aberto à Informação Científica (2005) e o Manifesto de Acesso Aberto a Dados da Pesquisa Brasileira para Ciência Cidadã (2016), ambos do IBICT; a Declaração de Salvador, sobre a importância do acesso aberto para países em desenvolvimento, a Carta de

¹⁵ Ver em: <<https://gitlab.com/declaraciondepanama/declaracion-de-panama/commit/2a6cb1120832bd5ce255242227e47da529912717>>

¹⁶ Ver em: <<http://amelica.org/>>

¹⁷ Ver em: <<https://www25.senado.leg.br/web/atividade/materias/-/materia/101006>>

¹⁸ Ver em: <<https://www.scielo.br/?lng=pt>>

São Paulo, em 2005, e a Declaração de Florianópolis, em 2006, para expressar seu apoio ao acesso aberto à literatura revisada por pares; a Universidade de São Paulo (USP)¹⁹, com o seu Repositório de Reprodução da USP, que é a Biblioteca Digital da Produção Intelectual da Universidade de São Paulo (BDPI), inaugurada em 22 de outubro de 2012, e o Portal de Revistas, em 2008; a Fundação Oswaldo Cruz²⁰, com sete revistas científicas em acesso aberto e o Repositório Arca, lançado como repositório institucional em 2011, que abriga diversificada tipologia de documentos, além de artigos, teses e dissertações, estes de forma mandatória, e o Portal de Periódicos Fiocruz, lançado em 2015, ambos em consequência direta da Política de Acesso Aberto ao Conhecimento²¹, de 2015; o Compromisso n. 3 do 4º Plano de Ação Nacional em Governo Aberto, lançado em outubro de 2018, para governança de dados científicos e o avanço da Ciência Aberta no Brasil, com a participação do IBICT, Fiocruz e Embrapa; e de novo o IBICT, com acesso aberto a periódicos científicos e à Biblioteca Digital Brasileira de Teses e Dissertações (BDTD)²², lançada em 2002, que integra sistemas de informação de teses e dissertações de instituições de ensino e pesquisa brasileiras. A BDTD conta com um acervo de mais de 126 mil teses e dissertações de mais de 100 instituições de ensino brasileiras.

2.4 Dados de pesquisa: (in)definições

Dados não fluem como petróleo, grudam como cola, ou criam fogo por fricção como os fósforos. O seu valor reside no seu uso, que motivou o relatório Bits do Poder²³. A questão não declarada é: “o que são dados?”. O único acordo sobre definições é que nenhuma definição será suficiente [...] (BORGMAN, 2015, p. 3-4, tradução nossa)

¹⁹ Ver em: <<https://repositorio.usp.br/>> e em <<https://www.revistas.usp.br/wp/>>

²⁰ Ver em: <<https://www.arca.fiocruz.br/terms/sobre.jsp>> e em <<https://periodicos.fiocruz.br/pt-br/about>>

²¹ Ver em: <https://portal.fiocruz.br/sites/portal.fiocruz.br/files/documentos/portaria_-_politica_de_acesso_aberto_ao_conhecimento_na_fiocruz.pdf>

²² Ver em: <<http://bdtd.ibict.br/vufind/>>.

²³ Ver em: <<https://www.nap.edu/catalog/5504/bits-of-power-issues-in-global-access-to-scientific-data>>

Robert K. Merton (1942), um dos principais precursores da sociologia da ciência, já defendia na década de 1940 que os resultados de pesquisa deveriam ser livremente acessíveis a todos e que os pesquisadores abrissem mão dos direitos autorais de seus achados. Contudo, o conceito de acesso aberto a dados científicos disponíveis em formato digital é atribuído à formação do sistema World Data Center, criado para arquivar e distribuir dados coletados dos programas observacionais do Ano Geofísico Internacional de 1957-1958²⁴.

Christine L. Borgman, professora catedrática em Estudos da Informação na Universidade da Califórnia (UCLA), é autora de artigos e livros sobre dados de pesquisa. Sem ser contrária à abertura de dados de pesquisa, a pesquisadora prima por um posicionamento crítico sobre esse movimento e seus desdobramentos (WALLIS; BORGMAN, 2012). No já mencionado artigo “The conundrum of sharing research data” (O enigma do compartilhamento de dados de pesquisa), de sua autoria, a autora explora o caráter complexo do compartilhamento de dados e critica a sua promoção exagerada, pois há “espessas camadas de complexidade sobre a natureza dos dados, pesquisa, inovação e academia, incentivos e recompensas, economia e propriedade intelectual e políticas públicas” a serem superadas. Portanto, diz ela, compartilhar dados de pesquisa é complexo e difícil, ou um “enigma”, como afirma no título (BORGMAN, 2012, *on line*).

Em relação à definição de dados, a pesquisadora da UCLA revela que o termo “dados” ainda não adquiriu uma definição de consenso. Após analisar, em seu livro *Big data, little data, no data* (BORGMAN, 2015), diversas definições de dados construídas na academia, no Conselho Nacional de Ciência dos EUA e na Nasa, ela afirma que “O resumo mais abrangente é dizer que os dados são representações de observações, objetos ou outras entidades usadas como evidência de fenômenos para fins de pesquisa ou estudos acadêmicos” (p. 28). Assinala que o termo “entidade” “pode ter uma existência material, como textos em papel ou papiros, ou podem ser digitais, como sinais de sensores ou formulários preenchidos de uma pesquisa *on line*” (BORGMAN, 2015, p. 28). Para a autora,

[...] a incapacidade de ancorar o conceito de maneira a esclarecer o que são e o que não são dados em uma determinada situação contribui muito para a confusão sobre assuntos como planos de gerenciamento de dados, políticas de dados abertos e curadoria de dados (BORGMAN, 2015, p. 28, tradução nossa).

²⁴ Ver em: <<https://access.okfn.org/2015/04/30/open-access-to-research-data-timeline/>>

Instituições políticas, organismos internacionais e atores da comunicação científica, tais como periódicos científicos e agências de financiamento de pesquisa, a partir de meados da década de 2000, passaram a apresentar documentos sobre regras para disponibilização de dados de pesquisa contendo versões da definição de dados de pesquisa. Nessa época, intensificavam-se os esforços para dispor em acesso aberto resultados de pesquisa financiada com recursos públicos ou doados.

Um desses documentos é o “Principles and Guidelines for Access to Research Data from Public Funding” (Princípios e diretrizes para acesso a dados de pesquisa de financiamento público) (OCDE, 2007). Publicado pela OCDE, destina-se a orientar como disponibilizar dados de pesquisa coletados mediante uso de recursos públicos, preferencialmente em formato digital e legível por máquina. Estabelece que dados de pesquisa são

[...] registros factuais (pontuações numéricas, registros textuais, imagens e sons) usados como fontes primárias de pesquisa científica e que são comumente aceitos na comunidade científica como necessários para validar resultados da pesquisa (OCDE, 2007, p. 13, tradução nossa).

Segundo esse mesmo documento, “Um conjunto de dados de pesquisa constitui uma representação parcial e sistemática do tema que está sendo pesquisado” (OCDE, 2007, p. 13). Entretanto, o termo “dados de pesquisa”, no entendimento da OCDE, não abrange cadernos de laboratório, análises preliminares e rascunhos de artigos científicos, planos para pesquisas futuras, avaliação por pares ou comunicações pessoais com colegas ou objetos físicos (por exemplo, amostras de laboratório, cepas de bactérias e animais de teste como ratos).

A Comissão Europeia (2017, *on line*) também não se absteve de trazer a sua contribuição. O documento “Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020” (Diretrizes para as Regras de Acesso Aberto a Publicações Científicas e Acesso Aberto a Dados de Pesquisa no Horizonte 2020), publicado pela Comissão Europeia (2017, *on line*), reúne regras de um projeto piloto sobre acesso aberto a publicações científicas revisadas por pares e a dados de pesquisa. Dirigido a beneficiários de recursos para pesquisas financiadas ou cofinanciadas no âmbito do projeto europeu Horizonte 2020 (COMISSÃO EUROPEIA, 2014), enuncia que dados de pesquisa referem-se

a informações, em particular fatos ou números, coletadas para serem examinadas e consideradas como base para raciocínio, discussão ou cálculo. Em um contexto de pesquisa, exemplos de dados incluem estatísticas, resultados de experimentos, medições, observações resultantes de trabalho de campo, resultados de pesquisas, gravações de entrevistas e imagens. O foco está nos dados de pesquisa disponíveis em formato digital. Os usuários normalmente podem acessar, extrair, explorar, reproduzir e disseminar dados de pesquisa de acesso aberto gratuitamente (COMISSÃO EUROPEIA, 2017, p. 4).

Neste mesmo documento, ao mencionar os tipos de dados abrangidos pelas regras publicadas, a Comissão Europeia também define dados subjacentes, um dos focos desse trabalho, como “os dados necessários para validar os resultados apresentados em publicações científicas [...]” (COMISSÃO EUROPEIA, 2017, p. 9, tradução nossa).

Outra definição de dados abertos que merece atenção é a fornecida pelo projeto Open Definition, da Open Knowledge Foundation. Baseada na definição de *software* livre, a “definição aberta” estabelece inicialmente princípios definidores de “abertura”, buscando tornar preciso o significado de “aberto” nos termos “dados abertos” e “conteúdo aberto”. Resumidamente, portanto, segundo a OFK (2020, *on line*), “Aberto significa que qualquer pessoa pode acessar, usar, modificar e compartilhar livremente para qualquer finalidade (sujeita, no máximo, a requisitos que preservem a proveniência e a abertura)”. Partindo desta definição, a OKF (2020, *on line*) ainda informa que “Dados e conteúdo abertos podem ser usados, modificados e compartilhados livremente por qualquer pessoa para qualquer fim.”

No entanto, Jorge e Albagli (2017) sublinham diferenças de significados entre abertura e compartilhamento de dados. Concordam com a definição da OFK para abertura, mas entendem que o compartilhamento pauta-se por acordos de acesso, que estabelecem as condições de uso, o que pode restringir o compartilhamento, o que o distingue do espírito irrestrito dos dados abertos.

Editores que integram o ecossistema da publicação científica também cuidaram de definir dados de pesquisa. A Springer Nature, ao valorizar o reuso de dados, declara que “Os dados de pesquisa²⁵ se referem à coleção de arquivos que dão suporte ao seu projeto, estudo ou publicação de pesquisa, como planilhas, documentos, imagens, vídeos ou áudio”. A PLoS, em sua página de diretrizes sobre como disponibilizar dados de pesquisa, defende que o pesquisador deve compartilhar um “conjunto mínimo de dados”, definido como “dados

²⁵ Ver em: <<https://www.springernature.com/gp/authors/research-data>>

necessários para replicar todas as descobertas do estudo relatadas no artigo, bem como metadados e métodos relacionados” (PLOS MEDICINE, 2014, *on line*).

Por seu turno, a Fundação Bill & Melinda Gates (2017b, *on line*, tradução nossa), agência de financiamento à pesquisa estadunidense, define dados de pesquisa subjacentes como aqueles que “englobam todos os dados primários, metadados associados e quaisquer dados adicionais relevantes necessários para entender, avaliar e replicar os resultados do estudo relatados na totalidade”. Afirmar ainda que os dados subjacentes “podem ser compilados em qualquer tipo de arquivo, incluindo as instruções de acesso necessárias, código ou arquivos de informações de suporte, para garantir que os arquivos possam ser acessados e utilizados por terceiros” (BILL & MELINDA GATES FOUNDATION, 2017b, *on line*), evidenciando uma preocupação em fornecer amplo acesso aos dados de pesquisa.

Na área da saúde, o National Institutes of Health (NIH), principal agência de financiamento de pesquisa em saúde dos EUA (ver análise sobre o NIH no capítulo 4), utiliza a terminologia “Dados finais de pesquisa” ao divulgar a sua definição. Um dos maiores financiadores de pesquisa médica e comportamental daquele país, fundado em 1887 como Laboratório de Higiene, com dotação orçamentária para o ano fiscal de 2020 de US\$ 41,68 bilhões²⁶ para financiamento de suas atividades, o NIH define assim dados de pesquisa²⁷:

Dados finais de pesquisa são o material factual registrado comumente aceito na comunidade científica como necessário para documentar e fundamentar os resultados da pesquisa. Isso não significa estatísticas ou tabelas resumidas; ao contrário, significa os dados nos quais as estatísticas e tabelas resumidas são baseadas. Para os fins desta política, os dados finais da pesquisa não incluem cadernos de laboratório, conjuntos de dados parciais, análises preliminares, rascunhos de artigos científicos, planos para pesquisas futuras, relatórios de revisão por pares, comunicações com colegas ou objetos físicos, tais como géis ou amostras de laboratório (NIH, 2003b, *on line*, tradução nossa).

Sayão e Sales (2014) recuperam um elemento importante para a definição e compreensão do que sejam dados de pesquisa, ao afirmarem que a noção de dados varia entre pesquisadores e entre áreas do conhecimento. Tal constatação levou-os ao raciocínio de que dados são gerados para diferentes propósitos, por diferentes comunidades acadêmicas e

²⁶ Ver em: <<https://www.niaid.nih.gov/grants-contracts/budget-appropriation-fiscal-year-2020>> e no documento do Senado americano em <https://www.appropriations.senate.gov/imo/media/doc/H1865PLT_44.PDF>

²⁷ Ver em: <https://grants.nih.gov/grants/policy/data_sharing/data_sharing_guidance.htm#data>

científicas, e ainda por distintos processos, o que acarreta a intensificação da percepção da diversidade existente.

Em 2005, o National Science Board (NSB), conselho que estabelece as políticas da National Foundation Science (NFS), agência governamental dos EUA que financia pesquisa básica de faculdades e universidades norte-americanas nos campos da matemática, ciência da computação e ciências sociais, entre outros, afirmava que “dados digitais são a moeda do universo de coleta de dados, que, como a moeda na esfera financeira, vem de várias formas diferentes” (NATIONAL SCIENCE BOARD, 2005, p. 19). O NSB identificou, nesse estudo, três formas de dados de pesquisa, quanto à sua natureza, a seguir adaptadas e resumidas:

a) *Observacionais*: São os que resultam do reconhecimento, observação ou registro de fatos ou ocorrências de fenômenos, geralmente com instrumentos, por exemplo, observações da temperatura do oceano em uma data específica ou a atitude dos eleitores antes de uma eleição. Nas ciências sociais, incluem indicadores econômicos e entrevistas. Como constituem registros históricos que não podem ser coletados uma outra vez, são considerados os mais importantes a preservar;

b) *Computacionais*: São os resultados da execução de um modelo ou simulação de computador, cujas informações abrangentes dos dados de entrada, do hardware e do software podem facilitar a reprodução dos dados no futuro, sugerindo a conveniência de arquivamento do próprio modelo e dos metadados. Comuns nas ciências físicas e da vida, também podem ser encontrados nas ciências sociais; e

c) *Experimentais*: São os resultados de procedimentos em condições controladas para testar ou estabelecer hipóteses ou descobrir ou testar novas leis. Originam-se, por exemplo, de medições de padrões genéticos ou de taxas de reação química, experimentos físicos num colisor linear e experimentos de psicologia controlados em laboratório. Os que são projetados para serem replicáveis precisamente não necessitam ser armazenados indefinidamente. Na prática, no entanto, pode não ser possível reproduzir com precisão todas as condições experimentais, donde se torna necessária a preservação dos dados (BORGMAN, 2015; NATIONAL SCIENCE BOARD, 2005).

Borgman cita ainda uma quarta categoria de dados: os registros, que englobam formas de dados que não se encaixam facilmente em categorias de observação, experimento ou computação ou que resultam de qualquer uma dessas categorias. Incluem documentação de atividades governamentais, comerciais, públicas e privadas; livros e outros textos; materiais

de arquivo; documentação na forma de gravações de áudio e vídeo, que podem vir a ser tratados como dados de pesquisa (BORGMAN, 2015).

A falta de uma definição de consenso para dados levou Borgman (2015, p. xvii, tradução nossa) à seguinte reflexão:

Um ponto de partida para entender os papéis dos dados na academia é reconhecer que os dados raramente são coisas. Não são objetos naturais com uma essência própria. Em vez disso, os dados são representações de observações, objetos ou outras entidades utilizadas como evidência de fenômenos para fins de pesquisa ou estudo. Essas representações variam de acordo com o estudo, circunstância e ao longo do tempo. Através das ciências, das ciências sociais e das humanidades, os estudiosos criam, usam, analisam e interpretam dados, muitas vezes sem concordar com esses dados. Conceitualizar algo como dados é em si um ato acadêmico. Estudos giram em torno de evidências, interpretações e argumentos. Os dados são um meio para um fim, que geralmente é o artigo do periódico, livro, documento de um congresso ou outro produto digno de reconhecimento acadêmico. Raramente uma pesquisa é feita com o reuso de dados em mente.

2.5 O acesso aberto a dados de pesquisa

Muitas políticas de agências que financiam (e exigem) o acesso aberto a dados estão vinculadas a políticas de acesso aberto a publicações (BILL & MELINDA GATES FOUNDATION, 2017a; 2017b; COMISSÃO EUROPEIA, 2017; NIH, 2019; WELLCOME TRUST, 2018). A política do Reino Unido²⁸ é explícita quanto a essa relação, pois se compromete a garantir que resultados de pesquisa sejam livremente acessíveis, exigindo que os autores especifiquem como os dados associados à publicação podem ser acessados, mas reconhecendo que é possível haver motivos para proteger o acesso a dados. O NIH exige que as publicações resultantes de seus fundos de doações sejam depositadas no repositório de acesso aberto PubMedCentral, bem como planos de gerenciamento que detalhem de que forma os dados poderiam ser compartilhados (BORGMAN, 2015; NIH, 2019).

²⁸ Ver "RCUK Policy on Open Access and Supporting Guidance" em: <https://www.psa.ac.uk/sites/default/files/RCUK%20Policy%20and%20Revised%20guidance%20on%20Open%20Access.pdf>

Todavia, o acesso aberto a artigos científicos e o acesso aberto a dados subjacentes a esses artigos diferem. O acesso aberto a artigos tem um conceito simples, porém geralmente mal compreendido, em virtude de interesses concorrentes das partes interessadas. Para melhor entendimento, Borgman (2015) evoca Peter Suber, um dos signatários e principal redator da Declaração da Conferência de Budapeste (2002), pioneiro e ativo batalhador pelo acesso aberto à literatura científica já antes dessa conferência. Suber define-o sucintamente: “A literatura de acesso aberto é digital, *on line*, gratuita e livre da maioria das restrições de direitos autorais e licenciamento” (SUBER, 2012, *on line*, tradução nossa), sustentando-o em dois princípios, a seguir descritos.

Considerando que o acesso aberto à literatura científica opera em um domínio diferente do acesso aberto a outras formas de conteúdo, o primeiro princípio de Suber (2012) diz respeito à base na qual essa literatura se estabelece: os autores são os detentores dos direitos autorais de seus trabalhos, pelo menos até o momento em que cedam seus direitos a um editor. O segundo princípio é que raramente os pesquisadores são pagos para escrever artigos científicos, o que lhes permite distribuir amplamente seu trabalho sem perda de receita, o que não é o caso de autores de outros domínios distintos da ciência. Autores acadêmicos escrevem artigos para causar impacto com seus achados ou para influenciar, mas não para obter remuneração. Também têm interesse em atingir um grande público que o cite. Suas principais fontes de financiamento para pesquisa são os salários e as bolsas de pesquisa (BORGMAN, 2015; SUBER, 2012).

Em comparação com os dados de pesquisa abertos, enquanto os autores são os detentores dos direitos autorais dos artigos científicos, pelo menos inicialmente, o mesmo raramente ocorre nos dados. A autoria é um tópico muito debatido dentro e entre campos da ciência. E determinar quem se qualifica como “autor” dos dados é um tópico pouco explorado na maioria das obras em colaboração (WALLIS, 2012, apud BORGMAN, 2015). Mesmo quando pessoas e grupos atribuem permissão para dados, os direitos e responsabilidades podem permanecer obscuros. Muitas formas de dados são criadas e controladas por pesquisadores, mas a propriedade é uma outra questão. Algumas formas de dados são consideradas fatos, que não são protegidos por direitos autorais. As políticas de direitos autorais sobre dados podem variar de acordo com a instituição, agência financiadora, contrato, jurisdição e outros fatores (BOGMAN, 2015).

Da mesma forma, no segundo princípio de Suber (2012), a diferença é evidente. Autores acadêmicos escrevem artigos em periódicos científicos para obter impacto, e não remuneração. E os financiadores, empregadores e os próprios autores estão interessados em divulgar o mais amplamente possível o trabalho que descreve a pesquisa. Artigos científicos são mais fáceis de ser disseminados para o público, enquanto a liberação de dados geralmente requer investimentos além da realização da pesquisa (como a curadoria) e da redação dos artigos. Há dados que se tornam ativos valiosos para pesquisadores, pois foram acumulados ao longo de toda uma carreira. Portanto, exigem cuidado na divulgação, se forem divulgados (BORGMAN, 2015).

Como parte de um ecossistema de pesquisa acadêmica em constante evolução, o movimento de acesso aberto se ampliou e agregou a abertura de dados de pesquisa. Assim, a uma avaliação realística sobre dificuldades na abordagem de dados de pesquisa, que permeia a obra de Borgman, cabe citar os benefícios que Tennant et al. (2016) resumem em artigo de revisão no qual analisam os impactos acadêmicos, econômicos e sociais do acesso aberto.

Para Tennant e colaboradores, o compartilhamento de dados é fundamental para o progresso científico, mas também ganhou corpo porque se tornou um requisito comum, juntamente com o acesso aberto, para financiamento e publicação de pesquisas. Esses autores afirmam e, neste estudo, no capítulo 4, veremos que periódicos (*PLoS Medicine*, *Nature Medicine* e *theBMJ*) e agências de financiamento (Bill & Melinda Gates Foundation, NIH e Wellcome Trust) inseriram em suas políticas uma diretriz que obriga autores a compartilharem dados de pesquisa, sob pena de o artigo submetido ou a solicitação de financiamento serem reprovados (TENNANT et al., 2016). Por outro lado, o acesso aberto tem importante papel quando se trata de compartilhar gratuitamente dados de pesquisa em tempos de emergência de saúde pública, como no caso do vírus Zika (JORGE, 2018) e, atualmente, da Covid-19 (OPAS, 2020).

Os alegados benefícios dos dados abertos são diversos e incluem uma vantagem de citação. Assim, combinada com a vantagem de citação de artigos de acesso aberto, a disponibilização de dados de pesquisa juntamente com as publicações pode aumentar as citações, em média, 30% (PINOWAR; VISION, 2013) e até 69% (PINOWAR; DAY; FRIDSMA, 2007), dependendo fundamentalmente do campo (DORCH; DRACHEN; ELLEGAARD, 2015).

Tennant et al. (2016, *on line*), por sua vez, apontam ainda mais sete benefícios oriundos da abertura de dados, a seguir adaptados e resumidos:

a) o aprimoramento da reprodutibilidade, derivado do acesso total aos dados, metadados e códigos usados para produzir os resultados finais, fornecidos juntamente com a publicação;

b) o estímulo a inovações, que podem gerar novos métodos analíticos;

c) a possibilidade de novas perguntas de pesquisa serem feitas, a partir dos dados subjacentes, mediante análise de grandes volumes de dados e consequentes descobertas novas e talvez surpreendentes;

d) a percepção de que dados compartilhados e visíveis para a comunidade incentiva pesquisadores a garantirem seu envolvimento em melhor documentação de dados e métodos de pesquisa;

e) fornecimento de *backup* digital para conjuntos de dados, protegendo valiosos recursos científicos, além de uma quantidade considerável de dados produzidos todos os dias, que não é publicada, sair da gaveta e não permanecer oculta;

f) redução do custo da realização da pesquisa, com o desincentivo de coleta de dados redundantes não acessíveis ao público e chance de pesquisadores se aproximarem mais e melhor do que está acontecendo em seus campos; e

g) dados abertos têm valor econômico e podem contribuir para o valor agregado de produtos e serviços nos principais setores.

Na esfera dos dados subjacentes de pesquisa, a questão dos direitos autorais impõe-se ao trazer algumas novidades relativamente ao processo de abertura e compartilhamento. Para serem totalmente abertos, os dados de pesquisa precisam estar disponíveis, acessíveis, inteligíveis, avaliáveis e utilizáveis, preconiza o relatório da The Royal Society (2012). Contudo, de acordo com Paul Uhlir (2015), para que isso aconteça, são necessárias licenças de uso comum, renúncias de direitos, termos novos que ainda precisam ser melhor definidos nesse contexto, visando a uma necessária mudança cultural. Para tanto, não descartada a consideração de que a atribuição de propriedade para dados não é tarefa fácil, Uhlir apoia, além da adoção das licenças e do instrumento CC0 para renúncia total a direitos autorais, da Creative Commons, a criação de um *status* de domínio público, que permitiria e garantiria a

reutilização e a re-disseminação irrestritas, além da interoperabilidade legal de dados de pesquisa. O autor advoga que o “[...] domínio público pode ser criado formalmente por leis públicas por meio de legislação nacional que exclua certas categorias de dados e informações da proteção de direitos autorais ou proíba a imposição de restrições ao seu uso” (UHLIR, 2015, p. 19).

2.6 O papel dos editores científicos na abertura: falta padrão?

Instados por governos, organizações internacionais e agências de financiamento à pesquisa, editores e periódicos cada vez mais consideram e desenvolvem práticas e políticas em torno da abertura de dados subjacentes aos artigos que publicam (PLOS MEDICINE, 2014; NATURE, 2016a; THEBMJ, 2017; 2019; ALEIXANDRE-BENAVENT, 2019; ICMJ, 2018).

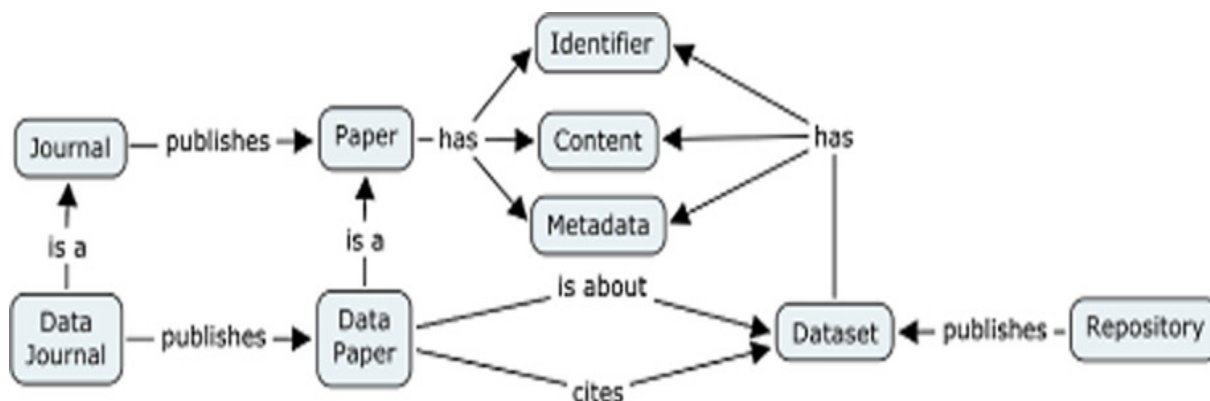
Alguns editores e periódicos já fizeram movimentos significativos nessa direção: exigência de depósito dos dados subjacentes aos artigos em repositórios digitais, de preferência temáticos (de uma área), institucionais ou, quando estes não existirem, generalistas ou multidisciplinares (aceitam qualquer tipo de dados), visando ao acesso e reuso; vinculação (*linkage*) entre publicações e dados (MAYERNIK; PHILIPS; NIENHOUSE, 2016; HRYNASZKIEWICZ, 2019) e colaboração com repositórios de dados de pesquisa; citação de dados, com a criação do DataCite²⁹; licenciamento para reuso; uso de identificadores persistentes, como DOI e ORCID (Open Researcher and Contributor ID); mineração de texto e dados; e revisão por pares de dados. Um dos movimentos mais marcantes nessa direção foi a criação dos *data journals*, periódicos que publicam artigos sobre dados e não sobre a pesquisa como um todo.

Os *data journals* foram criados como um novo meio de comunicação científica, numa tentativa de encontrar uma primeira solução para a demanda por abertura de dados de pesquisa (CANDELA et al., 2015). São publicações em meio digital, revisadas por pares, para descrição de conjuntos de dados científicos por meio dos *data descriptors*, um tipo de artigo que combina a característica de conteúdo narrativo de manuscritos de periódicos tradicionais com metadados estruturados e curadoria. Descrevem fluxos de trabalho experimentais e

²⁹ DataCite é uma organização sem fins lucrativos que fornece identificadores persistentes (DOIs) para dados e outros resultados de pesquisa a comunidades científicas, tornando-os detectáveis e promovendo o compartilhamento e a citação de dados. Ver em: <<https://datacite.org/>>

apontam para registros de dados arquivados publicamente. Na Figura 1, vemos um mapeamento do artigo de dados comparado ao do artigo de periódico comum. Há *data journals* que cobram taxas de processamento de artigos para publicação em acesso aberto³⁰ (SPINAK; PACKER, 2014; CANDELA et al., 2015; OREGON, 2019).

Figura 1. Mapa conceitual do artigo de dados



Fonte: Candela, 2015.

Pfeiffenberger e Carlson (2011) e Candela e colaboradores (2015) identificam o *Earth System Science Data* como um dos primeiros periódicos de dados. Lançado em 2008, é um periódico internacional interdisciplinar da área de Ciências da Terra, concebido especificamente para a publicação de artigos sobre conjuntos de dados originais de pesquisa.

Em 2009, o *Journal of Robotics Research* passa a solicitar e publicar artigos de dados, como um novo gênero de artigo de periódico, juntamente com os artigos tradicionais, modelo misto que persistiria, como veremos adiante. Seu objetivo “era facilitar e incentivar o lançamento de conjuntos de dados de alta qualidade e revisados por pares para a comunidade de robótica” (NEWMAN, 2009, *on line*, tradução nossa). Em 2013, o Nature Publishing Group lançou o *Scientific Data*, periódico de acesso aberto que publica descrições de conjuntos de dados (SCIENTIFIC DATA, 2018).

³⁰ Ver APCs de: *Scientific Data* <<https://www.nature.com/sdata/about/oa>>; *Data in Brief* <<https://www.elsevier.com/journals/data-in-brief/2352-3409/open-access-journal>> e *Earth System Science Data* <https://www.earth-system-science-data.net/about/article_processing_charges.html>.

Segundo Candela e colaboradores (2015), o número de periódicos de dados vem crescendo, mas ainda não se nota uma linhagem expressiva de periódicos de dados “puros”. Dos 116 periódicos que esse pesquisador inventariou em pesquisa publicada em 2015, 109 misturavam artigos tradicionais com artigos de dados, enquanto apenas 7 publicavam unicamente artigos de dados. O tema que aparece mais representado neste tipo de publicação é a Medicina, seguida de Bioquímica, Genômica e Biologia Molecular.

Todavia, apesar da mobilização de alguns editores e da criação dos *data journals*, ainda há periódicos e editores que cumpriram poucos ou nenhum movimento significativo em direção à abertura de dados de pesquisa, como veremos adiante. E, entre aqueles que o fizeram, há relativamente pouca coisa em comum, o que pode evidenciar a falta de padrões na implantação dessa nova cultura. Ademais, alguns editores identificam certa relutância de pesquisadores em aderir à abertura, pois eles alegam não quererem se antecipar demais ao que percebem como comportamentos e culturas dominantes da comunidade de pesquisa (THE ROYAL SOCIETY, 2012; OPEN RESEARCH DATA TASK FORCE, 2017).

Em 2014, a pedido da PLoS, organização fundada em 2003 por editores científicos, que publica revistas científicas eletrônicas de medicina e interdisciplinares e cobra taxas de processamento de artigos para oferecer acesso aberto à informação científica, Lin e Strasser (2014, *on line*) reuniram um grupo de líderes de gestão de dados, incluindo gestores de repositório, bibliotecários, financiadores, construtores de infraestrutura, diretores de programa, desenvolvedores e pesquisadores. A missão era discutir sobre “O que os editores podem fazer para promover o trabalho de bibliotecas e instituições no avanço do acesso e disponibilidade de dados?”. A premissa era a de que “os periódicos desempenham um papel crítico [...], sejam eles comerciais, sem fins lucrativos, de sociedades científicas, de acesso aberto, institucionais etc.” (LIN; STRASSER, 2014, *on line*, tradução nossa), em um futuro ecossistema em que os dados de pesquisa são considerados parte integrante das comunicações acadêmicas. Os resultados foram a uma consulta pública da comunidade e redundaram em oito recomendações para os editores aumentarem o acesso a dados:

1. Estabelecer e fazer cumprir uma política mandatória de disponibilidade de dados.
2. Contribuir para o estabelecimento de padrões comunitários para gerenciamento e compartilhamento de dados.

3. Contribuir para o estabelecimento de padrões comunitários para preservação de dados em repositórios confiáveis.
4. Fornecer canais formais para compartilhamento de dados.
5. Trabalhar com repositórios para agilizar o fluxo de submissão de dados.
6. Exigir a citação apropriada para todos os dados associados a uma publicação – produzidos e usados.
7. Desenvolver e informar indicadores que darão suporte a dados como um resultado acadêmico de primeira classe.
8. Incentivar o compartilhamento de dados promovendo o seu valor. (LIN; STRASSER, 2014, *on line*, tradução nossa).

Neste mesmo sentido, os autores brasileiros Costa e Leite (2017) e Dudziak (2016) também propõem um conjunto de princípios e recomendações para a comunicação de dados de pesquisa, que vão desde a adoção de um caráter mandatório das políticas até a proteção dos direitos de propriedade intelectual e sugestão de licenças para dados.

Vasilevsky et al. (2017) revisaram as “instruções aos autores” de 318 periódicos biomédicos com o objetivo de analisar os requisitos e características de compartilhamento de dados de cada revista, além de recomendações de licenciamento de dados de pesquisa. Concluíram que muitos periódicos ainda não implementaram políticas de abertura de dados e os requisitos variam amplamente entre os periódicos, não havendo garantia de que os dados subjacentes estejam disponíveis e reutilizáveis ao máximo.

O estudo de Womack (2015, *on line*, tradução nossa) sobre práticas de compartilhamento de dados subjacentes dos 10 principais periódicos nas disciplinas de biologia, química, matemática e física também constatou que “práticas de compartilhamento de dados ainda são relativamente raras nos principais periódicos dessas disciplinas”, provavelmente indicando o estado de transição para uma cultura de abertura.

No segmento de publicação científica da área da saúde, um dos focos desta pesquisa, o Comitê Internacional de Editores de Revistas Médicas³¹ (International Committee of Medical Journal Editors – ICMJE) revela-se importante fonte a ser observada, pois tem ajudado a disseminar a cultura da abertura e compartilhamento de dados de pesquisa clínica entre

³¹ Criador das Normas Vancouver e conhecido como "Grupo Vancouver", o ICMJE é um tradicional comitê editorial, nascido em 1978. Composto por editores de periódicos médicos, o comitê fornece orientações rígidas sobre normalização e políticas editoriais, incluindo uma política de compartilhamento de dados de pesquisa clínica (ICMJE, 2018).

centenas de filiados espalhados pelo mundo. Em 2016, publica sua proposta inicial em editorial, replicado em várias revistas médicas, no qual considera uma obrigação ética a abertura de dados de ensaios clínicos (TAICHMAN et al., 2016).

Em 2017, publica novo editorial, também replicado em várias revistas médicas³², com os termos da proposta do ICMJE detalhados para implementação (TAICHMAN et al., 2017; ICMJE, 2018). Em ambos os editoriais constam o requisito que condiciona a publicação do artigo a uma declaração de concordância de compartilhamento de dados assinada pelo pesquisador e a exigência de apresentação de um plano de gestão de dados.

Em pesquisa que teve o objetivo de testar a eficácia da abertura e compartilhamento de dados de ensaios clínicos randomizados publicados nos periódicos *theBMJ* e *PLOS Medicine* entre 2013 e 2016, Naudet e colaboradores (2018, *on line*) acabaram testando também a política de abertura de dados de ensaios clínicos do ICMJE. Dezesete entre 37 ensaios clínicos selecionados satisfaziam a definição de disponibilidade de dados. Destes 17, 14 foram totalmente reproduzidos em todos os seus resultados primários. Contudo, o grupo de autores, formado por pesquisadores dos EUA e da Europa, não considerou a disponibilidade de dados ótima e concluiu que “as práticas de compartilhamento de dados precisam tornar-se mais difundidas e otimizadas para permitir análises significativas e reutilização de dados”.

Três manuscritos de diferentes épocas – 2010, 2015 e 2019 – oferecem uma amostra de um lento processo de abertura e compartilhamento de dados. O primeiro manuscrito é de Pisani e AbouZahr (2010, *on line*). Ambos sustentam que financiadores de pesquisa em saúde pública não enfrentavam adequadamente os obstáculos de compartilhamento de dados, pois ainda predominavam “estruturas profissionais que recompensam a publicação de análises, mas não dados, e fluxos de financiamento e planos de carreira que continuam subestimando o trabalho crítico de gerenciamento de dados”, além de “Questões práticas [...]: como e onde os dados devem ser armazenados a longo prazo, quem controlará o acesso e quem pagará por esses serviços?”

Cinco anos depois, Carr e Littler (2015, *on line*), consultores da Wellcome Trust, agência de financiamento de pesquisa, sediada em Londres, no Reino Unido, criada em 1936

³² O editorial foi publicado simultaneamente em: *Annals of Internal Medicine*, *British Medical Journal*, *Canadian Medical Association Journal*, *Chinese Medical Journal*, *Deutsches Ärzteblatt (German Medical Journal)*, *Ethiopian Journal of Health Sciences*, *JAMA (Journal of the American Medical Association)*, *Nederlands Tijdschrift voor Geneeskunde (The Dutch Medical Journal)*, *New England Journal of Medicine*, *New Zealand Medical Journal*, *PLOS Medicine*, *Revista Médica de Chile*, *The Lancet*, and *Ugeskrift for Laeger (Danish Medical Journal)* (TAICHMAN, 2017).

pelo magnata farmacêutico *Sir Henry Wellcome* para financiar pesquisas em saúde humana e animal, se empenhavam em apresentar um esforço dos financiadores de pesquisa para mudar o quadro referido por Pisani e AbouZahr (2010): a criação do Fórum de Dados de Pesquisa em Saúde Pública³³, que reuniu financiadores globais de pesquisa em saúde, com a presença de editores, que trabalhavam juntos para coordenar suas atividades e “aumentar a disponibilidade de dados de pesquisa em saúde pública e epidemiologia de maneira equitativa, ética e eficiente.”

Apesar da criação desse Fórum, o terceiro manuscrito, de Hajduk e colaboradores (2019, *on line*, tradução nossa), confirma a imposição de obrigações contratuais de grandes financiadores. Além disso, eles concordam que compartilhar dados de saúde “não é tão fácil quanto parece” e sugerem que “isso se deve, ao menos parcialmente, à falta de padrões claros e diretrizes estabelecidas que expliquem onde, quando e como compartilhar dados”, mas também a “preocupações sobre a sensibilidade dos dados e a privacidade do paciente, bem como os aspectos técnicos do processamento de dados antes que os dados possam ser compartilhados” e informações sobre questões éticas e legais claras e precisas.

Ainda na área da saúde, especialmente no segmento de ensaios clínicos, um constante dilema permanece insolúvel e parece ser um dos desafios da abertura (TUCKER et al., 2016): como compartilhar dados pessoais oriundos de pesquisa em saúde, em um nível em que se consiga uma contribuição efetiva aos estudos para melhoria da saúde das pessoas sem macular o seu direito à privacidade, que inclui os dados pessoais e os dados sensíveis³⁴? De acordo com Petersen (2018, *on line*, tradução nossa), “os profissionais de saúde precisam pensar de maneira mais ampla sobre como preservar a privacidade em um ambiente de saúde impulsionado pelo compartilhamento de dados”.

Popescu et al. (2018, *on line*, tradução nossa) pesquisaram a correlação entre dados pessoais e direito de propriedade intelectual no campo da pesquisa médica. Os autores entendem que “A divulgação de informações médicas coletadas por meio de trabalhos científicos está sujeita ao fato de que o progresso em qualquer campo deve ser incentivado, a fim de aumentar a qualidade de vida”, mas destacam que é preciso “criar um equilíbrio entre os interesses do pesquisador e o interesse público”.

³³ Ver em: <<https://wellcome.ac.uk/what-we-do/our-work/public-health-research-data-forum>>

³⁴ Os dados pessoais são aqueles que revelam a identidade do indivíduo. São considerados dados pessoais sensíveis os relativos à condição racial, sexual, social ou política do indivíduo (BRASIL, 2018).

Iain Hrynaszkiewicz, editor e pesquisador, com passagens pelo Biomed Central, F1000, Springer Nature e PLoS, é um ativo defensor da abertura de dados de pesquisa subjacentes. Em 2009, Hrynaszkiewicz e Altman (2009) identificavam, como editores da revista *Trials*, da BioMed Central, a “propriedade” e a privacidade como problemas para publicação de dados brutos de ensaios clínicos em periódicos biomédicos revisados por pares, frutos de políticas de financiadoras da pesquisa que não abordavam tais preocupações em suas orientações. Em 2010, Hrynaszkiewicz e colaboradores (2010, *on line*, tradução nossa), com base na pouca orientação sobre privacidade, tanto de periódicos quanto de financiadoras de pesquisa, propuseram “um padrão mínimo para desidentificação de conjuntos de dados para garantir a privacidade do paciente ao compartilhar dados de pesquisas clínicas”.

Em 2012, preocupados com o outro “problema” – propriedade (que é abordada no capítulo 3) –, Hrynaszkiewicz e Cockerill (2012) sugeriram à BioMed Central uma proposta que denominaram “*open by default*” (ver mais detalhes na seção 3.2). Em virtude de constatarem a publicação de dados de pesquisa com licenças para reuso inapropriadas, o ponto central era propor um padrão de licenciamento combinando uma licença e um instrumento jurídicos da Creative Commons. Após consulta pública, por maioria de 29 votos a favor, 5 contra e 8 abstenções, o contrato de licença de direitos autorais da BioMed Central para artigos de acesso aberto foi alterado, de modo que todos os dados subjacentes a artigos publicados e/ou em arquivos adicionais adotassem por padrão o instrumento CC0 de Dedicção ao Domínio Público, enquanto os artigos científicos publicados deveriam adotar por padrão uma licença CC BY, a mais aberta da CC (ver sobre licenciamento nas seções 3.4 e 3.5 e no Quadro 7). A BioMed adota esse protocolo até hoje³⁵. A proposta da mudança é encontrada em Hrynaszkiewicz e Cockerill (2012), e o relatório da consulta pública, em Hrynaszkiewicz; Busch; Cockerill (2013).

Em 2017, agora atuando pela Spring Nature, Hrynaszkiewicz e colaboradores (2017) desenvolveram uma estrutura comum e padronizada para as políticas de dados de pesquisa de todos os periódicos desse conglomerado editorial. Um grupo de cerca de 30 editores, que cobriam todas as disciplinas, auditaram e identificaram as características comuns das políticas de dados de pesquisa dos periódicos publicados pela Spring Nature e ainda ouviram editores acadêmicos, bibliotecários e financiadores. O trabalho resultou na definição de quatro tipos de

³⁵ Ver em: <<https://www.biomedcentral.com/about/policies/open-data>>

política de compartilhamento de dados (Quadro 3), que funcionam como modelos, pois podem ter pequenos ajustes para se adequarem ao escopo de cada periódico da Springer Nature.

Em janeiro de 2017, mais de 700 periódicos, dentre os 2.500 que a Springer Nature publica, passaram a adotar uma dessas quatro políticas padrão. Para ampliar o potencial de padronização e harmonização da política de dados entre financiadores, instituições, repositórios, sociedades e outros editores, a estrutura comum e padronizada para as políticas foi disponibilizada sob uma licença CC BY 4.0 (HRYNASZKIEWICZ et al., 2017; GRANT; HRYNASZKIEWICZ, 2018).

Quadro 3. Tipos resumidos de políticas de dados dos periódicos Springer Nature

Tipos de Política	Resumo da Política	Exemplos de Periódico
Tipo 1	Compartilhamento de dados e citação de dados são incentivados	<i>Photosynthesis Research</i> (ver 'Instruções aos autores')
Tipo 2	Compartilhamento de dados e evidência de dados são incentivados	<i>Plant and Soil</i> (ver 'Instruções aos autores')
Tipo 3	Compartilhamento de dados é incentivado e declarações de disponibilidade de dados são uma exigência	<i>Palgrave Communications</i> (ver 'Políticas editoriais')
Tipo 4	Compartilhamento de dados, evidência de compartilhamento de dados e avaliação por pares de dados são uma exigência	<i>Scientific Data</i> (ver 'Políticas de dados')

Fonte: Estes resumos e a íntegra das políticas podem ser encontrados em:

<<https://www.springernature.com/gp/authors/research-data-policy/research-data-policy-types>>

Nota: A lista de periódicos que adotaram uma política padronizada de dados de pesquisa, incluindo informações sobre qual tipo de política de dados eles implementaram, encontra-se em

<<https://www.springernature.com/gp/authors/research-data-policy/springer-nature-journals-data-policy-type/12327134>>.

Em 2019, ainda na mesma Springer Nature, Hrynaszkiewicz (2019) discutia sobre como os editores poderiam ajudar na promoção da qualidade e reprodutibilidade de dados, a fim de promover transparência e abertura. Ele lista seis áreas para se alcançar essas metas:

1. Compreender os problemas e motivações dos pesquisadores.
2. Aumentar a conscientização e mudar comportamentos.
3. Melhorar a qualidade, transparência e objetividade do processo de revisão por pares.
4. Aperfeiçoar a infraestrutura de comunicação acadêmica e inovação.
5. Aumentar os incentivos.
6. Tornar a publicação de pesquisas o mais aberta e acessível (HRYNASZKIEWICZ, 2019, *on line*, tradução nossa).

Para além das demonstrações singulares de práticas de abertura de dados de pesquisa que apresentamos, as narrativas reunidas de três manuscritos de diferentes momentos sobre essa mesma temática – de Pisani e AbouZahr (2010), Carr e Littler (2015), e Hajduk e colaboradores (2019) – e o árduo trabalho de Hrynaszkiewicz e colaboradores em três grupos editoriais de ponta, que atravessou um período que vai de 2009 a 2019, podem ser reveladores da falta de práticas e padrões consistentes na abertura de dados de pesquisa subjacentes.

Christine Borgman vai mais longe. Segundo a professora do departamento de Estudos da Informação da UCLA, os pesquisadores têm produzido um “dilúvio” de dados sem precedentes, usando novos métodos e instrumentos, mas os dados não estão prontamente disponíveis, como nos campos da astronomia e da genômica, que já possuem tradição de compartilhamento. A análise de Borgman confirma a tendência indicada em nossa sumária revisão de literatura, ao indicar que em outros campos as práticas de abertura e compartilhamento variam amplamente, além de os dados assumirem várias formas e serem tratados de várias maneiras, utilizando-se “abordagens difíceis de se interpretar quando esses dados são removidos de seu contexto inicial” (BORGMAN, 2012, *on line*, tradução nossa).

Ademais, a autora acrescenta que há desincentivos que precisam ser resolvidos para a abertura e o compartilhamento aumentarem. São eles: a falta de recompensa ou crédito; a quantidade substancial de trabalho para documentar dados em formatos reutilizáveis;

preocupações com o uso indevido ou mal interpretado dos dados; controle sobre propriedade intelectual [inclusive direito autoral] e a necessidade de restringir o acesso ou desidentificar dados sobre seres humanos ou espécies ameaçadas (BORGMAN, 2012; SAYÃO; SALES, 2016b).

Para a autora, em resumo, o compartilhamento de dados de pesquisa é, além de um enigma, desafiador para se “entender quais dados podem ser compartilhados, por quem, com quem, sob quais condições, por que e com quais efeitos”. Somente com muita pesquisa sobre o tema teremos “as respostas a essas perguntas [que] informarão a política e a prática de dados”, acredita Borgman (2012, *on line*, tradução nossa).

2.7 A percepção dos pesquisadores

Segundo alguns estudiosos do tema da abertura na ciência, há pouco consenso sobre o entendimento do que é e como incentivar a abertura e o compartilhamento de dados de pesquisa (BORGMAN, 2012; WALLIS; ROLANDO; BORGMAN, 2013; GRAND et al., 2016). Em consequência, existem questões em suspenso que se refletem no entendimento dos pesquisadores: pouca clareza sobre como implementar e aplicar ciência aberta; as políticas de abertura de dados de pesquisa subjacentes de revistas têm termos e requisitos diferentes para os pesquisadores; as instituições oferecem diferentes infraestruturas para repositórios e bancos de dados; e as comunidades acadêmicas têm distintos compromissos e objetivos, e ainda se discutem que recompensas adotar para estimular pesquisadores a aderirem à abertura. Com base nesse cenário, de que forma os pesquisadores têm percebido a abertura e o compartilhamento de dados de pesquisa subjacentes a artigos científicos?

Em uma de suas tantas provocações acerca das discussões sobre benefícios, riscos e recompensas associados ao uso de dados de pesquisa, Borgman (2015) diz que os pesquisadores, muitas vezes, têm grande dificuldade em reutilizar seus próprios dados. Segundo a autora, tornar esses dados úteis para outros desconhecidos, para fins imprevistos, é ainda mais difícil. O compartilhamento de dados é a norma em apenas alguns campos porque é muito difícil de fazer. Ademais, os incentivos são mínimos, enquanto investimentos massivos em infraestruturas de conhecimento são exigidos.

De acordo com resultados preliminares de nossa pesquisa bibliográfica, a abertura e o compartilhamento de dados de pesquisa ainda não são práticas rotineiras entre pesquisadores

que publicam resultados de suas pesquisas em revistas científicas. Entre outras, a questão dos direitos autorais dos dados, da privacidade, da citação como incentivo, da complexidade de uma mudança cultural, a falta de ferramentas de coleta, análise e gerenciamento de dados, a necessidade de treinamento figuram como motivadores importantes dessa retração.

Em um estudo empírico sobre abertura e compartilhamento de dados de pesquisa subjacentes a artigos por autores que publicaram nas revistas de saúde *PLoS Medicine* ou *PLoS Clinical Trials*, Savage e Vickers (2009) solicitaram dados a 10 pesquisadores. Mesmo lembrados de que a revista exige o compartilhamento de dados, 7 não compartilharam seus dados. Entre as razões para não compartilhar, além das preocupações com a privacidade do paciente (para áreas médicas) e com futuras oportunidades de publicação, o desejo de manter direitos exclusivos como autor de dados, que levaram muitos anos para serem produzidos, foi decisivo para o não compartilhamento.

Apesar de a abertura e o compartilhamento de dados de pesquisa subjacentes estarem associados ao aumento da taxa de citação (PINOWAR; DAY; FRIDSMA, 2007), o que significa uma moeda de valor para os pesquisadores, pois as contagens de citações são usadas para financiamento de pesquisa e promoções na carreira, as práticas reais ainda estão atrasadas e nem sempre em conformidade com regulamentos de agências de financiamento, repositórios e editoras.

Em pesquisa realizada entre setembro de 2013 e fevereiro de 2014, estimulada por políticas pioneiras em ciência aberta e acesso aberto promovidas pelo governo britânico à época, Levin et al. (2016) efetuaram 22 entrevistas em profundidade, semiestruturadas, com pesquisadores britânicos, entre 35 e 60 anos, das áreas de biologia de sistemas, biologia sintética e bioinformática, que ocupam cargos sêniores em 11 instituições de ensino superior no Reino Unido. Por meio de uma análise das transcrições das entrevistas, Levin et al. (2016) identificaram 7 temas principais, que caracterizam a compreensão dos entrevistados sobre a abertura na ciência, além de 9 fatores que moldam a prática da abertura na pesquisa (Quadro 4).

Quadro 4. Pesquisa sobre percepção de ciência aberta com pesquisadores britânicos da área de biologia. Visão geral da análise temática (2013-2014).

<i>Entendimentos dos pesquisadores biomédicos sobre abertura na ciência</i>	<i>Fatores que afetam a prática da abertura na ciência</i>
1. Doação e acesso oportuno a componentes de pesquisa	1. A existência de repositórios e bancos de dados para dados, materiais, software e modelos
2. Padrões para o formato e a qualidade dos componentes de pesquisa	2. A competitividade dos campos acadêmicos
3. Metadados e anotação [de dados e adição de metadados a conjunto de dados]	3. A natureza digital da pesquisa
4. Colaboração e cooperação com colegas e comunidades	4. Sistemas de crédito na pesquisa acadêmica
5. Liberdade de escolha de locais e estratégias para a divulgação de pesquisas	5. Estruturas de carreira na pesquisa acadêmica
6. Sistemas de revisão por pares transparentes [anonimato como barreira para honestidade e qualidade]	6. Colaborações com parceiros industriais, bem como tentativas de comercialização
7. Acesso a componentes de pesquisa em contextos não ocidentais e/ou não acadêmicos [fornecer acesso a dados e publicações em locais com menos recursos ou infraestruturas, físicas, econômicas ou intelectuais]	7. Modelos e diretrizes para propriedade intelectual
	8. Visões governamentais sobre o <i>status</i> e o papel social desempenhado pelas universidades
	9. A existência de várias, e às vezes conflitantes, políticas governamentais de ciência aberta

Fonte: Levin et al., 2016, tradução nossa.

Segundo Levin (2016), os pesquisadores entrevistados destacaram que diferentes abordagens de propriedade intelectual – incluindo os direitos autorais – de código aberto e fechado afetaram o compartilhamento e a disseminação de ideias e recursos. Alguns afirmaram que licenças de código aberto como Creative Commons, LGPL ou GNU (licenças de software livre) beneficiaram o desenvolvimento de recursos e tecnologias compartilhadas, enquanto outros enfatizaram que a comercialização continuava a depender de patentes e segredos comerciais. A maioria declarou, no entanto, que os contratos de transferência de materiais constituíam uma barreira significativa ao compartilhamento e disseminação de materiais de pesquisa. Confirmando a tendência já vislumbrada neste estudo, o nono item da

categoria “Fatores que afetam a prática da abertura na ciência” (Quadro 4) revela a falta de padrão em políticas governamentais para a ciência aberta.

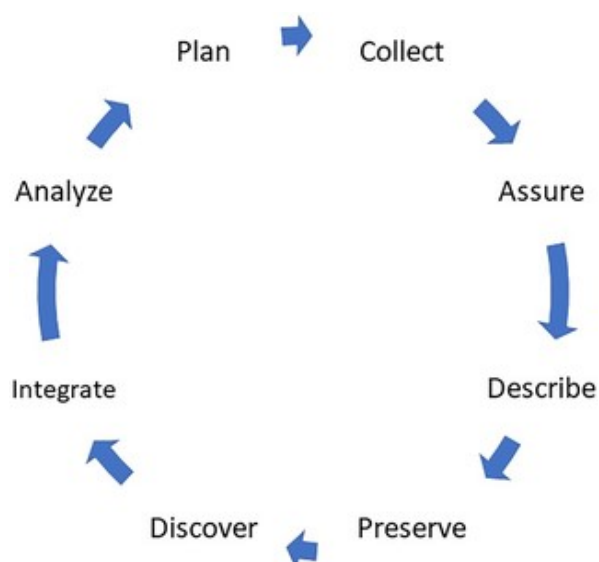
A pesquisadora Carol Tenopir, da Universidade do Tennessee e da Universidade do Havaí, é uma das precursoras em estudos sobre o impacto do acesso aberto (à literatura e aos dados científicos) na comunicação científica. Em duas pesquisas lideradas por ela e planejadas para momentos diferentes, com amostras abrangentes, ambas projetadas pelo DataONE³⁶ e financiadas pela NSF, Tenopir et al. (2011; 2015) demonstram a complexidade da mudança para uma cultura de compartilhamento de dados. Na primeira, de 2011, inquiriram 1.329 cientistas de diversas áreas sobre práticas e percepções de compartilhamento de dados, suas barreiras e facilitadores. Na categoria “Motivos para não disponibilizar dados eletronicamente”, os pesquisados citam como principal motivo de seus dados não estarem disponíveis eletronicamente para reuso o ‘tempo insuficiente’ (603/53,6%), seguido pela ‘falta de recursos financeiros’ (445/39,6%). No tocante à propriedade intelectual como barreira, 24,1% (271) responderam que ‘não têm direito para divulgar dados’. No quesito legal, as percepções sobre compartilhamento e reutilização de dados variam entre os participantes mais jovens e mais velhos. A condição de obtenção de permissão legal para o uso de dados foi selecionada por uma proporção significativamente maior entre as idades de 22 a 39 e uma proporção significativamente menor nas pessoas com 50 anos ou mais (TENOPIR et al., 2011).

Na segunda, Tenopir et al. (2015, *on line*, tradução nossa) notaram um ambiente mais favorável, mas ressaltaram que “ainda existem riscos e barreiras que podem estar retardando o movimento de compartilhamento de dados”, entre eles a falta de ferramentas de gerenciamento de dados fáceis de usar, incorporadas às ferramentas padrão de coleta e análise de dados e que ajudam a resolver a principal barreira percebida no estudo anterior liderado por essa autora – a falta de tempo para o compartilhamento de dados. Também mencionam: necessidade de treinamento para pesquisadores; auxílio de gerentes de dados bem treinados para ajudar nas tarefas para descrever e compartilhar dados; e incorporação de incentivos, como citação e coautoria, em políticas de gerenciamento de dados para pesquisadores que compartilham dados, como por exemplo em repositórios de dados que definiriam padrões de citação para crédito adequado a autores cujos dados forem utilizados (TENOPIR et al., 2015).

³⁶ Programa que reúne uma rede de repositórios de dados operados por centros de pesquisa, universidades, organizações sem fins lucrativos, iniciativas de ciência cidadã, organizações governamentais e não governamentais, fornecendo acesso a dados sobre a vida na Terra e o meio ambiente que a sustenta. Ver em: <<https://www.dataone.org/>>

O novo estudo de Tenopir et al., publicado em 2020, explora as atitudes e práticas de compartilhamento de dados dos cientistas, concentrando-se em três estágios do ciclo de vida dos dados do DataONE: *describe* (descrição), *preserve* (preservação) e *discovery* (descoberta). Além desses três, descritos a seguir, o ciclo de vida dos dados do DataONE (Figura 2) fornece uma visão geral dos estágios do gerenciamento e da preservação dos dados para uso e reúso em mais cinco componentes: *Plano*: descrição dos dados que serão compilados e como os dados serão gerenciados e disponibilizados ao longo de sua vida útil; *Coleta*: as observações são feitas manualmente ou com sensores ou outros instrumentos e os dados são colocados em um formato digital; *Garantia*: a qualidade dos dados é garantida através de verificações e inspeções; *Descrição*: os dados são descritos de maneira precisa e completa usando os padrões de metadados apropriados; *Preservação*: os dados são enviados para um arquivo de longo prazo apropriado (ou seja, um *data center*); *Descoberta*: dados potencialmente úteis são localizados e obtidos, juntamente com as informações relevantes sobre os dados (metadados); *Integração*: dados de fontes diferentes são combinados para formar um conjunto homogêneo de dados que pode ser facilmente analisado; *Análise*: os dados são analisados.

Figura 2. Ciclo de vida dos dados³⁷



Fonte: Tenopir et al., 2020, tradução nossa. Disponível em:

<<https://www.dataone.org/data-life-cycle>>. Acesso em: 30 out. 2020.

³⁷ Ver em : <<https://www.dataone.org/data-life-cycle>>, tradução nossa.

Desenvolvida com base nas duas mencionadas pesquisas anteriores, de 2011 e 2015, lideradas por essa mesma pesquisadora, a atual pesquisa foi realizada em 2017-2018, sob o patrocínio da American Geophysical Union. Recebeu 2.184 respostas, sendo 1.372 de geofísicos e 812 respostas de pesquisadores de diferentes organizações dos EUA, do Oriente Médio, Europa Oriental, Central e do Norte. O estudo constatou que a maioria dos entrevistados exibiu práticas de dados de alto e médio risco, já que armazena seus dados em seus computadores pessoais, servidores departamentais ou unidades de *drives* externos.

Essa maioria também considerou que a falta de acesso aos dados gerados por outros pesquisadores ou instituições era um grande impedimento ao progresso da ciência em geral. Embora as atitudes em relação ao compartilhamento e à reutilização de dados sejam positivas, a prática nem sempre oferece suporte a essas duas ações. Como fora descoberto nas duas pesquisas anteriores, o atual estudo conclui que é claramente necessária a assistência a pesquisadores por meio de gerentes de dados, repositórios de dados disponíveis para armazenamento de longo e curto prazos e programas educacionais para conscientizar e ajudar a gerar boas práticas de dados. Tal fato evidencia uma lenta evolução em uma década, durante a qual foram realizadas três pesquisas sobre a percepção do pesquisador quanto à abertura de dados de pesquisa. Nesse último estudo de 2020, os autores não fazem menção a dificuldades encontradas por pesquisadores em relação a direitos autorais ou outras espécies de propriedade intelectual (TENOPIR et al., 2020).

Com a justificativa de haver uma lacuna entre o que políticos e tomadores de decisão consideram um futuro de dados abertos de pesquisa gerenciados de forma padronizada e a realidade de como os pesquisadores estão se comportando em resposta a essa visão, o conglomerado editorial Elsevier e o Centro de Estudos de Ciência e Tecnologia (Centre for Science and Technology Studies - CWTS), órgão da Universidade de Leiden, na Holanda, publicaram o relatório *Open Data: The Researcher Perspective* (WOUTERS; HAAK; 2017), resultado de uma pesquisa global *on line* com 1.200 pesquisadores.

Em geral, as atitudes em relação à abertura e compartilhamento de dados de pesquisa se mostraram positivas, mas os dados abertos ainda não são uma realidade para a maioria dos pesquisadores. Apesar de declararem que: “ter acesso aos dados de pesquisa publicados beneficiaria a sua própria pesquisa” (73%); “compartilhar dados de pesquisa é importante para fazer pesquisas em seu campo” (69%); “estão dispostos a permitir que outras pessoas

acessem seus dados de pesquisa” (64%), pouco mais de um terço dos entrevistados (n=1200) não publica seus dados e não compartilhou dados de seu último projeto, além de menos de 15% compartilharem dados em um repositório de dados (WOUTERS; HAAK; 2017, *on line*).

Em consonância com os indicativos da falta de padrão na implementação da abertura e compartilhamento de dados de pesquisa, obtidos mediante a nossa pesquisa sumária sobre como editores lidam com essa nova narrativa da ciência aberta (ver seção 2.6 deste capítulo), o relatório mostrou que há uma divisão quase uniforme entre os pesquisadores que acreditam que não há padrões claros para citar dados publicados (45%) e aqueles que acreditam que existem padrões claros (41%).

À pergunta “Como e por que os pesquisadores estão compartilhando dados?”, a maioria deles respondeu que prefere armazenar seus dados perto de sua “casa”, isto é, em arquivos departamentais ou institucionais (37% e 34%, respectivamente), o que contraria os propalados princípios FAIR (Quadro 5), “um guia para editores e administradores de dados para ajudá-los a avaliar se suas escolhas [...] estão tornando artefatos de pesquisa digital localizáveis, acessíveis, interoperáveis e reutilizáveis [*findable, accessible, interoperable e reusable*]” (WILKINSON, 2016, *on line*, tradução nossa). Vinte por cento armazenaram dados sob a forma de arquivamento pessoal e instalações em nuvem, e um em cada oito (12%) disse que não arquiva seus dados (WOMACK, 2015).

Quadro 5. Os princípios orientadores do FAIR

<p>Para ser localizável:</p> <p>F1. (meta) dados são atribuídos a um identificador globalmente único e persistente</p> <p>F2. os dados são descritos com metadados ricos (definidos por R1, abaixo)</p> <p>F3. metadados incluem clara e explicitamente o identificador dos dados que descreve</p> <p>F4. (meta) dados são registrados ou indexados em um recurso pesquisável</p> <p>Para ser acessível:</p> <p>A1. (meta) dados são recuperáveis por seu identificador usando um protocolo de comunicação padronizado</p> <p>A1.1 o protocolo é aberto, gratuito e universalmente implementável</p> <p>A1.2 o protocolo permite um procedimento de autenticação e autorização, quando necessário</p> <p>A2. metadados são acessíveis, mesmo quando os dados não estão mais disponíveis</p> <p>Para ser interoperável:</p> <p>I1. (meta) dados usam uma linguagem formal, acessível, compartilhável e amplamente aplicável para a representação do conhecimento</p> <p>I2. (meta) dados usam vocabulários que seguem os princípios FAIR</p> <p>I3. (meta) dados incluem referências qualificadas a outros (meta) dados</p> <p>Para ser reutilizável:</p> <p>R1. meta (dados) são ricamente descritos com uma pluralidade de atributos precisos e relevantes</p> <p>R1.1. (meta) dados são liberados com uma licença de uso de dados clara e acessível</p> <p>R1.2. (meta) dados estão associados à proveniência detalhada</p> <p>R1.3. (meta) dados atendem aos padrões da comunidade relevantes para o domínio</p>

Fonte: Wilkinson, 2016, *on line*, tradução nossa.

Ainda de acordo com o relatório, os pesquisadores preferem publicar seus dados ao lado de um artigo de pesquisa em um *data journal* (misto) do que em um repositório, porque, como autores, recebem benefícios, tais como mais possibilidades de colaboração, maior reprodutibilidade da pesquisa, maior probabilidade de serem citados e incentivo a outros a retribuir e disponibilizar seus dados. O estudo mostra que a conformidade com os requisitos de periódicos ou editores ou com os mandatos de compartilhamento de dados não costuma ser considerada importante (WOUTERS; HAAK; 2017).

A afirmativa do relatório da Elsevier-Universidade de Leiden quanto à preferência dos pesquisados por *data journals* em detrimento da Via Verde, contudo, não se coaduna com o resultado da pesquisa efetuada pelos pesquisadores Kratz e Strasser (2015). Ambos promoveram um estudo sobre as perspectivas do pesquisador quanto aos *data journals* (e revisão por pares de dados), sob a hipótese de que ainda não havia um consenso sobre o que

deveria ser uma “publicação de dados”. E, sobretudo, por serem desconhecidas as percepções e expectativas de pesquisadores quanto a esse tipo de publicação.

Kratz e Strasser (2015) perguntaram a 250 pesquisadores das áreas de ciências e ciências sociais quais eram suas expectativas no que se refere ao que poderia ser útil para avaliar a confiabilidade, o impacto e aumentar o prestígio de uma “publicação de dados”. As expectativas centraram-se na disponibilidade justamente em banco de dados ou repositório aberto, mas poucos esperavam que os dados fossem revisados por pares, embora considerassem importante a avaliação tanto dos dados quanto da sua documentação e a contagem de *downloads* no mesmo nível da contagem de citações para impacto.

O relatório Elsevier-Universidade de Leiden (WOUTERS; HAAK; 2017) quis saber “Por que os pesquisadores são reticentes em compartilhar seus próprios dados abertamente?”. Com esta pergunta, desvelou o tema por ele denominado “barreiras legais”. Estas são entendidas pelos participantes da pesquisa como preocupações com privacidade, questões éticas e direitos de propriedade intelectual, incluindo os direitos autorais, um dos focos deste estudo. No fundo, diz o relatório, tudo se resume à questão básica de propriedade, responsabilidade e controle de dados (WOUTERS; HAAK; 2017), sobre a qual há uma falta de acordo, segundo Borgman (2015).

Contudo, os pesquisadores participantes da pesquisa possuem crenças claras sobre quem possui os dados, demonstrando um afastamento da realidade sobre o *status* de propriedade de dados de pesquisa. Dois terços (64%) mencionam “eu mesmo” como o proprietário dos dados *antes* da publicação, seguido por seus colegas e colaboradores, indicando que uma pessoa tem mais propriedade sobre os dados do que um instituto, departamento ou financiador. Quase a metade (47%) dos pesquisados sente – incorretamente – que *após* a publicação dos dados a propriedade é transferida para o editor. Apesar de “sociedade” não ter sido apresentada como opção de escolha na pesquisa, um número substancial de pesquisadores respondeu que a sociedade em geral ou o contribuinte são os proprietários dos dados (WOUTERS; HAAK; 2017).

As preocupações legais (e éticas) são citadas como razões para a não publicação de dados de pesquisa ao lado de um artigo. Segundo o relatório Elsevier-Universidade de Leiden, uma proporção substancial das respostas da pesquisa sobre este tópico menciona que os dados são proprietários ou que os pesquisadores não têm consentimento para compartilhar dados.

Além disso, os entrevistados responderam que não gostam da ideia de que outras pessoas possam abusar ou interpretar mal seus dados (muito menos levar crédito por eles).

Leis, políticas e práticas aplicáveis à propriedade, controle e liberação de dados de pesquisa variam de acordo com jurisdições, governos e agências de fomento à pesquisa, governo. Para Borgman (2015, p. 75, tradução nossa), “Confusão sobre propriedade, direitos de liberação de dados, responsabilidade de liberação de dados e as condições sob as quais liberar os dados são restrições significativas à sua disponibilidade”.

No próximo capítulo, adentramos no tema dos direitos autorais aplicáveis a dados de pesquisa subjacentes, discorrendo sobre questões legais que eles encerram, que incluem conceito, princípios e limitações do direito de autor, sua face internacionalizada, proteção a bases de dados, a essencialidade do licenciamento de dados, domínio público e considerações sobre autoria e propriedade de dados.

3 IMPLICAÇÕES DO DIREITO AUTORAL NA DISPONIBILIZAÇÃO PÚBLICA DE DADOS DE PESQUISA SUBJACENTES

Neste capítulo, apresentamos, resumidamente, um histórico, a conceituação, o conteúdo e o objeto dos direitos autorais, os princípios, o seu caráter internacional, as limitações a esses direitos, autoria e propriedade de dados, entre outras questões que são afetas ao direito autoral.

Também tratamos dos específicos direitos *sui generis* de bases de dados e o licenciamento de dados, de modo a subsidiar uma análise da eventual aplicabilidade dos direitos autorais aos dados de pesquisa em um contexto de abertura ou disponibilidade pública desses dados.

Priorizamos nesta tese a análise de algumas das questões legais referentes ao regime de direitos autorais que podem incidir sobre dados de pesquisa, o que não significa dizer que sempre incidirão *a priori* sobre essa categoria de dados. Essa etapa é fundamental para aqueles que desejam compartilhar publicamente dados de pesquisa e os que queiram acessá-los para reuso saberem exatamente o que podem ou não fazer com eles. A falta de clareza ou a desinformação sobre eventual incidência de direitos autorais podem ter o mesmo resultado que proibir a reutilização dos dados.

3.1 Direito autoral: breve histórico, conceituação, conteúdo e objeto, princípios

Há um certo consenso de que foi com Gutenberg e a descoberta da imprensa, no século XV, que surge a questão da proteção jurídica do direito autoral. Todavia, a consolidação dessa proteção só ocorreu no início do século XVIII, após um longo arco de tempo em que editores recebiam privilégios de monarcas para exploração econômica da obra escrita por determinado prazo, obtendo ganhos pecuniários mediante a apropriação dos direitos de reprodução, distribuição e utilização dos autores. Em reação a esse monopólio, os autores desencadearam um processo reivindicatório de seus direitos que culminou com o Copyright Act, de 1710, ato legislativo editado pela rainha Anne, da Inglaterra, considerado o marco do estabelecimento dos direitos autorais como prerrogativa dos autores. Nos EUA, a norma foi inserida na Constituição de 1783 e, posteriormente, em lei própria, o Federal Copyright Act, de 1790. A consolidação do direito autoral também ganhou força na França, especialmente em razão da Revolução Francesa, em 1789, que acrescentou ao conceito inglês

– mais voltado ao aspecto patrimonial – a primazia do autor sobre a obra (o *droit d’auteur*). Diferente da inglesa, a doutrina francesa enfatizava aspectos morais como o ineditismo, a paternidade e a integridade da obra, considerados inalienáveis e irrenunciáveis. Em 1793, a lei francesa inseria o direito de exclusividade do autor sobre a utilização econômica de sua obra, inclusive a sucessão desses direitos a herdeiros e cessionários por dez anos após a morte do autor (GANDELMAN, 2007; BITTAR, 2008).

De acordo com Gandelman (2007, p. 29), “essa formatação legal, de origem inglesa e francesa, em grande parte perdura até os dias atuais, abrangendo [...] a proteção de textos em geral [...], software de computadores e [...] as publicações digitalizadas (portais e *sites*)”.

Segundo Bittar, os EUA e a Inglaterra desenvolveram um sistema legislativo sobre direito autoral que esse autor chama de “comercial”, mais voltado para a proteção da obra do que do autor – daí *copyright*, direito de controle sobre a cópia da obra, que exige registro para ser exercido. Por outro lado, com influência doutrinária francesa, a Europa adotou o sistema da Convenção de Berna (1886), que Bittar denomina “individual”, de caráter subjetivo, dirigido à proteção do autor, centrado na exclusividade que lhe é outorgada, permitindo diversas possibilidades de utilização econômica, independente de registro de autoria da obra (BITTAR, 2008).

No Brasil, a primeira disposição legal a respeito de direito do autor deu-se em lei de 1827, que instituiu os cursos jurídicos no país. A edição de lei específica em nosso país sobre a matéria encontrava dificuldade na influência doutrinária francesa, a qual dispunha que ideias não poderiam ser objeto de propriedade – entendimento que permanece até hoje nas diversas legislações nacionais. Em 1898, com amparo em artigo da primeira Constituição republicana brasileira, de 1891, que tratava de direitos do autor, surge a lei nº 496/1898³⁸, a denominada Lei Medeiros Albuquerque, que definia o direito autoral sobre obras literárias, científicas e artísticas. Vale mencionar, ainda, a citação a direito de autor na última Constituição brasileira (BRASIL, 1988), no capítulo destinado aos Direitos Fundamentais do Cidadão, a lei nº 5.988/1973³⁹, revogada, e a que lhe sucedeu – a lei nº 9.610/1998 (BRASIL, 1998) – e que se encontra em vigor atualmente em nosso país (GANDELMAN, 2007; BITTAR, 2008).

³⁸ Ver em: <<https://www2.camara.leg.br/legin/fed/lei/1824-1899/lei-496-1-agosto-1898-540039-publicacaooriginal-39820-pl.html#:~:text=1%C2%BA%20Os%20direitos%20de%20autor,ou%20de%20qualquer%20outro%20modo>>

³⁹ Ver em: <http://www.planalto.gov.br/ccivil_03/leis/L5988compilado.htm>

Em busca de uma conceituação, Bittar (2008, p. 8) declara que “o Direito de Autor ou o Direito Autoral [...] regula as relações jurídicas, advindas da criação e da utilização econômica de obras intelectuais [...] compreendidas na literatura, nas artes e nas ciências”, enquanto a Organização Mundial de Propriedade Intelectual – OMPI (World Intellectual Property Organization – WIPO⁴⁰), dispõe que “é um termo legal usado para descrever direitos que criadores têm sobre suas obras literárias e artísticas” (OMPI, 2020, *on line*, tradução nossa). A Convenção de Berna, sobre a qual nos estendemos mais adiante, deixa claro, em seu artigo 2, item 1, que “a expressão ‘obras literárias e artísticas’ inclui toda a produção no domínio literário, *científico* e artístico, qualquer que seja o modo ou a forma de expressão [...]” (CONVENÇÃO DE BERNA..., 1886, *on line*, tradução nossa, grifo nosso).

Gandelman (2007, p. 33) acentua os dois aspectos fundamentais que compõem o conteúdo do direito autoral

[...] o [direito] *moral*, que garante ao criador o controle à menção de seu nome na divulgação de sua obra e o respeito à sua integridade, além dos direitos de modificá-la ou retirá-la de circulação; e o [direito] *patrimonial*, que visa regular as relações jurídicas da utilização econômica das obras intelectuais.

Os direitos morais são inalienáveis e irrenunciáveis, dizem respeito à paternidade e integridade da obra, “compreendidos como sendo direitos pessoais do autor” (SOUZA, 2013, p. 6), enquanto os direitos patrimoniais são transferíveis total ou parcialmente. Polêmico por sua característica imaterial, o direito autoral, um tipo de propriedade intelectual, ora é definido como direito de propriedade, ora como direito da personalidade. Todavia, possui princípios que o instruem e o distinguem (Quadro 6).

⁴⁰ Organização criada em 1967, cuja missão declarada em seu *site* é “liderar o desenvolvimento de um sistema de propriedade intelectual internacional (PI) equilibrado e eficaz que permita inovação e criatividade para o benefício de todos”. Ver em: <<https://www.wipo.int/about-wipo/en/>>

Quadro 6. Lista não exaustiva de princípios que informam o direito autoral

Esta lista não exaustiva destaca alguns dos princípios essenciais que caracterizam o direito autoral:

- a proteção refere-se à expressão de uma ideia, e não à ideia em si, que é entendida como um bem comum;
- os fatos ou eventos descritos em uma expressão criativa (em um livro ou jornal) não são protegidos por direitos autorais;
- a criação deve estar inserida em suporte material, a fim de fazer jus à proteção legal, à exceção da comunicação oral (aula, palestra);
- a proteção de direitos autorais é automática, não exigindo nenhum registro formal;
- a qualidade da expressão intelectual (obra) não constitui critério para obtenção de proteção legal;
- a originalidade e a criatividade, que correspondem à necessidade de haver na expressão da ideia componentes individualizadores e traços próprios, que não a confundam com outra já existente;
- o direito autoral é limitado no tempo. A convenção de Berna sugere como prazo mínimo de proteção a direitos patrimoniais a vida do autor mais 50 anos, após a sua morte. No Brasil, a lei prevê 70 anos (BRASIL, 1998). Findo esse prazo, a obra cai em domínio público.
- o domínio público ocorre quando, passado o prazo de proteção, os direitos exclusivos deixam de existir e a obra pode ser utilizada livremente pela coletividade, sem necessidade de autorização ou qualquer tipo de pagamento;
- sem prévia e expressa autorização do detentor dos direitos autorais, qualquer utilização da obra é ilegal, salvo as exceções em lei.
- as limitações e exceções à exclusividade do direito autoral.

Fonte: GANDELMAN, 2007; BITTAR, 2008; CARROLL, 2015; OCDE, 2015b; DOLDIRINA et al., 2018.

Os direitos concedidos pelas leis de direitos autorais em nível internacional são bastante amplos. Para fins deste estudo, são relevantes o *direito de reprodução* da obra de qualquer maneira ou forma; o *direito de distribuição de cópias* da obra mediante venda ou outra transferência de propriedade (artigos científicos, bancos de dados); o *direito de comunicação da obra ao público* (fazer o *upload* de uma obra – um artigo científico ou um banco de dados – e oferecê-la para *download* a usuários da internet é um ato de disponibilização ao público, que precisa ser autorizado pelo titular do direito) (OCDE, 2015b).

3.1.1 A face internacional do direito autoral

Em sua evolução histórica, ressalta o caráter internacionalizado do direito autoral, expresso nos tratados internacionais. Segundo Gandelman (2007), a explosão tecnológica dos meios de comunicação do mundo moderno e a obsessão por imprimir uma uniformização do direito autoral nas diversas legislações nacionais deram origem aos tratados internacionais. Seu objetivo é conferir aos autores e titulares de direitos a mesma proteção legal que cada país garante a seu autor ou titular nacional – o princípio da reciprocidade.

Três tratados internacionais destacam-se: a Convenção de Berna (1886)⁴¹, revista pela última vez em Paris, em 1971, que estruturou os princípios básicos dos direitos autorais e influenciou o direito interno de inúmeros países, transformando-se no instrumento internacional de referência para proteção de direitos autorais; o Acordo sobre Aspectos dos Direitos de Propriedade Intelectual Relacionados ao Comércio (Agreement on Trade-Related Aspects of Intellectual Property Rights)⁴², conhecido como Acordo Trips, promovido pela OMC, realizado em Montevideu, em 1994, no âmbito do Acordo Geral de Comércio e Tarifas (General Agreement on Tariffs and Trade - GATT), que eleva os padrões do nível mínimo de proteção, além de estender a proteção a todos os campos tecnológicos, inclusive aos novos possíveis campos (GANDELMAN, 2007; BITTAR, 2008; PERALTA; SILVA; TERUYA,

⁴¹ Convenção de Berna para a Proteção de Obras Literárias e Artísticas, de 9 de setembro de 1886 (a seguir denominada Convenção de Berna), concluída em Paris em 4 de maio de 1896, revisada em Berlim em 13 de novembro de 1908, com revisão concluída em Berna, em 20 de março de 1914; revisada em Roma em 2 de junho de 1928, em Bruxelas em 26 de junho de 1948, em Estocolmo em 14 de julho de 1967 e em Paris em 24 de julho de 1971, e emendada em 28 de setembro de 1979. Ver em: <http://www.wipo.int/treaties/en/text.jsp?file_id=283698>.

⁴² Ver em: <<https://www.wipo.int/edocs/lexdocs/laws/pt/br/br039pt.pdf>> e <<http://www.mdic.gov.br/index.php/comercio-externo/negociacoes-internacionais/1885-omc-acordos-da-omc>>

2011; ALBAGLI; MACIEL, 2012); e, na esteira do Acordo Trips, o Tratado de Direitos Autorais da OMPI (WIPO Copyright Treaty - WCT), realizado em 1996, conhecido pela sigla WCT (OMPI, 1996), um acordo especial da Convenção de Berna, que trata da proteção de obras e dos direitos de seus autores no ambiente digital, em resposta ao desenvolvimento das tecnologias da informação e comunicação. Entrou em vigor em 2002 e estabelece a proteção por direitos autorais para programas de computador e bases de dados; confere aos autores o direito de distribuição, de locação e um direito mais amplo de comunicação ao público; e incorpora o chamado “teste de três etapas”, que veremos adiante, para criação de limitações e exceções aos direitos autorais. O Acordo Trips e o Tratado WTC “contribuíram para expandir a estrutura de Berna para ‘novos’ temas, como software e bancos de dados, entre outros, e novas formas de proteção, como medidas de proteção tecnológica [Technological Protection Measure - TPM]” (GUIBAULT; WIEBE, 2013, p. 19, tradução nossa).

A Convenção de Berna e o Acordo Trips foram incorporados à legislação brasileira (GANDELMAN, 2007). Quanto ao WCT, não constava até agosto de 2020 nem ratificação nem adesão na página de tratados administrados pela OMPI relativa ao Brasil⁴³.

Já em meados da década de 1990, ao comentar sobre proteção legal na era digital, Negroponte afirmou que, em tese, a “lei do direito autoral está totalmente ultrapassada”, pois era como um “artefato gutenberguiano”, que provavelmente sucumbisse “inteiramente antes que se possa corrigi-la” (NEGROPONTE, 1995, p. 57, apud GANDELMAN, 2007, p. 139). Negroponte chamava a atenção, naquela época, sobre a facilidade de se fazer cópias digitais de obras intelectuais, bem como a dificuldade de se exercer controle sobre essa prática.

Em 2012, com um olhar na ciência e nas colisões que ela faz com os direitos autorais, Reichman e Okedije (2012, p. 1.369, tradução nossa) afirmavam que,

[...] desde a década de 1990, tem havido uma extensão sem precedentes da lei de direitos autorais e direitos relacionados, protegendo tanto a literatura quanto as coleções de dados no domínio da ciência básica, sem exceções adequadas para a pesquisa como tal.

David (2004) confirma Negroponte e Reichman e Okedije. Em um artigo em que defende a necessidade de se pensar uma variedade de medidas corretivas para restaurar o

⁴³ Ver em: <https://www.wipo.int/treaties/en/ShowResults.jsp?country_id=23C>

equilíbrio adequado aos direitos de propriedade intelectual (o que inclui os direitos autorais), ele afirma que há

[...] uma “concatenação de forças que têm impulsionado a privatização do domínio público de dados e informações, e os rumos recentes da evolução do regime de direitos da propriedade intelectual [...] para construir proteções compensatórias para a busca do conhecimento por meio da “ciência aberta”. (DAVID, 2004, p. 19).

3.1.2 Limitações e exceções ao direito autoral

As limitações excluem a incidência do direito autoral em algumas utilizações de obras literárias, artísticas ou científicas protegidas, abrindo exceções ao princípio monopolístico conferido ao detentor de direitos autorais, para atender a interesses de ordem pública, como o direito à informação, incluindo a científica, o acesso à cultura e ao ensino isentos de autorização (BITTAR, 2008).

Em geral, os países estabelecem uma lista de usos permitidos por lei, mas essas listas variam e os usos listados podem com frequência ser descritos de forma bastante específica e restrita. O Reino Unido estabeleceu uma nova exceção de direitos autorais que permite que os pesquisadores façam cópias de qualquer material protegido por direitos autorais para fins de análise computacional (mineração de texto e dados), sem precisar de permissão do detentor dos direitos autorais, porque o Parlamento não tinha certeza se as limitações e exceções existentes permitiriam a cópia para fins de mineração de texto e dados (INTELLECTUAL PROPERTY OFFICE, 2014; CARROLL, 2015, *on line*; OECD, 2015b).

Como as limitações e exceções não são harmônicas em nível internacional, aduz Carroll (2015, *on line*, tradução nossa), “a liberdade de usar a camada protegida por direitos autorais de um conjunto de dados – por exemplo, copiando todo o conjunto – sem permissão depende do país em que a cópia ocorre”.

Apesar de Doldirina et al. (2018, p. 15, tradução nossa) admitirem que limitações a direitos autorais e direitos *sui generis* de bases de dados “sejam relevantes para muitos usos de dados de pesquisa”, eles acreditam que, pelo fato de sua aplicabilidade ser determinada caso a caso, elas “não representam uma solução juridicamente adequada para implementar os

amplios requisitos de reutilização e interoperabilidade para dados públicos de pesquisa em geral”. Por outro lado, Carroll (2015) não faz qualquer comentário contrário à sua adoção.

Conforme Reichman e Okedije (2012, p. 1.378, tradução nossa), “as primeiras limitações e exceções que se referiam à ciência foram espremidas no artigo 10 (2) da Convenção de Berna (1886)”. Até 1948 permitia-se aos Estados fornecerem “Excertos de obras literárias ou artísticas em publicações educacionais ou *científicas*” (REICHMAN; OKEDIJE, 2012, p. 1.378, tradução nossa, grifo nosso), desde que a inclusão se justificasse por sua finalidade. Em 1967, em uma de suas revisões, a Convenção de Berna (1886) incorporou um direito exclusivo de reprodução junto com um “teste de três passos” para permitir exceções, que confinou a legislação nacional sobre exceções ao direito de reprodução a três requisitos indefinidos: 1) “certos casos especiais”, que 2) “não entravam em conflito com a exploração normal da obra” e que 3) “não prejudicavam injustificadamente os interesses legítimos do autor” (REICHMAN; OKEDIJE, 2012, p. 1.379, tradução nossa).

Segundo os dois autores, a referência expressa à ciência foi excluída. Uma versão truncada do artigo 10 (2), referindo-se apenas a trechos de ensino, foi incorporada à Convenção de Berna em 1971. Desde então, “a decisão de lidar com a exceção principal para ciências ao abrigo do ‘teste dos três passos’, e não por meio de uma disposição separada, introduziu novos níveis de incerteza sobre o escopo das atividades científicas lícitas” (REICHMAN; OKEDIJE, 2012, p. 1.380), pois não há uma disposição específica que favoreça a pesquisa científica. O “teste dos três passos” foi, em seguida, incorporado ao WCT (OMPI, 1996) para determinação de extensão por interpretação das já existentes ou criação de novas limitações e exceções em ambiente digital⁴⁴ (CABRERA BLÁZQUEZ et al., 2017).

3.1.3 Autoria e propriedade de dados de pesquisa

O direito autoral concede ao autor, ou aos autores, de uma obra original os direitos exclusivos de distribuir cópias da obra publicamente, exibi-la e executá-la publicamente ou, de outra forma, comunicar a obra ao público e fazer adaptações nela. Mas quem detém esses direitos? É uma questão que deve ser bem respondida porque se torna relevante quando é necessário, por exemplo, obter permissão para fazer o uso pretendido dos dados de pesquisa de terceiros.

⁴⁴ Ver em: <https://www.wipo.int/treaties/en/ip/wct/summary_wct.html>

A princípio, os direitos autorais pertencem ao autor, ou aos autores, de uma obra protegida e tipificada em leis do direito autoral. De acordo com Carroll, “para fins de direitos autorais, o autor é a pessoa [ou pessoas] que toma as decisões criativas ou editoriais sobre como expressar os fatos e ideias subjacentes”. Esta, contudo, é uma versão de autoria mais restrita do que aquela aplicada na ciência, cuja lacuna entre ambas pode ser identificada na forma como a autoria é estabelecida para uma publicação científica (CARROLL, 2015, *on line*).

Ao descrever a forma de comunicar ciência na física experimental de altas energias, que pode envolver milhares de colaboradores em torno de um mesmo projeto científico, Knorr-Cetina (1999) nos informa que os cientistas dessa área se baseiam no comunitarismo, cujo primeiro mecanismo de comunicação é “a eliminação do físico individual enquanto autor, através de publicações estritamente colectivas em ordem alfabética”, pois reconhecem que os resultados surgem do esforço da equipe (KNORR-CETINA, 1999, p. 389). Como em qualquer campo da ciência, a física e os físicos desenvolveram convenções sobre quem será listado como autor e em que ordem, sinalizando esse reconhecimento para a comunidade em geral.

No entanto, para fins de direitos autorais, apenas os membros da equipe que se expressaram escrevendo as palavras, desenhando as figuras ou, de outra forma, criando uma expressão original são autores com direitos autorais (CARROLL, 2015).

Assim, se houver uma camada de direitos autorais em um conjunto de dados ou banco de dados, “o(s) proprietário(s) dos direitos autorais associados a esta camada é(são) aquele(s) que escolheu(eram) como organizar, combinar, anotar ou visualizar os dados, e não o(s) envolvido(s) em sua geração ou coleta”, diz Carroll (2015, *on line*, tradução nossa).

Contudo, quando o criador da obra protegida é um funcionário dentro do âmbito do seu emprego, em alguns países, como os EUA, o empregador é tratado como autor, e o empregado não tem direitos. A regra gera divergências quando se aplica a materiais de pesquisa e ensino criados por funcionários de uma universidade. Alguns argumentam que a regra não se aplica aos resultados da pesquisa. No resto do mundo, o(s) criador(es) individual(is) inicia(m) como proprietário(s) dos direitos, mas eles podem ser automaticamente transferidos para o empregador se o contrato de trabalho assim estipular (CARROLL, 2015).

Wallis e Borgman (2012) publicaram os resultados de uma pesquisa realizada entre 2002 e 2006, com 42 cientistas do Center for Embedded Networked Sensing (CENS), órgão da National Science Foundation, nos EUA. Preocupadas com a definição de responsabilidade pelos dados que cada vez mais são depositados em repositórios, isolados ou em conjunto com o depósito de artigos, as duas pesquisadoras partiram de uma análise de Foucault (1969) sobre a função do autor e perguntaram aos cientistas do CENS se eles se consideravam os autores de um conjunto de dados e quais critérios determinavam a autoria dos dados. O resultado traz relevantes contribuições sobre o tema, embora não tenha atingido o ponto perquirido, deixando clara a necessidade de pesquisas adicionais em razão do entendimento pouco claro dos participantes tanto sobre responsabilidade pelos dados quanto sobre o que lhes determina a autoria. Grande parte das respostas dava como autores de dados os colaboradores intelectuais, o próprio equipamento usado para coletar dados, o pesquisador principal, pois era ele quem trazia financiamento, ninguém e a instituição que apoiou a coleta de dados, que também era encarada como critério de determinação de autoria.

3.2 A aplicabilidade do direito autoral em dados de pesquisa

Os direitos que podem ser aplicáveis a dados de pesquisa são: segredos industriais, direitos autorais (foco deste estudo), que incluem direitos sobre bancos de dados e programas de computador, e direitos de patentes para algumas formas de dados. Aqueles que buscam reutilizar dados podem se deparar também com direitos protegidos por contratos, cujas cláusulas poderão conter inclusive alguma restrição de uso (CARROL, 2015). Além desses direitos, são protegidos os direitos à privacidade, sobretudo na pesquisa em saúde, com destaque para a proteção aos dados pessoais, particularmente aos dados sensíveis da pessoa, mais sujeitos a usos discriminatórios ou lesivos, pois dizem respeito à origem racial ou étnica, convicção religiosa, opinião política, filiação a sindicato ou a organização religiosa, filosófica ou política, à saúde ou à vida sexual, a dado genético ou biométrico do titular do dado (GUANAES et al., 2018; DONEDA, 2019, p. 142-143).

No Brasil, o domínio das ciências está contemplado no art. 7, § 3º da lei n. 9.610/1998 (Lei dos Direitos Autorais brasileira). Nele o legislador assegura que “a proteção [do direito autoral] recairá sobre a *forma literária* ou artística, não abrangendo o seu *conteúdo científico* ou técnico, sem prejuízo dos direitos que protegem os demais campos da propriedade imaterial” (BRASIL, 1998, *on line*, grifos nossos).

Embora seja nova, a temática sobre direitos autorais na perspectiva da abertura de dados de pesquisa subjacentes a artigos, em particular, e sobre propriedade intelectual, em geral, já vem sendo abordada e desenvolvida por editores e especialistas desse campo.

Em uma política de periódicos sobre dados de pesquisa subjacentes, os editores têm o importante papel de orientar os autores na adoção de uma licença para dispor seus dados abertamente em um repositório, mas também de fornecer informações basilares sobre direitos de propriedade intelectual que, eventualmente, possam incidir sobre os dados a serem disponibilizados e eventualmente reutilizados.

Fiel ao escopo de concentrar este estudo no campo dos direitos autorais relacionados a dados de pesquisa, a seguir, mapeamos e dialogamos com três trabalhos que se preocupam em apontar dificuldades, contradições e, ao mesmo tempo, em sugerir recomendações que pavimentem o arcabouço legal da abertura de dados de pesquisa.

Em primeiro lugar, nesta perspectiva, Hrynaszkiewicz e Cockerill (2012), pesquisadores a serviço do BiomedCentral (BMC), plataforma pioneira na publicação em acesso aberto, que conta com cerca de 300 periódicos revisados por pares em acesso aberto, relataram a implementação, em 2013, de uma nova abordagem jurídica relacionada à abertura de artigos e dados de pesquisa da qual participaram ativamente, mencionada na seção 2.6 deste estudo.

Normalmente, os autores aplicam a licença de atribuição CC BY para artigos, a mais aberta entre as seis da Creative Commons (Quadro 7). No entanto, com frequência, contrariando os termos jurídicos da licença, periódicos de “acesso aberto” impõem restrições à reutilização comercial de artigos publicados e à criação de trabalhos derivados, incluindo a mineração de texto em algumas ou até mesmo em quaisquer jurisdições, maculando o espírito do acesso aberto (MURRAY-RUST; MOLLOY; CABELL, 2014, p. 20-23).

No que se refere aos dados de pesquisa, os direitos autorais e o licenciamento são complexos em nível internacional, pois apresentam barreiras legais ao compartilhamento, integração e reutilização de dados, o que acarreta a publicação sob licenças nem sempre apropriadas, restringindo “a transferência e a descoberta mais eficientes de conhecimento científico” (HRYNASZKIEWICZ; COCKERILL, 2012, *on line*, tradução nossa).

Com base nestas questões levantadas, entre outras, Hrynaszkiewicz e Cockerill (2012) reportaram o processo com que o BMC instaurou o “modo acesso aberto *por padrão*” para

seus artigos e dados de pesquisa. Quanto aos artigos, o BMC exige que o autor adote, “por padrão”, a licença Atribuição CC BY em sua plenitude, oferecendo acesso aberto irrestrito ao conteúdo e concedendo os direitos de distribuir, adaptar e criar a partir do trabalho, sem restrições para fins comerciais, desde que seja atribuído o crédito pela criação original.

Em relação aos dados de pesquisa, em casos como a subsistência de direitos autorais de dados digitais derivados e integrados de diferentes fontes, diferenças legais internacionais, “empilhamento” de licenças de escopos distintos em conjuntos de dados que possivelmente resultarão em novo conjunto, a CC BY pode não funcionar, pois o detentor ainda tem o direito de impor algum nível de restrição, como proibir pesquisas futuras em grandes coleções de conteúdo, o que pode acontecer na pesquisa de mineração de texto e dados. Desta forma, para evitar entraves à integração e reutilização e para ajudar a permitir a interoperabilidade de longo prazo, o BMC adota e recomenda o instrumento de renúncia CC0 1.0 Dedicção ao Domínio Público, da Creative Commons (Quadro 7), que permite a reutilização de dados sem a necessidade de permissão especial de seus criadores originais, que renunciam aos direitos autorais e direitos relacionados aos dados publicados (HRYNASZKIEWICZ; COCKERILL, 2012).

Preocupa a comunidade de pesquisadores, no entanto, o fato de o instrumento de renúncia CC0, ao contrário das seis licenças Creative Commons, não obrigar a quem reutiliza dados fazer a devida citação. Para esta dificuldade, os dois autores ressaltam que as próprias normas da comunidade científica oferecem a solução: a prática de não citação no meio científico é, em geral, identificada e resolvida por meio da revisão por pares, da ética e das políticas editoriais (HRYNASZKIEWICZ; COCKERILL, 2012, *on line*).

Assim, após um processo de readaptação, todos os artigos em acesso aberto do BMC, a partir de 2013, após consulta pública e alterações em seus contratos padrão de direitos autorais, ostentam a seguinte declaração de licença:

Este artigo está licenciado sob uma Licença Internacional Creative Commons Atribuição 4.0, que permite o uso, compartilhamento, adaptação, distribuição e reprodução em qualquer meio ou formato, desde que você dê o devido crédito ao(s) autor(es) original(is) e à fonte, forneça um *link* para a licença Creative Commons e indique se foram feitas alterações. As imagens ou outro material de terceiros neste artigo estão incluídos na licença Creative Commons do artigo, a menos que indicado de outra forma em uma linha de crédito para o material. Se o material não estiver incluído na licença Creative Commons do artigo e seu uso pretendido não for permitido por regulamentação legal ou

exceder o uso permitido, você precisará obter permissão diretamente do detentor dos direitos autorais. Para ver uma cópia desta licença, visite <<http://creativecommons.org/licenses/by/4.0/>>. *O instrumento de renúncia Dedicção ao Domínio Público, da Creative Commons (<http://creativecommons.org/publicdomain/zero/1.0/>), aplica-se aos dados disponibilizados neste artigo, a menos que de outra forma indicado em uma linha de crédito aos dados (grifos nossos).*⁴⁵

Em segundo lugar, em mais um trabalho que atinge a problematização central deste estudo – “Como as revistas científicas, particularmente as da área da saúde, estão lidando com a questão da abertura de dados de pesquisa subjacentes aos artigos que publicam em face do direito autoral?” –, Michael Carroll (2015), professor de Direito do Washington College of Law, diretor do Programa de Justiça da Informação e Propriedade Intelectual e coordenador da Creative Commons nos EUA, contribui com um panorama amplo sobre como o depósito de dados em conexão com um artigo publicado pode provocar questões de direito autoral e outros tipos de proteção legal para pesquisadores, seus empregadores, financiadores e outros pesquisadores que reutilizam dados de pesquisa.

Carroll dedica-se a examinar também a polêmica questão sobre a propriedade dos dados (AL-KHOURI, 2012; BORGMAN, 2012; KOSTOVA, 2016) e a diferença entre autor e detentor de direitos (ver mais a respeito na subseção 3.1.3). Além disso, excluindo os direitos que protegem a privacidade e a segurança nacional, que eventualmente podem ser reivindicados, Carroll (2015) lista e analisa os diferentes tipos de direitos legais para dados: segredo comercial, patente, direito *sui generis* de bases de dados, direito autoral, este o regime de interesse deste trabalho, e contratos, que se sobrepõem a todos esses direitos, mas também podem restringir o reuso simplesmente com uma cláusula ajustada entre o detentor de direitos autorais e o pesquisador usuário.

Os direitos autorais, segundo Carroll (2015), se aplicam ao compartilhamento de dados de pesquisa na medida em que “o criador, o proprietário ou o repositório que abriga os dados queiram limitar direitos como a cópia, a distribuição ou outros reusos de dados”, concedidos ao autor de uma obra original literária, artística ou científica. Tais direitos garantem ao autor a atribuição da autoria, característica que mais interessa ao contexto de reuso de dados, segundo Carroll. Se uma fonte de dados – o pesquisador ou repositório – dá

⁴⁵ Ver a declaração original de licença em: <<https://bmcbiol.biomedcentral.com/articles/10.1186/s12915-020-00797-1#rightslink>>

permissão para reutilização dos dados e o uso pretendido se encaixa nos termos da permissão, não há com o que se preocupar, pois o usuário terá uma base legal para protegê-lo.

Mas nem sempre. Carroll (2015) exemplifica. Se alguém for buscar dados no Laboratório Europeu de Biologia Molecular-Instituto Europeu de Bioinformática (European Molecular Biology Laboratory-European Bioinformatics Institute - EMBL-EBI), encontrará nos termos de uso que “[os] bancos de dados públicos do EMBL-EBI estão disponíveis gratuitamente para qualquer indivíduo e para qualquer finalidade”⁴⁶ (tradução nossa).

O que pareceria uma permissão para copiar e reutilizar os dados à vontade logo se transforma em uma incerteza jurídica, quando os mesmos termos do Instituto alertam o usuário de que terceiros podem reivindicar direitos autorais ou outros direitos de propriedade intelectual sobre os dados originais, cabendo ao pesquisador não infringir esses direitos. Isto ocorre por dois motivos: primeiro, a fonte dos direitos autorais é a legislação nacional. Certos tratados internacionais harmonizam os direitos dos proprietários de direitos autorais, mas permitem que os direitos dos usuários variem de país para país; segundo, certos países adicionam tipos de proteção além do que os tratados estabelecem, como a criação de um direito de bancos de dados específico [direito *sui generis*], aplicável somente dentro de suas fronteiras (CARROLL, 2015, *on line*).

Carroll advoga a adoção das licenças Creative Commons e a ferramenta CC0 Dedicção ao Domínio Público para reuso dos dados, considerando esta como a principal ferramenta para afastar todas as possibilidades legais de impedimento ao reuso.

O autor estadunidense destaca ainda um aspecto de vivo interesse da ciência: os direitos autorais não impõem restrições ao compartilhamento de dois blocos de construção básicos do conhecimento – fatos e ideias –, que pertencem ao domínio público, afirma Carroll (2015). Os pesquisadores rotineiramente contam com essa liberdade de copiar e parafrasear em sua prática diária, fornecendo o devido crédito ao autor original. “Da mesma forma”, segundo o autor, “dados observacionais e experimentais [ver a respeito na seção 2.4] brutos são ‘fatos’, para fins de direitos autorais, que podem ser compartilhados e reutilizados sem restrição” desse regime jurídico (CARROLL, 2015, *on line*).

Em terceiro lugar, os autores do estudo “Abordagens legais para acesso aberto a dados de pesquisa”, Doldirina, Eisenstadt, Onsrud e Uhler (2018), defendem que existe uma

⁴⁶ Ver em: <<http://www.ebi.ac.uk/about/terms-of-use/>

tendência global para adoção de políticas de dados abertos. Revelam que um “grande impulso para esse aumento no compartilhamento internacional de dados de pesquisa é o seu papel crítico na abordagem dos desafios globais” (p. 4-5). Exemplos disso são as situações de emergência em saúde pública, como no caso do vírus Zika (JORGE, 2018), declarada em 2016, e no da pandemia de Covid-19⁴⁷ (Corona Virus Disease 2019), em 2020. Em ambos os casos, a disponibilidade aberta e rápida de informações e dados de pesquisa para compartilhamento, bem como o rompimento temporário de formalidades acadêmicas (PRIMEZ; BRANDÃO; MOMEN, 2016) são fatores fundamentais para o combate à pandemia⁴⁸. Salvam vidas. São causas humanitárias, de combate a doenças infecciosas, em que caem também as barreiras do direito autoral, pelo menos enquanto a crise durar (WELLCOME TRUST, 2016; THE LANCET, 2020).

No entanto, Doldirina e colaboradores (2018) ressaltam que, sobretudo em tempos com ausência de emergências públicas, para efetivar todo o potencial de compartilhamento de dados de forma livre e aberta, é necessário superar barreiras legais, além das técnicas, de gerenciamento institucionais, econômicas ou socioculturais, assertiva corroborada por Borgman (2015, p. 75, tradução nossa), que acrescenta o entrelaçamento de “leis, políticas e práticas aplicáveis à propriedade, controle e liberação de dados [que] variam de acordo com a agência de fomento, governo e jurisdição”.

A par disso, conforme Doldirina e colaboradores (2018, p. 6, tradução nossa), pode haver ainda restrições legítimas ao acesso e uso de dados gerados publicamente, como “a proteção da segurança nacional, segurança pública, privacidade pessoal, confidencialidade, interesses proprietários de exploração comercial no setor privado e termos definidos de uso exclusivo pelo pesquisador”, com o que concorda Carroll (2015). Esses autores referem ainda a possibilidade de retenção de dados quando sua liberação puder provocar prejuízos a espécies ameaçadas ou artefatos arqueológicos, ou à proteção de direitos de povos indígenas.

Com base nessa premissa, Doldirina e colaboradores (2018, p. 12) advogam que “é altamente desejável que os conjuntos de dados produzidos por ou usados em pesquisa pública tenham termos e condições de reutilização irrestritos”, o que afastaria a incerteza para

⁴⁷ Ver em: <https://www.paho.org/bra/index.php?option=com_content&view=article&id=6101:folha-informativa-novo-coronavirus-2019-ncov&Itemid=875>

⁴⁸ A Fundação Oswaldo disponibilizou informações em âmbito nacional e internacional a pesquisadores sobre o Covid-19: <<https://portal.fiocruz.br/coronavirus-2019-ncov-informacoes-para-pesquisadores>>.

reutilizadores e beneficiaria a atividade de pesquisa. No entanto, os dados produzidos ou utilizados em pesquisa podem estar sujeitos a direitos autorais (e outros direitos de propriedade intelectual, fora do raio deste estudo), regulados por legislações nacionais desconhecidas.

Embora ofereça uma estrutura internacional comum de padrões mínimos para a proteção de obras literárias, artísticas e científicas e seja reconhecida como uma fonte que codifica princípios fundamentais de proteção de direitos autorais, a Convenção de Berna (1886) não cria proteção legalmente aplicável a esses direitos. Não existe um regime internacional de proteção automática a direitos de conteúdo criativo de forma idêntica em todo o mundo. Portanto, a proteção nacional de direitos autorais varia entre jurisdições e depende das leis de cada país e da interpretação dada por tribunais nacionais (DOLDRINA et al., 2018).

Em consequência, as normas internacionais convencionadas e relacionadas aos princípios de direitos autorais, alguns deles mencionados no Quadro 6, geraram um conjunto de requisitos mínimos em todo o mundo, em quase todos os países. Porém, ao serem implementadas, segundo Doldirina e colaboradores (2018), geraram também áreas de diferença entre as diversas legislações nacionais, já que os critérios de proteção apresentados na Convenção de Berna (1886) podem ser interpretados de modo diferente, conforme a jurisdição.

A primeira área diz respeito à criatividade, cujo limite mínimo, para determinação de existência de direito autoral em uma obra, não é interpretado de forma homogênea. Nos EUA, os direitos autorais protegem apenas a expressão, não se estendendo a ideias, enquanto na Alemanha o direito autoral protege o autor em suas relações intelectuais e pessoais com a obra e o uso que faz dela. A segunda área é a variação do prazo de proteção. A Convenção estabelece que é o tempo de vida do autor mais 50 anos, mas nos Estados Unidos, por exemplo, obras contratadas têm prazo de 95 anos de proteção e são relevantes para a produção de dados comerciais e não comerciais. No Brasil, é o tempo de vida do autor mais 70 anos após sua morte. Tais diferenças podem acarretar situações em que a mesma obra ainda esteja protegida em um estado, mas já seja de domínio público em outro. A terceira, que aumenta o risco de proteção diferente do mesmo objeto, é a introdução de novas categorias de propriedade intelectual, como o direito *sui generis* de bases de dados da União Europeia, que pode criar barreiras para intercâmbio e compartilhamento (ver subseção 3.3.2). A quarta área

que, com frequência, contém diferenças entre fontes reguladoras em variadas jurisdições e é relevante para usuários de material protegido por direitos autorais, incluindo conjunto de dados de pesquisa, é a de exceções e limitações aos direitos dos autores (ver subseção 3.1.2) (DOLDIRINA et al., 2018).

A título de exemplificação da quarta área, o principal conjunto de limitações dos EUA é o denominado *fair use*⁴⁹ (uso justo). Ele limita direitos exclusivos concedidos pela lei de direitos autorais para uso de uma obra protegida para fins de crítica, comentário, reportagem de notícias, ensino (incluindo várias cópias para uso em sala de aula), pesquisa acadêmica ou científica. A lei americana estabelece parâmetros definidores para determinação do uso justo. No Reino Unido, há uma limitação semelhante, porém menos permissiva, conhecida como *fair dealing*⁵⁰ (tratamento justo), relacionada à pesquisa e estudo privado, crítica ou revisão ou reportagem de notícias. O que a torna menos permissiva e acentua a diferença de tratamento legislativo para praticamente os mesmos assuntos, trazendo-lhe certo nível de incerteza, é que na lei inglesa não há uma definição estatutária para tratamento justo, cuja determinação é feita caso a caso (CARROLL, 2015; DOLDIRINA et al., 2018).

As diferenças apontadas suscitam questões, especialmente na internet ou em um ambiente *on line*, onde se dá a recuperação e o uso de dados na atualidade. Uma questão que pode se tornar crucial nestes dois casos é a correta identificação da lei e da respectiva jurisdição a serem aplicadas quando dados são criados em uma jurisdição e usados em outras. A regra padrão é que prevalece “a jurisdição na qual os direitos de propriedade intelectual [inclusive os direitos autorais] são criados [...] na ausência de um acordo em contrário” (DOLDIRINA et al., 2018, p. 16, tradução nossa). A situação se complica quando o agrupamento de dados origina-se de muitos usuários em diferentes jurisdições e o que é gerado é distribuído em todo o mundo. São questões, no entanto, que podem ser pacificadas em um contrato ou licença indicada pelo detentor dos direitos, segundo esses autores.

Uhlir e Clement (2016), endossados por Doldirina e colaboradores (2018), defendem, todavia, que algumas regras padrão claras devem ser estabelecidas e cumpridas pelos vários atores, a fim de evitar problemas relacionados ao que denominam de *interoperabilidade legal dos dados*, que é

⁴⁹ Ver em: <<https://www.law.cornell.edu/uscode/text/17/107>>

⁵⁰ Ver em: <<https://www.gov.uk/guidance/exceptions-to-copyright#fair-dealing>>

[...] a capacidade de combinar dados de duas ou mais fontes sem conflitos entre restrições impostas pelos provedores de dados (ou seja, o suporte a uma restrição nega inerentemente o suporte a outra) e sem a necessidade de solicitar autorização dos provedores de dados caso a caso,

a qual ocorre quando:

- as condições legais de uso são claras e prontamente determináveis para cada um dos conjuntos de dados, normalmente por meios automatizados;
- as condições legais de uso impostas em cada conjunto de dados permitem a criação e o uso de produtos combinados ou derivados; e
- os usuários podem acessar e usar legalmente cada conjunto de dados sem solicitar autorização dos criadores de dados, caso a caso, supondo que as condições de uso acumuladas para cada um e todos os conjuntos de dados sejam atendidas (DOLDIRINA et al., 2018, p. 8, tradução nossa).

3.2.1 Aplicação de direitos autorais em dados factuais

Observações relatadas por cientistas e outros pesquisadores de características físicas e naturais estão mais próximas de serem fatos descobertos do que fruto de uma criação intelectual. É o caso das medidas, contagens, descritores padrão, como sexo, idade e altura, que muitas vezes estão contidos em conjuntos de dados digitais “que indicam *o que é* ou *o que aconteceu*. Mesmo havendo discordância significativa sobre a precisão das observações, os fatos observados e descobertos não são protegidos por direitos autorais”, diz Dodirina e colaboradores (2018, p. 17, tradução nossa), com o que concorda Carroll (2015).

Contudo, a renderização estilística dos fatos, procedimento pelo qual se obtém o produto final de um processamento digital qualquer, é passível de proteção de direitos autorais. Mas ante a existência de um número limitado de formas de expressão de um conceito ou ideia e ainda da exígua distinção entre a ideia e a expressão, a *doutrina da fusão* pode ser aplicada. A doutrina impede que um autor monopolize uma ideia simplesmente com direitos autorais de algumas expressões dela. Consiste na impossibilidade de se manter “o monopólio da descoberta sob a proteção de direitos autorais de forma e dimensões”, por exemplo, “de uma célula microscópica ou de um prédio desenterrado de um sítio

arqueológico”, já que “as expressões ilustrativas não interpretativas do tamanho e da forma provavelmente são *as mesmas dos fatos*” (DOLDIRINA et al., 2018, p. 17, tradução nossa).

Fatos que foram acessados de modo legal podem ser usados, reusados ou divulgados por qualquer pessoa, para qualquer fim, sem proteção legal. Mas, para uma resposta definitiva, o resultado de uma avaliação vai depender das circunstâncias de cada caso. Por outro lado, conjuntos de dados ou coleção de fatos podem estar sujeitos à proteção de direitos autorais, mesmo sendo fatos não passíveis de proteção dos direitos autorais. Basta para isso que sua “seleção, disposição ou organização” esteja impregnada de alguma criatividade humana ou originalidade. Nestes casos, porém, “os direitos autorais corresponderão tão somente ao arranjo criativo ou seleção dos elementos de um conjunto de dados. Suas ‘partes’ não são protegidas por direitos autorais” (Doldirina et al., 2018, p. 18, tradução nossa).

Com sua importante função transportada para os dados de pesquisa, os metadados resumem as informações básicas sobre um conjunto de dados e frequentemente se organizam por meio de uma estrutura formal padrão, o que facilita muito a pesquisa e a localização de dados. As estruturas de campo de metadados padronizadas não são, em geral, cobertas pela proteção dos direitos autorais, pois são desprovidas dos ingredientes criatividade e originalidade. São comuns. Todavia, se uma seleção e coleta específicas de elementos de metadados, feitas para atender a um objetivo específico, alcançar o nível de “criatividade”, poderá justificar a proteção de direitos autorais. “Essa falta de clareza jurídica onera substancialmente fornecedores e usuários de dados”, diz Doldirina e colaboradores (2018, p. 18, tradução nossa).

3.3 A proteção legal a bancos de dados

Por mais nobre que seja a sua meta, o movimento pela abertura de dados de pesquisa financiada por recursos públicos levanta várias questões jurídicas e políticas distintas daquelas relativas à publicação de artigos científicos. Para os especialistas em direito autoral Guibault e Wiebe (2013, p. 6, tradução nossa), o acesso aberto a dados de pesquisa é “um objetivo de política relativamente novo, por isso menos atenção foi dada às características específicas dos dados de pesquisa”. Com patrocínio e encomenda do OpenAire⁵¹, projeto da Comissão Europeia para fornecer acesso aberto ilimitado e sem barreiras aos resultados de

⁵¹ Ver em: <<https://www.openaire.eu/mission-and-vision>>

pesquisa financiada por recursos públicos na Europa, esses dois pesquisadores da Universidade de Direito de Göttingen realizaram um estudo panorâmico sobre proteção legal a dados de pesquisa na Europa, no qual dedicaram especial atenção ao direito *sui generis* de bases de dados – parte que exploraremos mais do estudo de ambos –, que culmina com recomendações legais para acesso e reuso de dados de pesquisa em uma infraestrutura de acesso aberto, o OpenAIREplus⁵². Guibault e Wiebe (2013, p. 3, tradução nossa) entendem que “falta uma análise do *status* legal de tais dados, de como disponibilizá-los sob os termos de licença corretos e um entendimento completo do significado da abertura – especialmente de suas implicações legais”.

3.3.1 A proteção de bancos de dados sob a lei de direitos autorais

A proteção dos bancos de dados por direitos autorais está prevista no Tratado de Direitos Autorais da OMPI, conhecido como WCT (OMPI, 1996), que é um acordo especial da Convenção de Berna (1886), realizado em Genebra, em 1996 (ver subseção 3.1.1). Segundo Uhlir e Clement (2016, p. 35, tradução nossa), “No sentido jurídico, uma base de dados é uma compilação organizada de dados que goza de proteção legal aplicável, desde que cumpra determinados critérios que estão previstos nas legislações nacionais”.

Segundo o WCT, em seu art. 5, a proteção de direitos autorais se aplica a bases de dados (compilações de dados ou coleções de dados), “sob qualquer forma, que, em razão de seleção ou organização de seus conteúdos, constituam criações de natureza intelectual”, originais de seus autores, implicando um nível de criatividade e originalidade nessa organização. Ainda de acordo com art. 5, do WCT, “Esta proteção não cobre os dados ou materiais em si e é entendida sem prejuízo de quaisquer direitos autorais que subsistam com relação aos dados ou materiais contidos na compilação” (OMPI, 1996, tradução nossa).

Se o conteúdo do banco de dados foi simplesmente “montado” de forma metódica, é provável que não se enquadre como uma obra original, nem estará elegível para proteção de direitos autorais. Assim, a proteção se aplica à originalidade e criatividade da expressão da organização somente e não adere ao conteúdo do banco de dados. Alguns desses conteúdos,

⁵² OpenAIREplus (2ª Geração de Infraestrutura de Acesso Aberto para a Investigação na Europa) foi um projeto de 30 meses, iniciado em 2011, criado para facilitar o acesso via OpenAire a toda a produção científica em acesso aberto do Espaço Europeu de Pesquisa, o que inclui uma rede de repositórios institucionais e disciplinares, centros de dados, editores e coleções agregadas, e proporcionar estruturas técnicas para gestão e interligação de dados científicos. Ver mais informações em: <<https://www.openaire.eu/openaireplus-project>>

no entanto, podem estar sujeitos à proteção individual de direitos autorais (por exemplo, um artigo, uma foto) ou podem não ser protegidos como tal por propriedade intelectual (por exemplo, dados brutos) (EUROPEAN IP HELPDESK, 2020).

Quando é aplicável, de acordo com o European IP Help Desk (2020), órgão de consulta sobre propriedade intelectual da União Europeia, a proteção de direitos autorais proíbe qualquer reprodução, distribuição, comunicação, exibição ou *performance* não autorizadas para o usuário do banco de dados. É essencial notar que a proteção surge automaticamente para o criador do banco de dados, ou para qualquer pessoa designada pela legislação nacional como titular do direito, no momento de sua conclusão, expirando 70 anos após a morte do autor. Essa proteção tem um caráter internacional. Em consequência, as condições de proteção de direitos autorais podem variar de jurisdição para jurisdição, como já visto neste capítulo.

No Brasil, de acordo com o art. 7, item XIII, da lei n. 9.610/1998 (Lei dos Direitos Autorais), são obras intelectuais protegidas as “bases de dados [...] que, por sua seleção, organização ou disposição de seu conteúdo, constituam uma criação intelectual” (BRASIL, 1998, *on line*). Ainda em relação à base de dados, o § 2º do art. 7 dessa lei estabelece que a mencionada proteção “não abarca os dados ou materiais em si mesmos”, sem prejuízo de quaisquer direitos autorais que continuem a existir nos dados ou materiais contidos nas obras (BRASIL, 1998, *on line*).

3.3.2 A proteção de bancos de dados pelo direito *sui generis*

Na União Europeia, ao promulgarem a diretiva 96/9/CE (COMISSÃO EUROPEIA, 1996) para proteção dos bancos de dados por direitos autorais, prevista no WTC, os legisladores europeus acrescentaram um direito novo nesse tipo de proteção, complementar aos direitos autorais – o direito *sui generis*, um direito criado pelo legislador europeu para oferecer proteção a bancos de dados não criativos, que, como o direito autoral, nasce automaticamente. Alguns países (Coreia do Sul, Rússia e México) adotaram legislações similares. Assim, a diretiva europeia, posteriormente transposta para a legislação nacional dos estados-membros da União Europeia, fornece uma dupla camada de proteção para bancos de dados:

(a) direitos autorais, quando a seleção ou organização do banco de dados representa a criação intelectual do próprio autor; a proteção cobre apenas a estrutura do banco de dados, não o seu conteúdo; e

b) direito *sui generis* [de “espécie única”] do banco de dados, que protege o conteúdo do banco de dados, quando houver um investimento substancial na obtenção, apresentação ou verificação dos dados, de atos de extração (por exemplo, cópia) e reutilização (por exemplo, redistribuição, comunicação para o público etc.) do todo ou de uma parte substancial do conteúdo desse banco de dados [ou ainda de extração sistemática de partes não substanciais] (GUIBAULT; WIEBE, 2013, p. 23, tradução nossa).

Essas duas formas de proteção são cumulativas quando os pré-requisitos para ambas são atendidos. O “investimento substancial” deve ser entendido como um investimento financeiro e/ou profissional, que pode consistir na aplicação de recursos financeiros e no dispêndio de tempo, esforço e energia na obtenção e coleta dos conteúdos. Na maioria dos países não europeus, EUA e Brasil inclusive, os bancos de dados são protegidos apenas se eles (ou certas partes ou características) se qualificarem como “obras” dentro do significado do direito autoral (GUIBAULT; WIEBE, 2013; CARROLL, 2015; UHLIR; CLEMENT; 2016; EUROPEAN IP HELPDESK, 2020).

Segundo Guibault e Wiebe (2013), os bancos de dados exercem um papel central no campo da pesquisa científica e a proteção legal que lhes é dada certamente influencia as ações de cientistas e pesquisadores quanto aos dados disponíveis e a serem criados. Doldirina e colaboradores (2018, p. 21) afirmam, no entanto, que essa específica legislação de proteção de banco de dados “confere direitos estatutários adicionais aos titulares de bancos de dados, que podem ser usados para impor proteções adicionais” nas jurisdições que endossaram tal legislação, enquanto o professor Carroll (2015) assim se pronuncia sobre o direito *sui generis*:

Por mais frustrante que isso possa ser para uma comunidade de pesquisa globalizada, em uma classe restrita de casos, esse direito poderia se aplicar ao *download* de uma quantidade substancial de dados que ocorrer em um computador conectado à Internet na Europa ou Coreia do Sul, mas não em outro lugar (Carroll, 2015, *on line*).

Hugenholtz (2019), professor do Instituto de Informação Legal da Universidade Van Amsterdam e da Kluwer Law International, que reúne especialistas em legislação europeia sobre direitos autorais, ao discorrer sobre as vantagens do uso da mineração de texto e dados,

que geralmente se refere à análise baseada em computador de grandes corpos de dados para obter conhecimento na ciência, na indústria e no comércio, frisa que hoje ela é uma prática padrão em pesquisa farmacêutica, jornalismo, recuperação de informações e outras áreas. No entanto, o autor deixa claro que o direito *sui generis* de bancos de dados pode criar conflitos potenciais com a mineração de texto e dados, ao impedir a extração de parte substancial do conteúdo de um banco de dados. O direito *sui generis* é um bom exemplo da criação de um direito específico, no caso pela União Europeia, que causa divergências por não ser abrigado em grande parte da legislação de países que seguem de modo essencial as regras ditadas pela Convenção de Berna (1886) e, pior, que pode atrasar a abertura de dados de ou para pesquisa.

De fato, em um universo em disputa, em que se usam argumentos legítimos para às vezes alavancar qualquer tipo de interesse, pouca ou nenhuma atenção se dá às necessidades da ciência aberta. Senão, vejamos o que David (2004, p. 14-15, tradução nossa) tem a dizer sobre a proteção *sui generis* a bancos de dados:

A Diretiva de Bancos de Dados da UE aplica-se igualmente a bancos de dados não eletrônicos e eletrônicos, embora, [...] tenha se originado como uma resposta estratégica de “política industrial” ao desenvolvimento comercial de bancos de dados (eletrônicos) *on line* na América.

Fazendo coro às críticas à diretiva 96/9/CE (COMISSÃO EUROPEIA, 1996), Doldirina e colaboradores (2018) afirmam que, tal como se apresenta, a lei protege informações compiladas nos bancos de dados, mas também até “mesmo os meros fatos que formam mais do que uma ‘parte insubstancial’ do banco de dados”, definida quantitativa ou qualitativamente, desde que tenha havido investimento substancial na obtenção, verificação e apresentação de seu conteúdo. Com isso, a proteção vai muito além dos direitos autorais ao proteger o conteúdo do banco de dados e conferir ao titular do banco o direito de impedir a extração e/ou a reutilização do todo ou de parte substancial do conteúdo (DOLDIRINA et al., 2018, p. 20, tradução nossa).

Em abril de 2019, o Conselho da União Europeia aprovou uma nova Diretiva de Direitos Autorais (Diretiva 2016/0280). Uma das preocupações do texto é a insegurança jurídica que o direito *sui generis* provoca. A Diretiva reconhece que a mineração de texto e dados pode beneficiar a comunidade científica e a inovação, ao permitir tratamento de

grandes quantidades de dados. Mas também reconhece, no considerando (8), que “organismos e instituições [científicos] são confrontados com a insegurança jurídica por não saberem até onde podem levar a mineração de texto e dados de conteúdos digitais”. A própria Diretiva, no considerando (11), fornece a solução para afastar a insegurança jurídica, que “deverá ser eliminada através da previsão de uma exceção obrigatória para as universidades e outros órgãos de pesquisa”.⁵³

3.4 Remover restrições: Licenciamento de dados de pesquisa e domínio público

Um dos países que vem dando atenção à questão ligada aos direitos aplicáveis a dados, o que inclui os direitos autorais, tem sido a Austrália. O primeiro passo foi criar o Serviço Nacional de Dados da Austrália⁵⁴ (Australian National Data Service - ANDS), uma infraestrutura de dados de pesquisa nascida da parceria entre duas universidades, a agência nacional de pesquisa científica e um órgão de infraestrutura de pesquisa do governo que financia o ANDS. Essa infraestrutura tem uma ampla lista de parceiros⁵⁵ e é adotada por instituições internacionais, como o The Agricultural Information Management Standards Portal (AIMS), órgão da Food and Agricultural Organization (FAO)⁵⁶, da Organização das Nações Unidas (ONU).

Sua abordagem sobre direito autoral de dados está concentrada no guia “Research Data Rights Management Guide” (ANDS, 2018), que apoia os princípios FAIR (Quadro 5, capítulo 2) e é dirigido a detentores de direitos, usuários e fornecedores de dados que trabalham com direitos sobre dados de pesquisa e licenças. Seu propósito é “aumentar a conscientização sobre as implicações de não se adotar licenças em dados de pesquisa” (ANDS, 2018, p. 3, tradução nossa). Segundo o documento, a questão de determinar

⁵³ Ver: <<https://data.consilium.europa.eu/doc/document/PE-51-2019-INIT/pt/pdf>>

⁵⁴ O ANDS é uma bem-sucedida iniciativa da comunidade científica australiana, formada em 2008 pela Universidade Monash, líder do projeto; a Universidade Nacional da Austrália; a Commonwealth Scientific and Industrial Research Organisation (CSIRO), agência nacional de pesquisa científica; e a National Collaborative Research Infrastructure Strategy (NCRIS), órgão do governo que financia o ANDS. Em 1º de julho de 2018, o ANDS, o Nectar (National eResearch Collaboration Tools and Resources) e o Research Data Services (RDS) se juntaram para criar o Australian Research Data Commons (ARDC). Fonte: <https://en.wikipedia.org/wiki/Australian_Research_Data_Commons> e <<https://www.ands.org.au/>>

⁵⁵ Ver em: <<https://www.ands.org.au/partners-and-communities/partners>>

⁵⁶ Ver em AIM- FAO: <<http://aims.fao.org/activity/blog/articulating-value-data-australian-national-data-service-ands>>

[...] se há direitos autorais em dados parece [...] ainda não bem compreendida no setor de pesquisa [...] principalmente porque há alguma complexidade na avaliação da subsistência dos direitos autorais nos dados, em oposição a livros e outras obras criativas (ANDS, 2018, p. 4).

De acordo com o guia, a falta de uma licença ou declaração similar cria incerteza para aqueles que querem fazer reuso, o que limitará o uso potencial dos dados de pesquisa. O documento afirma que quando uma licença não é fornecida, a lei da maioria dos países presume que o material protegido por direitos autorais é do tipo “todos os direitos reservados”. Desta forma, fica subentendido que, embora possa ser acessível, o detentor dos direitos não permitiu a reutilização de seus dados.

É importante alertar que a reutilização dos dados com a restrição “todos os direitos reservados”, isto é, sem a devida permissão (estatutária ou sob uma licença) constituirá, de um modo geral, uma violação dos direitos autorais, podendo expor o infrator ou seu empregador a sérias penalidades de acordo com a lei de direitos autorais (ANDS, 2018).

As licenças devem ser compatíveis com outras licenças. Segundo a OKF (2020), elas só serão abertas se atenderem às permissões necessárias, tais como: uso, redistribuição, modificação, separação, compilação, não discriminação, propagação, aplicabilidade para qualquer finalidade, sem cobrança.

Uma licença não deve limitar ou tornar incertas ou diminuir as permissões necessárias, mas algumas exceções ou condições são aceitáveis, como por exemplo: atribuição de crédito (a colaboradores, detentores de direitos, patrocinadores e criadores, desde que não onerosas); integridade (obras modificadas indiquem por nome ou número ser diferente da original ou indiquem as alterações feitas); compartilhamento semelhante; aviso de retenção de direitos autorais e identificação da licença; proibição de restrição técnica a exercício de direitos; e não agressão (modificação para permitir usos adicionais) (OKF, 2020).

Em artigo apresentado no XX Encontro Nacional de Pesquisa em Ciência da Informação (Enancib), em 2019, Guanaes e Albagli (2019) ressaltam que as seis licenças Creative Commons e o instrumento CC0 Dedicção ao Domínio Público (Quadro 7) têm favorecido a circulação do conhecimento, ao promoverem o reuso legalmente informado de artigos e dados de pesquisa. Entre outras possibilidades, tais instrumentos oferecem aos

autores ou detentores de direitos autorais que queiram compartilhar uma criação intelectual (conjuntos de dados de pesquisa incluídos) uma escala ampla de opções que abrange a manutenção de todos os direitos e a renúncia a eles, à exceção dos direitos morais (BRANCO; BRITTO, 2013).

No entanto, a Creative Commons, organização não governamental, criada em 2001, com sede na Califórnia, voltada para expandir o acesso a obras artísticas, literárias e científicas, por meio de suas licenças públicas e gratuitas (Quadro 7) que permitem a cópia e o compartilhamento com menos restrições do que o tradicional “todos direitos reservados”, decorrentes da cessão total dos direitos patrimoniais, não recomenda o uso das Não Comerciais (NC) e Não Derivativas (ND) (Quadro 7) em bancos de dados destinados ao uso acadêmico ou científico. A primeira, por reduzir a visibilidade da obra intelectual ao restringir o uso comercial. A segunda, por não permitir compartilhamento da obra de forma adaptada (CREATIVE COMMONS, 2018a).

As licenças CC não são revogáveis, isto é, pode-se interromper a qualquer momento a disponibilização da obra, ou dos dados, sob licença, mas isso não afetará direitos associados a cópias que já estejam em circulação. Elas duram o mesmo prazo conferido ao direito de autor e permitem que ele autorize ou não o uso comercial ou trabalhos derivados do seu (CREATIVE COMMONS, 2018a).

Além das seis licenças, a Creative Commons disponibiliza o instrumento de renúncia de direitos autorais CC0 Universal Dedicção ao Domínio Público e a ferramenta Marca de Domínio Público (Quadro 7), um rótulo que ‘sinaliza’ que determinado trabalho está livre de direitos autorais e em domínio público (GUANAES, 2020, p. 83).

Uma característica comum às seis licenças é a obrigatoriedade de atribuir crédito ao criador original da obra. No entanto, o mesmo não ocorre com o CC0. Neste caso, os pedidos de atribuição de crédito não são vinculativos por meio de requisitos legais. De qualquer forma, é possível solicitar atribuição de crédito a quem for reusar dados e, no limite, no campo científico, os pesquisadores que reutilizam dados estão sujeitos a princípios e normas éticos amplamente adotados, como é o caso dos padrões de citação na comunidade acadêmica para artigos e dados de pesquisa (ANDS, 2018).

É preciso sublinhar que o CC0 Dedicção ao Domínio Público não é uma licença. As seis licenças CC permitem que o detentor mantenha seus direitos autorais. O CC0, por outro lado, permite a cientistas, educadores, artistas e outros criadores e proprietários de conteúdo

protegido por direitos autorais ou direitos de banco de dados renunciarem a todos os seus direitos, à exceção dos direitos morais, respeitado o limite permitido por lei sobre a matéria em cada país (BRANCO; BRITTO, 2013). O CC0 é apropriado para uso em repositórios de dados por maximizar a reutilização (DRYAD, 2020; FIGSHARE, 2020a; ZENODO, 2020a), pois dá direito ao usuário de copiar, modificar, distribuir e executar o trabalho, mesmo para fins comerciais, sem pedir permissão (CREATIVE COMMONS, 2018b), atributos característicos da condição jurídica denominada domínio público.

Segundo Souza (2011), o domínio público, inclusive na legislação brasileira, deriva do esgotamento do prazo de proteção dos direitos patrimoniais, que dão ao autor de uma obra o direito de explorá-la economicamente por toda a sua vida, e os herdeiros por 70 anos a contar de janeiro do ano seguinte à morte do autor, conforme a legislação brasileira. Quem utiliza obra que “caiu” em domínio público tem apenas que respeitar os direitos morais (direito à autoria e à integridade da obra), não havendo para si qualquer pagamento ou remuneração.

Vale mencionar ainda outros tipos de licenças usadas para dados. Ball (2014) cita a licença sob medida, escrita para direitos específicos que um autor queira disponibilizar, mas ressalva que não é trivial escrevê-la e enfrentar intrincados caminhos legais quando se têm as licenças padrão, como as já citadas CC. As três licenças Open Data Commons (ODC) são um projeto da Open Knowledge Foundation, criado em 2007 por Jordan Hatcher. Já a licença Open Government License (OGLv3) foi criada pelo National Archives e é usada no Reino Unido. As licenças ODC e a OGLv3 são mais voltadas a bancos de dados (Quadro 7).

Quadro 7. Licenças para permissão de reúso de dados de pesquisa em bases legais

Licenças	Sigla	Características
Creative Commons¹		
Atribuição CC BY	CC BY	Distribuir, remixar, adaptar e criar a partir do seu trabalho, mesmo para fins comerciais, desde que lhe atribuam crédito pela criação original. Quem reusa deve informar quais alterações efetuou. Maximiza a disseminação e uso dos materiais licenciados.
Atribuição Compartilhamento Igual	CC BY-SA	Remixar, adaptar e criar a partir do seu trabalho, mesmo para fins comerciais, desde que lhe atribuam crédito e que licenciem as novas criações sob termos idênticos. É comparável às licenças de software livre e <i>copyleft</i> , em que trabalhos derivados terão a mesma licença.
Atribuição Sem Derivações	CC BY-ND	Reúso para qualquer finalidade, inclusive comercial, mas o trabalho não pode ser compartilhado com outras pessoas de forma adaptada.

		Deve-se atribuir o crédito pela criação original.
Atribuição Não Comercial	CC BY-NC	Remixar, adaptar e criar a partir do seu trabalho para fins não comerciais. Deve-se atribuir o crédito pela criação original, mas quem reusa não tem que licenciar os trabalhos derivados sob os mesmos termos.
Atribuição Não Comercial Compartilhamento Igual	CC BY-NC-SA	Remixar, adaptar e criar a partir do seu trabalho para fins não comerciais, mas o licenciamento das novas criações deve ser feito sob termos iguais aos do licenciante. Deve-se atribuir o crédito pela criação original.
Atribuição Não Comercial Sem Derivações	CC BY-NC-ND	É a mais restritiva das seis licenças “Atribuição”. Só permite <i>downloads</i> do trabalho original e o compartilhamento, desde que deem o devido crédito, mas não podem alterá-lo ou usá-lo para fins comerciais.
CC0 Universal Dedicção ao Domínio Público ²	CC0 1.0	Copiar, modificar, distribuir e executar o trabalho, mesmo para fins comerciais, sem pedir autorização. O detentor dos direitos autorais que opta por esse instrumento posiciona sua obra no domínio público, isto é, renuncia a todos os seus direitos sob as leis de direito de autor e/ou de direitos conexos referentes à obra, em todo o mundo, na medida permitida por lei.
Marca de Domínio Público ²		É um rótulo inserido na obra que “sinaliza” que ela está livre de direitos autorais e em domínio público.
Open Data Commons		
Licença e Dedicção ao Domínio Público	PDDL	Dedica o banco de dados e seu conteúdo ao domínio público, gratuitamente, para que todos usem como acharem melhor.
Licença de Atribuição	ODC-By	Os usuários são livres para usar o banco de dados e seu conteúdo de maneiras novas e diferentes, desde que forneçam a atribuição à fonte dos dados e/ou do banco de dados.
Licença de Banco de Dados Aberta	ODC-ODbL	Uso subsequente do banco de dados deve fornecer atribuição de autoria, uma versão irrestrita do novo produto deve estar sempre acessível e quaisquer novos produtos feitos com material ODbL devem ser distribuídos usando os mesmos termos. É a mais restritiva de todas as licenças ODC.
Open Government License	OGLv3	Esta licença aplica-se a provedores de informação do setor público do Reino Unido. Protege, originalmente, direitos autorais e de banco de dados ante o licenciamento de informações da Coroa britânica.

Fonte: Elaboração própria, com base em CREATIVE COMMONS, 2018a; CREATIVECOMMONS, 2018b; ODI, 2019; UNIVERSITY LIBRARIE, 2019.

Notas: (1) Nas seis licenças “Atribuição” da Creative Commons, o usuário deve dar o crédito pela criação original; (2) Não são considerados licenças, mas instrumentos jurídicos, com finalidades distintas, para informar a opção do detentor de direitos autorais pelo domínio público.

Ball (2014) é o autor do guia adotado pelo Digital Curation Center (DCC), um centro de curadoria digital que apoia pesquisadores, instituições científicas, repositórios e editores a decidirem como aplicar uma licença a dados de pesquisa e qual seria a mais adequada. É um guia prático, dirigido ao Reino Unido, mas alguns aspectos podem ser aplicados internacionalmente. Segundo o autor, somente o detentor dos direitos ou alguém legalmente autorizado por ele pode conceder uma licença, devendo os direitos de propriedade serem estabelecidos antes do licenciamento, que também pode restringir o uso se for apresentada dentro de um contrato. O autor ressalta que três condições são comumente encontradas nas licenças: a atribuição do crédito, o requisito *copyleft*, em que qualquer trabalho derivado deve ser liberado sob a mesma licença, e a adoção de licença não comercial, para que o licenciado não explore o trabalho comercialmente.

3.5 CC-BY é a licença mais aberta, mas não está livre de controvérsias

Entre as seis licenças CC, a CC BY é considerada pela própria Creative Commons e por grande parcela de editores como a menos restritiva de direitos em relação às outras cinco (Quadro 7), permitindo maior possibilidade de reuso. Mas uma polêmica acompanha a CC BY: ela permite a comercialização do conteúdo que licencia. É ético comercializar conteúdo (artigo científico, coleções de periódicos ou conjuntos de dados de pesquisa) proveniente de pesquisa financiada com recursos públicos e destinado originalmente ao acesso aberto gratuito? No Reino Unido, a adoção da CC BY é incentivada por agências financiadoras de pesquisa, como a Wellcome Trust⁵⁷ e o The Research Councils UK (RCUK)⁵⁸. Mas... ela não está livre de controvérsias.

Em 2016, acadêmicos, administradores, editores e profissionais do direito reuniram-se na Universidade de Cambridge para debater se a CC BY pode de fato vir a ser um problema. Pesquisadores de áreas que lidam com informações confidenciais – como as áreas médicas e criminais – e se preocupam em apresentar e distribuir seus achados de forma apropriada tinham um bom argumento. A questão com a CC BY era que os achados poderiam ser mal interpretados, o que seria prejudicial ao pesquisador e poderia afetar os participantes de uma

⁵⁷ Ver em: <<https://wellcome.ac.uk/funding/guidance/open-access-guidance/open-access-policy>>

⁵⁸ Ver em: <<https://www.ukri.org/files/legacy/documents/rcukopenaccesspolicy-pdf/>>

pesquisa. Assim, os resultados alterados da pesquisa podem afetar não apenas a pesquisa, mas também os participantes (UNIVERSITY OF CAMBRIDGE..., 2016).

Outro tópico levantado na reunião foi a comercialização de obras disponíveis gratuitamente em acesso aberto. A CC BY autoriza o reuso para fins comerciais. Quem estaria sendo prejudicado na exploração de obras acadêmicas? Em geral, o pesquisador não tem interesse em ganhar dinheiro com o que publica. A preocupação era que as pessoas pudessem estar sendo exploradas ao pagarem por coisas que deveriam estar disponíveis gratuitamente (UNIVERSITY OF CAMBRIDGE..., 2016).

Em dezembro de 2017, uma comunidade científica comprometida com a ciência aberta, com o movimento de abertura de artigos científicos e dados de pesquisa subjacentes a artigos, reunida na cidade do México, formada pelo Sistema Regional de Informação Online para Revistas Científicas da América Latina, do Caribe, Espanha e Portugal (Latindex), Rede de Revistas Científicas da América Latina e do Caribe, Espanha e Portugal (Redalyc), Conselho Latino-Americano de Ciências Sociais (Clacso) e Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT), assinou uma declaração conjunta recomendando o uso da licença Creative Commons Atribuição-Não-Comercial-Compartilha-Igual (CC BY-NC-SA, na sigla em inglês). Esta licença é considerada restritiva. Permite que o usuário remixe, adapte e crie a partir da obra original para fins não comerciais, mas o licenciamento das novas criações deverá ser feito em termos iguais ao do licenciante, conforme o Quadro 7.

A ideia da comunidade era impedir que resultados da pesquisa científica e acadêmica financiados com recursos públicos sejam comercializados e encerrados em sistemas proprietários fechados, barrando a qualquer pessoa o acesso aberto e gratuito à produção do conhecimento científico (DECLARAÇÃO DO MÉXICO, 2017).

Com essa decisão, embora sem fazer qualquer menção na declaração, as entidades reunidas na capital mexicana estavam desaconselhando seus filiados e seguidores a adotarem a decantada e adotada licença CC BY, que também permite que o usuário remixe, adapte e crie a partir da obra original, mas não proíbe o reuso para fins comerciais e não inclui a exigência de licenciamento de novas criações nos mesmos termos da licença originária. É a licença de caráter mais aberto entre as seis da Creative Commons. Só perde para o instrumento CC0 Dedicção ao Domínio Público, que retira os direitos patrimoniais da obra protegida, mantendo apenas os direitos morais do titular dos direitos autorais.

A denominada “Declaração do México” em favor do ecossistema latino-americano de acesso aberto não comercial, em seus considerandos para justificar a adoção da CC BY-NC-SA, afirma que tem consciência de

[...] que o obstáculo ao acesso ao conhecimento internacional não é tecnológico, mas sim ser capaz de pagar o acesso às bases de dados comerciais, razão pela qual criaram-se consórcios nos países para enfrentar o aumento contínuo dos preços, mas o custo e as restrições ao uso de documentos aumentaram e atualmente se somam ao custo de pagar para publicar em acesso aberto (DECLARAÇÃO DO MÉXICO, 2017, tradução nossa),

aludindo às taxas de processamento de artigos (em inglês, Article Processing Charges - APC), cobradas por periódicos científicos híbridos (BJÖRK; SOLOMON, 2014), e ao fato de que essas bases de dados comerciais se tornaram matéria-prima da avaliação da pesquisa. Em outro trecho da declaração, seus signatários se dizem frustrados porque

[...] as revistas publicadas pelos grandes monopólios editoriais negam o depósito nos repositórios nacionais e institucionais da versão final dos artigos [PDFs diagramados] – artigos que são financiados pelas instituições – e, quando permitem o depósito, exigem uma licença comercial (DECLARAÇÃO DO MÉXICO, 2017, tradução nossa).

Um grande problema, portanto.

Contudo, em seu *site* pioneiro no Brasil, cienciaaberta.net, Abdo (2018) sintetiza a opinião de uma corrente do movimento pela ciência aberta, criticando a postura das entidades que assinaram o documento, favorável ao uso de licenças CC com cláusula NC (não comercial, no caso a CC BY-NC-SA) para publicações em acesso aberto, como forma de combater os grandes conglomerados editoriais.

Abdo (2017; 2018) sustenta, entre outros argumentos, que uma licença NC apenas aumenta a motivação econômica de se adquirir a revista ao permitir formas de exclusão e, portanto, com isso agregar-lhe valor como mercadoria. Dessa forma, a licença NC faz com que o mais forte exclua o mais fraco da concorrência, reforçando a tese dos direitos autorais como fator de exclusão e mercantilização do conhecimento, conclui o pesquisador do

laboratório LISIS-IFRIS, em Paris. A par desse viés econômico, Abdo (2014, *on line*) associa a opção pela NC a entraves para a circulação do conhecimento, já que

[...] licenças ‘não comerciais’ dificultam ou impedem a difusão de ideias em diversos meios solidários, assim como dificultam a pessoas solidárias de polinizar ideias em outros meios. Ao mesmo tempo, criam um nicho incompatível com os conteúdos livres, diminuindo o valor multiplicativo das redes de produção do comum, sem nenhuma contrapartida.

3.5.1 SciELO adota a CC BY

Abdo (2014) está na companhia de um *player* de respeito, a Scientific Electronic Library Online, de onde saiu a conhecida e prestigiada sigla SciELO, é um dos mais importantes programas de acesso aberto à informação científica da América Latina, reconhecido mundialmente. O projeto, pioneiro no Brasil, foi lançado em 1998 pela Fapesp, Centro Latino-Americano e do Caribe de Informação em Ciências da Saúde/Organização Pan-Americana da Saúde/Organização Mundial da Saúde (Bireme/OPAS/OMS), com apoio do CNPq e de periódicos de sociedades e instituições acadêmicas (PACKER et al., 1998). Seu objetivo é “implementar uma biblioteca virtual eletrônica, proporcionando acesso completo a uma coleção de títulos de periódicos, uma coleção de fascículos de títulos de periódicos individuais, bem como ao texto completo dos artigos.”⁵⁹.

Atualmente, a Rede SciELO⁶⁰ fornece acesso aberto ao conteúdo de 1.375 periódicos ativos⁶¹, organizados em coleções nacionais e temáticas. A Coleção SciELO Brasil abriga 375 periódicos, com 295 ativos e 80 não ativos⁶².

Assim como Abdo (2014), esse peso pesado latino-americano da comunicação científica em acesso aberto demonstra total discordância do posicionamento dos signatários da declaração do México, pois adota desde junho de 2015 a licença CC BY, tão criticada

⁵⁹ Ver em: <<http://www.scielo.br/>>

⁶⁰ Participam atualmente da rede SciELO os seguintes países: África do Sul, Argentina, Bolívia, Brasil, Chile, Colômbia, Costa Rica, Cuba, Espanha, México, Paraguai, Peru, Portugal e Uruguai. Candidatos à participação na Rede: Equador, Venezuela e Índias Ocidentais. Fonte: SciELO, dados recebidos por e-mail, em 14 fev. 2020.

⁶¹ Dados coletados na página “Periódicos” <<https://scielo.org/pt/periodicos/listar-por-ordem-alfabetica>>, na seção “Periódicos ativos”, da SciELO, em 12 fev. 2020.

⁶² Ver em: <http://www.scielo.br/scielo.php?script=sci_alphabetic&lng=en&nrm=iso>

naquele documento, para periódicos e artigos científicos da Coleção SciELO. A partir de 2016, essa licença foi adotada como padrão para todas as coleções da Rede SciELO. Até 2013, a licença padrão era a CC BY NC, e a CC BY era oferecida como alternativa (SANTOS; SALES; PACKER, 2015).

No documento “Declaração de acesso aberto”, o SciELO (2019, *on line*) reafirma a adoção da CC BY, endossando o reúso de artigos, periódicos e coleções para fins comerciais. Ressalva, no entanto, que “Os periódicos indexados antes de 2015 podem manter sua licença original”, porém os incentiva a aderir à CC BY. Quanto aos periódicos que ingressaram após 2015, a biblioteca eletrônica determina que “é necessário” que adotem a licença CC BY.

A Tabela 1 apresenta a distribuição das licenças CC adotadas em 2014 entre os 287 periódicos integrantes da coleção SciELO Brasil naquele ano. A atribuição CC BY-NC fora definida como padrão até 2013 e era adotada por 80% (229) dos periódicos. Observe-se que 17% (49) dos periódicos já adotavam a licença CC BY e um número reduzido de periódicos (9), a licença mais restritiva CC BY-NC-ND. A expectativa do SciELO era a de que mais de 90% dos periódicos da coleção SciELO Brasil adotassem a licença CC BY já a partir de 2016 (SANTOS; SALES; PACKER, 2015).

Tabela 1. Distribuição das licenças CC no SciELO Brasil, em 2014

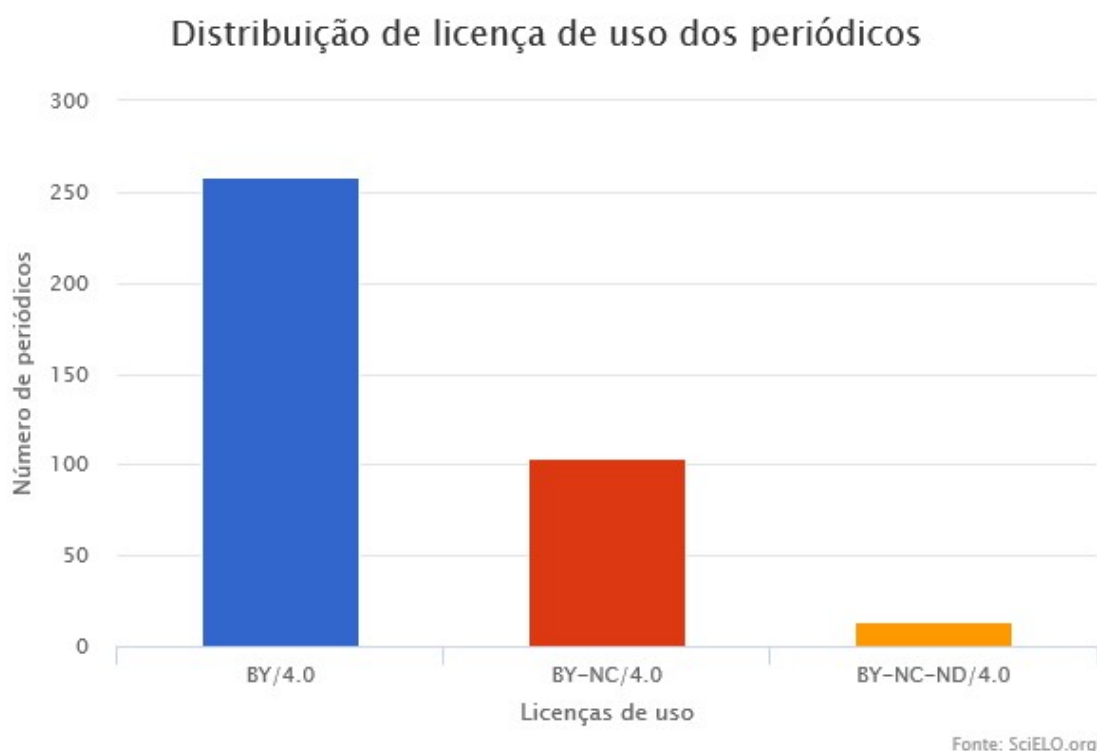
Licença CC em 2014	N. Títulos
CC-BY-NC	229
CC-BY	49
CC-BY-NC-ND	9
Total	287

Fonte: Adaptado de SANTOS; SALES; PACKER, 2015.

A Figura 3 espelha o número de periódicos por licença Creative Commons na Coleção SciELO Brasil, em fevereiro de 2020. Registre-se que o número de periódicos saltou de 287 para 375 no período de 2014 a 2020. De acordo com a nota da Figura 3, o número de licenças

de uso CC BY (“BY”, na nota) dos periódicos aumentou mais de cinco vezes em seis anos, passando de 49 em 2014 para 258 em 2020.

Figura 3



Fonte: SCIELO, 2020b.

Nota: Número de periódicos por licença Creative Commons na Coleção SciELO Brasil, em fev. 2020: BY/4.0 = 258; BY-NC/4.0 = 103; BY-NC-ND = 14. **Total:** 375 (295 títulos ativos e 80 não ativos).

Significa e chama a atenção que, atualmente, os artigos de mais de duas centenas e meia de revistas científicas da Coleção SciELO Brasil, publicadas sob a licença CC BY, bem como os dados publicados nos próprios artigos (como arquivos de apoio, que herdam a licença CC BY) podem, a princípio, ser reutilizados para fins comerciais, sem qualquer pedido de autorização ao detentor dos direitos autorais.

A licença CC BY-NC, que proíbe a comercialização, cresceu um pouco mais do dobro na comparação de 2014 (49) com 2020 (103). Embora o número percentual de licenças CC BY-NC-ND/4.0 tenha crescido cerca de 64% na comparação de 2014 com 2020, em valores absolutos (9 para 14), respectivamente, não nos parece expressivo. Neste caso, o que chama a

atenção é o fato de que 14 revistas de um dos maiores programas de acesso aberto dourado da América Latina, embora exerçam seu direito de escolha de licença, adotam a mais restritiva da Creative Commons – CC BY-NC-ND/4.0 –, pois é uma licença que apenas permite *downloads* do trabalho original e o compartilhamento, desde que se dê o devido crédito, porém é proibido alterá-lo ou usá-lo para fins comerciais (Quadro 7).

Depreende-se do exposto a seguinte questão: os periódicos indexados no SciELO, que são financiados com recursos públicos e cujas instituições editoras adotam políticas de acesso aberto aos resultados de pesquisa (artigos e dados), ao pactuarem com autores externos uma cessão total não exclusiva de direitos patrimoniais não comerciais de utilização de uma obra, portanto sem finalidade de lucro por quem a disponibiliza e utiliza, têm o direito de aderir à licença CC BY, conforme determina o SciELO?

A nova fronteira da ciência aberta apresenta um estágio avançado em termos de proposição teórica, mas ainda não oferece, na prática, um padrão universal de requisitos para o processo de implementação ou um conjunto normativo que exclua dados de pesquisa da proteção de direitos autorais. Além disso, a exigência de agências de financiamento de pesquisa de depósito imediato de resultados de pesquisa em repositórios pode estar dando início a uma nova dinâmica na comunicação científica. Ela envolve revistas científicas, repositórios de dados de pesquisa e as próprias agências. Como esses atores da cena científica estariam lidando com a abertura de dados de pesquisa subjacentes aos artigos e as repercussões do direito autoral sobre esses dados? O que suas políticas dizem a respeito?

4 ABERTURA DE DADOS DE PESQUISA E O DIREITO AUTORAL NAS POLÍTICAS DE REVISTAS DE SAÚDE, REPOSITÓRIOS E AGÊNCIAS

A operação de publicação e disponibilidade pública dos dados de pesquisa subjacentes a artigos envolve três importantes atores do ecossistema da ciência aberta: a revista científica, os repositórios de dados de pesquisa e as agências de financiamento à pesquisa (BORGMAN, 2012; ALBAGLI, 2015; CHAN; OKUNE; SAMBULI, 2015; NATIONAL ACADEMIES OF SCIENCE, 2018).

Neste capítulo, em primeiro lugar, fazemos uma revisão de literatura sobre como a abertura de dados de pesquisa subjacentes vem afetando e sendo afetada por um novo processo de comunicação científica, mediante a publicação desses dados, que envolve diferentes perspectivas dos atores envolvidos, tais como autores, editores, gestores de repositórios e financiadores. Espera-se, assim, mostrar um retrato crítico sobre a prática desse procedimento.

Em segundo lugar, apresentamos os resultados de uma pesquisa em que é feita uma análise das políticas de dados e de direitos autorais de três periódicos, três repositórios de dados e três agências de financiamento à pesquisa, discriminados mais adiante. A pesquisa foi realizada por meio de avaliação observacional diretamente nos *sites* desses atores cruciais para o empreendimento científico entre fevereiro e maio de 2020.

O objetivo principal foi identificar e analisar o que suas políticas e normativas para a abertura de dados dizem sobre a mediação do direito autoral nas ações de depósito, publicação, acesso, compartilhamento e reúso de dados de pesquisa. Há conflitos entre as políticas? Espera-se assim contribuir com o conhecimento sobre como esses três segmentos lidam com a questão do direito autoral aplicável a dados de pesquisa subjacentes.

No que que concerne às revistas científicas, relatamos os resultados da investigação sobre políticas de abertura executada nos *sites* das revistas *PLoS Medicine*, *theBMJ* e *Nature Medicine*, adotadas, respectivamente, em 2014, 2015 e 2016. A escolha das três publicações deve-se à sua inserção no campo da saúde, incluindo a pesquisa clínica, à adoção das recomendações editoriais e de compartilhamento de dados de pesquisa da mesma fonte, o ICMJE (ver seção 2.6), e por já estarem vivenciando o novo tempo da abertura de dados de pesquisa e implicações com direitos autorais e licenciamento de dados.

Quanto aos repositórios, são avaliadas as diretrizes das políticas constantes dos *sites* dos seguintes repositórios generalistas: Dryad, Figshare e Zenodo, selecionados por exemplificarem o novo papel que os repositórios vêm desempenhando na comunicação científica, ao oferecerem acesso aberto e formas inovadoras de visualização de resultados de pesquisa, especialmente de dados vinculados a artigos científicos, como será visto adiante, utilizando-se de tecnologia de ponta. Seu reconhecimento foi constatado pelo aumento de citações a esses três repositórios generalistas, o que indicaria um “crescimento exponencial dos conjuntos de dados que estão sendo disponibilizados” (DIGITAL SCIENCE, 2018, p. 3, tradução nossa) por essas plataformas.

Em relação às agências de fomento à pesquisa, analisamos, com base em observação de seus *sites*, as políticas do Nacional Institute of Health, Wellcome Trust/Wellcome Open Research e Bill & Melinda Gates Foundation. O NIH (2003a), por seu pioneirismo tanto no estímulo à abertura de dados quanto na percepção da necessidade de novas práticas de pesquisa em decorrência disso, como a adoção de plano de compartilhamento de dados, ainda no início deste século. O Wellcome Trust, por sua liderança exercida na emergência mundial do vírus Zika, em 2016, quando organizou uma reunião de órgãos globais de saúde, solicitando que “todos os dados de pesquisa coletados durante o surto do vírus Zika e futuras emergências de saúde pública fossem disponibilizados o mais rápida e abertamente possível” (WELLCOME TRUST, 2016, tradução nossa), fato que se repetiu na pandemia de Covid-19 (WELLCOME TRUST, 2020). E também por sua iniciativa inovadora de criar a plataforma Wellcome Open Research, objeto de nossa análise, que introduziu uma série de práticas incomuns na comunicação científica clássica, em consonância com a ciência aberta, mas também por se constituir em um canal próprio da Wellcome Trust para publicação de artigos vinculados a dados subjacentes de pesquisas que a agência financia. Quanto à Bill & Melinda Gates Foundation (2017a), a seleção ocorreu por uma atuação que tem apontado para caminhos inovadores que incomodam a estrutura reinante de difusão da produção científica. Uma diretriz da fundação Gates determinava que os artigos e conjuntos de dados subjacentes das pesquisas que financiava deveriam ser disponibilizados em acesso aberto imediatamente após a publicação, sob a licença CC BY, que permite reutilização abrangente, inclusive para fins comerciais. Foi um abalo na cultura do embargo, que causou reação imediata das principais revistas acadêmicas (ver seção 4.4, neste capítulo). Estaria nascendo um novo jeito de divulgar a ciência?

4.1 O desafio das revistas de saúde

As primeiras revistas científicas surgiram em meados do século XVII, mediante o esgotamento do regime de circulação da informação da época sobre descobertas e resultados de pesquisa: as cartas (MEADOWS, 1999).

O advento das revistas científicas está ligado a duas publicações: *Journal de Sçavans*, mais tarde *Journal de Savants*, criada em Paris por Denis de Sallo, que teve seu primeiro número publicado em janeiro de 1665; e *Philosophical Transactions*, editada em Londres por Henry Oldenburg, da recém-fundada Royal Society (1662). A criação da *Philosophical* foi influenciada pelo primeiro número da revista francesa e lançada em março de 1665 (MEADOWS, 1999).

A revista francesa e a inglesa tinham conteúdo e intenções nitidamente distintos. A *Journal de Savants* publicava notícias sobre resumos de livros editados na Europa, necrológicos, descrição de progressos científicos, decisões jurídicas e cobertura de eventos de interesse de “homens letrados”, podendo ser considerada “a primeira revista em sentido moderno” (MEADOWS, 1999, p. 6). Já a *Philosophical Transactions* se interessava apenas por “estudos experimentais” e seu conteúdo era revisado por membros do Conselho da Real Society, características que a credenciam como precursora da revista científica tal qual a conhecemos atualmente (MEADOWS, 1999, p. 6-7).

No livro *Conhecimento público*, John Ziman (1979) apresenta sua versão sobre as origens da revista científica. Ele acreditava que a criação da revista científica era um passo decisivo para o desenvolvimento do método científico, por constituir-se em “um mecanismo por meio do qual os resultados de minuciosas pesquisas podem ser publicados parceladamente”. Além disso, também reconhecia que alterou a forma de comunicação de informações científicas entre pares, feita por correspondência particular ou publicações esparsas de livros ou panfletos até meados do século XVII (ZIMAN, 1979, p. 117).

Para Ziman (1979), a comunicação da pesquisa à comunidade científica, o manuscrito que garante a prioridade da pesquisa, as críticas, as citações de outros autores são partes tão importantes quanto o “germe da ideia que deu origem a tudo” (p. 116).

Durante o século XX, as revistas científicas consolidaram sua posição como os principais meios de difusão dos resultados de pesquisa, especialmente nas ciências naturais e

médicas, a maioria delas publicada por sociedades científicas antes da II Guerra Mundial (LARIVIÈRE; HAUSTEIN; MONGEON, 2015).

As revistas de saúde, mais especificamente as médicas gerais, surgiram na Era do Iluminismo, como, por exemplo, a *Lancet*, que começou em 1823, e a *The New England Journal of Medicine*, em 1812. No Brasil, merece registro a *Memórias do Instituto Oswaldo Cruz*⁶³, da Fundação Oswaldo Cruz, publicada em 1909. Elas influenciam a prática clínica, a política de saúde, a saúde pública e a pesquisa biomédica, informam Ioannidis, Belbasis e Evangelou (2010), ao estudarem o impacto a longo prazo das revistas médicas publicadas entre 1959 e 2009. O crescimento dessas revistas, incluindo as científicas, foi exponencial ao longo do século XX, até meados da década de 1990, quando o modelo de negócios tradicional – venda de assinaturas principalmente para instituições – não funcionava mais (SMITH, 2006; LARIVIÈRE; HAUSTEIN; MONGEON, 2015).

Conforme já revelado no capítulo 2, nas últimas décadas do século XX, os preços dos periódicos aumentaram exponencialmente, inviabilizando o tradicional sistema em que a comunicação científica era impressa em papel. Em consequência, nasce o movimento pelo acesso aberto à informação científica que começou a se desenvolver pelas mãos dos cientistas da computação, que já dispunham seus resultados de pesquisa gratuitamente na internet, com acesso a pequenas comunidades. Em 1991, com o servidor de *preprints* ArXiv⁶⁴, especializado em física de alta energia, Paul Ginsparg inaugura a prática de autorquívamento com o depósito de trabalhos científicos dessa comunidade (SWAN, 2016).

“Na mesma época em que os repositórios se desenvolviam como locais para o arquivo de materiais de Acesso Aberto, também surgia o veículo alternativo de disseminação de AA – os periódicos científicos de Acesso Aberto”, revela Swan (2016, p. 14), um novo tipo de revista científica, que disponibiliza seus conteúdos *on line* gratuitamente. Um movimento que permite a um pesquisador disseminar a sua própria pesquisa via internet, dar acesso a ela para outros pesquisadores e acessar pesquisas de outros colegas (WILLINSKY, 2006; SUBER, 2012; LAAKSO; BJÖRK, 2012). Harnad (1995) e Odlyzko (1994) sugerem que a intensificação da passagem do suporte impresso para o eletrônico date de meados da década

⁶³ *Memórias do Instituto Oswaldo Cruz* foi criada por decreto em 1907, mas começou como revista científica em 1909. Em 2016, foi avaliada pela Thomson Reuters como o periódico científico mais citado entre os países latino-americanos (BRANDÃO, 2019; EXAME, 2016).

⁶⁴ Ver em: <www.arxiv.org>

de 1990. Meadows (2001, p. 5) previu que “um dos desenvolvimentos mais significativos [do periódico eletrônico] será uma crescente diversificação do processo de publicar”, além de questões de direitos autorais, como a cópia eletrônica, que seriam de difícil controle.

Revistas científicas ainda hoje são o veículo mais prestigiado para disseminação da informação científica, sobretudo por meio de artigos, mas os dados relacionados às pesquisas relatadas em artigos científicos – os dados subjacentes – ganharam protagonismo na ciência aberta e os pesquisadores têm sido incentivados a disponibilizá-los publicamente. No entanto, o direito autoral, se não dificulta, constitui-se em uma preocupação, em geral em virtude do descompasso de um regramento legal ante a dinâmica do desenvolvimento das práticas científicas, beneficiadas por tecnologias cada vez mais refinadas, facilitadoras do empreendimento científico, como a produção científica colaborativa (MURRAY-RUST, 2008; CHAN, 2014; SAYÃO; SALES, 2014).

Nesse novo cenário, como as revistas lidam com a abertura de dados de pesquisa e as repercussões do direito autoral sobre esses dados? Uma nova sistemática é criada: agências de financiamento de pesquisa exigem que os pesquisadores depositem os dados resultantes de pesquisas que elas financiam preferencialmente em repositórios temáticos ou, caso não existam no campo da pesquisa realizada, em repositórios generalistas ou multidisciplinares, o que pode estar dando início a uma nova dinâmica na comunicação científica (NIH, 2003b; NATIONAL RESEARCH COUNCIL, 2003; OECD, 2007; 2015; COMISSÃO EUROPEIA, 2017; NATIONAL ACADEMIES OF SCIENCE, 2018).

No entanto, diferente do que ocorre com os artigos científicos, as questões do direito autoral sobre dados de pesquisa podem ser mais complexas e representam uma novidade, conforme assinalam Guibault e Wiebe (2013) e Carroll (2015). Os financiadores de pesquisa contam com o apoio de editores e seus periódicos, considerando-os fundamentais para o estabelecimento da cultura de abertura de dados de pesquisa. As revistas, por sua vez, conforme seus campos de inserção, também elaboram políticas que abordam as questões do direito autoral. Para saber um pouco mais sobre como as revistas se comportam ante esse novo trecho da ciência, observamos as políticas de abertura de dados de pesquisa de três delas – *PLoS Medicine*, *theBMJ* e *Nature* –, a partir de seus *sites*, para uma análise não exaustiva das orientações sobre compartilhamento de dados e o que dizem a respeito do direito autoral sobre dados de pesquisa subjacentes.

4.1.1 *PLOS Medicine*

Em 2004, os fundadores da Public Library of Science (PLOS), entre eles o Nobel de Medicina, Harold Varmus, lançaram a revista *PLOS Medicine*, revisada por pares, com o objetivo de abranger todo o espectro das ciências médicas. Em uma “Mensagem dos fundadores da PLoS” (EISEN; BROWN; VARMUS, 2004), publicada na edição número 1, eles forneciam o escopo da nova publicação da PLoS, uma editora de acesso aberto, sem fins lucrativos, criada em 2001 por um grupo de médicos liderados por Varmus. Sua primeira publicação foi a revista *PLOS Biology*, lançada em 2003.

Em pleno raiar do movimento pelo acesso livre à informação científica, o intuito do grupo era lançar uma revista médica geral pela internet, na contramão de suas congêneres da época que já utilizavam esse meio de distribuição do século XXI, mas “deixaram seu modelo de negócios no século 19, continuando a cobrar dos leitores pelo acesso, assim como faziam com suas revistas impressas. Isso tem sido bom para os negócios – a publicação médica nunca foi tão lucrativa – mas tem um custo enorme” (EISEN; BROWN; VARMUS, 2004, *on line*, tradução nossa).

Nascia assim um dos primeiros modelos de negócio do movimento de acesso aberto à informação científica: os autores dos artigos de pesquisa aceitos pagam “uma taxa de publicação para cobrir os custos de revisão por pares, supervisão editorial e produção”; em contrapartida, tudo “o que é publicado na *PLOS Medicine* está disponível gratuitamente *on line* em todo o mundo, sem restrições de distribuição, cópia, impressão ou uso legítimo” (EISEN; BROWN; VARMUS, 2004, *on line*, tradução nossa). Os fundadores da PLoS alegam que não esperavam que os autores cobrissem individualmente os custos de publicação em acesso aberto, mas sim que as agências governamentais, empresas, fundações, instituições de pesquisa, hospitais ou universidades que patrocinam a pesquisa paguem a taxa (EISEN; BROWN; VARMUS, 2004).

O modelo suscita polêmicas. Por exemplo, poderia ser bom para os países do Norte do planeta, mas e para os países do Sul? Como fazer para convencer governos e estimular instituições científicas destes últimos a incluírem a taxa de processamento de artigos em suas magras dotações orçamentárias destinadas à ciência? (GUANAES; GUIMARÃES, 2012). Mas esta é uma outra história, sobre a qual não cabe nos estendermos aqui.

A *PLoS Medicine*, em setembro de 2020, cobrava US\$ 3.000⁶⁵ por artigo de pesquisa que reporte ensaios clínicos, revisões/meta-análises sistemáticas, epidemiologia, testes de diagnóstico, modelagem preditiva e explicativa, pesquisa de custo-efetividade, estudos qualitativos e pesquisa translacional com foco clínico.

Em março de 2014, a PLoS publicou a sua política de dados, a qual abrange todas as revistas da editora (BLOOM; GANLEY; WINKER, 2014). Desde então, a *Plos Medicine* dedica uma página específica, “Data Availability” (Disponibilidade de Dados) (PLOS MEDICINE, 2014), para informar suas diretrizes sobre a abertura e compartilhamento de dados subjacentes. Logo na Introdução, a revista destaca quatro diretrizes, recorrentes nas políticas das outras duas revistas analisadas: a) a exigência de que os autores disponibilizem publicamente todos os dados necessários à replicação da pesquisa, sem restrições no momento da publicação, e indicação de como obter acesso público aos conjuntos de dados, caso haja restrições legais ou éticas; b) a exigência de uma declaração de concordância de disponibilidade dos dados, descrevendo a conformidade com a política de dados da PLoS, que será publicada como parte do artigo; c) a publicação dos dados está condicionada ao cumprimento da política da revista; e d) os autores devem depositar os dados relevantes (dados necessários para replicar todas as descobertas do estudo relatadas no artigo, bem como metadados e métodos relacionados) em repositórios públicos de dados.

A revista apenas “incentiva” o autor a produzir um plano de gestão de dados, para atender financiadores de pesquisa, e a citar qualquer tipo de dado disponível publicamente, devendo incluir as informações mínimas recomendadas pelo DataCite⁶⁶, de modo que se tornem detectáveis e os metadados sejam disponibilizados para a comunidade. Nesta questão específica, também apoia o roteiro prático de citação de dados desenvolvido para editores acadêmicos pelo Publishers Early Adopters Expert Group, como parte de um projeto do FORCE11.org e do programa NIH BioCADDIE (COUSIJN, 2018), uma iniciativa para atender uma demanda de pesquisadores. Tal posicionamento de apenas “incentivar” não atende por completo às diretrizes do ICMJE (mais detalhadas na próxima revista analisada) sobre o tema, que considera fundamental para a disseminação da cultura de abertura de dados tanto a citação robusta quanto a elaboração de um plano de gestão de dados, desde o início do

⁶⁵ Ver em: <<https://plos.org/publish/fees/>>

⁶⁶ Ver em: <<https://datacite.org/>>

ciclo da pesquisa, e sua conseqüente apresentação no momento da submissão do artigo a uma revista de saúde (ICMJE, 2018).

Ainda segundo a página específica sobre abertura de dados, “Todos os dados e metadados subjacentes às descobertas relatadas devem ser depositados em repositórios públicos de dados apropriados” (PLOS MEDICINE, 2014, *on line*, tradução nossa), abrindo a possibilidade de opção por repositórios temáticos, generalistas ou institucionais, que devem estar registrados na declaração de concordância de abertura de dados. A revista pede atenção à página “Recommended Repositories” (Repositórios recomendados)⁶⁷, na qual orienta os autores a escolherem como primeira opção repositórios adequados ao seu campo de estudo, sugerindo a adoção dos princípios FAIR (Quadro 5) para dados subjacentes das áreas de ciências ambientais e biomédicas, o re3data.org⁶⁸ como fonte de escolha de repositórios em todas as áreas, aliado ao FAIRsharing⁶⁹, um serviço de curadoria que fornece informações que envolvam licenciamento, certificados e padrões de dados e metadados, bancos de dados, políticas de dados etc. Entre outros repositórios generalistas recomendados, a *PLoS Medicine* indica o Dryad, o Figshare e o Zenodo, sobre os quais apresentamos um breve estudo de suas abordagens relativas a direito autoral aplicável a dados de pesquisa ainda na seção 4.2 deste capítulo.

A *PLoS Medicine*, todavia, reconhece que em alguns casos o autor não poderá disponibilizar publicamente o conjunto de dados por razões éticas ou legais, o que chama de “restrições de acesso a dados aceitáveis”, informando que sua política não substitui regulamentos, legislação ou estruturas éticas locais. Neste caso, a revista informa que os autores devem esclarecer tais limitações na declaração de disponibilidade no momento do envio. Por outro lado, a revista não considerará manuscritos cujos autores não abram seus dados em virtude de interesses pessoais, como patentes ou publicações futuras (PLOS MEDICINE, 2014).

⁶⁷ Ver em: <<https://journals.plos.org/plosmedicine/s/recommended-repositories>>

⁶⁸ Ver em: <<http://re3data.org/>>

⁶⁹ Ver em: <<https://fairsharing.org/>>

4.1.2 Direitos autorais: incompatibilidades

Assim como a sua política de dados, publicada em 2014 (BLOOM; GANLEY; WINKER, 2014, *on line*), a política de licenças e direitos autorais, apresentada na página “Licenses and Copyright” (Licenças e direitos autorais)⁷⁰ da *Plos Medicine*, aplica-se “a todos os periódicos PLoS, a menos que indicado de outra forma”, significando que os editores podem fazer ajustes específicos à sua revista.

Também valendo para todos os periódicos PLoS, o documento constante da página “PLoS *Terms of Use*” (PLoS Termos de Uso), que regula a relação entre um usuário e o *site* da editora, estipula claramente os termos de cessão de direitos autorais no que diz respeito a dados de pesquisa subjacentes e outros resultados de pesquisa:

[...] ao submeter *qualquer conteúdo* a um *site* da PLoS, como comentários, revisões por pares, manuscritos e *outros resultados de pesquisa*, o remetente concede à PLoS um direito irrevogável, não-exclusivo, isento de pagamento de direitos autorais, de publicar, reproduzir, distribuir, transmitir, exibir, armazenar, traduzir, criar trabalhos derivados. E também de usar seu conteúdo (no todo ou em parte) e/ou incorporá-lo a outros trabalhos de qualquer forma, formato ou mídia, que sejam conhecidos ou desenvolvidos a partir de agora, em todo o mundo e em qualquer idioma, por toda a duração de tal direito, além de permitir terceiros, mediante sublicenciamento, a realizar todo ou qualquer um dos itens descritos (grifos nossos)⁷¹.

Em consonância com esses termos e de volta à página “Licenses and Copyright”, a revista declara que adota para todos os seus conteúdos a licença Creative Commons Attribution (CC BY), a qual permite copiar, distribuir, exibir e executar o trabalho, fazer trabalhos derivados e até mesmo uso comercial da obra, desde que se dê crédito ao autor original. Caso o manuscrito contenha fotos, imagens, figuras, tabelas, arquivos de áudio, vídeos etc., a revista solicita que o autor forneça uma prova de que o proprietário desses materiais autorizou o uso naquele manuscrito. Havendo dúvidas sobre a falta de permissão, a revista se reserva o direito de remover o material.

Merece comentar que a revista mistura, portanto, diretrizes tanto para artigos quanto para outros trabalhos, incluindo conjuntos de dados subjacentes, o que não é aconselhável

⁷⁰ Ver em: <<https://journals.plos.org/plosmedicine/s/licenses-and-copyright>>

⁷¹ Ver em: <<https://plos.org/terms-of-use/>>

sobretudo em tema pouco familiar ao pesquisador, como o direito autoral, o que pode suscitar retração deste em abrir seus dados publicamente (TENOPIR, 2011; CARBON et al., 2019).

A *PLoS Medicine* frisa, ainda na página “Licenses and Copyright”, que “se algum dado relevante for enviado aos repositórios com políticas de licenciamento declaradas, as políticas [regramento do reuso] não deverão ser mais restritivas do que a CC BY” (tradução nossa). Com essa diretriz, a *PLoS Medicine*, ainda que de maneira indireta, sem mencioná-los, indica que aceita instrumentos que declarem em domínio público um conjunto de dados subjacentes. Isto porque instrumentos como a CC0, da Creative Commons, ou a PDDL, da Open Data Commons (ver Quadro 7), são de cunho ainda mais permissivo do que a adotada CC BY, pois retiram todos os direitos patrimoniais do autor ou detentor de direitos autorais do referido conjunto de dados de pesquisa.

Contudo, declarar que admite apenas a licença CC BY para todos os seus conteúdos e recomendar o repositório Dryad, por exemplo, que adota somente o instrumento CC0 Dedicção ao Domínio Público (admite raras exceções) para depósito de conjunto de dados de pesquisa subjacentes, pode acarretar incompatibilidades entre os direitos concedidos pela licença e o instrumento e causar dúvida no autor que deseja abrir seus dados. Afinal, que licença legal deverá seguir para disponibilizar publicamente seus dados? A falta de informação padronizada sobre licenciamento e direito autoral e o desconhecimento sobre o teor das licenças públicas, como as Creative Commons, por exemplo, podem ser um fator inibidor para o autor abrir seus dados subjacentes publicamente (TENOPIR, 2011; PANHUIS, et al., 2014; VASILEVSKY et al., 2017; CARBON et al., 2019).

4.1.3 *theBMJ*

O londrino *theBMJ* é um periódico médico publicado sem interrupção desde outubro de 1840 pelo BMJ Group, conglomerado editorial ligado à Associação Médica Britânica (British Medical Association), que publica mais de 70 revistas médicas e de ciências afins e presta apoio ao Serviço Nacional de Saúde (National Health Service - NHS) britânico em capacitação profissional. Hoje, adota acesso aberto e gratuito a todos os artigos de pesquisa original, cobrando taxas de processamento dos artigos aos autores que desejarem fornecer acesso aberto a eles. Para outro tipo de conteúdo, adota o modelo de assinatura, demonstrando assim sua inclinação pelo modelo híbrido de publicação em acesso aberto (BJÖRK; SOLOMON, 2014; WEITZEL, 2014). Publica, além de resultados originais de pesquisa,

artigos de revisão, artigos comentando os fatores clínicos, científicos, sociais, políticos e econômicos que afetam a saúde, cartas, jornalismo investigativo e ajudam os médicos a executar a pesquisa clínica e pública, de acordo com sua página sobre “modelo de publicação”. Além da versão *on line*, tem ainda duas versões impressas semanais e uma mensal, que totalizam 122 mil exemplares (THEBMJ, 2016).

O periódico adota as seguintes políticas editoriais: (a) a publicação contínua e a revisão por pares aberta desde 2014 para trabalhos de pesquisa e análise, cujo modelo inclui a publicação do histórico da avaliação ao lado do artigo, mediante a concordância de pareceristas em ter seus comentários assinados publicados; (b) o *fast track* (publicação acelerada, em geral até 24 horas após a submissão) para artigo ou número especial, quando há uma razão de política pública para publicação urgente, como epidemias e pandemias, caso do vírus Zika e, mais recentemente, do Covid-19; (c) a cobrança de taxa de processamento de artigo, para disponibilizar o texto em acesso aberto, no valor de US\$ 4800 na aceitação do manuscrito; e (d) depósito dos artigos para acesso aberto no repositório PubMed Central, também em atendimento a exigências de financiadores de pesquisa (THEBMJ, 2016; GUANAES; ALBAGLI, 2019).

Em relação ao modelo de direitos autorais adotado para artigos científicos, há três possibilidades: a licença CC BY Não Comercial (NC), que adota como padrão; a CC BY, para resultados de pesquisa financiada com recursos públicos pelo NIH, Research Councils UK, Wellcome Trust e agências similares que exigem a aplicação dessa licença da Creative Commons, em cumprimento à política de acesso aberto dessas agências; e uma licença própria exclusiva (ou não exclusiva, para funcionários públicos que não podem concedê-la). A CC BY NC, licença padrão da revista, dá as mesmas permissões da CC BY, exceto a permissão de uso comercial, o que pode vir a ser feito se o *theBMJ* consentir (BMJ, 2018; THEBMJ, 2020).

No que se refere à abertura e ao compartilhamento de dados de pesquisa subjacentes, o *theBMJ* (2015, *on line*, tradução nossa) compartilha dados desde 2015. Aplica-se “a todos os ensaios clínicos submetidos, não [mais] apenas àqueles que testam medicamentos ou dispositivos médicos”, conforme política de compartilhamento de dados iniciada em 1º de janeiro de 2013 (THEBMJ, 2012). O periódico segue o conjunto de diretrizes do ICMJE para compartilhamento de dados, que publicou editorial em 2017, no *Annals of International*

Medicine, no qual detalha como as revistas que seguem a sua cartilha devem proceder para abrir dados provenientes de ensaios clínicos (TAICHMAN et al., 2017).

Corroborando as políticas de compartilhamento de dados dos periódicos PLoS (PLOS MEDICINE, 2014) e dos periódicos Nature (2016a), a do *theBMJ* também exige dos autores uma declaração de compartilhamento para todos os trabalhos de pesquisa, incluindo ensaios clínicos e excluindo relato de um estudo, além da inclusão de um plano de compartilhamento de dados. A declaração deve indicar: se os dados anonimizados⁷² ou não dos participantes serão compartilhados; quais dados em particular serão compartilhados; se os documentos suplementares relacionados estarão disponíveis; quando os dados ficarão disponíveis e por quanto tempo acessíveis; por quais critérios de acesso os dados serão compartilhados (inclusive com quem, para quais tipos de análises e por qual mecanismo) (ICMJE, 2018; THEBMJ, 2019).

Ademais, o ICMJE (2018) consigna em suas Recomendações que não deve haver violação ao direito de privacidade dos participantes da pesquisa. Informações que possam identificar o sujeito de pesquisa não devem ser publicadas, a menos que seja essencial para fins científicos específicos, mas com o consentimento livre e esclarecido por escrito do paciente para a publicação. Esta postura frente à proteção da privacidade de seres humanas participantes de pesquisa em saúde é detalhada em uma página de ajuda a pesquisadores “BMJ Author Hub”⁷³ e em um artigo publicado na revista *BMJ Open* (OHMANN et al., 2017).

O *theBMJ*, a par das diretrizes do ICMJE, tem a sua própria política, que “se aplica a qualquer artigo de pesquisa que relate os principais pontos finais de um estudo controlado randomizado” (THEBMJ, 2019, *on line*, tradução nossa). Incentiva os autores a vincularem seus artigos aos dados brutos de seus estudos e a citarem os dados na lista de referências, de acordo com as orientações do DataCite⁷⁴. Quando se tratar de ensaios clínicos, os autores devem se comprometer a disponibilizar os dados anonimizados mediante um pedido. Porém, se o artigo não relata um ensaio clínico, não é exigido que os autores concordem com o

⁷² Dado anonimizado é o dado relativo a titular que não possa ser identificado, considerando a utilização de meios técnicos razoáveis e disponíveis na ocasião de seu tratamento (BRASIL, 2018).

⁷³ A “BMJ Author Hub” é uma página de apoio ao pesquisador, com informações sobre como enviar o artigo, escrita e formatação, promoção do trabalho, políticas e acesso livre. Disponível em: <<https://authors.bmj.com/>>

⁷⁴ Ver em: <<https://datacite.org/value.html>>

compartilhamento dos dados, mas informem se desejam fazê-lo. Se uma solicitação de réuso de dados for recusada, os autores do artigo devem explicar por quê. Os solicitantes de dados devem estar preparados para fornecer aos autores do trabalho solicitado um protocolo detalhado para o estudo proposto e para fornecer informações sobre o financiamento e recursos para realização do estudo, além de poder convidar o autor original para participar da reanálise (OHMANN et al., 2017; THEBMJ, 2019).

Para as operações de depósito, curadoria, acesso, compartilhamento e réuso de dados subjacentes, o *theBMJ* recomenda a ajuda de re3data.org⁷⁵ e [FAIRsharing](https://fairsharing.org/)⁷⁶, que fornecem uma lista de repositórios com curadoria, e indica: para dados clínicos, repositórios de acesso controlado, como clinicalstudydatarequest.com, the YODA project, ou Vivli; para dados pré-clínicos, o GenBank; para dados em geral, os repositórios Dryad, Figshare, Zenodo (THEBMJ, 2019). Oferece a pesquisadores, por meio de política própria, a mineração de textos e dados.

4.1.4 Direito autoral: diretrizes para artigos e dados se confundem

Por adotar a política de abertura de dados do ICMJE, é razoável concluir que o *theBMJ* também adota suas diretrizes para fins de regulação dos direitos autorais. Contudo, a política de direito autoral do ICMJE quanto a dados de pesquisa subjacentes não é detalhada, pois seu objetivo é dar orientações gerais, cabendo aos editores filiados desenvolverem uma política que atenda às especificidades de cada campo. A única menção direta a direitos autorais feita por esse comitê de editores médicos internacionais sobre dados em suas recomendações é a seguinte:

Os periódicos devem esclarecer o tipo de direitos autorais sob o qual o trabalho será publicado, e se o periódico retiver os direitos autorais, deve *descrever detalhadamente* como será realizada a *transferência desses direitos* de todos os tipos de conteúdos, *incluindo* áudios, vídeos, protocolos e *bases de dados* (ICMJE, 2018, tradução nossa, grifos nossos).

⁷⁵ Ver em: <<http://re3data.org/>>

⁷⁶ Ver em: <<https://fairsharing.org/>>

O *theBMJ*, no entanto, não divulga em local de fácil acesso – como estão as seções de seu *site* “Modelo de publicação” (THEBMJ, 2016), “Compartilhamento de dados” (THEBMJ, 2019) e “Direitos autorais, acesso aberto e permissão de reuso” (THEBMJ, 2020) – uma “descrição detalhada” sobre como será realizada a “transferência de direitos autorais” de bases de dados ou de conjuntos de dados, ignorando a citada recomendação do ICMJE (2018).

Só iremos encontrar uma referência do *theBMJ* à sua gestão de direitos autorais e outros direitos de propriedade intelectual para bases ou conjuntos de dados e dados, na página “BMJ Author Hub”⁷⁷, na seção “Policies/Copyright and authors’ rights”, no item “Work” (Obra) do documento em PDF “Author License BMJ Publishing Group Limited (BMJ): BMJ Journals – BMJ wholly owned or co-owned journals” (BMJ, 2018). Tal documento descreve a licença própria do BMJ, misturada a cláusulas idealizadas para um contrato de cessão de direitos autorais cujo objeto principal é o artigo.

A denominada “Author License BMJ”, que, para efeitos legais, estabelece as condições em que os direitos autorais são cedidos pelo autor ao periódico e os usos que lhe são permitidos, assim define “Obra”:

É o trabalho que você [autor] enviou para publicação no periódico. A Obra incluirá todo o texto, áudio, vídeo e material audiovisual, resumos, *bancos de dados*, tabelas, *dados*, [...] incluindo todos os rascunhos da Obra, a versão da Obra aceita para publicação pelo BMJ e a versão do registro” [a que é efetivamente publicada] (BMJ, 2018, tradução nossa, grifos nossos).

Assim, ao se referir no citado documento a ‘dados’ e ‘base de dados’ como partes integrantes, que estão incluídas na Obra que lhe é submetida, depreende-se que a licença própria do BMJ que dispõe sobre direitos autorais também abrange os dados de pesquisa subjacentes.

Nessa “licença de autor do BMJ”, o autor “concede uma licença mundial, irrevogável e isenta de *royalties* ao BMJ [...] em perpetuidade”, para o periódico exercer diversos direitos, tais como editar, adaptar, publicar, distribuir, exhibir, reproduzir, traduzir e armazenar a obra, para proteger violação de direitos da propriedade intelectual, inclusive direitos autorais. O

⁷⁷ Ver em: <<https://authors.bmj.com/>>

documento frisa que a licença concedida é exclusiva, o que significa que o autor só poderá fazer usos da obra determinados no contrato (BMJ, 2018).

A obra em acesso aberto, portanto os artigos e os bancos ou conjuntos de dados subjacentes que os integram, pode ser reutilizada sob os termos da licença Creative Commons Atribuição Não Commercial (CC BY NC 4.0), que, como já dito, é a licença padrão da revista, dá as mesmas permissões da CC BY, mas não permite uso comercial. Em casos de artigos financiados pelo Wellcome Trust, UK Research and Innovation, NIH ou qualquer outro financiador que exija o uso da licença CC BY, ou onde o BMJ concordou expressamente que a licença CC-BY deverá ser aplicada, a obra poderá ser reusada sob os termos da licença CC BY 4.0 ou sua versão subsequente.

Desta forma, o *theBMJ* adota a licença CC BY NC como padrão, a CC BY para agências que exigem essa licença da Creative Commons e uma dura licença exclusiva própria (BMJ, 2018). Como visto, as diretrizes sobre direitos autorais para artigos se confundem com as ditadas para bases ou conjuntos de dados e dados. No entanto, para serem depositados no Dryad, por exemplo, repositório com o qual o *theBMJ* tem parceria, todos os conjuntos de dados enviados devem obedecer aos termos do instrumento CC0 Dedicção ao Domínio Público, da Creative Commons, ferramenta muito mais liberal do que as licenças citadas. Assim, na prática, a relação *theBMJ*-Dryad pode estar funcionando, mas em tese é bem provável que cause dúvidas expressivas na área legal em autores que queiram disponibilizar seus conjuntos de dados em acesso aberto.

4.1.5 *Nature Medicine*

Lançada em 1995, a revista *Nature Medicine* pertence ao portfólio da NatureResearch, uma divisão internacional do conglomerado editorial Springer Nature. Por isso, utiliza, além de poucas diretrizes de seus editores, as políticas editoriais e de abertura de dados de pesquisa dessa divisão. É um periódico mensal de acesso pago (compra ou aluguel de artigos e assinatura⁷⁸), revisado por pares, que publica artigos de pesquisa originais, que variam desde novos conceitos em biologia humana e patogênese de doenças até bases pré-clínicas robustas para novas modalidades terapêuticas e desenvolvimento de medicamentos a todas as fases do

⁷⁸ Preços vigentes em 01/04/2020: Assinatura – Instituições (sob consulta); Individual: Impressa&on line (US\$225.00); On line (US\$62.00). Aluguel ou compra de artigo (a partir de US\$8.99). Fonte: <https://www.nature.com/nm/subscribe>

trabalho clínico. Interessa-se também por tecnologias inovadoras que serão usadas em futuro próximo, destinadas a melhorar a saúde humana, de acordo com a página “About the Journal”⁷⁹ do seu *site*.

A NatureResearch não exige que os autores de artigos de pesquisa original de suas revistas, como a *Nature Medicine*, cedam seus direitos autorais a ela. Os autores concedem uma licença exclusiva para a NatureResearch publicar. Em troca, eles podem reutilizar seus artigos em futuros trabalhos impressos sem pedir permissão ao editor da revista. No entanto, para artigos sob encomenda, os direitos autorais são retidos pela Nature Research⁸⁰.

Em algumas circunstâncias excepcionais, como uma emergência em saúde pública, revistas dessa divisão da Spring Nature, que cobram pelo acesso (compra ou aluguel de artigos e assinatura), podem publicar tipos específicos de artigos sob uma licença Creative Commons, a fim de favorecer o reúso. Em tempo de Covid-19, por exemplo, a *Nature Medicine* se soma a outros periódicos ao oferecer acesso aberto e gratuito à publicação semanal “Covid-19 Research in Brief”. Em destaque logo na primeira página do *site* da revista, traz um resumo de todas as pesquisas recentes realizadas pela ciência em resposta à pandemia. Os editores autorizam inclusive a mineração de textos e dados.

Ainda na citada página “About the Journal”, na seção “Clinical Research”, *Nature Medicine* informa que segue as recomendações do ICMJE (2018) e da Rede de aprimoramento da qualidade e transparência da pesquisa em saúde (EQUATOR), para registro e notificação de estudos clínicos.

Para a abertura e compartilhamento de dados clínicos, seguindo políticas do ICMJE e da NatureResearch, exige uma declaração de disponibilidade dos dados e, a partir de janeiro de 2019, passou a requerer a inclusão de um plano de compartilhamento de dados no registro do ensaio clínico (NATURE MEDICINE, 2019).

Em relação à abertura e ao compartilhamento de dados subjacentes aos demais artigos, *Nature Medicine* adota, como todos os periódicos que exibem “Nature” no título, as políticas da NatureResearch, divisão da Springer Nature à qual pertence. Desta forma, na página “Políticas Editoriais” (Editorial Policies), na seção “Padrões de relato e disponibilidade de

⁷⁹ Ver em: <<https://www.nature.com/nm/about>>

⁸⁰ Ver em: <<https://www.nature.com/nature-research/editorial-policies/self-archiving-and-license-to-publish>>

dados, materiais, código e protocolos” (Reporting standards and availability of data, materials, code and protocols)⁸¹, a primeira orientação que o autor recebe é a seguinte:

[...] Uma condição para publicar em um periódico da NatureResearch é a exigência que fazemos aos autores para que disponibilizem prontamente materiais, dados, código e protocolos associados aos leitores sem restrições indevidas (grifos do site, tradução nossa).

Este aviso, encontrado também nas outras duas revistas analisadas (*PloS Medicine* e *theBMJ*), não poderia ser considerado um desvio ético, por seu caráter autoritário, impositivo?

Desde 2016, a NatureResearch passou a exigir uma declaração de disponibilidade de dados para todos os seus periódicos (GRANT; HRYNASZKIEWICZ, 2018; NATURE NANOTECHNOLOGIE, 2020), incluindo aqueles que cobram pelo acesso, a qual deve constar nos manuscritos de pesquisas originais com informação clara sobre quais dados estão disponíveis, quando e para quem, e como acessá-los, indicando ainda quaisquer restrições ao acesso a dados. A alegação para essa declaração é a de que outros devem ser capazes de encontrar os dados e materiais de pesquisa, de replicar e construir sobre as afirmações publicadas pelos autores (NATURE, 2016a; NATURE MEDICINE, 2019).

A política determina que os dados subjacentes aos artigos devem ser disponibilizados aos editores e revisores no momento da submissão para avaliação e que os autores devem incluir citações para conjuntos de dados nas listas de referências de artigos, após terem recebido DOIs, atribuídos por um repositório de dados, obedecendo a normas mínimas do DataCite⁸².

O periódico é favorável ao compartilhamento de grandes conjuntos de dados por meio de repositórios públicos, que coordenem a liberação pública dos dados com a data da publicação da revista. O site informa que é de responsabilidade do autor se comunicar com o repositório para esse fim. Para tanto, publica uma lista de repositórios recomendados no site da revista *Scientific Data* (2016). O periódico indica que repositórios não estruturados como Figshare, Zenodo e Dryad são alternativas adequadas, se não existirem os estruturados. A

⁸¹ No original: “[...] A condition of publication in a Nature Research journal is that *authors are required to make materials, data, code, and associated protocols promptly available to readers without undue qualifications.* [...]”. Fonte: <<https://www.nature.com/nature-research/editorial-policies/reporting-standards#availability-of-data>>

⁸² Ver em: <<https://datacite.org/value.html>>

Nature Medicine incentiva os autores a considerarem a publicação de um *data descriptor* no *data journal Scientific Data*, da própria Nature⁸³.

Por fim, a *holding* Spring Nature oferece a pesquisadores e agências de financiamento o “Research Data Support”⁸⁴, uma cesta de serviços pagos de apoio a dados de pesquisa, que incluem curadoria, aconselhamento de editor profissional de dados de pesquisa, coordenação da publicação e embargos etc. Os conjuntos de dados são publicados no repositório Springer Nature Figshare e continuam sob gerenciamento da equipe de apoio a dados de pesquisa da Springer Nature. O preço do serviço, em três diferentes moedas nacionais, sai a € 315 / £ 285 / US\$ 370, mais IVA ou impostos locais⁸⁵. O serviço de suporte a dados de pesquisa não reivindica direitos autorais dos dados enviados e recomenda a adoção da licença Creative Commons CC BY ou do instrumento CC0 Dedicção ao Domínio Público.

4.1.6 Direitos autorais: poucas orientações

As diretrizes da política de direitos autorais para dados de pesquisa da *Nature Medicine* são aquelas explicitadas pela *holding* Spring Nature e indicadas por sua subsidiária NatureResearch, de cujo portfólio o periódico faz parte. Elas estão na página “Research data policies”, do *site* da *holding*, abrigadas na seção “Research data policies FAQs”⁸⁶. Em uma pergunta sobre direitos autorais e licenças para dados depositados, a resposta é que os repositórios onde os dados são mantidos orientarão essas questões “pelos termos da licença ou ferramenta legal que permita o potencial máximo de reutilização, como a CC0 Dedicção ao Domínio Público, da Creative Commons” (tradução nossa). Em seguida, nesta mesma resposta à pergunta, para tirar outras dúvidas, a *holding* indica um “conjunto de perguntas frequentes” sobre licenciamento aberto de dados de pesquisa de outra subsidiária sua, a BioMed Central. A resposta finaliza com a afirmação de que “A Springer Nature não reivindica direitos de propriedade intelectual nos conjuntos de dados enviados” (tradução nossa).

⁸³ Ver em: <<https://www.nature.com/nature-research/editorial-policies/reporting-standards#availability-of-data>>

⁸⁴ Ver em: <<https://www.springernature.com/gp/authors/research-data/research-data-support>>

⁸⁵ Ver em: <<https://www.springernature.com/gp/authors/research-data/research-data-support/pricing-for-research-data-support>>

⁸⁶ Ver em: <<https://www.springernature.com/gp/authors/research-data-policy/data-policy-faqs>>

A FAQ da Biomed Central⁸⁷ sobre dados abertos é mais completa. Admite que restrições legais e incertezas em torno dos dados científicos são uma barreira ao compartilhamento e à reutilização, indicando os direitos autorais como os mais problemáticos para os dados (HRYNASZKIEWICZ, 2019). Às vezes, sinaliza a FAQ, não há clareza se os dados são protegidos por direitos autorais e a lei difere muito internacionalmente. A Biomed Central adota: a) Licença de Atribuição CC BY 4.0 para todos os seus artigos; b) Ferramenta CC0 Dedicção ao Domínio público para dados de pesquisa; c) Incentiva a citação da fonte; e d) permite a mineração de textos e dados (HRYNASZKIEWICZ; COCKERILL, 2012; HRYNASZKIEWICZ; BUSCH; COCKERILL, 2013)

Em uma pergunta sobre de quem é a responsabilidade de verificar se os autores têm permissão para compartilhar dados como material suplementar ou depositar em um repositório de dados, a resposta é que os pesquisadores só podem compartilhar dados para os quais têm permissão legal para compartilhar ou publicar. Segundo a FAQ, em geral é responsabilidade do depositante (autor) garantir que eles tenham os direitos ou permissões necessárias para enviar dados para um periódico ou repositório.

O objetivo de analisar como um periódico que cobra por acesso a seus artigos lida com a questão do direito autoral na abertura de dados de pesquisa subjacentes era para identificar eventuais diferenças em relação a periódicos de acesso aberto e gratuito ante essa questão nova. Note-se que a *Nature Medicine* cobra por acesso a seus artigos, mas exige que o pesquisador declare que concorda com a disponibilização pública dos dados que fundamentam sua pesquisa para fins de abertura e compartilhamento, o que poderia indicar de certa forma a influência do movimento pela abertura de dados de pesquisa. Além disso, destacamos a confusão de fontes de diretrizes para abertura e compartilhamento de dados (Spring Nature, NatureResearch, BioMed Central) e as magras orientações sobre direito autoral em dados científicos fornecidas pelo próprio periódico, o que confirma parte da literatura sobre o tema, que sugere a falta de padrão, conforme visto na seção 2.6 deste estudo.

No Quadro 8, elencamos alguns requisitos de políticas para disponibilização de dados de pesquisa subjacentes a artigos científicos identificados em cada revista pesquisada.

⁸⁷ Ver em: <<https://www.biomedcentral.com/about/policies/open-data>>

Quadro 8. Requisitos das políticas de dados de pesquisa de cada revista

	<i>Plos Medicine</i>	<i>theBMJ</i>	<i>Nature Medicine</i>
Pede declaração de concordância do autor para compartilhar dados	Sim, no momento da publicação	Sim	Sim, em caso de recusa, envia caso à agência financiadora
Rejeita artigo se não concordar em compartilhar	Sim	Sim	Sim
Tipo de licença ou instrumento jurídico que recomenda para reuso	CC BY, no mínimo	CC BY-NC (padrão), CC BY (agências), licença própria exclusiva	Licença do repositório; Serviço de apoio pago sugere CC BY e CC0
Plano de gestão ou de compartilhamento de dados	Incentiva	Exige	Exige
Citação de dados	Incentiva	Recomenda	Exige
Liberar dados imediatamente para publicação	Sim	Só para agências financiadoras	Sim
Privacidade de dados	Sim, encorajam uso de dados anônimos ou não identificados	Sim, dados desidentificados	Sim
Recomenda repositório	Incentiva autor a depositar	Sim; e para dados clínicos, repositórios de acesso controlado	Incentiva autor a depositar; repositório temático
Acesso aberto	Sim	Sim; e acesso controlado para dados clínicos	Sim
Libera dados após publicação	Sim, envia-os para repositório Figshare	Sim	Sim
Princípios FAIR	Sim	Sim	Sim
Curadoria	--	--	Fornece como serviço opcional pago
Revisão por pares de dados	Sim	--	Sim

Fonte: Elaboração própria (2020).

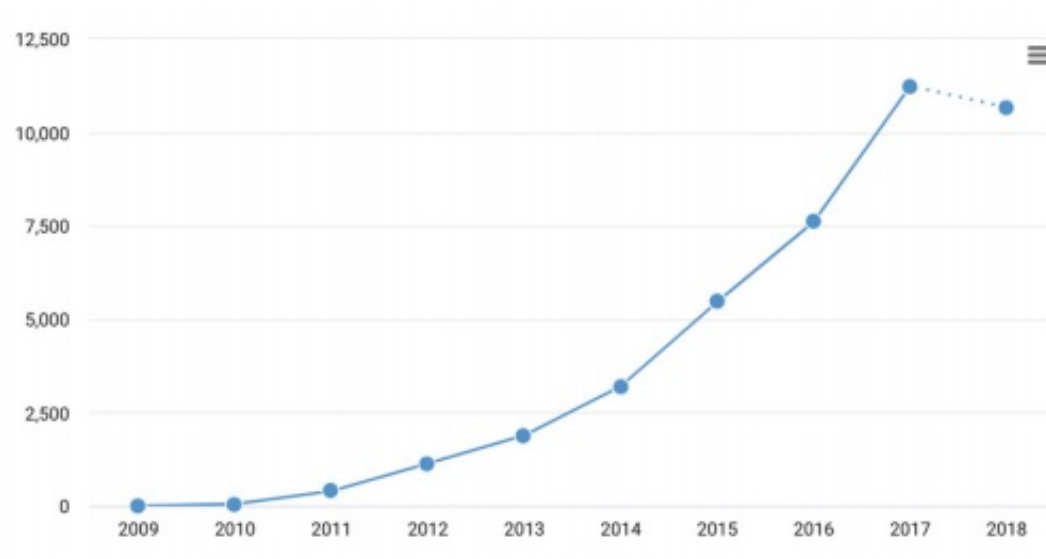
4.2 Os repositórios se reinventam

Provavelmente, um dos pontos de inflexão de um novo modelo de comunicar resultados de pesquisa é a exigência que agências de financiamento à pesquisa e revistas científicas passaram a fazer ao pesquisador, a fim de que deposite os dados e conjuntos de dados resultantes do ciclo de vida de suas pesquisas financiadas com recursos públicos em repositórios de acesso aberto específicos para dados de pesquisa. O intuito declarado é contribuir para o reuso o mais amplo possível de interessados (PLOS MEDICINE, 2014; NATURE, 2016a; BILL & MELINDA GATES FOUNDATION, 2017a; 2017b; ICMJE, 2018; THEBMJ, 2019; NIH, 2020).

Essa prática vem se generalizando, conforme mostrado na Figura 4, publicada no relatório *The State of Open Data 2018* (DIGITAL SCIENCE, 2018), produzido pelo repositório Figshare, que analisa as atitudes globais em relação aos dados abertos. A Figura 4 indica crescimento dos conjuntos de dados que estão sendo disponibilizados. Em uma entrevista que o relatório realizou com pesquisadores, 64% dos entrevistados revelaram que disponibilizaram dados abertamente em 2018, correspondendo a um aumento de 4% em relação a 2017 e 7% a 2016. Outro fator relevante para pesquisadores se decidirem a depositar dados de pesquisa em acesso aberto, de acordo com Strasser (2014), são as citações de dados⁸⁸. Segundo o relatório, elas motivam mais os entrevistados a disponibilizarem abertamente os dados, e a prática da citação de dados aumentou de 7% em 2017 para 46% em 2018 (DIGITAL SCIENCE, 2018, p. 8).

⁸⁸ A Joint Declaration of Data Citation Principles estabeleceu em 2014 que os dados da pesquisa são parte integrante do registro científico e acadêmico. Por isso, devem ser citados e referenciados. As citações e referências de dados de pesquisa seguem padrões semelhantes às de textos científicos, devendo inclusive lhes ser atribuído DOI. Ver em: <<https://www.force11.org/datacitationprinciples>>

Figura 4. Evolução de conjuntos de dados de pesquisa disponibilizados em repositórios de dados, 2009-2018.



Fonte: DIGITAL SCIENCE, 2018.

Desta forma, o repositório de dados tornou-se uma infraestrutura-chave para a publicação de dados de pesquisa subjacentes a artigos, conforme afirma Assante et al. (2016), em pesquisa que analisou as soluções oferecidas por repositórios de dados generalistas, entre eles o Dryad e o Figshare, recomendados a autores por periódicos como *theBMJ*, *Plos Medicine* e pelo periódico de dados *Scientific Data* (SCIENTIFIC DATA, 2016; DIGITAL SCIENCE, 2018; GUANAES; ALBAGLI, 2019).

Se, por um lado, a indicação de um repositório por parte do editor aparentemente o distancia do complexo processo operativo de disponibilização pública dos dados de pesquisa, resguardando-o das atividades diretas de armazenamento, curadoria, acesso, compartilhamento, reúso e eventuais discussões sobre direitos autorais e licenciamento de conjuntos de dados, por outro lado, a medida confere ao repositório um papel essencial na infraestrutura mundial de pesquisa, que vai além dos tradicionais armazenamento e preservação de documentos (ASSANTE et al., 2015; BANZI et al., 2019; KINDLING et al., 2017).

Por protagonizar esse novo papel, o repositório de dados dá indícios de introduzir uma nova dinâmica na comunicação científica, além de ser um dos possíveis fatores de emergência de um novo regime de direito autoral, dois temas caros à presente pesquisa (ASSANTE et al., 2015; 2016; SAYÃO; SALES, 2016a; COALITION S, 2018).

Um dos projetos mais valiosos para a abertura de dados de pesquisa foi a criação do re3data.org⁸⁹, um registro global de repositórios de dados de pesquisa que reúne repositórios de diferentes disciplinas acadêmicas e permite uma escolha informada do mais adequado à natureza da pesquisa (KINDLING et al., 2017). A Universidade de Harvard disponibiliza gratuitamente o Dataverse⁹⁰, *software* que cria um repositório de dados de pesquisa personalizado para pesquisadores, revistas e instituições. Existem outras opções, como o sistema de gerenciamento de dados CKAN⁹¹, um *software* livre de código aberto, cuja sede legal é na Open Knowledge Foundation, mas que opera como um órgão autônomo.

4.2.1 Repositórios de dados generalistas e o direito autoral

Os repositórios de dados de pesquisa generalistas, que estão abertos para publicar qualquer tipo de conjunto de dados de pesquisa, de qualquer comunidade específica (ASSANTE, 2016; VISION, 2010), têm dado uma significativa contribuição no quesito infraestrutura para disponibilizar dados abertos (ASSANTE et al., 2015). Estão modificando o tradicional modo de fazer comunicação científica, ao

[...] fornecerem gerenciamento a longo prazo dos dados [...] a aplicação de identificadores [DOI, ORCID] a todo o material, para que seja citado com precisão, e os criadores dos dados originais sejam devidamente reconhecidos quando os dados são reutilizados, [e] ajudam na preparação de dados para compartilhamento e no gerenciamento de processos às vezes complexos de solicitação e concessão de acesso (BANZI, 2019, *on line*, tradução nossa),

instaurando a publicação ampliada (artigo mais descrição de dados de pesquisa) e nos dando a pista de novos rumos para a comunicação científica (SAYÃO; SALES, 2016b; COAR, 2017).

⁸⁹ Ver em: <<http://re3data.org/>>

⁹⁰ Ver em: <<https://dataverse.org/>>

⁹¹ Ver em: <<https://ckan.org/>>

Para conhecer as políticas desses repositórios relativas a direitos autorais aplicáveis a dados de pesquisa, analisaremos, diretamente em seus *sites*, três repositórios do universo generalista: Dryad, Figshare e Zenodo.

Com o objetivo de manter certa simetria na referida análise, tomamos como objeto preferencial de pesquisa os “Terms of Service” ou “Terms of Use” dos repositórios, sem prejuízo de obtermos informações em outras seções dos respectivos *sites* e blogues. São tópicos apresentados nos *sites* dos repositórios em questão que estipulam condições para orientar e regular juridicamente as relações entre repositório, autor ou detentor dos direitos autorais e terceiros que fazem o reuso dos dados postos em disponibilidade. Portanto, a meta é mapear tanto as políticas de proteção aos direitos autorais como o modelo de licenciamento de dados de pesquisa explicitados nesses termos e, eventualmente, em outras seções, buscando contribuir para apresentar um retrato dessa instância da abertura de dados de pesquisa.

4.2.2 Dryad

A versão original do repositório de dados de pesquisa Dryad⁹² foi lançada em 2009 nos EUA. Seu desenvolvimento foi realizado com base no software de repositório de código aberto DSpace (DRYAD, 2020, *on line*). Desde então, selecionou, publicou e arquivou quase 30 mil conjuntos de dados subjacentes a artigos acadêmicos, abrangendo as áreas de ciências da vida e biomédica, além das de ecologia e biologia evolutiva, que lhe deram origem (CDL, 2019, *on line*).

Em 2019, o Dryad se fundiu ao Dash, um serviço de publicação de dados desenvolvido no Centro de Curadoria da Universidade da Califórnia (UC3), um programa da California Digital Library (CDL). Desta fusão resultou uma “plataforma de publicação de dados aberta e com curadoria para os pesquisadores compartilharem e publicarem seus dados” (CDL, 2020, *on line*). Além de publicar e fazer a integração de conjuntos de dados de pesquisa subjacentes a artigos científicos com as revistas em que são publicados, a plataforma passou a receber o depósito de conjuntos de dados enviados por pesquisadores de maneira independente, a oferecer estatísticas padronizadas de uso e citação de dados atualizadas e exibidas para cada conjunto e aumentou o tamanho do limite de arquivos para 300 GB (CDL, 2019; 2020, *on line*).

⁹² Ver em: <<https://datadryad.org/stash>>

O repositório observa os princípios FAIR (Quadro 5) (WILKINSON, 2016), usa identificadores persistentes, como o DOI, para cada conjunto de dados, e o ORCID, para a citação do autor original dos dados, e aceita artigos com embargos.

Organização sem fins lucrativos, o Dryad recebe dados de qualquer área do conhecimento, em qualquer formato e cobra uma taxa de publicação por envio de dados no valor de US\$ 120, que deve ser paga quando o conteúdo⁹³ é aprovado pelo pessoal da curadoria, atendendo a requisitos de políticas de financiadores de pesquisa e editoras, dentre eles Wiley, The Royal Society, PLoS e BMJ (VISION, 2010; Dryad, 2020).

De acordo com Greenberg et al. (2009), a versão inicial do repositório Dryad foi projetada para a preservação, acesso e reutilização de dados científicos subjacentes à pesquisa publicada no campo da biologia evolutiva, ecologia e disciplinas relacionadas. Foi apoiado por instituições norte-americanas, como a NSF, que financiou o projeto, e o seu desenvolvimento se deu por meio de colaboração conjunta que reuniu o National Evolutionary Synthesis Center (NESCent); o Metadata Research Center at the School of Information and Library Science, a University of North Carolina at Chapel Hill (UNC/CH), a North Carolina State University, a University of New Mexico e a Yale University.

O empreendimento teve como parceiros adicionais as principais sociedades científicas e periódicos do campo da biologia evolutiva. Uma das motivações dessa comunidade era se antecipar ao requisito futuro – que poderia se tornar mandatório – de depósito e abertura de dados de pesquisa financiada com recursos públicos, para estar em conformidade com a Política Conjunta de Arquivamento de Dados (em inglês, Joint Data Archiving Policy - JDAP), que era adotada pelos periódicos do campo da biologia evolutiva (GREENBERG et al. (2009).

4.2.3 Direitos autorais na perspectiva do Dryad

Para se precaver de uma forma consistente no que concerne à aplicabilidade de direitos autorais aos conjuntos de dados de pesquisa subjacentes a artigos científicos que abriga, o repositório Dryad solicita de pronto ao autor o consentimento para liberação dos dados ao domínio público. Em suas diretrizes, na seção “Terms of Service” de seu *site*, o

⁹³ Por ‘conteúdo’, entenda-se [o conjunto de] dados e outros materiais (por exemplo, código de software, arquivos *readme*, multimídia) (DRYAD, 2020, *on line*).

subitem 2.1, “Content Criteria”, informa que só aceita o envio de conjuntos de dados “por um indivíduo que declare e garanta que é o criador e o proprietário do Conteúdo ou que possua direitos suficientes para disponibilizar o Conteúdo sob uma Renúncia CC0” (instrumento CC0 Dedicção ao Domínio Público, da Creative Commons). O Dryad salienta, no entanto, nesse mesmo item, que se reserva o direito de disponibilizar o conteúdo sob outros termos de licenciamento (DRYAD, 2020, *on line*).

Tal posição pela adoção do instrumento CC0 Dedicção ao Domínio Público como regra ainda é reforçada no subitem 3.1, “Curation”, o qual estabelece que o repositório poderá, mas não é obrigado a revisar o Conteúdo para “obter informes sobre direitos autorais e licença que sejam potencialmente incompatíveis com uma Renúncia CC0”, bem como no subitem 5.1, “Permission”, quando explicita que ao “enviar o conteúdo para o repositório, o remetente concede permissão irrevogável ao Dryad para disponibilizar o conjunto de dados ao público sob uma renúncia CC0”. Tal não é diferente no subitem 5.2, “Representations and Warranties”, no qual o Dryad estabelece que

O remetente declara e garante que a publicação do conteúdo não viola a propriedade intelectual ou os direitos de privacidade de qualquer pessoa ou entidade (incluindo coautores ou cocriadores) e que a distribuição do conteúdo sob uma renúncia CC0 está em conformidade com todas as leis e diretrizes para editores, financiadores e empregadores e/ou restrições contratuais (Dryad, 2020, *on line*, tradução nossa).

Já no subitem 7.1, “Content Usage”, em outra menção à renúncia de direitos autorais sobre dados, o Dryad incentiva o reuso do Conteúdo do repositório “de qualquer maneira” [...], sendo que “Na medida do possível por lei, os remetentes renunciaram a todos os direitos autorais e direitos relacionados ou vizinhos a esses dados” (DRYAD, 2020, *on line*, tradução nossa).

No item 2.3, “Expressions of Concern, Retraction and Removal of Data Files” (Manifestações de preocupação, retração e remoção de arquivos de dados), o Dryad esclarece como age em uma disputa de direitos autorais. Quem desejar que um Conteúdo seja retirado, por violação de direitos autorais, questões de privacidade ou outra causa legítima, o repositório disponibiliza um *e-mail* para contato; pede a identificação do trabalho ou material questionado, para localização e eventual remoção provisória ou permanente do Conteúdo; dados do reclamante para contato; e uma declaração do reclamante de que o uso do material

não está autorizado pelo proprietário, seu procurador ou pela lei. O repositório informa também como procede se receber uma declaração do editor que expresse preocupação, retração ou algo similar associada ao Conteúdo. Neste caso, continua publicando o Conteúdo, mas com um aviso e um *link* para a declaração (DRYAD, 2020).

Os subitens 5.1 e 5.2 (DRYAD, 2020) suscitam a questão do controle sobre os direitos autorais, problema potencialmente controverso, apontado por Uhlir (2006). Para este autor, os pesquisadores são os criadores da literatura científica e têm os interesses proprietários originais e morais sobre seus trabalhos, mas muitas vezes a instituição empregadora é quem detém os direitos autorais, decorrentes das cláusulas de um contrato de trabalho. De acordo com pesquisa de Ginsburg (2004, apud UHLIR, 2006), cerca de um terço das universidades dos EUA reivindicou direitos autorais sobre os trabalhos produzidos por seus funcionários, sem que muitos percebessem o fato ou o ignorassem ao lidar com seus editores. Todavia, para Uhlir (2006), estudioso do tema que integra o *staff* do Dryad, pouco importa quem é o proprietário dos direitos autorais, pois um e outro são os únicos que podem fazer a cessão dos direitos autorais a terceiros.

Uhlir (2006) advoga que a transferência dos trabalhos [dados ou artigos] para os repositórios abertos atende tanto ao interesse privado dos autores quanto ao interesse público, se o autor se mantiver como detentor dos direitos autorais, “[...] para ter a flexibilidade de fazer a cessão de direitos de distribuição não exclusivos ao editor de registro, ao repositório e a todos os usuários *downstream*” (UHLIR, 2006, p. 102, tradução nossa).

No contexto do ano de 2006, Uhlir identificava uma grande confusão por parte dos autores sobre a retenção de direitos autorais. Eles desconfiavam da importância de deterem tais direitos, uma vez que a cessão de seus direitos autorais os impedia de depositar cópias de suas obras em repositórios abertos, mesmo sendo eles os donos dos direitos autorais das obras que produzem.

4.2.4 Por que o CC0?

A inclinação do Dryad pelo uso como regra do CC0 Dedicção ao Domínio Público, instrumento – não licença – de renúncia a direitos autorais (na extensão permitida por lei) da Creative Commons, encontra uma explicação mais ampla no documento “Why does Dryad use CC0?” (Por que o Dryad usa CC0?) (BLOG DRYAD, 2011), no qual o repositório

procura esclarecer a sua opção. Nele, argumenta que o CC0 foi criado para reduzir impedimentos legais, intencionais ou não, à reutilização de dados. Ressalva que o CC0 não isenta os pesquisadores de seguirem as normas de suas comunidades para comunicação científica, sobretudo a obrigação de citar os autores dos dados originais. E crê que “[...] essas expectativas são melhor articuladas e aplicadas pela própria comunidade mediante o emprego de processos como a revisão por pares” (BLOG DRYAD, 2011, *on line*, tradução nossa).

No tocante especificamente à escolha do instrumento CC0 Dedicção ao Domínio Público em detrimento das demais licenças, nesse documento o Dryad explica que “segue as recomendações da Science Commons [ramo da Creative Commons que visa fazer com que a *web* funcione para a ciência da maneira que atualmente funciona para a cultura⁹⁴], que desencoraja os pesquisadores de presumirem direitos autorais e usarem licenças que incluam condições de ‘atribuição’ [BY] e ‘compartilhar igual’ [*share-alike*] para dados científicos” (BLOG DRYAD, 2011, *on line*, tradução nossa). As licenças ‘BY’ exigem a atribuição ao criador da obra intelectual como a única condição para reutilizar o material. A “compartilha igual” também requer crédito ao criador, mas o maior problema é que as adaptações do material têm que ser liberadas sob a mesma licença.

No que se refere ao primeiro caso, o repositório afirma que “especificar como a ‘atribuição’ deve ser realizada pode colocar um usuário em desacordo com as práticas de citação aceitas”. Por exemplo, “quando um pesquisador compilar uma consulta a partir de 50.000 bancos de dados, estará sujeito a uma ação judicial caso não atribua autoria a 50.000 proprietários?”. Em relação ao segundo caso, de acordo com o Dryad, impor a condição a produtos derivados de adotarem licença idêntica cria “problemas significativos de ‘compatibilidade de licença’ [...]”, e “No contexto dos dados, os problemas de compatibilidade de licenças provavelmente criarão barreiras significativas para a integração e reutilização de dados, tanto para provedores quanto para usuários de dados” (BLOG DRYAD, 2011, *on line*, tradução nossa).

O Dryad conclui, nesta justificativa publicada em seu *blog*, que, em virtude do potencial de consequências negativas, intencionais ou não, do uso de direitos autorais, do alcance do domínio público e do poder das normas da ciência, “acreditamos que licenças de direitos autorais e restrições contratuais são simplesmente a ferramenta errada [para dados], mesmo que essas licenças e contratos sejam usados com a melhor das intenções.” (BLOG

⁹⁴ Ver em: <<http://www.dcc.ac.uk/resources/briefing-papers/legal-watch-papers/science-commons>>

DRYAD, 2011, *on line*, tradução nossa). Esta posição do Dryad de não recomendar o uso de licenças de direitos autorais para dados subjacentes à pesquisa publicada, como as da Creative Commons, vai ao encontro do que preconiza os Panton Principles (MURRAY-RUST et al., 2010), criados por pesquisadores dos EUA e do Reino Unido. Eles são contrários à adoção das licenças CC e defendem a dedicação de tais dados ao domínio público, seja via licença PDDL, seja pelo instrumento CC0 (Quadro 7).

O repositório Dryad observa ainda que a adoção do CC0 confere três vantagens relevantes: a interoperabilidade, por meio da qual “outras pessoas e serviços de indexação poderão determinar termos de uso”, uma vez que o CC0 é legível tanto por homem quanto por máquina; a universalidade, pois o CC0 “é um mecanismo único, global e universal, que abrange todos os dados e todos os países” (BLOG DRYAD, 2011, *on line*, tradução nossa), além de ser amplamente conhecido; e a simplicidade, já que dispensa que humanos façam ou respondam solicitação de dados individuais, o que significa mais cientistas se dedicando a fazer ciência.

4.2.5 Figshare

O Figshare⁹⁵ é um repositório *on line* de acesso aberto, voltado para instituições, editores e pesquisadores. Foi fundado em 2011 por Mark Hahnel, que desenvolveu inicialmente a plataforma como uma solução pessoal com o intuito de organizar e publicar produtos de pesquisa gerados em seu doutorado em biologia de células-tronco no Imperial College of Science, Technology and Medicine, em Londres. Com o conceito de permitir que pesquisadores possam armazenar, publicar, gerenciar e disseminar livremente qualquer tipo de resultado de pesquisa, incluindo figuras, conjuntos de dados, imagens e vídeos, disponibilizando-os ao público, o repositório foi relançado em Londres, em 2012, financiado pela Digital Science, divisão global de tecnologia da editora Macmillan Science & Education. Entrou como parceiro no lançamento do ORCID e teve a F1000Research como primeiro cliente editor (ASSANTE, 2016; WIKIPEDIA, 2019). Oferecia como novidade uma capacidade ilimitada de armazenamento público de dados de pesquisa, *links* de citação e licenciamento sob os termos da Creative Commons, entendendo que este era um dos fatores para incentivar a reutilização (MERRIL, 2012).

⁹⁵ Ver em: <<https://figshare.com/>>

A versão também abria a possibilidade de os pesquisadores publicarem resultados negativos de seus trabalhos, na crença de que “[...] dados publicamente disponíveis levam a mais citações” (PINOWAR; DAY; FRIDSMA, 2007, *on line*, tradução nossa) e, com isso, “[...] os pesquisadores podem acumular citações por seus esforços de outros pesquisadores que se baseiam no trabalho” (MERRIL, 2012, *on line*, tradução nossa).

Em 2013, PLoS e Figshare fecharam uma parceria, permitindo que os autores de revistas PLoS hospedassem seus conjuntos de dados no Figshare, aos quais é atribuído um DOI para torná-los citáveis. E também se beneficiassem dos recursos de acesso e visualização de conjunto de dados, arquivos de texto, documentos, animações, vídeos e arquivos de apresentação integrados ao artigo, ou seja, a visualização do material suplementar é feita na própria página do artigo sem a necessidade de baixar arquivos, em razão do uso de um *widget*, elemento de interação em uma interface gráfica (CREATIVE COMMONS, 2013; PLOS, 2013). Nesse mesmo ano, passou a ser rastreado pela Altmetrics, que mede o impacto da produção científica fora da academia, e lançou o repositório institucional, concentrado no aspecto de gerenciamento de dados de pesquisa, em resposta a “diferentes financiadores exigindo diferentes níveis de abertura, mas, acima de tudo, solicitando que os acadêmicos tenham um plano de gerenciamento de dados de pesquisa” (HAHNEL, 2013, *on line*, tradução nossa).

Em 2015, firma parceria com o DataCite (FIGSHARE, 2020a), consolidando uma questão muito discutida e reivindicada pelos pesquisadores, motivo de objeção a abrir dados de pesquisa, encontrada em pesquisas sobre o tema: a devida citação dos autores originais dos dados de pesquisa subjacentes a artigos que são reutilizados (KRATZ; STRASSER, 2014; TENOPIR, 2011; 2015).

Desde 2016, o Figshare publica o relatório *The State of Open Data*. Por meio de análises, artigos e pesquisas junto à comunidade científica, o relatório visa contribuir para se entender como a comunidade de pesquisa vê as oportunidades e os desafios dos dados abertos, a extensão da conscientização sobre dados abertos (ver, por exemplo, Figura 4, neste capítulo), os incentivos ao seu uso e as perspectivas que os pesquisadores têm sobre como abrir seus próprios dados de pesquisa. Sobretudo os que forem financiados pelo erário público.

O Figshare oferece serviços básicos gratuitos⁹⁶ para usuários: *upload* de arquivos até 5 GB de qualquer formato; 20 GB de espaço particular para armazenar pesquisa até a decisão de torná-la pública; espaço público ilimitado para cumprir mandatos de financiadores, editores e institucionais. O repositório ressalva que se reserva o direito de alterar tais serviços a qualquer momento. Também oferece “serviços *premium*”, com vários níveis de custo para os usuários, dependendo do volume de artigos, tamanho dos dados, integração com outros sistemas e complexidade da implementação. Oferece ainda um *hot site* para editores com serviços, tais como um portal onde dados suplementares associados a artigos podem ser exibidos e rastreados e um fluxo de trabalho contínuo de transferência de dados para revisão por pares, curadoria e publicação⁹⁷.

4.2.6 Direitos autorais: além das licenças CC

De acordo com a página “Copyright and Licence Policy”, o Figshare adota o instrumento CC0 Dedicção ao Domínio Público como padrão para os pesquisadores que desejarem compartilhar publicamente os seus conjuntos de dados. O repositório justifica a opção pela CC0 para conjunto de dados porque “Entendemos que é difícil para os pesquisadores saber se os dados de pesquisa estão sujeitos à lei de direitos autorais” (FIGSHARE, 2020c, *on line*, tradução nossa). Desta forma, segundo o Figshare, qualquer dúvida legal que os pesquisadores possam ter é removida, no entanto alerta que a adoção da CC0 desobriga quem reúsa de citar o autor original do conjunto de dados, mas frisa que não está afastada a “responsabilidade moral” de dar crédito. Para garantir a citação do autor do conjunto dos dados, o repositório recomenda a opção pela licença CC BY, que julga mais adequada para figuras, mídia, pôsteres, artigo acadêmico, conjuntos de arquivos (FIGSHARE, 2020c).

Cabe ressaltar que, além de quem reúsa ser obrigado a fazer a citação ao autor original, tanto por observância à licença quanto às normas acadêmicas, a CC BY ainda estabelece que aquele que fizer o réuso dos dados deve fornecer um *link* para a licença e ainda informar se foram feitas alterações, tendo o cuidado de não fazer de modo que sugira que o licenciante endossa o resultado daquele réuso. Apesar de ser, de fato, uma das licenças de

⁹⁶ Ver em: <<https://figshare.com/features>>

⁹⁷ Ver em: <<https://knowledge.figshare.com/publishers>>

espírito mais aberto, a CC BY “[...] pode não lhe dar todas as autorizações necessárias para o uso pretendido. [...] tais como direitos de imagem, de privacidade ou direitos morais, [que] podem limitar o uso do material” (CREATIVE COMMONS, 2020, *on line*, tradução nossa).

Ainda em relação ao instrumento CC0, o Figshare considera-o ideal para disponibilizar dados tabulares e banco de dados. Considera também que o CC0 é especialmente recomendado para dados científicos (FIGSHARE, 2020c). Tal entendimento é corroborado tanto por periódicos quanto por financiadores de pesquisa (PLOS MEDICINE, 2014; NATURE, 2016a; THEBMJ, 2017; COMISSÃO EUROPEIA, 2017), que aconselham a adoção do CC0 Dedicado ao Domínio Público principalmente para dados científicos por sua natureza simplificadora da relação repositório-(re)usuário. Na medida em que o licenciante dos dados de pesquisa subjacentes a artigos científicos escolhe o CC0, dedica “[...] o trabalho ao domínio público, renunciando a todos os seus direitos sob as leis de direito de autor e/ou de direitos conexos referentes ao trabalho, em todo o mundo, na extensão permitida por lei”. Tal fato pode significar uma maximização do reuso, pois quem reutilizar um banco de dados ou um conjunto de dados sob o instrumento CC0 “[...] pode copiar, modificar, distribuir e executar o trabalho, mesmo para fins comerciais, tudo sem pedir autorização” (CREATIVE COMMONS, 2018b, *on line*, tradução nossa).

Apesar da preferência pelo instrumento CC0 e pela licença CC BY, o Figshare põe à disposição de seus usuários as outras cinco licenças Creative Commons, inclusive as mais restritivas (ver Quadro 7). Além do pacote CC, oferece o “Modelo de licença restritiva”, desenvolvido para licenciar material que contenha informações pessoais ou confidenciais com alto risco de segurança, e a licença de atribuição Open Data Commons (ODC-By), para bancos de dados (FIGSHARE, 2020b).

É relevante frisar que a decisão sobre qual licença será usada cabe exclusivamente ao licenciante (autor ou detentor dos direitos autorais), o que não quer dizer que os editores e os repositórios de dados não possam recomendar a adoção dessa ou daquela licença (OPEN AIRE, 2019). A Creative Commons (2018c) incentiva os repositórios de dados de pesquisa a apresentarem alternativas de licenças, acompanhadas das explicações sobre cada conteúdo delas, para uma escolha consciente e informada.

Na página “Terms and Conditions” de seu *site* (FIGSHARE, 2020d), que funciona como um contrato entre as partes, o repositório explicita deveres do usuário nos quais são mencionados alguns aspectos procedimentais a respeito dos direitos autorais: exigência de

uma afirmação de boas práticas na vigência do contrato do procurador de um usuário e da sua capacidade jurídica; a ciência do usuário ou de seu procurador de que poderá indenizar o Figshare por uso indevido de acesso ao *site*, serviço ou conteúdo, bem como dos termos de uso e por violação a propriedade intelectual; as regras procedimentais em uma eventual disputa de direitos autorais; e orientações sobre como agir em caso de necessidade de relatar casos de violação de direitos autorais.

4.2.7 Zenodo

Com vistas a aprofundar a sua política de apoio à ciência aberta, em 2017 a Comissão Europeia estendeu o projeto piloto criado para incentivar o acesso aberto a artigos científicos, originados de pesquisas financiadas pela Comissão, para o acesso aberto a dados de pesquisa. Ao estender o piloto, a Comissão firmou o entendimento de que o acesso aberto se tornava a configuração padrão para os dados de pesquisa gerados no âmbito do programa Horizon 2020 (COMISSÃO EUROPEIA, 2017).

Os tipos de dados elegíveis para abertura são:

- 1- “dados subjacentes” (os dados necessários para validar os resultados de pesquisa apresentados em publicações científicas), incluindo os metadados associados (ou seja, metadados que descrevem os dados de pesquisa depositados); e
- 2- quaisquer outros dados (por exemplo, dados com curadoria não diretamente atribuíveis a uma publicação ou dados brutos), incluindo os metadados associados, conforme especificado e dentro dos prazos estabelecidos no plano de gestão de dados, isto é, de acordo com o planejamento de cada projeto/pesquisador financiado (COMISSÃO EUROPEIA, 2017, *on line*, tradução nossa).

Entre as incompatibilidades admitidas pela Comissão Europeia para a não participação de projetos financiados por ela na abertura dos dados, constavam a proteção de resultados que possam ser explorados comercial ou industrialmente; a confidencialidade por questões de segurança; e as regras de proteção a dados pessoais.

O repositório Zenodo⁹⁸ iria contribuir com esse projeto para dados abertos.

⁹⁸ Ver em: <<https://zenodo.org/>>

O Zenodo (2020a, *on line*) é um repositório generalista *on line*, criado pelo projeto OpenAir, que foi contratado pela Comissão Europeia para apoiar sua política de dados abertos, fornecendo um repositório abrangente. Está hospedado e é operado pelo CERN (Organisation Européenne pour la Recherche Nucléaire, em francês), laboratório localizado na Suíça, pioneiro em código aberto, acesso aberto e dados abertos, o qual por sua vez integra o programa europeu de infraestrutura European Open Science Cloud (EOSC)⁹⁹. O EOSC oferece um ambiente virtual com serviços abertos e integrados para armazenamento, gerenciamento, análise e reutilização de dados a pesquisadores e profissionais em ciência, tecnologia, humanidades e ciências sociais.

Lançado em 2013 e capacitado pelas ferramentas desenvolvidas pelo CERN para gerenciamento de *big data*, o repositório foi projetado para atender a “cauda longa” da ciência produzida por pesquisadores de instituições menores, de todos os campos da ciência, auxiliando-os a depositar e compartilhar gratuitamente arquivos digitais (dados incluídos) que resultem de pesquisa, financiada ou não pela União Europeia, representados por uma ampla variedade de formatos (SICILIA; GARCIA-BARRIOCANAL; SÁNCHEZ-ALONSO, 2017), entre eles publicações (livro, capítulos de livros, artigo de conferência, artigo de revista, patente, *preprints*, relatório, tese, nota técnica, documento de trabalho etc.), pôsteres, apresentações, conjuntos de dados, imagens (figuras, gráficos, desenhos, diagramas, fotos), software, vídeos/áudio e materiais interativos, como aulas (ZENODO, 2020b).

O Zenodo tem a singularidade de apoiar a criação de comunidades acadêmicas, cujo funcionamento fica a critério dos usuários do repositório. São altamente heterogêneas, incluindo editores, periódicos, eventos e projetos, além de coleções que usam critérios de inclusão baseados em tecnologia, como os dados vinculados, que é o caso da comunidade de dados que compartilha conjuntos de dados (SICILIA; GARCIA-BARRIOCANAL; SÁNCHEZ-ALONSO, 2017). Em janeiro de 2020, contava com 4.123 comunidades¹⁰⁰.

Repositório bastante abrangente, não impõe requisitos de formato, tamanho, restrições de acesso ou licença. De acordo com a apresentação no *site* do repositório,

⁹⁹ Ver em: <<https://www.eosc-portal.eu/about/eosc>>

¹⁰⁰ Ver em: <<https://zenodo.org/communities/>>

[...] dados, software e outros arquivos digitais de apoio a publicações podem ser o núcleo, mas igualmente bem-vindos são os materiais associados às conferências, projetos ou às próprias instituições, todos necessários para entender o processo acadêmico (ZENODO, 2020a, *on line*, tradução nossa).

O Zenodo também abriga conteúdo fechado e restrito, desde o início do ciclo de vida da pesquisa, que muitas vezes pode durar meses ou anos. Tal medida facilita a vida do pesquisador quando, ao final do trabalho, os dados e arquivos digitais forem disponibilizados publicamente. O repositório fornece *links* para revisores, ajuda os pesquisadores a receberem crédito, tornando os resultados da pesquisa citáveis mediante o uso de DOIs. E aceita conteúdo embargado, o qual será aberto automaticamente quando houver a decisão por sua abertura (ZENODO, 2020a).

Por fazer parte do OpenAire, o Zenodo abre a possibilidade de os pesquisadores do Horizonte 2020 se beneficiarem dos serviços que essa infraestrutura europeia para acesso aberto oferece. Entre eles, diversos guias que orientam o pesquisador em vários temas, como a escolha da licença mais apropriada para dados a serem publicados e disponibilizados publicamente para reuso e acesso a recursos tecnológicos, como a ferramenta Amnesia¹⁰¹, para anonimização de dados pessoais (OPEN AIRE, 2020). Fornece também ajuda a esses mesmos pesquisadores a reportarem suas publicações ao Portal de Participantes da Comissão Europeia e a cumprirem a Política de Acesso Aberto da Comissão Europeia e o Piloto de Dados de Pesquisa, demonstrando o cumprimento de exigências de quem lhes financiou a pesquisa (ZENODO, 2020b).

4.2.8 Direitos autorais: *download* não transfere direitos

De acordo com a página “General Policies”, o uso e reuso de dados estão sujeitos à licença sob a qual os arquivos de dados foram depositados. Na página “Terms of use” de seu *site*, o repositório estabelece que “os metadados do Zenodo podem ser reutilizados livremente sob a renúncia CC0”, salvo indicação em contrário (ZENODO, 2020a, *on line*, tradução nossa). A recomendação desse instrumento da Creative Commons, por meio do qual o autor ou detentor dos direitos autorais renuncia a seus direitos autorais, na extensão permitida por

¹⁰¹ Ver em: <<https://www.openaire.eu/item/amnesia-data-anonymization-made-easy?highlight=WyJhbm9ueW1pemF0aW9uIl0=>>

lei, dá mostras não só de seu compromisso com o movimento de acesso aberto a dados de pesquisa como confirma também a tendência à adoção da CC0 por repositórios generalistas de dados de pesquisa. Entretanto, como já dito, o Zenodo não impõe restrições quanto ao uso de licenças.

Outra disposição dessa mesma seção do *site* que merece atenção é a que não permite o *download* em massa de endereços de *e-mails* extraídos do Zenodo. Esta cláusula poderia estar baseada em um direito específico previsto na legislação da União Europeia e, por conseguinte, na legislação dos países-membros. Trata-se da lei que criou o direito *sui generis*, que protege o investimento do criador da base de dados, no caso o repositório Zenodo. Segundo o artigo 7º, n. 1, é proibida a extração da totalidade ou de uma parte substancial do seu conteúdo, caso “[...] representem um investimento substancial do ponto de vista qualitativo ou quantitativo” (COMISSÃO EUROPEIA, 1996, *on line*, tradução nossa).

Ainda na seção “Terms of use”, o Zenodo (2020a, *on line*) regulamenta tanto o *upload* quanto o *download* de dados, estabelecendo que o remetente deve garantir que o conteúdo é adequado para divulgação em acesso aberto e que atende os termos prescritos pelo repositório e leis aplicáveis. Esse compromisso inclui privacidade, proteção de dados e direitos de propriedade intelectual, o que abrange direitos autorais.

Como já dito, os usuários do conteúdo – todos os tipos de artefatos de pesquisa – devem respeitar as condições das licenças aplicáveis. São informados de que o fato de fazerem o *download* e o reuso do conteúdo do Zenodo não transfere qualquer direito de propriedade intelectual sobre o conteúdo aos usuários. Por fim, ao frisar que o usuário é o único responsável pelo uso e *download* do conteúdo, isentando o CERN de qualquer responsabilidade, o repositório dispõe que esse laboratório, sem aviso prévio, pode a seu critério alterar ou excluir conteúdo inapropriado e restringir ou remover o acesso do usuário que julgue estar interferindo em suas operações ou violando leis aplicáveis.

No Quadro 9, destacamos alguns requisitos das políticas para disponibilização de dados de pesquisa subjacentes a artigos científicos, ou de dados independentes, identificados em cada repositório pesquisado.

Quadro 9. Requisitos das políticas de dados de pesquisa de cada repositório

Repositórios			
	Dryad	Figshare	Zenodo
Licenciamento de dados subjacentes	CC0 padrão	CC0 padrão; admite CCBY e demais licenças CC; licença própria; e ODC-By	Sugere CC0; aceita outras licenças
Taxa de publicação de dados	US\$ 120	Gratuito e a combinar	Gratuito
Privacidade de dados	Sim	Sim	Sim, nos termos do Open Aire
Citação de dados	Sim	Sim	Sim
Acesso aberto	Sim	Sim	Sim
Aceita dados com embargo	Sim	Sim	Sim
Aceita dados não subjacentes a artigos	Sim	Sim	Sim

Fonte: Elaboração própria, 2020.

4.3 As agências de financiamento batem pé: acesso aberto a dados de pesquisa já!

As agências que financiam a pesquisa já demonstraram a sua intenção de fortalecer políticas de abertura e compartilhamento de dados, a fim de fazer com que os profissionais das diversas áreas do conhecimento, entre eles os da saúde pública, como epidemiologistas e pesquisadores da pesquisa clínica, trilhem o caminho de seus pares geneticistas, que já detêm uma boa prática, inclusive com questões sobre direito autoral e propriedade intelectual (NIH, 2003a; 2019; OCDE, 2007; WELLCOME TRUST, 2010; OMS, 2011; HAHNEL, 2015). Essas intenções estão formalizadas em diversos documentos, entre os quais uma declaração da OMS, de 2015, e uma declaração conjunta da OMS, de 2017, com o apoio de alguns financiadores de pesquisa internacionais, dentre eles Wellcome Trust e Fundação Bill & Melinda Gates, sobre divulgação pública de resultados de ensaios clínicos¹⁰².

¹⁰² Declaração da OMS sobre a divulgação pública dos resultados de ensaios clínicos (2015) e Declaração conjunta sobre divulgação pública de resultados de ensaios clínicos (2017). Fonte: <<https://www.who.int/ictrp/results/en/>>

No entanto, Pisani e AbouZahr (2010) destacam que boas intenções não são o suficiente. Para eles, além de uma mudança na cultura da ciência, é preciso que os próprios financiadores estabeleçam e firmem compromissos concretos. Os autores exemplificam com o caso do Projeto Genoma Humano¹⁰³, em 1990, em que os dois maiores financiadores – o Wellcome Trust e o NIH – investiram maciçamente na infraestrutura necessária para compartilhar dados em larga escala a longo prazo, além de mudarem os mecanismos de financiamento, estimulando o trabalho em equipe e enfatizando o valor de funções negligenciadas como o gerenciamento de dados, em vez de apenas olhar para os registros de publicação e citação na hora de decidir a concessão de um financiamento.

Na esteira do lançamento de documentos, a OCDE (2007) lançou o “Principles and Guidelines for Access to Research Data from Public Funding”, considerado seminal, que inspirou e orientou financiadores, instituições e grupos editoriais a estabelecerem políticas de abertura e compartilhamento de dados de pesquisa. Nele, a OCDE preconiza o respeito a direitos autorais e reconhece que “[...] certos dados de pesquisa serão necessariamente limitados por vários tipos de requisitos legais, que podem incluir restrições” (OCDE, 2007, p. 16, tradução nossa) de segurança nacional, privacidade e confidencialidade, segredos comerciais e propriedade intelectual (direito autoral incluído), proteção a espécies raras ou ameaçadas de extinção e processo legal. Seguiram-se outros documentos de apoio à nova agenda da ciência aberta, tratando também de aspectos legais do acesso aberto ou de um balanço de atividades, como o relatório “Making Open Science a Reality” (OCDE, 2015a, p. 5, tradução nossa), que “analisa [sob a ótica da ciência aberta] o progresso dos países da OCDE em tornar os resultados da pesquisa financiada publicamente, em especial publicações científicas e dados de pesquisa, acessíveis ao pesquisador e inovador”.

Em 2013, o Governo Obama definiu a estrutura da política de acesso aberto a resultados de pesquisa financiada por verbas públicas nos EUA. Em memorando aos chefes de departamentos e de agências executivas, instruía-os a aumentar o acesso a publicações e dados científicos revisados por pares, oriundos de pesquisa financiada pelo governo federal. Afirmava que ambos os resultados “[...] devem estar disponibilizados e acessíveis ao público, à indústria e à comunidade científica” (OFFICE OF SCIENCE..., 2013, *on line*, tradução nossa). O NIH já se antecipara e publicara em 2003 a sua política de compartilhamento de dados de pesquisa, “[...] mesmo sabedor de que questões legais e de privacidade podem

¹⁰³ Ver em: <<https://genoma.ib.usp.br/sites/default/files/projeto-genoma-humano.pdf>>

ocorrer, e [já pedia a pesquisadores que] apresentem um plano de gestão de dados” (NIH, 2003a, *on line*, tradução nossa).

Por sua vez, a Comissão Europeia (2014; 2017) instituiu o projeto Horizonte 2020, iniciado em 2014 com previsão de duração de sete anos (2014 a 2020), com orçamento de 80 bilhões de euros. Ele promove a abertura e o compartilhamento tanto para os artigos provenientes de pesquisa financiada pelo programa quanto para os dados de pesquisa daí decorrentes. Desde 2014, o Horizonte 2020 publica versões atualizadas de documentos de apoio a essa agenda, entre eles “Guidelines on Implementation of Open Access to Scientific Publications and Research Data”¹⁰⁴ e “Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020”¹⁰⁵.

Com o intuito de contribuir com o conhecimento sobre como as agências de financiamento de pesquisa se alinham a essa nova etapa da ciência, por terem um papel central na produção do conhecimento, dado que grande parte da pesquisa acadêmica é predominantemente financiada por dinheiro público, por meio de agências públicas e privadas, e ainda por frequentemente determinarem a direção da pesquisa (MATTHEWS; HO, 2008), analisamos as políticas de abertura de dados de três delas – National Institutes of Health, Wellcome Trust e Bill & Gates Foundation –, a partir de seus *sites*, no que se refere a orientações quanto ao direito autoral sobre dados de pesquisa subjacentes.

4.3.1 National Institutes of Health (NIH)

O NIH, uma parte do Departamento de Saúde e Serviços Humanos dos Estados Unidos, é a agência de pesquisa médica dos EUA¹⁰⁶.

As raízes do NIH remontam a 1887, quando o pesquisador Joseph J. Kinyoun identificou o bacilo da cólera no Laboratório de Higiene do Marine Hospital Service (MHS), Staten Island, Nova York. Em 1901, o laboratório foi reconhecido por lei pelo Congresso. Em 1930, a Lei Ransdell mudou o nome para Instituto Nacional de Saúde (no singular) e

¹⁰⁴ Ver em: <https://ec.europa.eu/research/participants/data/ref/h2020/other/hi/oa-pilot/h2020-hi-erc-oa-guide_en.pdf>

¹⁰⁵ Ver em: <https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf>

¹⁰⁶ Ver em: <<https://www.nih.gov/institutes-nih>>

autorizou bolsas para pesquisa médica e biológica. Em 1948, o National Heart Act deu seu nome definitivo: National Institutes of Health (no plural). Em 1960, havia dez institutos.¹⁰⁷

Hoje, o NIH é formado por 27 instituições: 20 institutos, que cobrem o mesmo número de diferentes áreas da saúde; seis centros, que atuam em setores que vão da pesquisa clínica à tecnologia da informação; e a National Library of Medicine, que abriga os bancos de dados do PubMed, Medline e Medline Plus.¹⁰⁸

A sede burocrática do NIH e os institutos e centros estão localizados em Bethesda, Maryland, nos EUA, em edifícios reunidos em ambiente semelhante a um *campus*¹⁰⁹. O NIH recebe financiamento do Congresso dos EUA e cada componente administra o próprio orçamento. Seu investimento anual em pesquisa biomédica é de cerca de US\$ 41,7 bilhões. Mais de 80% vão para pesquisas fora de seus muros, por meio de quase 50.000 subsídios para mais de 300.000 pesquisadores em mais de 2.500 universidades, escolas médicas e outras instituições de pesquisa em todos os estados dos EUA. Cerca de 10% do orçamento do NIH apoiam projetos conduzidos por quase 6.000 cientistas em seus próprios laboratórios, a maioria deles no *campus* do NIH, em Maryland¹¹⁰.

Nos idos de 2003, o NIH lançava sua política de compartilhamento de resultados de pesquisa que patrocina, com uma “declaração final” na qual afirma que “o compartilhamento de dados é essencial para a tradução acelerada dos resultados da pesquisa em conhecimento, produtos e procedimentos para melhorar a saúde humana”. Já naquela época, como até hoje, exigia que pesquisadores que solicitassem bolsa de “US \$ 500.000 ou mais em custos diretos em um único ano incluam um plano para o compartilhamento de dados ou indiquem por que o compartilhamento de dados não é possível” (NIH, 2003a, *on line*, tradução nossa).

O NIH reúne alguns documentos que compõem a sua política de abertura e compartilhamento na página “NIH Data Sharing Policy” (NIH, 2019). Nela, entre outros, encontram-se a citada declaração fundamental, um quadro de diretrizes, políticas e regulações de abertura e compartilhamento de dados de pesquisa, e a diretiva “NIH Data Sharing Policy and Implementation Guidance” (NIH, 2003b), que apresenta os objetivos do

¹⁰⁷ Ver em: <<https://history.nih.gov/exhibits/history/index.html>>

¹⁰⁸ Ver em: <<https://www.nih.gov/institutes-nih/list-nih-institutes-centers-offices>>

¹⁰⁹ Ver em: <<https://www.nih.gov/about-nih/who-we-are/organization>>

¹¹⁰ Dados colhidos em abril/2020, em: <<https://www.nih.gov/about-nih/what-we-do/budget>>

compartilhamento de dados, a aplicabilidade da política e uma ampla gama de orientações para implementação da sua política de dados. Esse documento traz a definição de dados do NIH, os quais denomina de “dados finais de pesquisa”, aqueles eleitos para abertura e compartilhamento:

Material factual registrado, comumente aceito na comunidade científica como necessário para documentar e validar os resultados da pesquisa. Isso não significam estatísticas ou tabelas resumidas; ao contrário, significam os dados nos quais as estatísticas e tabelas de resumo são baseadas. Para os fins desta política, os dados finais da pesquisa não incluem cadernos de laboratório, conjuntos de dados parciais, análises preliminares, rascunhos de artigos científicos, planos para pesquisas futuras, relatórios de revisão por pares, comunicações com colegas ou objetos físicos, como géis ou amostras de laboratório (NIH, 2003b, *on line*, tradução nossa).

O NIH incentiva a liberação e o compartilhamento dos dados finais da pesquisa de estudos que apoia, para uso por outros pesquisadores, na aceitação da publicação das principais conclusões do conjunto de dados final. Incentiva também o compartilhamento de materiais e recursos de pesquisa, as chamadas ferramentas de pesquisa. Porém, o NIH reconhece que, devido à amplitude e variedade de ciências que apoia, tanto o conteúdo exato da documentação de dados, quanto a formatação ou o transporte dos dados não são estabelecidos com regras e prazos. Por isso, “o mais sensato seria que membros das várias disciplinas e suas sociedades profissionais discutissem o compartilhamento de dados, determinassem padrões e melhores práticas e criassem um ambiente social que apoie o compartilhamento de dados” (NIH, 2003b, *on line*, tradução nossa).

O equilíbrio entre a proteção da privacidade dos dados e sua utilização anonimizada ou não (com consentimento) na pesquisa médica é uma das áreas mais desafiadoras, que reclamam por padrões e práticas uniformes em nível internacional (BIETZ et al., 2016; MALIN, 2018; POPESCU et al., 2018). O NIH preconiza proteção o tempo todo para seres humanos que participam de pesquisas clínicas. Alerta para que a remoção de um conjunto de dados de itens que podem identificar participantes individuais vá além da anonimização por meio da retirada de identificadores diretos, como nome, endereço, números de telefone e números de CPF. Segundo a agência, os pesquisadores devem considerar a remoção de

identificadores indiretos e outras informações que possam levar à “divulgação dedutiva” das identidades dos participantes (NIH, 2003b).

Em apoio à sua política de disseminação e disponibilidade pública de dados de pesquisa, o NIH mantém e recomenda aos bolsistas repositórios para abrigo dos dados resultantes de pesquisas que patrocina. Apesar disso, o NIH (2020) informa que não endossa ou exige compartilhamento em nenhum repositório específico, deixando a cargo dos pesquisadores a seleção do repositório mais adequado ao seu tipo de dados e disciplina (essa especificação existe apenas em iniciativas especiais). Assim, para ajudar os pesquisadores a localizarem repositórios adequados a seu tipo de dados e disciplinas e recurso apropriado para compartilhar conjuntos de dados, o NIH divulga a lista de repositórios que apoia no *hot site* Trans-NIH BioMedical Informatics Coordinating Committee:

- Repositórios específicos de domínio financiados pelo NIH que abrigam dados de um tipo específico ou relacionados a uma disciplina específica;
- Outros recursos específicos de domínio financiados pelo NIH, incluindo repositórios e bases de conhecimento, que têm limitações no envio e/ou acesso a dados; e
- Repositórios generalistas que hospedam dados, independentemente do tipo, formato, conteúdo ou assunto (NIH, 2020, *on line*, tradução nossa).

O NIH (2003a) ainda reconhece que o compartilhamento de dados, além das regras de privacidade e confidencialidade dos dados, pode esbarrar em políticas institucionais, bem como em leis e regulamentos federais e estaduais, que dispõem sobre propriedade intelectual e direito autoral.

4.3.2 Direitos autorais: sem licenças Creative Commons

O NIH trata da questão dos direitos autorais aplicáveis a dados de pesquisa no documento “NIH Grants Policy Statement” (NIH, 2018), mais precisamente na subseção “Rights in Data (Publication and Copyrighting)”, item 8.2.1. Desde logo, com a ressalva de que, em geral, os pesquisadores por ele financiados são os detentores dos direitos sobre dados resultantes de um projeto apoiado por uma doação, o NIH declara que,

[...] salvo disposição em contrário nos termos e condições do subsídio, quaisquer publicações, *dados* ou outros trabalhos passíveis de proteção de direito autoral desenvolvidos sob uma concessão do NIH, podem ser protegidos por direitos autorais sem a aprovação do NIH (NIH, 2018, *on line*, tradução nossa, grifo nosso).

Para este fim, “dados” significam

[...] informações gravadas, independentemente da forma ou mídia em que possam ser gravadas, e incluem escritos, filmes, gravações sonoras, reproduções pictóricas, rascunhos, desenhos ou outras representações gráficas, manuais de procedimentos, formulários, diagramas, fluxogramas de trabalho (NIH, 2018, *on line*, tradução nossa).

O NIH declara ainda que tais direitos sobre dados se estendem a estudantes, bolsistas ou estagiários com subsídios cujo objetivo principal é educacional, incluindo participantes de consórcio de pesquisa. Assim, são livres para proteger seus trabalhos com direitos autorais sem a aprovação do NIH (2018). Mas em todos esses casos o NIH deve receber dos detentores de direitos autorais, oriundos de trabalhos que financia, “uma licença livre de *royalties*, não exclusiva e irrevogável para o governo federal reproduzir, publicar, ou usar o material de qualquer outra forma ou autorizar outros a fazê-lo para fins federais” (NIH, 2018, *on line*, *on line*, tradução nossa). Assim, como órgão financiador, negocia por meio de uma licença própria alguns direitos sobre os conjuntos de dados dos pesquisadores financiados, que detêm direitos autorais.

Um ponto a ser destacado é o modelo que o NIH adota para a gestão dos direitos autorais aplicáveis a dados de pesquisa subjacentes. A agência não menciona nem recomenda em nenhum dos documentos pesquisados que os bolsistas do NIH adotem as licenças públicas e gratuitas, como as da Creative Commons.

Do ponto de vista jurídico-administrativo, o modelo é legítimo, mas está na contramão da almejada padronização de práticas em ciência aberta, bem como da interoperabilidade legal dos dados de pesquisa abertos, tema tratado no capítulo 3 (UHLIR; CLEMENT, 2016). O NIH ignora as chamadas licenças alternativas gratuitas, como a Creative Commons, ou a Open Data Commons, mas, com o intuito de intensificar o reuso, oferece três linhas de repositórios, duas financiadas pela agência e uma composta por repositórios generalistas,

como já visto. Tais repositórios oferecem todo tipo de licença, desde algumas próprias até as licenças Creative Commons e o instrumento CC0 Dedicção ao Domínio Público, o que pode causar incerteza e dúvidas no pesquisador no momento de disponibilizar e compartilhar dados.

4.3.3 Wellcome Trust/Wellcome Open Research

A Wellcome Trust, que se autointitula “instituição de caridade”, é uma das mais importantes agências privadas de financiamento não governamental à pesquisa biomédica do mundo. Com sede em Londres, foi fundada em 1936, logo após a morte do empresário e seu provedor, Sir Henry Wellcome, dono da Wellcome Foundation. Em 1995, a agência vendeu a maior parte da sua participação para a Glaxo, criando a Glaxo Wellcome, que se fundiu com a SmithKlineBeecham em 2000 para criar a poderosa indústria farmacêutica GlaxoSmithKline. Até fins de 2019, o seu portfólio de investimentos no mercado financeiro, uma das formas que utiliza para se autofinanciar, alcançava 26,8 bilhões de libras esterlinas. Planeja aplicar 5 bilhões de libras esterlinas em seus projetos entre 2019 e 2023. Registra que, em geral, não recebe doações ou subsídios do governo¹¹¹.

A Wellcome Trust financia ou cofinancia projetos no Reino Unido, África, Ásia, entre eles o Projeto Genoma do Câncer, Iniciativa de Emergência do Ebola, Structural Genomics Consortium, além de programas de pesquisa clínica e medicina tropical em países africanos e asiáticos¹¹². Também apoia projetos de pesquisa no Brasil, como por exemplo sobre o surto do vírus Zika em 2015¹¹³.

Em 2020, organizou duas campanhas para arrecadar fundos para o combate à pandemia do coronavírus: a Covid-Zero¹¹⁴, uma coalizão global de parceiros, para apoiar o suprimento do déficit de financiamento inicial de gastos de US\$ 8 bilhões, que ela estimou ser necessário para vacinas, testes para diagnóstico e tratamentos; e o Acelerador de Acesso às

¹¹¹ Ver em: <<https://wellcome.ac.uk/about-us>> e <<https://wellcome.ac.uk/about-us/history-wellcome>>

¹¹² Ver em: <<https://wellcome.ac.uk/grant-funding/people-and-projects/grants-awarded>>

¹¹³ Ver em: <<https://wellcome.ac.uk/about-us>>

¹¹⁴ Ver em: <<https://wellcome.ac.uk/what-we-do/our-work/coronavirus-covid-19>>

Ferramentas Covid-19¹¹⁵ (ACT-Accelerator, em inglês), uma colaboração global para acelerar o desenvolvimento, a produção e o acesso equitativo aos testes, tratamentos e vacinas.

A Wellcome Trust (2017) tem desempenhado um papel relevante na implantação da cultura da ciência aberta. Em 2017, lançou sua política mandatória de acesso aberto a resultados de pesquisas que financia, particularmente aos dados de pesquisa subjacentes a artigos, a qual confirmava uma declaração apoiada pela agência sobre compartilhamento de dados de pesquisa para melhorar a saúde pública (WELLCOME TRUST, 2010). Uma mudança na política de acesso aberto já está marcada para janeiro de 2021¹¹⁶, com o objetivo de se alinhar às diretrizes do Plano S (COALITION, 2018), que visa o acesso aberto completo e imediato a publicações acadêmicas revisadas por pares e também a dados de pesquisa, resultantes de pesquisas financiadas por doações públicas ou privadas. Com isso, deixará de permitir um embargo de seis meses para a abertura pública dos artigos e dados subjacentes resultantes da pesquisa que financia.

Em 2016, foi uma das signatárias do Concordat on Open Research Data¹¹⁷, acordo destinado a “garantir que os dados de pesquisa coletados e gerados por membros da comunidade de pesquisa do Reino Unido sejam disponibilizados abertamente para uso por outros [...] de maneira consistente com as estruturas e normas legais, éticas [...]”, combinando a publicação de artigos com os dados subjacentes. O documento também foi assinado por Higher Education Funding Council for England (Hefce), Research Council UK e Universities UK, que, junto com a Wellcome, formam as quatro principais organizações de pesquisa do Reino Unido.

Naquele mesmo ano, anunciou o lançamento da plataforma Wellcome Open Research, um modelo de publicação rápida, em acesso aberto, de artigos vinculados a seus dados de pesquisa subjacentes, como alternativa apenas para pesquisadores financiados pela agência que não escolhem periódicos externos para publicar (NATURE, 2016b). O objetivo é oferecer a esses bolsistas uma forma ágil de publicarem “quaisquer resultados que eles considerem que

¹¹⁵ Ver em: <<https://www.who.int/initiatives/act-accelerator>>

¹¹⁶ Ver em: <<https://wellcome.ac.uk/node/1925>>

¹¹⁷ Ver em: <<https://www.ukri.org/wp-content/uploads/2020/10/UKRI-020920-ConcordatonOpenResearchData.pdf>>. Fonte: UK Research and Innovation (UKRI) <<https://www.ukri.org/about-us/policies-standards-and-data/good-research-resource-hub/open-research/>>

vale a pena”¹¹⁸. A iniciativa foi seguida pela Fundação Bill & Melinda Gates, que lançou no ano seguinte a Gates Open Research (NATURE, 2017c).

Esse movimento quase sincronizado de duas das maiores agências de financiamento à pesquisa em saúde do mundo lançarem seus canais próprios de divulgação de artigos com vinculação a dados de pesquisa subjacentes teve consequências (ver “Conflito aberto”, seção 4.4, neste capítulo). Em suma, fez com que a Comissão Europeia passasse a cogitar também de um canal próprio (NATURE, 2017c) e obrigou revistas científicas de modelos de acesso fechado (assinatura ou acesso pago a artigos) ou acesso aberto híbrido a reverem suas diretrizes de publicação para acesso aberto, uma vez que Wellcome e fundação Gates exigiam em suas políticas a publicação imediata de artigos e dados de pesquisa em acesso aberto logo após o aceite de uma revista, sob uma licença CC BY. Essa decisão contrariava a política de revistas de acesso fechado (NATURE, 2017b; BMJ, 2018; SPRINGER NATURE, 2018). Em 2018, essas iniciativas foram reforçadas pelo Plano S europeu, uma coalizão de 17 financiadores de pesquisa europeus que propõem acesso aberto para todas as publicações acadêmicas financiadas por doações públicas ou privadas, sem embargo (COALITION, 2018).

A plataforma Wellcome Open Research é objeto da pesquisa apresentada neste capítulo, por introduzir uma série de práticas incomuns na comunicação científica clássica, em consonância com a ciência aberta, mas também por se constituir em um canal próprio da Wellcome Trust para publicação de artigos vinculados a dados subjacentes de pesquisas que financia, o que pode vir a afetar a forma tradicional de publicação em revistas científicas. Adota os princípios FAIR (Quadro 5), as diretrizes do ICMJE e cobra taxa de processamento de artigos, que é paga pela própria agência e varia de 116 a 775 libras esterlinas, valor descontado do financiamento que concede ao seu pesquisador. A plataforma, que tem a F1000Research como provedora dos serviços editoriais e tecnológicos, exige uma declaração de concordância para a disponibilidade dos dados, e rejeita o artigo, caso não haja uma boa justificativa para esse fim. Contudo, a Wellcome Open Research reconhece que pode haver casos de restrição ética, legal, de segurança, de proteção a dados pessoais e grandes conjuntos de dados que inviabilize a sua abertura (WELLCOME OPEN RESEARCH, 2017). Essa experiência em ciência aberta da Wellcome Trust também se caracteriza por:

¹¹⁸ Ver em: <<https://wellcomeopenresearch.org/about>>

- ser exclusiva para pesquisadores da agência que não escolhem periódicos externos para publicar;
- publicar artigos sobre pesquisas originais, com a seção “Disponibilidade de dados”, que descreve onde o conjunto de dados subjacentes e *software* usados podem ser encontrados, com identificador e referência para citação do conjunto de dados;
- permitir que autores revisem, alterem e atualizem seus artigos publicando novas versões, que são adicionadas ao histórico do artigo;
- publicar artigo em uma semana, após o autor finalizá-lo, e recomendar repositório;
- adotar revisão por pares aberta após publicação do artigo na plataforma, com revisores indicados por autores ou selecionados;
- publicar relatórios de avaliação por pares junto com o artigo, com nomes dos avaliadores, respostas dos autores e comentários de usuários registrados;
- compartilhar dados pode não ser viável por questões éticas ou de segurança ou problemas de proteção de dados, que devem ser descritos, sob pena de não aceitação do artigo;
- adotar licença CC BY para artigos;
- disponibilizar relatórios de revisão por pares também sob uma licença CC BY;
- publicar dados associados aos artigos, em repositório recomendado, sempre que possível, sob o instrumento CC0 Dedicção ao Domínio Público, antes da submissão do manuscrito (WELLCOME OPEN RESEARCH, 2017, *on line*, tradução nossa).

4.3.4 Direito autoral: negociação justa (*fair dealing*) facilita a pesquisa

No Reino Unido, o direito autoral é regulado pela Lei de Direitos Autorais, Design e Patentes¹¹⁹, de 1988. Como na maioria das legislações, protege obras originais, que traduzam a expressão de uma ideia, mas não a ideia em si. Esta lei apresenta limites ao direito autoral, que compõem a chamada “negociação justa” (*fair dealing*). Uma dessas disposições legais favorece a pesquisa no Reino Unido, uma vez que

Nenhuma negociação justa com um trabalho original constituirá uma infração se for realizada para um dos seguintes propósitos: *pesquisa para fins não comerciais*; estudo privado; crítica ou resenha; e relato de eventos atuais (GUIBAULT; WIEBE, 2013, p. 44, tradução nossa, grifos nossos).

¹¹⁹ Ver em: <<https://www.legislation.gov.uk/ukpga/1988/48/contents>>

Todavia, como soe acontecer no incerto terreno legal dos dados, a lei não define que tipo de negociação de uma obra é justa ou não, cabendo aos tribunais determinar se é ou não é justa. Segundo o *site* oficial do governo do Reino Unido, os tribunais em geral consideram uma negociação justa se o uso da obra não fizer com que o dono do material protegido perca receita e a parte utilizada da obra foi razoável e apropriada¹²⁰.

A Wellcome Open Research considera imprescindível que outras pessoas possam acessar os dados para replicar um estudo, analisá-los e, em alguns casos, reutilizá-los. Porém, em seu *site*, nas diretrizes de disponibilização dos dados, na seção “Policies” (WELLCOME OPEN RESEARCH, 2017), reconhece que o compartilhamento aberto pode não ser viável, entre outras razões, porque os dados podem estar licenciados por terceiros. Neste caso, a orientação é incluir no artigo associado aos dados as informações necessárias para que um leitor ou revisor acesse os dados pelos mesmos meios que os autores. Em caso de dados governamentais, as restrições devem ser explicadas e fornecidas informações para um leitor ou revisor solicitar acesso ao proprietário dos dados, desde que não existam interesses de natureza comercial.

Outras restrições podem ser éticas, de segurança, grandes conjuntos de dados e problemas de proteção de dados, em que o acesso a dados humanos não deve ser compartilhado, caso não sejam devidamente desidentificados, a menos que os participantes tenham dado consentimento por escrito para dispor seus dados identificáveis ao público.

Relativamente a licenças, os artigos e os relatórios de revisão por pares da Wellcome Open Research são publicados sob uma licença CC BY, que permite uso irrestrito, distribuição e reprodução em qualquer meio, bem como o uso comercial, desde que o trabalho original seja citado e mantenha os direitos autorais do artigo com o detentor atual desses direitos, em geral o autor ou a instituição.

Quanto aos dados subjacentes aos artigos, são disponibilizados, sempre que possível, sob o instrumento CC0 Dedicção ao Domínio Público, da Creative Commons. Essa opção facilita e incentiva a reutilização e ajuda a evitar os problemas de “empilhamento” de atribuição em virtude da combinação de vários conjuntos de dados, com vários autores que podem usar várias licenças diferentes (WELLCOME OPEN RESEARCH, 2017).

¹²⁰ Ver em: <<https://www.gov.uk/guidance/exceptions-to-copyright#fair-dealing>>

4.3.5 Fundação Bill & Melinda Gates

Criada em 2000 mediante a fusão de duas fundações dirigidas pelo pai de Bill Gates, a Fundação Bill & Melinda Gates concentra suas doações em cinco programas que atendem diversos países: Desenvolvimento Global, Crescimento e Oportunidades, Saúde Global, Política Global e Advocacy e um programa específico para os EUA. A fundação mantém escritórios na Índia, China, Reino Unido, Alemanha, Etiópia, Nigéria e África do Sul.

Segundo consta em seu *site*, a Fundação Bill & Melinda Gates, sediada em Seattle, Washington, EUA, trabalha em parceria com países em desenvolvimento com o foco na melhoria da saúde das pessoas. Nos EUA, investe em educação. A fundação mantém-se principalmente de doações do casal Bill e Melinda e do magnata Warren Buffet, que em 2006 anunciou doações na casa dos US\$30 bilhões “nas próximas décadas”, segundo as pesquisadoras Matthews e Ho (2008, *on line*), em artigo no qual questionam como uma pessoa famosa e bilhões de dólares podem afetar gastos e foco de pesquisas de órgãos públicos, nos EUA. As doações a beneficiários diretos chegaram a US\$5 bilhões, em 2018¹²¹.

O alcance geográfico da fundação Gates abrange o apoio a donatários em 50 estados americanos e iniciativas em 138 países. No Brasil, a fundação mantém projetos sobre dengue, malária, apoio a eventos de saúde pública e a bibliotecas, que envolvem universidades e instituições científicas, entre elas a Universidade de Brasília e o Instituto Gonçalo Moniz, unidade da Fiocruz¹²² na Bahia, à qual está vinculado o Centro de Integração de Dados e Conhecimentos para Saúde¹²³ (Cidacs), que realiza estudos e pesquisas com base em projetos interdisciplinares originados na vinculação de grandes volumes de dados, alguns deles apoiados pela fundação Gates.

A política de acesso aberto da Fundação Bill & Melinda Gates apoia-se no “Acesso Global”¹²⁴ (*Global Access*), um conceito desenvolvido pela fundação para garantir que “produtos, serviços, processos, tecnologias, materiais, software, dados resultantes e outras inovações [...] terão impacto positivo nos beneficiários atendidos pelo trabalho da fundação”

¹²¹ Ver em: <<https://www.gatesfoundation.org/Who-We-Are/General-Information/Foundation-Factsheet>>

¹²² Ver em: <<https://www.gatesfoundation.org/search#q/k=FIOCRUZ>>

¹²³ Ver em: <<http://cidacs.bahia.fiocruz.br/>>

¹²⁴ Ver em: <<https://www.gatesfoundation.org/How-We-Work/General-Information/Global-Access-Statement>>

(tradução nossa). Com base nesse conceito, exige que o conhecimento e as informações obtidas por seus financiamentos sejam rápida e amplamente divulgados, particularmente para seus programas Saúde Global e Desenvolvimento Global, inclusive mediando negociações de direitos autorais e outros direitos, como veremos adiante, na seção “Direitos autorais”.

Na página específica sobre a sua política de acesso aberto (BILL & MELINDA GATES FOUNDATION, 2017a), que entrou em vigor em 2015, reforça o compromisso em divulgar resultados de pesquisas por ela financiadas logo após sua publicação, sem período de embargo, e em permitir “o acesso irrestrito e a reutilização de todas as pesquisas publicadas revisadas por pares, financiadas, no todo ou em parte, pela fundação, incluindo quaisquer conjuntos de dados subjacentes”. A política de acesso aberto é formada pelos seguintes elementos:

1. *As publicações são encontráveis e acessíveis on line.* Em *site* da revista escolhida pelo beneficiário para publicar seu artigo e no repositório PubMedCentral.
2. *A publicação estará em acesso aberto.* Estabelece para artigos a adoção da licença Creative Commons Attribution CC BY 4.0 ou equivalente, segundo a fundação para reuso intenso, inclusive comercial. A fundação declara que publicar em acesso aberto é um requisito inegociável.
3. *A fundação pagará as taxas necessárias.* Paga, por exemplo, a taxa de processamento de artigos, que varia entre US\$1.200 e US\$5.200 (sem, no entanto, abater da subvenção concedida, como faz a Wellcome Trust).
4. *As publicações serão acessíveis e abertas imediatamente.* Estarão disponíveis logo após sua publicação, sem período de embargo (isto é, sem assinatura ou pagamento de taxa para obter acesso à publicação). A fundação Gates permitiu, como compensação temporária às editoras comerciais, embargo de até 12 meses, entre 2015 e 2017.
5. *Os dados subjacentes aos resultados da pesquisa publicados serão acessíveis e abertos imediatamente.* Também permitiu embargo para dados subjacentes no período de transição da implantação de sua de abertura, de 2015 a 2017 (BILL & MELINDA GATES FOUNDATION, 2017a; 2017b; CARALEE, 2019).

Em outra página sobre a sua política de acesso aberto, “Open Policy FAQ”, na seção intitulada “Underlying data guidelines” (Diretrizes para dados subjacentes), a fundação reserva um considerável espaço para as diretrizes de publicação dos dados subjacentes aos

resultados da pesquisa publicada, elemento de interesse deste estudo. A destacar, a iniciativa de definir a categoria de “dados subjacentes”:

Os dados subjacentes englobam todos os dados primários, metadados associados e quaisquer dados adicionais relevantes necessários para entender, avaliar e replicar os resultados do estudo relatados na totalidade. [...] podem ser compilados em qualquer tipo de arquivo, incluindo as instruções de acesso necessárias, código ou arquivos de informações de suporte, para garantir que os arquivos possam ser acessados e utilizados por terceiros (BILL & MELINDA GATES FOUNDATION, 2017b, *on line*, tradução nossa).

Na aludida página, deduz-se também que no período da pré-publicação dos dados subjacentes é exigido do candidato ao financiamento a apresentação de um plano de gestão de dados, além de um detalhamento sobre coleta, análise, armazenamento, consentimento informado, interoperabilidade e uso de padrões de diferentes disciplinas. Para a fundação, fornecer acesso aberto a dados subjacentes permite: acesso oportuno e sem barreiras; reavaliação de interpretações e análises; verificação, reprodução e reutilização de formas novas; proveniência e preservação. A fundação recomenda o depósito em repositório já estabelecido no domínio de pesquisa do beneficiário ou nos indicados pela revista. Sugere ainda consulta ao re3data.org e repositórios recomendados pelas revistas PLoS e pela revista *Scientific Data*, incluindo os generalistas Dryad, Figshare e Zenodo (BILL & MELINDA GATES FOUNDATION, 2017b).

Como consequência do advento de sua política de acesso aberto e para acelerar a publicação de artigos e conjuntos de dados, a Fundação Bill & Melinda Gates lançou, em 2017, na internet, a plataforma Gates Open Research¹²⁵, seu canal próprio para publicação rápida de artigos (sob licença CC BY 4.0) e conjuntos de dados (sob licença CCO 1.0) de qualquer pesquisa científica, translacional, aplicada e clínica básica, mas somente de pesquisadores que financia (pelo menos um autor). Foi estruturada nos mesmos moldes da Wellcome Open Research (NATURE, 2017a), inclusive com *design* e conteúdo semelhantes. A plataforma permite também a publicação de resultados negativos, protocolos, relatórios de casos, descobertas incrementais, além de artigos de pesquisa tradicionais. Assim como a

¹²⁵ Ver em: <<https://gatesopenresearch.org/>>

Wellcome, a Gates Open Research também contratou os serviços editoriais e de tecnologia do provedor de serviços F1000Research.

4.3.6 Direitos autorais como estratégia para garantir acesso

No portal Global Access, hospedado em seu *site*, a Fundação Bill & Melinda Gates explica sua relação com os direitos autorais, alvo da nossa pesquisa, englobando-os dentro da propriedade intelectual. Assim, em sua política para essa área¹²⁶, informa que, em relação aos direitos de propriedade que surgem como resultado de atividades que financia, “A posição padrão da fundação é não se apropriar dos direitos de propriedade intelectual [...]” (tradução nossa), e não exigir licença para tais direitos, que incluem os direitos autorais sobre dados de pesquisa.

A fundação afirma que seu modelo de concessão de doações prevê que parceiros donatários sejam os donos dos resultados, mas reserva-se o direito de conversar sobre como o donatário administrará um ativo de propriedade intelectual para atingir objetivos específicos do projeto, que pode envolver fins comerciais, e como avançar na missão de caridade da fundação.

Tal posicionamento, portanto, pode fazer com que exija uma licença, em casos limitados, “[...] para garantir que a fundação ou um sublicenciado possua os direitos de propriedade intelectual necessários para o desenvolvimento, fabricação e distribuição de avanços financiados pela fundação” (BILL & MELINDA GATES FOUNDATION, 2003, *on line*, tradução nossa), segundo os princípios do já mencionado Acesso Global. O conceito foi criado em 2003 e exige dos donatários e parceiros um compromisso de disponibilizar amplamente os produtos e informações gerados pelo financiamento da fundação, por meio de um gerenciamento de patentes, direitos autorais, marcas comerciais, segredos comerciais e direitos sobre dados, visto pela fundação como uma oportunidade de pensar a propriedade intelectual como forma estratégica de alcance de seus beneficiários finais (BILL & MELINDA GATES FOUNDATION, 2003).

É um ponto a ser destacado, uma vez que os direitos autorais (ou de propriedade intelectual) são muitas vezes encarados como obstáculos para a colaboração ou parceria (DAVID, 2004). A fundação, contudo, sustenta que, quando a propriedade intelectual é

¹²⁶ Ver em: <<http://globalaccess.gatesfoundation.org/>>

estruturada desde o início do projeto financiado, de forma a garantir o acesso mais amplo possível aos mais necessitados, “as barreiras percebidas levantadas por esses direitos são reduzidas, permitindo parcerias únicas que podem beneficiar imensamente os mercados de caridade que a fundação procura atender” (BILL & MELINDA GATES FOUNDATION, 2003, *on line*, tradução nossa).

No quesito licenciamento de dados subjacentes aos resultados de pesquisa com a finalidade de torná-los acessíveis e abertos, a Fundação Bill & Melinda Gates informa que o repositório escolhido deve: dar acesso imediato aos dados subjacentes após a publicação do artigo; adotar licença não mais restritiva do que a “licença genérica” CC BY 4.0; atribuir identificador do tipo DOI, para facilitar vinculação e citação; e fornecer armazenamento e preservação a longo prazo.

4.4 Conflito aberto

As agências de financiamento à pesquisa talvez sejam o segmento do ecossistema da produção científica que mais advoga o acesso aberto imediato a artigos e dados de pesquisa subjacentes (WELCOME TRUST, 2017; 2018; BILL & MELINDA GATES FOUNDATION, 2017a; COALITION, 2018). Além de deixarem claro em suas políticas que exigem acesso aberto a artigos e dados de pesquisa que financiam logo após a publicação, a Wellcome Trust e a Fundação Bill & Melinda Gates lançaram seus canais próprios de publicação de artigos, com dados vinculados e depositados em repositórios de acesso público, contrariando interesses de conglomerados editoriais de publicações científicas, despertando o interesse da Comissão Europeia em considerar essa ideia (NATURE, 2016b; 2017c).

Um acontecimento recente envolvendo a Fundação Bill & Melinda Gates nos dá pistas de que uma possível mudança esteja em curso. A fundação Gates declara em seu *site* que suas diretrizes sobre abertura de conjuntos de dados subjacentes a artigos de pesquisa são alinhadas às melhores práticas de políticas exigidas por vários editores (BILL & MELINDA GATES FOUNDATION, 2017b), mas a revista *Nature* (2017a) afirmou que a pesquisa da fundação Gates não podia ser publicada nas principais revistas acadêmicas, como a própria *Nature*, *Science*, *New England Journal of Medicine (NEJM)* e *Proceedings of National Academy of Sciences (PNAS)*, porque a sua política de acesso aberto a artigos e conjuntos de dados subjacentes colidia com as dessas revistas.

O ponto central do conflito era a diretriz da fundação Gates que determinava que os artigos e conjuntos de dados subjacentes das pesquisas que financiava deveriam ser disponibilizados em acesso aberto imediatamente após a publicação (BILL & MELINDA GATES FOUNDATION, 2017b), sob a licença CC BY, que permite reutilização abrangente, inclusive para fins comerciais. Por um lado, para a fundação, significava o fim dos embargos impostos por editoras a artigos e conjuntos de dados de suas pesquisas, mas os pesquisadores da fundação Gates poderiam ficar sem publicar em revistas de prestígio, de modelo fechado. Por outro lado, o fim dos embargos poderia provocar um prejuízo econômico para as editoras que praticam um modelo fechado de acesso, com a venda de assinatura de suas revistas e de acesso permanente ou temporário a artigos (*paywall*)¹²⁷.

Todavia, em fevereiro de 2017, a revista *Science* e mais quatro revistas, editadas pela Associação Americana para o Avanço da Ciência (American Association for the Advancement of Science - AAAS), firmaram um contrato com a fundação Gates para publicação de artigos e conjuntos de dados de pesquisadores da agência americana, aceitando as diretrizes antes consideradas colidentes (NATURE, 2017b). Em julho de 2018, os editores da *Science* encerraram a experiência que durou 18 meses e 26 artigos de pesquisadores da fundação Gates, que pagou aproximadamente US\$ 100.000 pelo primeiro ano de contrato e um valor não revelado por mais seis meses, um “imposto de prestígio”, segundo Peter Suber (NATURE, 2018).

Este caso é emblemático das transformações por que passa o sistema de comunicação científica. O aumento da exigência das agências por acesso aberto imediato a resultados de pesquisas que financiam com recursos públicos ou privados e que culminou com o lançamento do Plano S (COALITION S, 2019), apoiado por Wellcome Trust e fundação Gates (WELLCOME TRUST, 2018), pode estar se refletindo nas revistas científicas de modelo de acesso fechado. Representantes desse modelo, as revistas *NEJM* e *PNAS* não fizeram acordos com a fundação Gates, mas mudaram suas políticas e publicaram alguns artigos e conjuntos de dados da fundação em acesso aberto. *Nature* também publicou artigos financiados pela fundação Gates, mas não fez acordo (NATURE, 2018).

¹²⁷ Preços praticados pela revista *Nature Medicine*, vigentes em 01/04/2020: Assinatura – Instituições (sob consulta); Individual: Impressa&on line (US\$225.00); On line (US\$62.00). Aluguel ou compra de artigo (a partir de US\$8.99). Fonte: <https://www.nature.com/nm/subscribe>

Em novembro de 2020, a *Nature* anunciou que irá cobrar, a partir de 2021, US\$ 11.390, por artigo, de autores que desejem adotar o acesso aberto. Segundo a revista, a decisão foi “motivada pelo movimento ‘Plano S’, no qual os financiadores exigem que seus beneficiários tenham os resultados de sua pesquisa em acesso aberto [a partir de 2021] assim que são publicados” (NATURE, 2020, *on line*, tradução nossa).

Observa-se, nesta pesquisa com revistas científicas, repositórios e agências de fomento à pesquisa, que o cenário da comunicação científica está em transformação, expondo uma disputa.

De um lado, os repositórios, que, além de publicar dados de pesquisa subjacentes (ou não) a artigos científicos, evoluíram tecnologicamente e também publicam artigos e toda espécie de trabalho acadêmico, recomendando a adoção de licenças abertas e ferramentas de dedicação ao domínio público, para maximizar o reuso. Em sintonia fina com as agências de financiamento, que exigem publicação e disponibilização pública imediatas de artigos e dados de pesquisa que financiam, tão logo sejam aceitos para publicação por revistas, o atual modelo de repositório oferece a almejada celeridade na publicação desses dois resultados de pesquisa, o que pode vir a causar prejuízos ao negócio das editoras comerciais.

Do outro lado, uma grande parcela de revistas científicas, ainda ligada a um modelo de negócio do século XX, cujas raízes comerciais oferecem resistência ao movimento pela ciência aberta, seja pela manutenção de um regime de direito autoral que não atende mais às expectativas de uma ciência impulsionada pela tecnologia, seja pela cobrança de altas taxas de processamento de artigos, muitas vezes combinadas com o modelo de assinaturas, ou ainda cobrança por acesso permanente ou temporário a artigos (*paywall*). As revistas ainda se apresentam como um agregador de conhecimento, que valida os resultados das pesquisas, mas, ante a permanência de um modelo de comunicação tradicional e fechado, cabe a pergunta: até quando? O que teriam a dizer sobre esse cenário os editores de revistas científicas e gestores de repositórios?

5 ABERTURA DE DADOS DE PESQUISA: UMA VISÃO DE EDITORES DO CAMPO DA SAÚDE E GESTORES DE REPOSITÓRIOS

Um dos objetivos deste estudo é verificar como as revistas científicas, particularmente as da área da saúde, estão tratando a questão da abertura e compartilhamento de dados de pesquisa subjacentes aos artigos científicos e suas repercussões no que diz respeito ao direito autoral. Neste capítulo, apresentamos uma visão prevalentemente nacional e latino-americana de editores científicos do campo da saúde e de gestores de repositórios institucionais e de dados, estes um segmento novo no processo de comunicação de científica, consubstanciada em uma pesquisa empírica, realizada em duas partes.

Com o emprego de dois instrumentos de coleta – questionário e entrevista –, a primeira parte consistiu no envio de dois questionários eletrônicos, um específico para editores e outro específico para gestores de repositório (Apêndices A e B), ambos com perguntas fechadas, abertas e de múltipla escolha (VIEIRA, 2009), produzidos na ferramenta Google Formulários. A segunda parte foi efetivada com entrevistas com editores e gestores, mediante o apoio de dois roteiros com perguntas semiestruturadas (Apêndices C e D), também específicas para cada um dos dois segmentos. As entrevistas foram feitas na plataforma digital de videoconferências Zoom. O objetivo foi aprofundar os pontos de discussão contidos no questionário eletrônico. As entrevistas foram gravadas e transcritas.

Buscou-se aquilatar, por meio de uma abordagem qualitativa, a implementação da abertura e disponibilização em acesso aberto de dados de pesquisa subjacentes a artigos de revistas científicas e quais as relações entre proposições e políticas destas com as dos repositórios de dados abertos. E também verificar como uma parcela de cada um desses dois segmentos do sistema de comunicação científica está lidando com as questões de direito autoral envolvidas, tema ainda não suficientemente explorado, que afeta ou afetará os envolvidos com a disseminação da abertura de dados de pesquisa em um ambiente de ciência aberta.

Além dos resultados dos questionários e das entrevistas, apresentamos também o resultado de um exame observacional feito nos *sites* institucionais dos periódicos dos editores científicos selecionados para a pesquisa empírica. A meta era identificar políticas de abertura e compartilhamento de dados de pesquisa, incluindo orientações sobre direito autoral aplicável a dados.

Apesar de o atual estudo necessitar de abordagem de pesquisa indireta com seres humanos, não importou qualquer prejuízo físico, moral, social ou intelectual aos participantes nas duas fases já mencionadas, também não envolveu a abordagem de temas sensíveis ou participação de grupos vulneráveis, motivos pelos quais optou-se por não se submeter o projeto à avaliação do Comitê de Ética da instituição, via Plataforma Brasil. Contudo, observaram-se as orientações dispostas pelo Comitê de Ética em Pesquisa do Centro de Filosofia e Ciências Humanas (CEP-CFCH)¹²⁸ da Universidade Federal do Rio de Janeiro (UFRJ), para a aplicação tanto dos questionários eletrônicos quanto da realização das entrevistas.

Nessa perspectiva, todos os participantes das duas partes da pesquisa foram convidados mediante o envio por *e-mail* de um Termo de Consentimento Livre e Esclarecido - TCLE (Apêndices E e F), redigido de forma clara e objetiva, visando ao mais completo esclarecimento sobre a pesquisa da qual se propuserem participar. Nele constou ainda, conforme obrigam as orientações CEP-CFCH, o contato particular do pesquisador responsável.

Com relação à concepção, aplicação, coleta e análise das repostas dos questionários eletrônicos de editores e gestores (Apêndices A e B), as perguntas foram divididas nas duas categorias de temas de interesse do estudo, em ambos os questionários, numa tentativa prévia de organização e facilitação da fase de análise do discurso produzido e compreensão das repostas: 1) “Abertura, acesso, compartilhamento e reúso de dados de pesquisa”; e 2) “Direito autoral aplicável a dados de pesquisa”.

A fim de cumprir com o anonimato prometido no TCLE, os editores que responderam ao questionário eletrônico são designados como Respondente E1, Respondente E2 etc., enquanto os gestores de repositório são designados Respondente G1, Respondente G2 etc. Nesse mesmo sentido, os editores que participaram das entrevistas são denominados Entrevistado E1, Entrevistado E2 etc., ao passo que os gestores são denominados Entrevistado G1, Entrevistado G2 etc.

¹²⁸ Ver em: <<http://www.cfch.ufrj.br/index.php/comite-etica-apresentacao>>

5.1 Editores científicos: Questionário e Entrevista

A escolha de editores científicos para participarem da pesquisa realizada por meio de questionários eletrônicos recaiu sobre os profissionais que editam o conjunto de 7 periódicos da centenária Fiocruz¹²⁹ (Quadro 10) e aqueles que trabalham no conjunto de 18 periódicos da Coleção SciELO Saúde Pública¹³⁰ (Quadro 11).¹³¹

As entrevistas com editores foram realizadas com profissionais da Fiocruz, da Coleção Scielo Saúde Pública e do periódico *Cadernos de Saúde Coletiva*¹³², editado pela UFRJ. Buscou-se, assim, extrair um posicionamento de uma parcela representativa do campo da saúde a respeito da abertura e compartilhamento de dados de pesquisa e suas implicações com o direito autoral.

A opção pelos editores das sete revistas da Fiocruz (Quadro 10) deveu-se: à representatividade que elas angariaram no cenário de publicações científicas do setor de saúde brasileiro; ao fato de a instituição e seus periódicos estarem envolvidos no projeto de construção de uma política de abertura de dados de pesquisa desde 2015; e, também, em alguma medida, às atividades que o autor deste estudo exerce como editor executivo de um periódico científico da Fiocruz e exerceu como membro do Grupo de Trabalho em Ciência Aberta da Fiocruz, como coordenador do Grupo de Trabalho em Ciência Aberta – Marcos Legais, conforme já mencionado na Introdução.

¹²⁹ Ver em: <<https://periodicos.fiocruz.br/pt-br/revistas>>

¹³⁰ Ver em: <<https://www.scielosp.org/journals/alpha?status=current>>

¹³¹ Nota: a revista *Cadernos de Saúde Pública*, editada pela Fiocruz, integra os dois conjuntos de periódicos. Ver em: <<https://www.scielosp.org/>>.

¹³² Periódico editado pelo Instituto de Estudos em Saúde Coletiva, da UFRJ, no Rio de Janeiro, não cobra taxas para submissão e avaliação de artigos, obtém cessão não exclusiva dos direitos autorais para primeira publicação do artigo e não retém direitos de autor. Adota a licença CC BY. Não traz orientação sobre compartilhamento de dados de pesquisa ou direito autoral aplicável a dados de pesquisa, nem em seu site institucional, nem em sua página na SciELO. Fonte: <<https://www.scielo.br/revistas/cadsc/pinstruc.htm>>

Quadro 10. Conjunto de periódicos científicos da Fiocruz

Periódico	Cidade	Política para depositar dados	Taxa de publicação	Direitos autorais: artigos
<i>Cadernos de Saúde Pública</i>	Rio de Janeiro	Não	Não	CC BY; não retém direitos de autor/reserva direito à 1ª publicação
<i>História, Ciências, Saúde: Manguinhos</i>	Rio de Janeiro	Não	Não	CC BY; retém direitos de autor
<i>Memórias do Instituto Oswaldo Cruz</i>	Rio de Janeiro	Sim. “Os dados que geraram o manuscrito devem ser abertos e disponibilizados publicamente sempre que solicitados”	Não	CC BY; não retém direitos de autor
<i>Reciis</i>	Rio de Janeiro	Sim. “Estimula o depósito dos dados da pesquisa em repositório de dados abertos” e solicita que URL de acesso seja informada no artigo	Não	CC BY NC; não retém direitos de autor
<i>Revista Fitos</i>	Rio de Janeiro	Não	Não	CC BY NC; não retém direitos de autor
<i>Trabalho, Educação e Saúde</i>	Rio de Janeiro	Não	Não	CC BY; não retém direitos de autor
<i>Visa em Debate</i>	Rio de Janeiro	Sim. “Estimula o depósito dos dados da pesquisa em repositório de dados abertos” e solicita que URL de acesso seja informada no artigo	Não.	CC BY NC; retém direitos de autor

Fonte: Elaboração própria, 2020.

A escolha da coleção temática SciELO Saúde Pública (Quadro 11), abrigada na Scientific Electronic Library Online (SciELO)¹³³, seção Brasil, e de seus editores científicos como participantes foi feita obedecendo aos seguintes critérios: tanto os editores quanto as revistas pertencem ao campo da saúde, tópico de interesse deste projeto; as revistas da coleção apresentam variedade de cultura científica e de países onde são publicadas; e a SciELO tem como uma de suas prioridades executar um plano de alinhamento das coleções que abriga aos dogmas da ciência aberta, que inclui a abertura de dados de pesquisa. Esta coleção temática é composta por 18 revistas do campo da Saúde Pública, com 9 nacionais e 9 internacionais, que são publicadas na Espanha, Brasil, Colômbia, Argentina, Peru, México, Cuba e EUA.

¹³³ Ver mais sobre esse programa internacional de comunicação de pesquisa em: <<https://scielosp.org/about/about-scielo>>

Esse alinhamento à ciência aberta, cuja execução está prevista pela SciELO para o período de 2019 a 2023, portanto em plena execução, consiste, em suma, na

[...] adoção progressiva das boas práticas de comunicação da ciência aberta, que privilegiam rapidez e transparência nos processos e a abertura dos conteúdos subjacentes aos artigos em prol do seu reuso e da reprodutibilidade dos resultados (PACKER et al., 2018, *on line*).

As mudanças que a SciELO pretende implantar ao longo do quinquênio 2019-2023 já se fazem sentir: em abril de 2020, anunciou o início da operação do servidor “SciELO Preprints”¹³⁴ (SCIELO, 2020a) e em agosto do mesmo ano lançou uma operação piloto do repositório de dados “SciELO Data”¹³⁵ (ROCHA, 2020), para promover o domínio progressivo das metodologias e tecnologias. A implantação do repositório está prevista para ocorrer até o final de 2020, em cooperação com os periódicos da coleção. Além dessas duas etapas rumo a uma estrutura voltada para a ciência aberta, a SciELO prevê ainda a adoção de artigos de dados; avaliação por pares aberta de dados; e atualização das instruções aos autores baseada em práticas de ciência aberta (PACKER et al., 2018; PACKER; SANTOS, 2019). Detectamos, portanto, uma boa oportunidade de pesquisa, pois é de se imaginar que os editores estejam mobilizados, ou se mobilizando, para o pretendido alinhamento à ciência aberta.

Quadro 11. Coleção temática de periódicos científicos da SciELO Saúde Pública

Periódico	Cidade	Política para depositar dados	Taxa de publicação	Direitos autorais: artigos
<i>Cadernos de Saúde Pública</i>	Rio de Janeiro	Não	Não	CC BY; não retém direitos de autor/reserva direito à 1ª publicação
<i>Ciência & Saúde Coletiva</i>	Rio de Janeiro	Não	Não	CC BY; retém direitos de autor
<i>Epidemiologia e Serviços de Saúde</i>	Brasília	Sim. Publica ensaio randomizado se autor concordar em compartilhar dados a pedido[...] ¹	Não	CC BY NC e CC BY, simultaneamente; não retém direitos de autor/reserva direito à 1ª publicação

¹³⁴ Ver em: <<https://preprints.scielo.org>>

¹³⁵ Ver em: <<https://data.scielo.org>>

<i>Gaceta Sanitaria</i>	Barcelona	Incentiva autor a compartilhar dados em banco de dados e repositórios públicos[...]²	Sim. 450 euros p/ artigos e 200 p/ outros textos. Seções curtas são isentas.	CC BY NC-ND; e CC BY, para atender a órgão financiador; retém direitos de autor
<i>Interface - Comunicação, Saúde, Educação</i>	Botucatu	Não	Sim. Artigo em acesso aberto p/ algumas seções: 5 mil palavras: R\$ 800; + de 5 mil, R\$900.	CC BY
<i>MEDICC Review</i>	Califórnia	Não	Não	CCBY 4.0, atual; CC BY 2.0, p/ artigos publicados até 2015; retém direitos de autor
<i>Physis: Revista de Saúde Coletiva</i>	Rio de Janeiro	Não	Sim. Artigo aprovado: R\$ 600; Outras seções: R\$ 200.	CC BY; reserva direito à 1ª publicação; retém direitos de autor
<i>Revista Brasileira de Epidemiologia</i>	São Paulo	Incentiva autor a disponibilizar dados publicamente para compartilhamento³	Não	CC BY; reserva direito à 1ª publicação; retém direitos de autor
<i>Revista Cubana de Salud Pública</i>	La Habana	Não	Não	CC BY-NC-SA 4.0
<i>Revista de Salud Pública</i>	Bogotá	Não	Não	CC BY 4.0; retém direitos de autor
<i>Revista de Saúde Pública</i>	São Paulo	Não	Sim. R\$ 2.400 p/ artigos e R\$ 1.600 p/ comunicações breves	CC BY 4.0; retém direitos de autor
<i>Revista Española de Salud Pública</i>	Madri	Sim. Política do ICMJE.	Não	CC BY- NC-ND 3.0; não retém direitos de autor
<i>Revista Panamericana de Salud Pública</i>	Washington	Não	Não	CC BY- NC-ND 3.0; retém direitos de autor
<i>Revista Peruana Med. Exper. y Salud Pública</i>	Lima	Não	Não	CC BY NC 4.0; direito à 1ª publicação; não retém direitos de autor
<i>Salud Colectiva</i>	Buenos Aires	Não	Não	CC BY NC 4.0; não retém direitos de autor
<i>Salud Pública de México</i>	Cuernavaca	Não	Não	CC BY NC-SA; retém direitos de autor
<i>Saúde e Sociedade</i>	São Paulo	Não	Não	CC BY 4.0
<i>Saúde em Debate</i>	Rio de Janeiro	Não	Não	CC BY; retém direitos de autor

Fonte: Elaboração própria, 2020.

Notas: 1) Ver "Compartilhamento de dados": <<https://www.scielosp.org/journal/ress/about/#instructions>>; 2) Ver "Disponibilidad de bases de datos y material para réplica": <<https://www.gacetasanitaria.org/es-guia-autores>>; 3) Ver "Compartilhamento de dados": <<https://www.scielo.br/revistas/rbepid/pinstruc.htm>>

Questionário eletrônico para Editores científicos - A elaboração, a aplicação, o pré-teste e a coleta do questionário eletrônico para editores aconteceram entre 22 de maio e 14 de agosto de 2020. As perguntas desse questionário podem ser consultadas no Apêndice A.

Primeiro, procedeu-se a um pré-teste do questionário em português, que envolveu duas editoras da Fiocruz. A Respondente E1 absteve-se de fazer comentários. A Respondente E2 informou que não entendeu, de início, o motivo de interrogar “por que” uma revista decide adotar a abertura de dados subjacentes a artigos (pergunta 2, ver Apêndice A). Contudo, em contato que mantivemos com a Respondente E2, ela esclareceu que raciocina como editora de Humanidades e que nesta área há mais dificuldades para o entendimento sobre e como publicar dados do que em outras áreas. Acredita inclusive haver a necessidade de uma definição sobre quais poderiam ser os dados objeto de abertura e compartilhamento em Humanidades, discussão que já vem sendo travada na redação da revista em que trabalha. E acrescenta que, quanto à dúvida acima referida, para ela, a abertura de dados “passa pela confiabilidade do que está sendo publicado, pela possibilidade de começar uma nova pesquisa ou endossar a própria pesquisa” que um pesquisador queira reusar. Além disso, afirma que a abertura de dados à sociedade está ligada ao lugar onde a revista e o editor estão estabelecidos. No caso desta editora, em uma instituição pública, o que, segundo ela, requer compromisso com o uso do dinheiro público e visão da ciência como um bem público. Dessa maneira, segundo a respondente, “a abertura de dados é também uma prestação de contas do investimento em pesquisa com dinheiro público”.

Superada a fase de pré-teste, em junho/2020, o TCLE, contendo um *link* para o questionário, foi enviado aos editores da Fiocruz (Tabela 2), os quais o responderam em sua totalidade, e também aos editores dos periódicos nacionais da coleção SciELO Saúde Pública. Houve um reenvio para esses dois segmentos. Por fim, ainda em junho/2020, foi enviada uma versão em inglês do TCLE e do questionário aos 9 editores internacionais da coleção SciELO Saúde Pública (Tabela 2). Sem receber nenhuma resposta, foram feitos três reenvios. No primeiro, o questionário foi remetido em inglês e o TCLE em espanhol, exceto para a revista *Meddic Review*, que adota o idioma inglês. No segundo, o TCLE e o questionário foram mandados em inglês. Uma integrante do grupo de editores internacionais da SciELO Saúde Pública respondeu a este questionário, mas fez críticas ao envio do material em inglês, acrescentando que

[...] las revistas de SciELO Salud Pública que reciben trabajos de toda Iberoamérica aún no han resuelto claramente el tema de los datos abiertos, porque los ministerios y organismos públicos de salud tampoco lo han hecho, lo que imposibilita tomar decisiones unilaterales por parte de las revistas, quizás esto pueda ser una razón para la falta de respuesta.

Considero que es un tema en el que hay mucho aún por discutir y sobre todo incluir en los posgrados como formación en “capacidades en comunicación científica” (Respondente E16, por e-mail, 07-07-2020).

Assim, no terceiro envio, o TCLE e o questionário foram remetidos em inglês a 1 (*Meddic Review*) e em espanhol a 7 editores internacionais da SciELO Saúde Pública, porém o cenário pouco mudou: apenas mais dois editores desse grupo responderam (Tabela 2).

Tabela 2. Envio dos questionários eletrônicos aos editores científicos

	Nº revistas	Nº editores convidados	Não responderam	Respondentes
Fiocruz	7	8	-	8
SciELO Saúde Pública	18	9 nacionais	4	5
		9 internac.	6	3
Totais	25	26	10	16

Fonte: elaboração própria, 2020.

Entrevista com Editores científicos - Todos os editores científicos foram convidados por meio de TCLE (Apêndice E) para participarem das entrevistas. O objetivo foi aprofundar pontos da discussão contida no questionário eletrônico, por meio de uma interação mais dinâmica entre o entrevistador e o entrevistado. As entrevistas com editores obedeceram a um roteiro de perguntas semiestruturadas (Apêndice C) e estenderam-se pelo período de 10/08/2020 a 08/09/2020. Foram realizadas por intermédio da ferramenta Zoom, gravadas e transcritas. A seleção privilegiou editores científicos da área de saúde pública dos dois conjuntos de revistas da Fiocruz e da coleção SciELO Saúde Pública. Assim, foram convidados 2 editores da Fiocruz; 1 aceitou e o outro declinou o convite, alegando desconhecer o tema sobre direito autoral e licenciamento de dados de pesquisa por meio de licenças públicas na perspectiva das revistas científicas; 4 da coleção SciELO Saúde Pública, sendo que 2 aceitaram o convite e 2 não responderam; e 1 da *Nature Medicine*, que também não respondeu ao convite, reiterado uma vez. Para validação do roteiro de perguntas semiestruturadas, elaborado para as entrevistas, convidamos 1 editor científico da revista *Cadernos Saúde Coletiva*, editada pelo

Instituto de Estudos em Saúde Coletiva da UFRJ, que aceitou participar, não havendo qualquer comentário acerca do conteúdo das perguntas.

5.1.2 Um exame observacional dos *sites* das revistas selecionadas

Antes de proceder à análise das respostas dos editores, publica-se a seguir o resultado de um exame observacional sumário feito nos periódicos da Fiocruz (Quadro 10) e da SciELO Saúde Pública (Quadro 11), selecionados para a pesquisa empírica. O objetivo principal foi buscar identificar, por meio de observação, políticas de abertura e compartilhamento de dados de pesquisa, e orientações sobre como os autores devem lidar com direitos autorais eventualmente aplicáveis aos dados, incluindo informações sobre licenças públicas para reuso de dados de pesquisa subjacentes.

A pesquisa observacional se desenvolveu nos *sites* institucionais das revistas da Fiocruz e nos das revistas que integram o *hot site* SciELO Saúde Pública. A observação se replicou nas páginas de revistas da Fiocruz abrigadas na SciELO e nas páginas das revistas que compõem o *hot site* SciELO Saúde Pública.

A pesquisa foi iniciada em novembro de 2019 e atualizada e finalizada em agosto-setembro de 2020.

Dentre os *sites* dos periódicos pesquisados, apenas em 3, da Fiocruz, e em 4, da SciELO Saúde Pública, há registro de alguma orientação aos autores em apoio à publicação de dados de pesquisa subjacentes a artigos em repositórios de dados de pesquisa públicos (Quadro 10 e no Quadro 11). Contudo, não há qualquer orientação objetiva sobre como os autores devem lidar com direitos autorais eventualmente aplicáveis aos dados ou informações detalhadas sobre licenças públicas, do tipo Creative Commons, para que autores ou detentores de direitos autorais escolham a mais adequada para o que deseja com os dados a serem depositados. Também não há a exigência de plano de gestão de dados; não há cunho mandatório; não existe a imposição de assinatura em declaração de compartilhamento de dados aos autores; não há aconselhamento para o uso de identificadores digitais persistentes, como o ORCID e o DOI, nem recomendações sobre a citação dos autores originais dos dados que porventura forem reusados, ou ainda lista de repositórios recomendados.

Esse cenário pode indicar que há um longo caminho a ser percorrido no Brasil, no setor de periódicos de saúde pública, para se chegar ao patamar, mesmo ainda com falta de

regras padronizadas, dos periódicos científicos de países centrais analisados no capítulo 4, em termos de adoção de políticas de abertura de dados de pesquisa.

Há outro possível indicador da quase total ausência de políticas para aberturas de dados de pesquisa nas revistas científicas pesquisadas, associado ao fato de a maioria delas estar abrigada na coleção SciELO Saúde Pública e no SciELO Brasil. Em entrevista concedida a nós por dois editores cujas revistas integram essas duas coleções, quando perguntados sobre qual é o papel dos periódicos científicos e seus editores na abertura de dados de pesquisa, responderam:

A grande liderança neste processo é da SciELO. E os editores são reativos [às decisões dela] em relação a isso (Entrevistado E6).

O tema da abertura de dados também traz junto com ele a operacionalização e custos adicionais. É um tema com pouca penetração na comunidade científica no Brasil, em que a discussão chega mais em determinadas bolhas (Entrevistado E3).

Como já visto, a SciELO vem executando um plano de alinhamento à ciência aberta que ainda está muito incipiente, tendo lançado recentemente o piloto “SciELO Dados”, que poderá vir a ser o seu repositório público de dados de pesquisa. A fala dos editores quer significar, possivelmente, que eles atuam em compasso de espera de iniciativas e decisões da SciELO nesse campo.

É digno de nota que a maioria das seções pesquisadas, “Instruções aos autores”, “Sobre a revista” e assemelhadas, ainda é voltada apenas para atender aos aspectos relativos à redação, submissão, produção editorial e cessão de direito autoral tão somente de artigo científico para reprodução e distribuição. Apesar das referências a favor da ciência aberta e do movimento pelo acesso aberto na maioria das revistas, sempre se referindo a artigos científicos, não identificamos sinais consistentes de política editorial para lidar com dados de pesquisa em um contexto mundial de abertura e compartilhamento, conforme preconiza o RDA-Codata Legal Interoperability Interest Group (UHLIR; CLEMENT, 2016), para tratar de questões legais, ou mesmo o ICMJE (2018), tradicional comitê editorial composto por editores de revistas de saúde, cujas recomendações sobre abertura e compartilhamento de dados de pesquisa são seguidas apenas pela *Revista Española de Salud Pública*.

Dessa maneira, as abordagens sobre o direito autoral, como transferência total ou parcial de direitos patrimoniais e/ou utilização de licenças do tipo Creative Commons, convergem todas para os artigos científicos e para outras tipologias de textos científicos. Embora ofereçam acesso aberto a seus conteúdos, do conjunto SciELO Saúde Pública, 10 periódicos retêm direitos de autor; 5 não retêm; 3 não informam. Do conjunto Fiocruz, 2 retêm direitos de autor; e 5 não retêm. É como se essas revistas tivessem se organizado para a época do movimento pelo acesso aberto a artigos científicos e permanecessem com essa mesma estrutura demandada para o final do século XX, um sistema “prejudicado por um modelo secular que não foi atualizado para aproveitar as vantagens da tecnologia do século XXI” (SPARC, 2019, *on line*, tradução nossa). O Entrevistado E2 tem uma posição semelhante à da SPARC:

[...] as revistas têm um papel importantíssimo nesse processo [de abertura e compartilhamento de dados de pesquisa]. Agora, para que isso aconteça, elas vão ter que se desfazer dessa memória do século XX e abraçar as tecnologias do século XXI (Entrevistado E2).

[...] o formato atual do artigo, aquela coisa muito fechada, muito padronizada etc., ele não é adequado para esse cenário em que os dados terão independência e poderão existir por si só (Entrevistado E2).

Em conclusão, a abertura e o compartilhamento de dados de pesquisa, portanto, ainda não foram objeto de políticas robustas dos dois conjuntos de revistas científicas pesquisadas, que compõem o *hot site* SciELO Saúde Pública e o grupo de revistas científicas editadas pela Fiocruz. A exceção fica por conta da *Revista Española de Salud Pública*, que adota as regras do ICMJE. Seis fazem uma tímida alusão a orientações para disponibilização pública de dados de pesquisa, não se podendo dizer que constitua uma política de dados de pesquisa subjacentes a artigos.

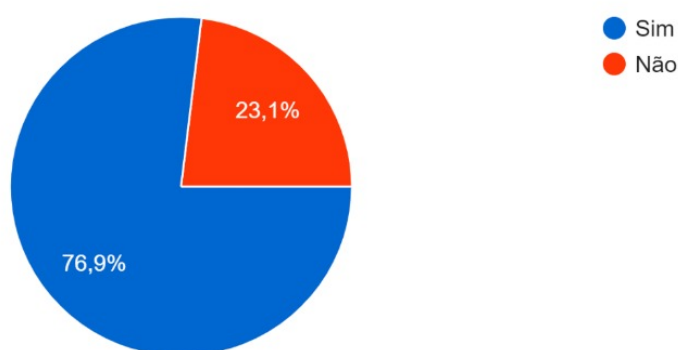
Cabe reiterar, no entanto, que a Fiocruz está atualmente em processo de elaboração de uma política de abertura de dados de pesquisa, e que a SciELO está em plena execução da implantação de um alinhamento de suas coleções de periódicos à ciência aberta, que abarca a adoção da abertura de dados de pesquisa (SANTOS; ALMEIDA; HENNING, 2017; GUANAES et al., 2018; PACKER et al., 2018; PACKER; SANTOS, 2019; SCIELO, 2020; ROCHA, 2020).

5.2 Abertura de dados e direito autoral: uma visão de editores do campo da saúde

A amostra de editores que responderam ao questionário eletrônico apresenta-se balanceada, com metade de respondentes da SciELO Saúde Pública e metade da Fiocruz (Tabela 2). Optamos por apresentar os resultados colhidos pelo questionário e pelas entrevistas, sistematizados em duas categorias. Relativamente à primeira **categoria**, “**Abertura, acesso, compartilhamento e reúso de dados de pesquisa**”, observa-se que uma maioria expressiva respondeu “sim” à **pergunta 1** sobre se é a favor da disponibilização pública de dados de pesquisa para fins de compartilhamento (Figura 5).

Figura 5. Disponibilização pública de dados de pesquisa

1. Você é favorável à disponibilização pública de dados de pesquisa subjacentes aos artigos científicos para fins de compartilhamento e reúso?



Ao responder por que é favorável à disponibilização de dados de pesquisa, o Respondente E2, da Fiocruz, foi abrangente: “Confiabilidade da pesquisa; verificação da metodologia usada; possibilidade de uso dos dados em outras pesquisas; pesquisas com financiamentos públicos devem ser entendidas como bem público”. De certa forma, talvez o alto percentual de “sim” possa representar, em pequena escala, o imaginário atual favorável à tese da abertura.

No entanto, as respostas negativas trazem aspectos a serem iluminados. A motivação da resposta negativa do Respondente E1 reflete uma preocupação comum na área da saúde, sobre privacidade e confidencialidade dos dados de participantes de pesquisa em saúde e integridade da pesquisa:

Existem vários dados que não devem ser disponibilizados ou que a mera informação de que serão disponibilizados pode produzir viés nas respostas de pesquisas qualitativas. [...] Portanto, um sim, sem condições explicitadas, me parece mais prejudicial do que benéfico às futuras pesquisas (Respondente E1),

enquanto as razões do “não” do Respondente E7 trazem a marca auspiciosa da dúvida e evidenciam questões fundamentais sobre a resistência de pesquisadores em compartilhar dados, além do inelutável problema ético da confidencialidade e privacidade:

Marquei o não porque eu tenho muitas dúvidas. Como tenho tradição de pesquisa qualitativa, há questões éticas e metodológicas que podem expor os sujeitos envolvidos na pesquisa. Há, por outro lado, em pesquisas de caráter documental ou discursivo, uma facilitação de pesquisadores futuros com dados de pesquisa já sistematizados a partir de arquivos e prontos para a análise. O trabalho intelectual também envolve a entrada no campo, o esforço de construção e realização de questionários e entrevistas ou o contato com um arquivo documental e não apenas análise de dados já sistematizados. Além disso, a marca autoral de um grupo de pesquisa ou de um pesquisador também está na construção metodológica e na pesquisa de campo (Respondente E7).

Estudos comprovam que há pesquisadores que retêm conjuntos de dados por causa das preocupações expressadas pelos Respondentes E1 e E7. De acordo com Hrynaskiewicz e Altman (2009), os problemas comuns para publicação de dados brutos incluem a privacidade do paciente e a propriedade, mas destacam que essas preocupações não parecem ser abordadas adequadamente nas orientações ou mandatos estabelecidos pelas revistas. Savage e Vickers (2009) observaram razões para a retração de abrir os próprios dados que incluem preocupações com a privacidade do paciente (para áreas médicas), as futuras oportunidades de publicação e o desejo de reter direitos autorais exclusivos de dados que levaram muitos anos para ser produzidos.

A prática do compartilhamento de dados também gera polêmicas no que concerne ao trabalho intelectual e esforço despendidos, conforme acentuou o Respondente E7. Em editorial publicado pelo *The New England Journal of Medicine*, Longo e Drazen (2016, *online*) introduziram o conceito de “pesquisadores parasitas”, o que gerou uma grande discussão

(o editorial obteve mais de 131 mil visualizações¹³⁶) sobre a reanálise e o reúso de dados de pesquisa. Ambos criticavam indivíduos que

[...] não têm nada a ver com o desenho e execução do estudo, mas usam os dados de outro grupo para seus próprios fins, possivelmente roubando a produtividade da pesquisa planejada pelos coletores de dados, ou mesmo usam os dados para tentar refutar o que os investigadores originais tinham postulado (LONGO; DRAZEN, 2016, *on line*, tradução nossa).

Em contrapartida, Greene et al. (2017, *on line*, tradução nossa), em artigo publicado na *Nature Genetics*, criaram o Prêmio de Parasita de Pesquisa, para homenagear os “pesquisadores parasitas”, ou seja, “indivíduos que praticam a arte de reanálise de dados para fins novos”, validando resultados e defendendo que “o processo pelo qual os cientistas retestam seletivamente suposições e descobertas anteriores constitui a base da descoberta científica em todas as disciplinas”. Porém, frisam que “Bons parasitas de pesquisa também têm a obrigação de dar crédito a seus hospedeiros de pesquisa” (GREENE et al., 2017, *on line*, tradução nossa).

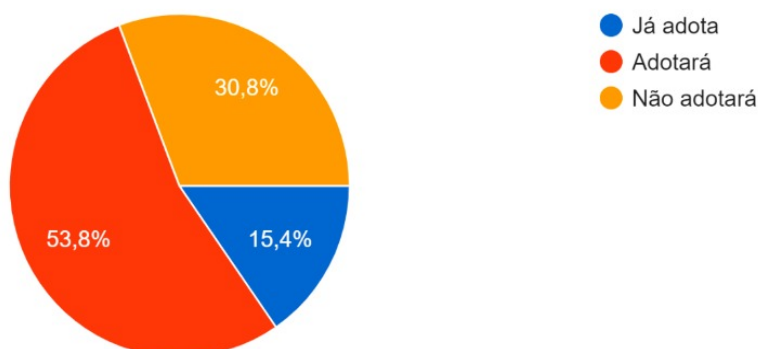
Apesar do percentual favorável de editores à abertura e compartilhamento de dados, ao se manifestarem, na **pergunta 2**, sobre a adoção da abertura, os respondentes que já a adotam totalizam o menor percentual (Figura 6), significando possivelmente que ele resulta do período de adequação pelo qual os dois conjuntos de periódicos pesquisados passam para se alinhar aos ditames da ciência aberta, principalmente à abertura e compartilhamento de dados de pesquisa.

As respostas “Não temos ainda um consenso sobre como funcionaria a abertura de dados de pesquisa qualitativa [...]” e “[...] tenho ainda muitas dúvidas. O debate ainda precisa ser maior” denotam, provavelmente, que ainda não há discussão suficientemente amadurecida nas instâncias editoriais das revistas e que há um processo de mudança em andamento.

¹³⁶ Ver em: <<https://www.nejm.org/doi/metrics/10.1056/NEJMe1516564>>

Figura 6. Adoção da abertura de dados de pesquisa

2- Em relação à abertura de dados de pesquisa subjacentes aos artigos, sua revista:



Por isso, talvez o cenário de incerteza e pouca discussão sobre o tema não permita aos editores respostas mais coerentes e se reflita no percentual negativo dos que “não adotarão” ser maior do que “os que não são favoráveis”. Tais escolhas também podem estar influenciadas pela preocupação daqueles que lidam com pesquisa qualitativa que, por sua vez, lidam com dados sensíveis, privacidade e confidencialidade. É o que se depreende a partir das seguintes respostas:

Estamos adaptando a estrutura tecnológica (Respondente E4).

Estamos implementando la publicación continua, los preprint y el acceso a los datos de investigación de los artículos que aceptamos para publicación (Respondente E14).

Aún no están claras las condiciones de apertura de muchos datos. Por ejemplo, si las grandes bases de datos de un Ministerio de Salud nacional o provincial no tiene políticas claras de apertura de datos, las investigaciones derivadas de esos datos no tendrían condiciones para abrir esos datos. Por otra parte, no hay una cultura instalada para el depósito de los datos en plataformas de datos abiertos, ni de reutilización de datos que están disponibles (Respondente E16).

Registre-se que dois respondentes da Fiocruz e dois respondentes da SciELO Saúde Pública afirmam que não vão adotar a abertura de dados. Eles só voltam a responder na sétima pergunta, uma vez que da pergunta 3 à 6 os temas só interessam àquele que “já adota” ou “adotará”.

Dois terços dos editores responderam, na **pergunta 3**, que exigirão do autor uma declaração de concordância em compartilhar dados de pesquisa, enquanto um terço não empregará essa prática. Seguindo uma tendência dos países centrais, conforme visto no capítulo 4 desta tese, na análise das revistas *Nature Medicine*, *theBMJ* e *Plos Medicine*, mais da metade dos respondentes condicionará a publicação do artigo à declaração de concordância assinada pelo pesquisador. No entanto, uma voz da terça parte negativa faz uma relevante ponderação, ligada à **pergunta 4**, sublinhando a diferença de realidades dos países latino-americanos, que talvez impeça a imposição de condições de aceitação de um trabalho, caso o pesquisador não assine a declaração de concordância:

Las realidades de los países de Iberoamérica son muy distintas como para imponer una norma que condicione la aceptación de un trabajo. Lo mismo sucede con las cuestiones éticas en investigación. No en todos los países las investigaciones en las que participan personas tienen la obligatoriedad de que el proyecto sea aprobado por un comité de ética. Por esa razón consideramos que la apertura de datos no puede “imponerse”, sino que debe ser sugerida como una de los lineamientos de la ciencia abierta (Respondente 16E).

À **pergunta 5**, sobre se exigirão plano de gestão de dados, mais da metade respondeu que sua revista não exigiria, sob argumentos diversos. Assim, o Respondente E2 afirmou que “Na história, não saberia como indicar esse plano de gestão dos dados”, e o Respondente E13 crê também ser de difícil aplicação em sua área por causa da “natureza das pesquisas em ciências humanas”. A razão do “não” do Respondente E4 prende-se ao fato de que “será exigido o acesso ao repositório de dados”, pois “O plano de gestão de dados deve ser uma exigência da instituição para a qual os autores trabalham”. Esta afirmativa é corroborada pela Respondente E9, para quem “Tais iniciativas devem partir dos financiadores, fazem sentido quando implementadas mais precocemente, ainda no planejamento da pesquisa, e não somente no momento da publicação”. Há revistas que ainda não trataram do tema. A Respondente E8 diz “não” ao plano de gestão porque “Nesse momento, como temos outras questões mais emergentes [...] ainda não é possível afirmar essa exigência [...]”, já o “não” do Respondente E12 significa ausência de decisão sobre a questão, pois “O Conselho editorial

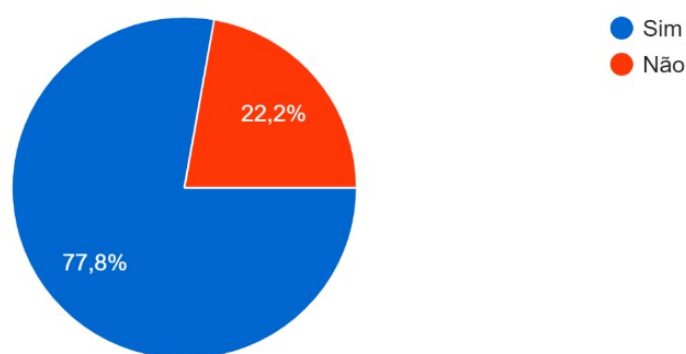
não decidiu ainda esse assunto [...]”. Segundo o Respondente E16, “El plan de gestión de datos ya es un requisito obligatorio del CONICET¹³⁷ en Argentina”.

Vale registrar que o ICMJE (2018), comitê internacional de normas para publicação de artigos sobre pesquisas em saúde, adotado por grande parte do conjunto de revistas pesquisadas, recomenda que “Os ensaios clínicos que começarem a inscrever participantes em ou após 1º de janeiro de 2019 devem incluir um plano de compartilhamento de dados no registro do ensaio”¹³⁸ (tradução nossa). O plano de gestão de dados de pesquisa vem sendo exigido por agências de apoio à pesquisa em todo o mundo, inclusive no Brasil, pela Fapesp, no momento da submissão de projetos¹³⁹ (NIH, 2003a; 2003b; COMISSÃO EUROPEIA, 2017; BILL & MELINDA GATES FOUNDATION, 2017a; 2017b; THBMJ, 2019).

Formularam-se três **perguntas (6, 7, 8)** sobre repositórios de dados. A maioria dos respondentes indicaria um ou mais repositórios, específicos de dados (**pergunta 6**), para depósito e disponibilização pública de dados de pesquisa (Figura 7).

Figura 7. Indicação de repositórios

6- Sua revista indica(rá) para o(s) autor(es) um ou mais repositórios, específicos de dados, para depósito e disponibilização pública de dados subjacentes ao artigo?



¹³⁷ O Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), autarquia do Ministério de Ciência, Tecnologia e Inovação Produtiva da Argentina, promove a ciência e a tecnologia naquele país. Ver em: <<https://www.conicet.gov.ar/>>

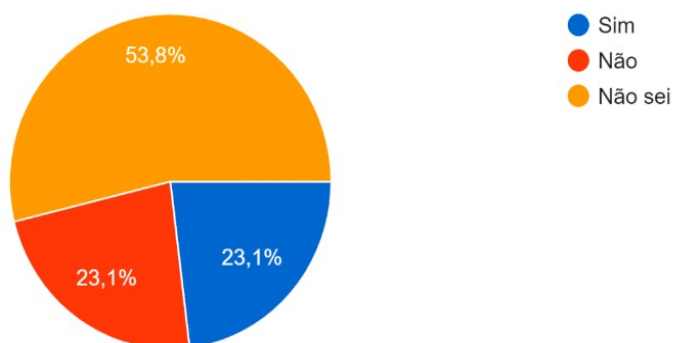
¹³⁸ Ver em: <<http://www.icmje.org/recommendations/browse/publishing-and-editorial-issues/clinical-trial-registration.html>>

¹³⁹ "Plano de gestão de dados se incorporam a projetos de pesquisa no Brasil". Pesquisa Fapesp. Ver em: <<https://revistapesquisa.fapesp.br/planos-de-gestao-de-dados-se-incorporam-a-projetos-de-pesquisa-no-brasil/>>. O Conselho Europeu de Pesquisa oferece um *template* de um plano de gestão de dados. Ver em: <https://ec.europa.eu/research/participants/data/ref/h2020/other/hi/oa-pilot/h2020-hi-erc-oa-guide_en.pdf>

No entanto, nova maioria formou-se em torno da resposta “não sei” à **pergunta 7**, que indagava aos participantes se indicariam repositório que aceita embargo (Figura 8). Essa dúvida pode expressar a pouca discussão sobre o tema em instâncias editoriais dessas revistas, parecendo reafirmar a incipiência da implementação da abertura de dados nesse setor.

Figura 8. Repositório com embargo

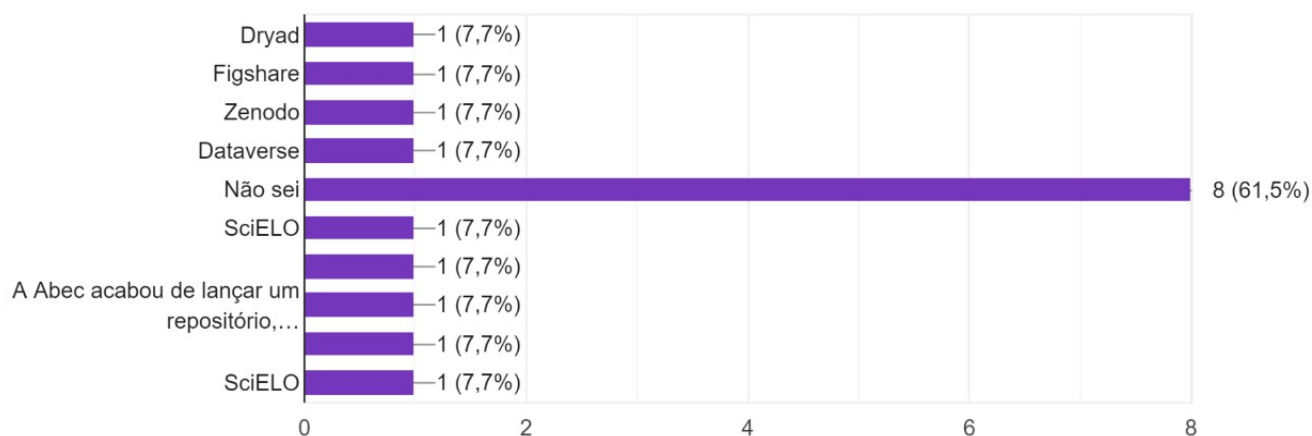
7- Sua revista indicaria repositório que aceita embargo que impeça a publicação imediata do conjunto de dados?



O mesmo “não sei” foi majoritário para a **pergunta 8**, de múltipla escolha, sobre que repositório a revista indicaria ao autor que desejasse compartilhar seus dados (Figura 9). Dataverse e SciELO foram repositórios citados em acréscimo aos submetidos à múltipla escolha.

Figura 9. Que repositório indicar?

8- Dentre os repositórios abaixo, qual a sua revista indicaria? Se quiser, marque mais de uma opção.



A resposta majoritária “Não sei” às **perguntas 7 e 8**, de novo, pode estar associada ao momento de transição para práticas de ciência aberta a que essas revistas estão se submetendo, tanto na Fiocruz quanto na SciELO. Ante esse cenário, é também possível inferir que essa discussão sobre a relação das revistas com as variadas nuances – editoriais, culturais, legais – que os repositórios de dados de pesquisa podem provocar ainda não está suficientemente amadurecida nas redações das revistas pesquisadas. Contudo, para um dos participantes,

La realidad es que, en la actualidad, las y los autores no tienen mucho conocimiento acerca de repositorios como Dataverse, cómo funcionan y cómo se deben depositar los datos, por lo que seguramente la revista deberá socializar información y generar guías informativas (Respondente E16).

As quatro últimas perguntas a seguir, cujos discursos inserem-se na **categoria “Direito autoral aplicável a dados de pesquisa”**, foram formuladas dentro do interesse de se verificar sumariamente como as revistas e seus editores se relacionam com o direito autoral.

Uma questão que certamente preocupa na abertura e compartilhamento de dados de pesquisa deriva de uma pergunta: “Quem, afinal, é o dono dos dados de pesquisa?”. Se o direito autoral for aplicável a determinado conjunto de dados de pesquisa, que, para tanto, deverá apresentar requisitos de criatividade e originalidade na sua criação intelectual para obter esse tipo de proteção, o autor é a pessoa física que criou a obra intelectual, no caso o conjunto de dados. Assim, inicialmente, é o autor o titular de direitos exclusivos sobre esse conjunto de dados: os direitos patrimoniais, que lhe conferem a prerrogativa de transferi-los para terceiros por meio de cessão ou licença, e os direitos morais, estes inalienáveis e irrenunciáveis (GANDELMAN, 2007; BITTAR, 2008; CARROL, 2015).

A Entrevistada G6 acrescenta uma carta que muda o jogo: o pesquisador que tem contrato de trabalho com uma instituição e que, nesse contrato, cede os direitos autorais de seus trabalhos para a instituição que é sua empregadora.

Na minha instituição, a Universidade da Califórnia, e acredito que em muitas instituições, a instituição é proprietária dos dados. Faz parte do acordo de ser pesquisador de uma instituição. É por isso que é melhor que os dados sejam licenciados [pelo instrumento] CC0, uma vez que os pesquisadores não possuem os dados e o editor certamente não deveria. Este é outro motivo pelo qual os dados não devem ser publicados no artigo como arquivos de apoio, pois herdariam a licença do artigo (Entrevistada G6, tradução nossa¹⁴⁰).

As **perguntas 9 e 10**, “Quem deve fazer o envio dos dados de pesquisa subjacentes a um artigo para fins de depósito em repositório?” e “Na sua opinião, a quem cabe decidir sobre qual será a licença pública que descreve os direitos para fins de reuso de dados subjacentes aos artigos?” foram formuladas com a intenção de avaliar se os editores tinham o conhecimento da diferença entre autor e detentor da titularidade de direitos autorais. Quanto à pergunta 9, os editores apontaram por unanimidade a resposta “O autor com capacidade para definir o licenciamento para reuso dos dados”, pois, além de autor, é informado nesta resposta que ele tem “capacidade” para agir, isto é, prerrogativa legal para tal. Portanto, é autor da obra e detentor dos direitos autorais. No entanto, na pergunta 10, apenas o Respondente E12

¹⁴⁰ No original: “At my institution, University of California, and I believe at many institutions, the institution owns the data. It is part of the agreement in being a researcher at an institution. That is why it is best that the data be licensed CC0 since the researchers do not own the data and the publisher certainly should not. This is another reason why data should not be published within the article as supporting files, it would inherit the article license.”

apontou a resposta “O detentor dos direitos autorais do conjunto de dados”, isto é, aquele que, por ser o titular dos direitos autorais, tem a prerrogativa de decidir sobre licenciamento. Tal fato poderia indicar um conhecimento insuficiente do grosso dos participantes sobre o que diferencia um autor de um detentor de direitos autorais.

A **pergunta 11** – “[...] Que licença a revista recomendaria para reuso legalizado dos dados de pesquisa subjacentes aos artigos disponibilizados publicamente?” – foi formulada para se tentar obter um retrato do conhecimento dos participantes sobre o licenciamento de dados de pesquisa subjacentes a artigos para compartilhamento e reuso, especificamente sobre as licenças Creative Commons (ver seção 3.4, no capítulo 3).

Dois resultados chamam a atenção: o primeiro é o alto índice de respostas que contemplam a opção por tipos de licença que contêm a limitação “não comercial” (NC). Dentre os 8 editores da Fiocruz, 4 cravaram a licença CC BY NC, que permite distribuir, remixar, adaptar, exceto comercializar o trabalho; 1 assinalou a licença CC BY NC SA, que se assemelha à CC BY NC, distinguindo-se dela apenas por exigir que os novos trabalhos sejam licenciados sob os mesmos termos; e 2 assinalaram a CC BY, que permite inclusive a comercialização do trabalho (Veja a polêmica CC BY NC *versus* CC BY na seção 3.5 do capítulo 3); e 1 não marcou nenhuma das opções oferecidas, manifestando-se pela opção “Outros”: “Gosto da ideia mais radical de abertura dos dados, assim como usamos no acesso aberto. Mas confesso que, quando afirmo isso, estou pensando na área de história [...]” (Respondente E2).

Em relação aos editores da coleção ScELO Saúde Pública, 3 editores internacionais manifestaram-se pela CC BY NC; de 5 editores nacionais desta coleção, 2 optaram pela CC BY NC; 2 pela CC BY; e 1 pela CC BY NC ND, a mais restritiva das seis licenças CC, pois permite *download* do trabalho e compartilhamento, mas sem que possa ser alterado de nenhuma forma ou utilizado para fins comerciais.

O segundo resultado que chama a atenção é que nenhum dos respondentes dos dois grupos de revistas optou pela hipótese mais aberta – a ferramenta CC0 Dedicção ao Domínio Público. Porém, Dryad, Figshare e Zenodo, repositórios de dados de pesquisa, e BioMed Central, editora científica de acesso aberto, a adotam como padrão para dados de pesquisa, na medida que a lei permitir, por defenderem que maximiza o potencial de descoberta e reutilização de dados e suporem que o licenciamento aberto pode aumentar o crédito

individual e as citações (DRYAD, 2001; HRYNASZKIEWICZ; BUSCH; COCKERILL, 2013).

Quando quase metade dos editores (Figura 10) responde “É indiferente” à **pergunta 12** – “Que repositório de dados de pesquisa a sua revista indicaria levando em conta o tipo de política de direitos autorais que ele adota?” –, significa, aparentemente, que ainda estão atrelados ao modelo no qual o autor/detentor dos direitos autorais de um artigo científico aceita fazer gratuitamente a cessão total dos seus direitos patrimoniais para o editor em troca da publicação do artigo.

O movimento pelo acesso aberto à informação científica e a ciência aberta trouxeram novos ares, mas, com base nessa resposta, é possível deduzir que o antigo regime ainda está bem vivo. O regime de direito autoral proposto para acesso aberto, de acordo com a conferência de Budapeste (2002), limita-se apenas a preservar a paternidade de quem criou a obra (artigo ou conjuntos de dados de pesquisa, no caso), por meio da devida citação.

Figura 10. Repositório e política de direitos autorais

12- Que repositório de dados de pesquisa a sua revista indicaria levando em conta o tipo de política de direitos autorais que ele adota?



A transferência exclusiva e em definitivo de direitos patrimoniais deve ser bastante ponderada. Do contrário, poderão existir situações tais como o autor original de um conjunto de dados ou de um artigo ter que pedir autorização a quem passou a deter os direitos autorais de seu trabalho para fazer qualquer uso dele. De posse dos direitos cedidos, o detentor pode impedir ou limitar a disseminação do trabalho do autor original em repositórios, blogs e

para fins comerciais. A indiferença quanto ao regime de direito autoral para dados de pesquisa só interessa à poderosa estrutura comercial, que não hesitará em se apropriar e transformar dados em mercadoria vendável.

Para quem professa os princípios da ciência aberta e do acesso aberto à informação científica, e está em processo de adequação para aderir ao movimento de abertura e compartilhamento de dados de pesquisa, a postura de indiferença mediante ao regime de direito autoral a adotar para dados de pesquisa parece incongruente.

É alvissareiro, contudo, que a outra quase metade tenha respondido “Indicaria um que obtenha um direito *não exclusivo* para publicar esses dados, mantendo os direitos patrimoniais com o(s) autor(es) ou detentor(es)”.

A Respondente E16 vai no âmago da questão:

Una de las premisas del acceso abierto es que los derechos patrimoniales de los datos y de las obras siempre quedes en manos de las y los autores, nunca ceden esos derechos (Respondente E16).

5.3 Gestores de repositórios: Questionário e Entrevista

A decisão de também ir a campo para fazer a escuta de gestores de repositórios de dados de pesquisa, seja por questionário, seja por entrevista, deu-se em razão do papel que esses atores assumiram no bojo de uma nova conformação da comunicação científica, que vem exigindo cada vez mais acesso aberto a artigos e dados de pesquisa (COMISSÃO EUROPEIA, 2017). Ademais, o repositório vem também incorporando em definitivo, na medida das possibilidades de países centrais e periféricos, as tecnologias de informação e comunicação, para atender à demanda básica de armazenamento e vinculação de dados provenientes de artigos de periódicos que reportam uma pesquisa original, num envolvimento com editores e com a questão do direito autoral em dados de pesquisa (ASSANTE et al., 2016).

Assim, selecionamos como participantes da pesquisa por meio de questionário eletrônico gestores dos seguintes repositórios de dados de pesquisa: Fiocruz; IBICT;

Metabuscador de dados da Fapesp¹⁴¹; os de quatro universidades brasileiras – Universidade Federal do Paraná (UFPR), da Universidade Federal de Santa Catarina (UFSC), Universidade Federal do Rio Grande do Sul (UFRGS) e Universidade de Brasília (UnB) –; e da Universidade do Minho (Tabela 3).

As entrevistas foram conduzidas com gestores do IBICT, da Rede Nacional de Pesquisa (RNP), do Metabuscador de dados da Fapesp, da UFPR e do repositório internacional de dados Dryad.

Como justificativas das escolhas dos participantes, têm-se: o Metabuscador, por seu pioneirismo no Brasil, ao reunir as universidades públicas de São Paulo em um só lugar para o depósito de dados de pesquisa que suas comunidades científicas produzem e pela contribuição ao fomento da cultura de compartilhamento de dados; as quatro universidades brasileiras, com o objetivo de obter uma sinalização, ainda que sumária, do setor universitário sobre a implementação de repositórios de dados de pesquisa, suas relações com os periódicos e como lidam com o direito autoral aplicável a dados de pesquisa; e o repositório DataRepositóriUM¹⁴², da Universidade do Minho, para se obter um olhar internacional sobre como os repositórios de dados estariam lidando com a questão.

A iniciativa paulista merece um registro à parte sobre a sua construção e lançamento. A Entrevistada G4 assim se pronunciou sobre o projeto de implantação:

O repositório é a forma que permite um novo tipo de comunicação mediado por dados. A decisão de implantação de um protótipo do metabuscador foi feita em 2013. Em 2017, o protótipo foi implantado. A estratégia descentralizada do projeto funcionou porque obrigava as sete universidades envolvidas a agir. Depois de várias intenções e protótipos, a Rede foi lançada em dezembro de 2019. Até a inauguração, o metabuscador não tinha problemas de direito autoral, ou propriedade intelectual, mas, após entrar em funcionamento, está começando a ter (Entrevistada G4).

¹⁴¹ Ver em: <<https://metabuscador.uspdigital.usp.br/>>

¹⁴² Ver em: <<https://www.uminho.pt/PT/investigacao-e-inovacao/cienciaaberta/Paginas/gestao-abertura-dados.aspx>>

Em 16 de dezembro de 2019, a Fapesp lançou a Rede de Repositórios de Dados Científicos do Estado de São Paulo, que ficou conhecida como “Metabuscador de dados”. Trata-se de um portal – um buscador de metadados – que foi desenvolvido pela USP e permite busca por instituição, autor, assunto, ano ou palavras-chave. A rede é composta pelas seis universidades públicas do Estado de São Paulo – Universidade Estadual de Campinas (Unicamp), Universidade de São Paulo (USP), Universidade Estadual Paulista (Unesp), Universidade Federal de São Carlos (UFSCar), Universidade Federal do ABC (UFABC) e Universidade Federal de São Paulo (Unifesp) –, o Instituto Tecnológico de Aeronáutica (ITA), Embrapa Informática Agropecuária (CNPTIA/Embrapa) e Fapesp Open Science, que, em meados de 2020, lançou o repositório “Covid-19 Data Sharing”, com três coleções de dados. Projeto que privilegia a descentralização, cada instituição desenvolveu o próprio repositório, criou grupos permanentes internos para sua gestão e meios para os repositórios se integrarem. Esta integração é viabilizada por um portal único, que diariamente busca informações sobre os dados de cada instituição e disponibiliza essas informações [metadados] de forma integrada. A iniciativa disponibiliza em uma plataforma aberta dados associados às pesquisas desenvolvidas em todas as áreas do conhecimento no Estado de São Paulo (FAPESP, 2019).

Apesar do valor da iniciativa, ao que parece o metabuscador ainda se mostra incipiente, tanto que em visita ao *site*, em meados de setembro de 2020, verificou-se que o ITA ainda não havia disponibilizado nenhuma coleção de dados; a Unifesp, apenas 1 coleção; a Universidade Federal do ABC, 3; a Unicamp e a Unesp, 5 cada uma; a UFSCar, 6; a USP, 26; e a Embrapa comparece com 270 imagens de uma coleção sobre doenças de um tipo de plantas, porém, “não é o repositório institucional de dados de pesquisa da Embrapa, que neste momento encontra-se em trabalho de desenvolvimento e implantação, que poderá ser disponibilizado em outubro de 2020, com o nome de AGdados”, de acordo com a Entrevistada G3.

É provável que o resultado apresentado pelo Metabuscador deva-se a fatores tais como: lançamento ainda recente (em dezembro de 2019); mudança de cultura, na qual o próprio pesquisador passa a depositar dados publicamente; e efeitos da pandemia de Covid-19 sobre as atividades das instituições que compõem a Rede, no ano de 2020.

A Universidade Federal do Paraná também merece ser destacada, pois, em 2018, viria a ser uma das pioneiras nesse setor da produção científica brasileira ao criar o repositório

“Base de Dados Científicos”¹⁴³, exclusivo para dados de pesquisa, que foi incorporado ao Sistema de Bibliotecas da universidade. No entanto, ainda apresenta limitações, pois o repositório não é público para *upload* de dados e atende apenas aos dados de pesquisa produzidos por pesquisadores e alunos de pós-graduação da universidade. Tais dados devem ter obrigatoriamente alguma vinculação com publicações da universidade, que podem ser teses, dissertações e artigos. Um dos motivos de não ser público é financeiro, pois a universidade teria de arcar com o custo dos DOIs atribuídos aos conjuntos de dados. Por outro lado, o acesso e o compartilhamento são abertos.

De acordo com o Entrevistado G5, do repositório Base de Dados Científicos, no atual estágio, ele apenas abriga dados brutos, autoarquivados por pesquisadores da instituição, mediante a verificação de bibliotecários, ainda não possui uma equipe fixa e a pouca divulgação interna sobre ele prejudica a sua utilização, que se expressa no número reduzido de dados armazenados.

Apesar de a UFSC e a UnB ainda não terem um repositório específico de dados de pesquisa, elas foram convidadas a dar sua visão sobre o tema. A Universidade Federal do Rio Grande do Sul respondeu por *e-mail* que “ainda não temos um repositório de dados de pesquisa. Estamos na fase inicial de criação do mesmo, assim sendo, possivelmente não teremos como responder seu questionário” (Mensagem e-mail, Gerência do Lume, 07/07/2020).

O gestor do DataRepositóriUM, da Universidade do Minho, embora tenha tido ciência do envio do TCLE e sido contatado pelo Messenger, não respondeu ao questionário. Assim, a visão internacional de repositório de dados de pesquisa coube à Entrevistada G6, do repositório Dryad, que respondeu a uma série de questões que lhe foram enviadas por escrito, a título de entrevista, e por *e-mail*. Inquirida sobre quais são as principais dificuldades e desafios para os repositórios adotarem as políticas de abertura de dados subjacentes aos artigos, ela respondeu:

Não acredito que os repositórios tenham muitos desafios no suporte às políticas de dados de periódicos, porque os repositórios e as comunidades de apoio ao gerenciamento de dados de pesquisa parecem estar mais avançados na adoção/promoção/implementação de práticas abertas de compartilhamento de dados. Parece que o mundo dos periódicos está tentando recuperar o atraso,

¹⁴³ Ver em: <<https://bdc.c3sl.ufpr.br>>

pois os recursos que oferecemos nos repositórios de dados de pesquisa são mais maduros ou avançados do que as políticas de dados dos periódicos. A maioria das políticas de dados de periódicos *recomenda*, mas não exige que os dados sejam abertos, e a maioria dos periódicos ainda aceita que os dados sejam publicados como arquivos de apoio dentro do artigo. Isso não é a melhor prática e realmente não torna os dados reutilizáveis. O desafio seria o que estou tentando defender, isto é, que os periódicos tenham políticas mais rígidas para garantir que os dados sejam mantidos em repositórios que apoiem as melhores práticas (Entrevistada G6, tradução nossa¹⁴⁴).

Questionário eletrônico para gestores - O pré-teste do questionário eletrônico foi realizado entre 22/05/2020 e 26/05/2020, com uma gestora de repositório da Fiocruz e com um gestor do IBICT, que não teceram qualquer comentário crítico. A consulta aos gestores por questionário eletrônico aconteceu de 22 de maio a 8 de julho de 2020. As perguntas do questionário eletrônico para gestores de repositórios podem ser consultadas no Apêndice B.

O questionário dos gestores de repositórios seguiu o mesmo modelo do dos editores, no que concerne à concepção, aplicação, coleta e análise das repostas. As perguntas foram divididas nas duas categorias de temas de interesse do estudo, com o objetivo de obter uma visualização organizada na fase de análise do discurso produzido: 1) “Abertura, acesso, compartilhamento e reuso de dados de pesquisa”; e 2) “Direito autoral aplicável a dados de pesquisa”. Elaborado no Google Formulários, o *link* do questionário eletrônico foi enviado por *e-mail*, juntamente com o TCLE, para os 16 gestores das instituições constantes da Tabela 3. Destes, 8 não responderam e 8 aceitaram participar respondendo às perguntas do questionário específico para gestores.

¹⁴⁴ No original: “I do not believe that repositories have many challenges in supporting journal data policies because repositories and the research data management supporting communities seem to be farther along in adopting/promoting/implementing open data sharing practices. It feels like the journal world is playing catch-up a bit in that the features we offer at repositories for research data are more mature or advanced than journal data policies. Most journal data policies *recommend* but do not require that data are open, and most journals still accept data to be published as supporting files within the article. That is not best practice and really does not make the data re-usable. The challenge would be that I am trying to advocate for journals to have stricter policies around making sure data are being held in repositories that support best practices.”

Tabela 3. Envio dos questionários eletrônicos aos gestores de repositórios

	Nº repositórios	Nº gestores convidados	Não responderam	Respondentes
Fiocruz	1	1	-	1
IBICT	1	1	-	1
Metabuscador Fapesp	9	9	5	4
Universidades brasileiras	4	4	2	2
Universidade Minho	1	1	1	-
Totais	16	16	8	8

Fonte: Elaboração própria, 2020.

Entrevista com gestores de repositórios - Da mesma forma que os editores, os gestores de repositórios foram convidados a participar das entrevistas por meio de TCLE. Foram conduzidas no período de 21/08/2020 a 23/09/2020. Todos os convidados aceitaram participar. Como o objetivo das entrevistas era também aprofundar pontos da discussão contida no questionário eletrônico para gestores, deu-se prioridade àqueles do grupo pré-selecionado, mas também foram convidados gestores do IBICT e da RNP, que atuam em projetos nacionais relevantes, como a “Parceria para Governo Aberto”¹⁴⁵ (do inglês, Open Government Partnership - OGP), uma iniciativa internacional composta, atualmente, por 75 nações, que prevê projetos de apoio à elaboração de políticas institucionais também para a ciência aberta.

A amostra compôs-se de: 2 gestores que participam do Metabuscador de dados da Fapesp; 1 da UFPR, que já opera um repositório de dados de pesquisa; 1 da RNP; 1 do IBICT; e, para oferecer um olhar internacional sobre o novo papel do repositório de dados de pesquisa e implicações com direito autoral, 1 do repositório internacional de dados Dryad,

¹⁴⁵ Ver em: <<https://wiki.rnp.br/pages/viewpage.action?pageId=107315238>>

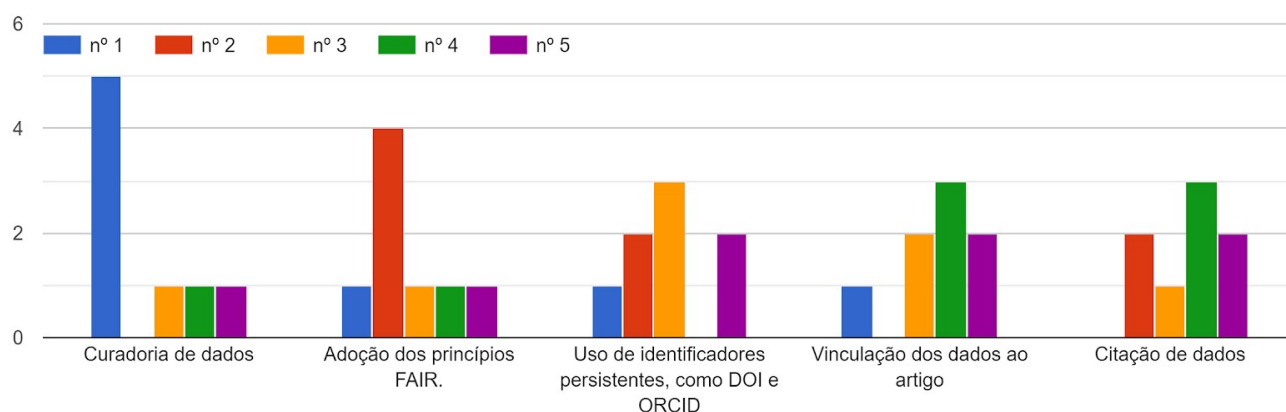
que aceitou o convite para responder a algumas perguntas, enviadas por *e-mail*. O roteiro de perguntas para gestores de repositórios podem ser acessadas no Apêndice D, e as perguntas dirigidas ao gestor do Dryad, no Apêndice G.

5.4 Abertura de dados de pesquisa e direito autoral: a perspectiva dos gestores de repositórios

Iniciamos a análise das respostas dos gestores de repositórios ao questionário eletrônico pelas perguntas que se inserem na **primeira categorização**, “**Abertura, acesso, compartilhamento e reúso de dados de pesquisa**”. Desta forma, a **pergunta 1** do questionário eletrônico convida o respondente a indicar uma ordem de importância dos serviços que um repositório de dados de pesquisa oferece.

Figura 11. Serviços de repositório: ordem de importância

1- O repositório de dados de pesquisa passou a ter um papel estratégico no sistema de comunicação científica. Indique a ordem de importância dos serviços que ele oferece...merando as opções de 1 a 5, sendo 1 a mais importante.



Mais da metade dos respondentes escolheu a ‘curadoria de dados’ como o serviço mais importante, entre as opções de múltipla escolha da primeira pergunta, seguida da ‘adoção dos princípios FAIR’, ‘uso de identificadores DOI e ORCID’, com as opções ‘vinculação dos dados ao artigo’ e ‘citação de dados’ empatadas (Figura 11). Ao indicarem a curadoria como serviço de maior importância de um repositório, pode significar que os respondentes encarem esse serviço também como o mais desafiador para eles. Segundo

Hardwicke et al. (2018), uma curadoria de dados praticada abaixo do ideal é um dos fatores que podem impedir a reprodutibilidade analítica e prejudicar a utilidade do compartilhamento de dados e a credibilidade das descobertas científicas. Hrynaszkiewicz (2019, *on line*, tradução nossa) concorda e acrescenta que editores podem ajudar a aumentar a qualidade dos dados e a reprodutibilidade da pesquisa se, entre outras áreas de atuação, “melhorar a infraestrutura de comunicação [...], promovendo o registro de estudos, parceria com repositórios de dados e fornecendo serviços que aperfeiçoem o compartilhamento e a curadoria de dados”.

A **pergunta 2** indaga sobre se o pesquisado é favorável à abertura e à disponibilização pública de conjuntos de dados de pesquisa subjacentes para compartilhamento e reuso, enquanto a **pergunta 3** questiona a exigência que algumas revistas fazem aos autores para que assinem uma declaração de concordância com o compartilhamento dos dados, chegando a considerar a não publicação em caso de discordância. A maioria das repostas à pergunta 2 converge para a primeira opção de resposta, que condiciona o “sim” à abertura e à disponibilização pública à situação de pesquisa financiada com recursos públicos. Na pergunta 3, também há convergência de opiniões positivas sobre a medida, que poderia conter questões de ética em pesquisa, já que o mérito da pesquisa, do artigo e dos dados estariam sendo subestimados em alguma medida.

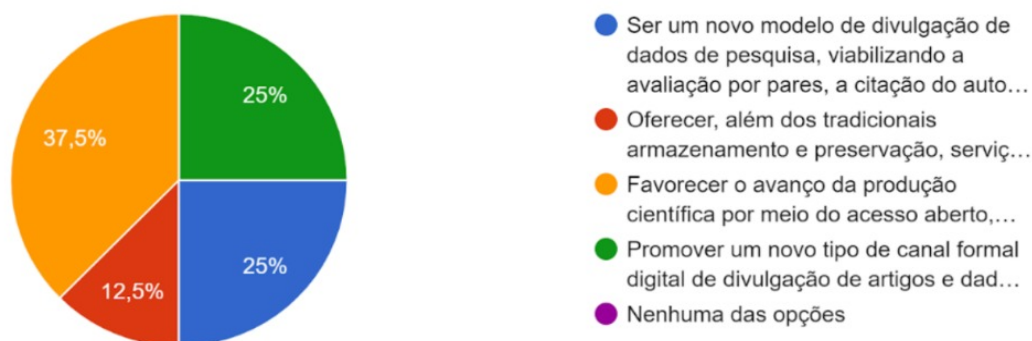
A Respondente G5, que integra o Metabuscador da Fapesp, faz a seguinte ponderação a favor da exigência de uma declaração assinada pelo autor, em concordância com o compartilhamento:

Porque a metodologia de coleta de dados e os dados coletados para o desenvolvimento da pesquisa é essencial para a atualização da temática em tempo posterior, possibilitando análise de séries temporais. Além disso, pesquisas efetuadas por pesquisadores de universidades (servidores públicos) deveriam seguir o preceito da publicidade e acesso aberto, quando possível (Respondente G5).

A resposta mais assinalada para a **pergunta 4** foi “Favorecer o avanço da produção científica por meio do acesso aberto, gratuito e sem barreiras a conjuntos de dados de pesquisa vinculados a artigos científicos”, que se coaduna com os princípios da ciência aberta.

Figura 12. O repositório na comunicação científica

4- Qual o papel do repositório de dados de pesquisa na comunicação científica?



Em segundo lugar, os respondentes se dividiram e marcaram duas respostas que obtiveram percentuais iguais (Figura 12): “Ser um novo modelo de divulgação de dados de pesquisa, viabilizando a avaliação por pares, a citação do autor original e a atribuição de identificadores, como DOI e ORCID [...]” e “Promover um novo tipo de canal formal digital de divulgação de artigos e dados de pesquisa, que represente uma mudança paradigmática na comunicação científica [...], sem vínculo com artigos e a formação de comunidades de pesquisa”. Essas duas respostas de múltipla escolha foram formuladas para se tentar aquilatar a percepção dos gestores sobre o estágio tecnológico dos repositórios de dados de pesquisa, que pode vir a pôr em questão o formato clássico (revista + artigo) de se comunicar e visualizar resultados de pesquisa, já que possibilitam iniciar a leitura de uma pesquisa pelo final dela, isto é, pelos dados de pesquisa.

Sobre como avalia o papel dos repositórios de dados no sistema de comunicação científica atual e na disseminação da abertura de dados de pesquisa, a Entrevistada G6 respondeu que

Está claro que os repositórios de dados são um componente essencial tanto da pesquisa científica quanto da comunicação científica. Existem milhares de repositórios que hospedam milhões de conjuntos de dados que, de outra forma, não estariam disponíveis. Também acredito que *os próprios dados são*

mais valiosos do que o artigo relacionado, pois representam as próprias descobertas. Não acredito que o público nem a maioria dos cientistas valorizem os repositórios da mesma forma que os periódicos/artigos, mas isso também ocorre porque o compartilhamento de dados de pesquisa é relativamente novo. Principalmente, porém, é porque não há um sistema de incentivos em vigor para creditar pesquisadores pela disponibilização de seus dados. Eu trabalho em uma iniciativa chamada Make Data Count que se concentra no desenvolvimento desse sistema, mas *vai demorar muitos anos* antes que os pesquisadores sejam promovidos ou creditados por sua publicação de dados de pesquisa (como são para publicar artigos) (Entrevistada G6, tradução nossa¹⁴⁶, grifos nossos).

O Entrevistado E2 também vê uma ascensão dos repositórios de dados na comunicação científica:

[...] se tudo o que preciso como pesquisador é que a informação ou o trabalho que eu fiz gerando esse dado seja divulgado rapidamente e de uma forma que todo mundo possa ter acesso, e eu posso fazer isso colocando num repositório de *preprint* e também num repositório de dados, por que é que eu vou mandar para uma revista esse trabalho? (Entrevistado E2).

O Entrevistado G3, do grupo Metabusador da Fapesp, acredita que o repositório é um serviço. Mas ressalta que os pesquisadores atuais não tiveram formação suficiente para lidar com o repositório como uma nova ferramenta digital de gestão da informação. E indica um caminho: a capacitação.

Eu vejo os repositórios como um serviço. Os pesquisadores estão cada vez mais demandados nesse novo contexto do sistema científico nacional, global. Eles têm precisado assimilar novas funções, têm que saber negociar, buscar, identificar fontes de recurso. Precisam saber lidar com pessoas, gerenciar projetos, times, equipes. E precisam, agora é uma necessidade muito evidente, ser capaz de gerenciar os dados que produzem, em atendimento às políticas

¹⁴⁶ No original: "It is clear that data repositories are an essential component of both scientific research and science communications. There are thousands of repositories hosting millions of datasets that would otherwise not be available. I also believe that the data themselves are more valuable than the related article, as they represent the findings themselves. I do not believe that the public nor most scientists value repositories like they do journals/articles but that is also because sharing research data is relatively new. Mostly though, it is because there is not an incentive system in place for crediting researchers for opening up their data. I work on an initiative called Make Data Count that is focused on developing this system, but it will take many years before researchers are promoted or credited for their research data publishing (like they are for publishing articles)."

que são estabelecidas sobretudo pelas agências de fomento e pelos periódicos científicos. E, nessa função, infelizmente, os pesquisadores da atualidade não receberam essa formação. Então, eu volto à noção do repositório como um serviço, porque me parece que uma instituição de pesquisa e desenvolvimento, ou de ensino e pesquisa, deve idealizar serviços voltados à gestão de dados científicos, dados de pesquisa. E deve se preocupar com a capacitação (Entrevistada G3).

Na **pergunta 5**, “Seu repositório tem uma diretriz específica para tratar da privacidade, dos dados pessoais, incluindo os dados sensíveis, da pessoa que participe de uma pesquisa em saúde pública?”, metade dos respondentes consideram que “Não, a privacidade dos dados é uma questão que deve ser tratada entre o pesquisador e o participante da pesquisa”. Todavia, os repositórios não devem se abster de divulgar orientações bem-informadas sobre a questão da privacidade dos dados. Segundo o NIH, o pesquisador deve ir além da anonimização de dados diretos (nome, cor, idade), conforme previsto em leis de proteção a dados pessoais e dados pessoais sensíveis, inclusive na lei brasileira, no art. 7, IV e no art. 11, c, respectivamente (BRASIL, 2018). Assim, a agência de saúde estadunidense recomenda que “os pesquisadores devem considerar a remoção de identificadores indiretos e outras informações que possam levar à ‘divulgação dedutiva’ das identidades dos participantes” (NIH, 2003b, *on line*, tradução nossa).

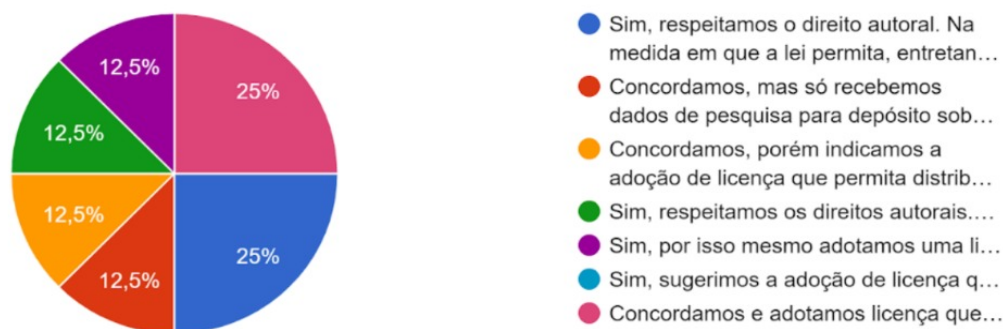
A partir da sexta pergunta, analisamos as respostas sobre como os gestores de repositórios lidam com as licenças e com os direitos autorais aplicáveis a dados de pesquisa, observando a **segunda categoria deste estudo “Direito autoral aplicável a dados de pesquisa”**.

Na **pergunta 6** – “Conjuntos de dados de pesquisa subjacentes a artigos são passíveis de proteção do direito autoral. Marque a opção de licenciamento Creative Commons que julga ideal para esse tipo de dados.[...]” –, como era de se esperar, um quarto dos gestores assinalou a licença CC BY. Porém, chama a atenção que também um quarto dos respondentes defendam a adoção da licença CC BY NC ND, considerada a mais restritiva das seis licenças Creative Commons (Figura 13). Quem a reusa não pode fazê-lo para fins comerciais e não pode compartilhar o trabalho na forma adaptada com outras pessoas. E a ciência aberta, onde fica? Esta licença não é recomendada para fins acadêmicos, inclusive pela própria Creative Commons¹⁴⁷.

¹⁴⁷ Ver: "Dados - Perguntas mais frequentes sobre dados e licenças CC" <<https://br.creativecommons.org/faq/>>

Figura 13. Qual é a melhor licença para reúso de dados?

6- Conjuntos de dados de pesquisa subjacentes a artigos são passíveis de proteção do direito autoral. Marque a opção de licenciamento Creative ...as exigem a citação do autor original dos dados.



Por outro lado, também causa certa estranheza que apenas um respondente assinale o instrumento CC0 Dedicção ao Domínio Público, que significa renúncia a todos os direitos patrimoniais (não aos direitos morais), na extensão que a lei permitir. Em consequência, surge a possibilidade de poder reusar dados de pesquisa sem precisar permissão especial de seus criadores originais¹⁴⁸ e a impossibilidade de reivindicação de propriedade sobre dados ou conjuntos de dados (CREATIVE COMMONS, 2018b).

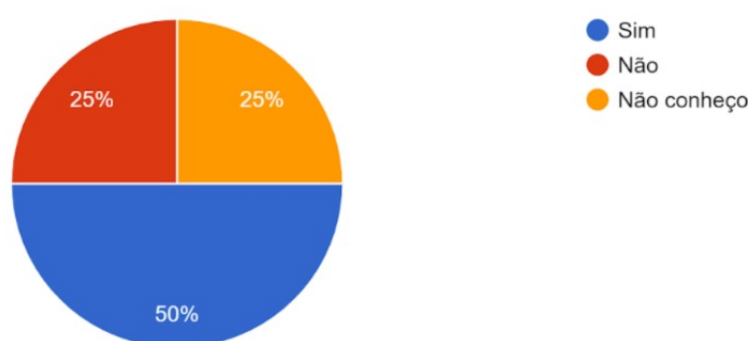
A **pergunta 7** refere-se especificamente ao instrumento CC0 Domínio Público – “Há uma tendência de os repositórios recomendarem aos pesquisadores a escolherem e os editores a apoiarem o instrumento CC0 [...], para licenciar o reúso de dados. Concorda [...]?”. A esta pergunta, houve quem respondesse que não conhece esse instrumento, mas a metade dos gestores (Figura 14) respondeu que concorda haver uma tendência entre editores e repositórios a recomendarem a adoção do CC0 Domínio Público (ver a respeito no capítulo 3). Na pergunta 6, contudo, apenas um respondente optou pelo CC0 como instrumento legal

¹⁴⁸ Ver em: BioMed Central <<https://www.biomedcentral.com/about/policies/open-data>>; e Dryad: <<https://blog.datadryad.org/2011/10/05/why-does-dryad-use-cc0/>>

para reuso de dados de pesquisa em seu repositório. Haveria algum vestígio de contradição ou desconhecimento do real alcance dessa ferramenta?

Figura 14. Domínio público para dados

7- Há uma tendência de os repositórios recomendarem aos pesquisadores a escolherem e os editores a apoiarem o instrumento CC0 Dedicã...eúso de dados. Concorde que há essa tendência?



Por que não considerar a CC0 se metade dos pesquisados percebe a tendência de recomendação em revistas e adoção em repositórios? Segundo o Escritório de Comunicação Acadêmica da Universidade da Califórnia, criado para apoiar seus pesquisadores a disseminarem sua produção científica, o que inclui orientação sobre direitos autorais e escolha de uma licença para dados de pesquisa,

O compartilhamento de dados com CC0 é cada vez mais encorajado [...] porque quanto mais fácil você tornar o reuso de seus dados para as pessoas – e mais claro você deixará que não planeja processar essas pessoas que os reutilizam – o mais provável é que seus dados sejam reutilizados [...] (FORTNEY, 2016, *on line*, tradução nossa).

No entanto, a Respondente G7 se contrapõe, indicando que, na sua experiência na gestão de um repositório que integra o Metabusador da Fapesp, verificou que depositantes opõem barreiras para depósito sob esse instrumento.

Apesar das recomendações de uso do CC0, nos estudos para implementação do RD institucional, identificamos em outros repositórios o uso de outras licenças que não a CC0, além de encontrar ainda barreiras por parte dos depositantes sobre a disponibilização em domínio público do conjunto de dados depositado (Respondente G7).

Porém, perguntado sobre qual é o conhecimento e a opinião que tem sobre a adoção de licenças públicas na disponibilização e reuso de dados de pesquisa, o Entrevistado G5 respondeu que,

[...] dentro da instituição [em que ele trabalha], para mim e para nossos bibliotecários, ainda é uma coisa nebulosa de entender o direito autoral, principalmente de arquivos digitais, de dados digitais, mas os nossos dados estão abertos para baixar, e o pesquisador tem que ter essa consciência (Entrevistado G5).

Na universidade em que o Entrevistado G5 trabalha, o autor escolhe que regime de direito autoral ele deseja para os dados que está disponibilizando no repositório. “Oferecemos um tutorial para ele explicando as etapas do autoarquivamento, que incluem as licenças, com explicações sobre elas, para ele escolher”, embora o Entrevistado G5 afirme que sempre direciona para a licença mais aberta.

A **pergunta 8**, “É favorável à padronização das normas internacionais de direito autoral para uso por todos os repositórios de dados de pesquisa?”, foi formulada com base em afirmações tais como “Os direitos autorais e o licenciamento de dados científicos, internacionalmente, são complexos e apresentam barreiras legais ao compartilhamento, integração e reutilização de dados [...] (HRYNASZKIEWICZ; COCKERILL, 2012; ASSANTE et. al., 2016).

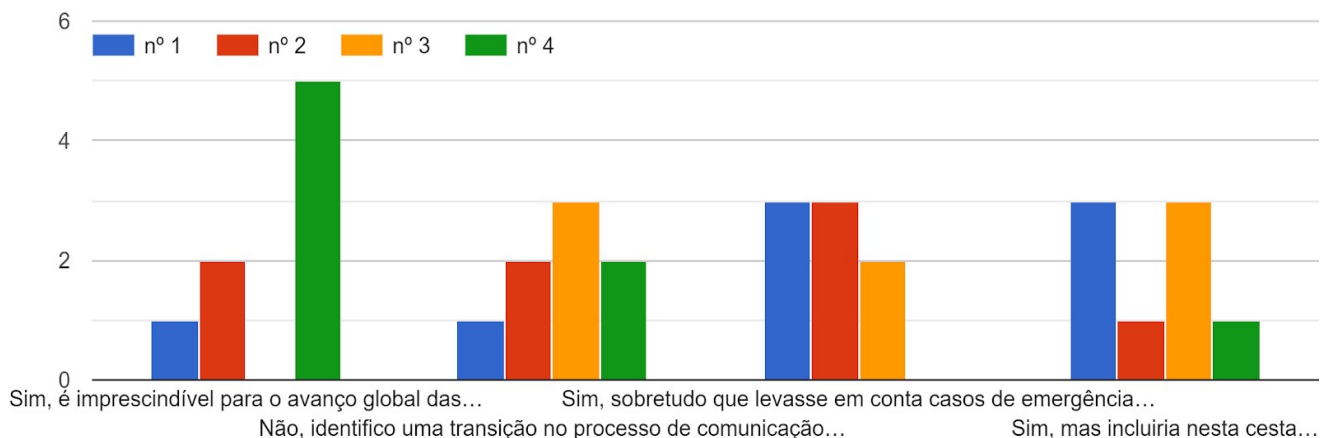
Mais da metade dos respondentes se mostra incrédula quanto a uma “harmonização” de normas para uso em repositórios de dados de pesquisa, em nível internacional. Em certa medida, essa posição poderia sugerir o desafio de se construir normas padrão para o relacionamento jurídico que pode advir da dinâmica da aplicação de direito autoral sobre bases de dados e conjunto de dados de pesquisa (OXENHAM, 2016; GILBERT, 2016).

A licença é um instrumento jurídico no qual o proprietário do conjunto de dados descreve a extensão da reutilização que ele permite (Ball, 2014). De acordo com Assante et al. (2016, *on line*), se a publicação acontecer em um repositório de dados, “existem dois tipos de licenças envolvidas: a acordada entre o repositório e o proprietário dos dados e aquela acordada entre o repositório e o consumidor de dados” (ver mais a respeito no capítulo 3).

Finalmente, a **pergunta 9**, “A fim de alcançar esse estágio de normas internacionais padrão para todos os repositórios, recomendaria que os legisladores criassem um direito novo, específico para a abertura de dados? Para indicar a sua ordem de prioridade, numere as opções de 1 a 4, sendo 1 a mais prioritária”. Mais da metade dos gestores respondeu ser prioritário que, “Sim, [um direito novo, específico para a abertura de dados] é imprescindível para o avanço global das pesquisas” (Figura 15).

Figura 15. Um direito novo para dados de pesquisa

9- A fim de alcançar esse estágio de normas internacionais padrão para todos os repositórios, recomendaria que os legisladores criassem um direito...e as opções de 1 a 4, sendo 1 a mais prioritária.



Em razão dessa resposta positiva, não há como não mencionar o direito *sui generis* de base de dados – um direito novo, criado e recomendado pela Comissão Europeia a seus estados-membros. Ele protege o conteúdo de uma base de dados e proíbe a extração substancial desse conteúdo, caso represente prejuízo ao investimento feito pelo dono da base. Note-se que é um acréscimo protetivo ao já existente sistema de proteção por direito autoral,

introduzido na lei europeia em legislação promulgada em 1996, que pode criar conflitos potenciais, por exemplo, com a mineração de texto e dados (COMISSÃO EUROPEIA, 1996; HUGENHOLTZ, 2019). Por que não criar um direito novo que favoreça a ciência aberta?

O material pesquisado foi além da temática dos direitos autorais, ao inserir a questão num contexto mais amplo de abertura e disponibilização de dados subjacentes a artigos. Por isso, comporta vários olhares. A título de discussão de alguns resultados da categoria “Direito autoral aplicável a dados de pesquisa”, ressaltamos que os pesquisados se mostram incrédulos quanto a uma “harmonização” de normas de diferentes legislações nacionais para uso no processo de publicação, abertura e compartilhamento de dados. Mas a complexidade de lidar com o caráter internacional dos direitos autorais pode representar uma barreira, sobretudo para a e-Science.

A indiferença como resposta a uma pergunta que propõe a escolha de um regime fechado, que transfere direitos exclusivos e em definitivo de conjuntos de dados para terceiros, poderia revelar que o regime que perdurou durante o século XX, no qual o autor/detentor cede totalmente seus direitos patrimoniais ao editor em troca da publicação do artigo, ainda está bem vivo. Porém, vai de encontro ao regime proposto ainda nos primórdios do movimento pelo acesso aberto à informação científica (BUDAPESTE, 2002), no qual o direito autoral limita-se apenas à preservação da paternidade da obra (artigo ou conjuntos de dados de pesquisa, no caso), por meio da devida citação ao autor original.

Afinal, a resposta positiva à criação de um direito novo e específico para a abertura e compartilhamento de dados de pesquisa, tal como o direito *sui generis* criado pela União Europeia especificamente para a proteção econômica de bases dados, sugere o estabelecimento de uma discussão sobre a criação de um direito a favor das práticas de ciência aberta, sem amarras legais. Um direito novo a ser aplicado exclusivamente no âmbito da produção científica, incluindo a abertura e compartilhamento de dados de pesquisa. Diante dos tempos em que vivemos, é urgente que o assunto entre nas pautas de governos, instituições de pesquisa, agências de financiamento à pesquisa e organizações político-econômicas intergovernamentais.

6 CONCLUSÃO

Neste capítulo final, derivamos algumas conclusões do estudo efetuado em torno da questão que o norteou: “Como as revistas científicas, particularmente as da área da saúde, estão lidando com a questão da abertura de dados de pesquisa subjacentes aos artigos que publica em face do direito autoral?”. Aqui, reiteramos que o debate sobre proteção de direitos de dados e conjuntos de dados de pesquisa é amplo, mas o foco deste trabalho foram as repercussões que os direitos autorais podem provocar sobre dados de pesquisa subjacentes a artigos. Cumpre dizer ainda que a abordagem do tema tem a ciência aberta em perspectiva, considerando-se a abertura de dados de pesquisa como um desdobramento do movimento de acesso aberto à informação científica.

Assim, como primeira conclusão, pudemos observar a provável emergência de um novo regime de direito autoral, ligado ao processo de publicação, abertura e compartilhamento de dados subjacentes a um artigo, disponibilizados em repositórios de dados. Durante esse processo podem ocorrer reivindicações sobre direito autoral de conjuntos de dados. Isto porque se o termo “dados” ainda não adquiriu uma definição de consenso, é menos trivial ainda estabelecer a propriedade de dados ou a existência de direitos autorais para conjuntos de dados que, para obterem tal proteção, têm que demonstrar um mínimo de originalidade, para não serem definidos como fatos, que não recebem proteção.

Estas e outras incertezas aliam-se à complexidade de lidar com diferentes legislações nacionais sobre direitos autorais e à possibilidade de “empilhamento” de licenças, que é a combinação de vários conjuntos de dados, com vários autores que podem usar várias licenças diferentes.

Tais fatos fizeram com que uma nova geração de repositórios de dados de pesquisa, dentre eles Dryad, Figshare e Zenodo, bem como as plataformas digitais de publicação de artigos e dados subjacentes, entre elas a Wellcome Open Research, objetos de pesquisa relatada no capítulo 4, passassem a recomendar aos autores ou detentores de direitos autorais que dessem preferência à adoção de um instrumento jurídico do tipo CC0 Dedicção ao Domínio Público, da Creative Commons, para regular o reuso desses dados. A editora BioMed Central, que faz parte da Spring Nature e tem cerca de 300 periódicos científicos em acesso aberto, também aderiu à CC0 para dados.

Conforme analisado no capítulo 3, por meio desse instrumento, o autor ou o detentor renuncia a todos os seus direitos patrimoniais relativos à obra (no caso, conjunto de dados de pesquisa), possibilitando que um pesquisador reutilize os dados sem pedir permissão. Na opinião de gestores de repositórios, essa decisão maximiza o reuso dos dados. Portanto, pode vir a se configurar em um novo tipo de regime de direito autoral, mais direcionado a dados, diferente do modelo clássico utilizado por periódicos científicos em que o autor faz cessão total de direitos patrimoniais em troca de publicação de seu trabalho.

Criadas por agências que financiam pesquisa científica, as plataformas constituem um modelo que pode vir a afetar o formato tradicional de publicação de artigos e dados de pesquisa em revistas. A plataforma Wellcome Open Research publica artigos avaliados por pares, sob a licença CC BY, e dados de pesquisa, sob o instrumento CC0 Dedicção ao Domínio Público, exclusivamente de pesquisadores financiados pela Wellcome Trust. Introduziu práticas incomuns na comunicação científica clássica, tais como a publicação em uma semana de artigos com a descrição de dados, a avaliação por pares *após* a publicação e a atualização constante dos artigos. É um canal próprio de comunicação de ciência da agência e causou polêmica com revistas de prestígio por impor publicação imediata de artigos e dados subjacentes sem possibilidade de embargos, conforme abordado no final do capítulo 4. A Fundação Bill & Melinda Gates seguiu o mesmo caminho e criou a Gates Open Research, nos mesmos moldes.

Concluimos em segundo lugar que, para efetivar todo o potencial de compartilhamento de dados, ainda é preciso superar barreiras legais estruturais. Entendemos, por exemplo, que seria pertinente a criação pelos governos de um direito novo, que pavimente o caminho da pesquisa científica e acadêmica, em vez do quadro atual, que restringe a sua evolução, ao pretender regular a disseminação do conhecimento. No limite, a ciência requer a criação de outras listas ou a atualização das já existentes listas de limitações e exceções ao direito autoral, com um olhar específico para o fazer científico.

Nessa mesma linha conclusiva, também seria bem-vinda, como alternativa à criação de um direito novo, a conversão de resultados de pesquisa científica – artigos e dados – a um *status* semelhante ao domínio público, que pode ser criada formalmente por leis públicas por meio de legislação nacional que exclua certas categorias de dados e informações científicos da proteção de direitos autorais ou proíba a imposição de restrições ao seu uso (UHLIR, 2015).

Sabemos, todavia, que não é tarefa fácil, pois existem altos interesses em jogo. Senão, vejamos: eventos como o Acordo TRIPS, de 1994, e o Tratado de Direitos Autorais da OMPI, o WCT, de 1996, impõem balizamentos sobre limitações e exceções aos direitos exclusivos, apesar de reconhecidas nas leis nacionais de direitos autorais, o que põe em risco a realização de pesquisas científicas. Tais eventos também podem ser vistos como uma resposta ao advento das tecnologias digitais e seu impacto na transmissão *on line* de obras protegidas por direitos autorais. Essa circunstância levou à adoção de regimes regulatórios, fundados no WCT, que podem impedir o uso da maioria das limitações e exceções existentes, e impedir que terceiros acessem fatos e ideias improprietários. À medida que essas tendências de privatização vão se infiltrando cada vez mais no campo da pesquisa científica, o acesso aos insumos do conhecimento básico torna-se cada vez mais complicado e potencialmente difícil ou caro de se obter (REICHMAN; OKEDIJI, 2012)

Isto posto, em terceiro lugar, concluímos que essa nova fronteira da ciência aberta apresenta um estágio avançado em termos de sua proposição como teoria, havendo o endosso de *players* de peso da comunicação científica, mas prescinde, na prática, de um padrão universal de requisitos para sua operacionalização em larga escala, de um sistema de recompensa que valorize os dados de pesquisa, além de um quadro normativo legal que favoreça o fazer científico, sobretudo a mineração de texto e dados, como já ocorre no Reino Unido e na União Europeia.

Assim, ao lado da questão do direito autoral, verificamos na literatura consultada e nas pesquisas desenvolvidas uma movimentação no sentido da abertura e compartilhamento de dados de pesquisa subjacentes a artigos científicos, mais intensa nos países centrais, mais lenta em países em desenvolvimento, particularmente os da América Latina.

Pudemos verificar na pesquisa observacional relatada no capítulo 4 que a revista científica e o repositório de dados de países centrais, mesmo em nova dinâmica de publicação de artigos vinculados a dados de pesquisa subjacentes, ainda não conseguiram lidar de modo consistente com a nova fronteira de abertura e compartilhamento desses dados. Quanto à origem da inconsistência, destacamos problemas como a falta de:

- compreensão de editores dos problemas e motivações dos pesquisadores em abrir seus dados e compartilhar;
- diretrizes padronizadas nas políticas de abertura e compartilhamento de dados subjacentes, que envolvem revistas, repositórios e agências;

- requisitos padronizados para compartilhamento de dados, alguns caros à comunidade científica, como a citação e a avaliação por pares; e

- padronização de orientações sobre o direito autoral e o licenciamento de dados.

Constatamos ainda que há um consenso sobre a existência de uma ampla variação de orientações, como as identificadas na área biomédica, sobre a incorporação de um plano de gestão de dados ou de declaração de concordância com compartilhamento, cuja negativa do autor pode valer a desaprovação da publicação.

Este cenário se repete nas disciplinas de biologia, química, matemática e física, mas não na astronomia e genômica, já habituadas ao compartilhamento de dados. Em pesquisas sucessivas sobre compartilhamento de dados, a especialista Carol Tenopir concluiu que é claramente necessária: a assistência a pesquisadores por meio de gerentes de dados; repositórios de dados disponíveis para armazenamento de longo e curto prazos; e programas educacionais para conscientização e para ajudar a gerar boas práticas de dados. Somam-se a isso uma insatisfatória disponibilidade de dados, além do predomínio de estruturas que recompensam a publicação de artigos, mas não de dados, e sistemas de incentivo que persistem em subestimar o trabalho crítico do gerenciamento de dados e pesquisadores que se propõem a compartilhar dados.

A par disso, na área da saúde permanece o constante dilema de como compartilhar dados pessoais oriundos de pesquisa em saúde, em um nível em que se consiga uma contribuição efetiva aos estudos para melhoria da saúde das pessoas sem macular o seu direito à privacidade, o que inclui os dados pessoais e os dados sensíveis.

Chegamos à quarta conclusão com base na pesquisa empírica realizada com editores científicos do campo da saúde do Brasil e, particularmente, de países em desenvolvimento da América Latina, reportada no capítulo 5. Esses editores atuam em revistas da Fiocruz, do *hot site* SciELO Saúde Pública e da UFRJ. Por meio da análise das respostas ao questionário eletrônico e às entrevistas e de um exame observacional nos respectivos *sites* de suas revistas, infere-se que o tema da abertura e compartilhamento de dados de pesquisa subjacentes a artigos e as questões de direito autoral daí decorrentes ainda não são uma prioridade.

Além de identificadas uma postura apenas reativa às iniciativas da SciELO Brasil em direção à ciência aberta e a falta de discussão interna sobre o assunto, o exame observacional realizado nos *sites* dos periódicos pesquisados identificou que apenas em 3, da Fiocruz, e em

4, da SciELO Saúde Pública, há registro de alguma orientação aos autores sobre publicação de dados de pesquisa subjacentes a artigos. Não foi encontrada qualquer orientação objetiva sobre como lidar com direitos autorais aplicáveis a esses dados ou informações detalhadas sobre escolha de licenças públicas para reuso de dados.

Cabe reiterar, no entanto, que a Fiocruz está atualmente em processo adiantado de elaboração de uma política de abertura de dados de pesquisa, assim como a SciELO Brasil está em plena execução da implantação de um alinhamento de suas coleções de periódicos à ciência aberta, que abarca a abertura e compartilhamento de dados de pesquisa. Ambas devem estar atentas às implicações do direito autoral sobre dados de pesquisa.

A quinta conclusão origina-se da pesquisa empírica efetuada com gestores de repositórios institucionais e de dados de pesquisa brasileiros, por meio de questionário eletrônico e entrevistas, referida no capítulo 5. Não obstante o reconhecimento do valor do Metabusador de dados da Fapesp, que reúne as universidades públicas paulistas, e do esforço pioneiro do repositório “Base de Dados Científicos”, da UFPR, a análise das respostas evidenciou que a situação ainda é a de uma implementação incipiente, cujo reflexo se vê no reduzido acervo de coleções de conjuntos de dados em ambos os repositórios. Percebeu-se, nestes dois casos, a necessidade de os profissionais envolvidos nesses projetos se apropriarem mais da questão dos direitos autorais sobre dados de pesquisa.

Uma das idealizadoras do projeto declarou em entrevista que houve uma execução impecável da estrutura tecnológica, a cargo da USP. No entanto, a questão dos direitos autorais e de licenciamento de dados não foi delineada ao tempo da construção do Metabusador, e está surgindo agora, após quase um ano de existência.

Como conclusão final, afirmamos que o direito autoral, naturalizado e ao mesmo tempo pouco pesquisado como elemento formador da área científica, revela-se provavelmente inadequado para regular relações jurídicas nesse campo. Talvez por um vício de origem. O direito autoral protege direitos econômicos de um criador de obra intelectual, como um livro, uma música, enquanto no universo da ciência os pesquisadores-autores, em geral, não têm interesse em ganhar dinheiro com suas produções intelectuais – artigos e conjuntos de dados. Seus ganhos advêm de seu trabalho em instituições de pesquisa e universidades. Querem apenas ser reconhecidos em sua comunidade – que é a via para acesso a recursos para a pesquisa e para a progressão na carreira –, ter a sua criação citada e atribuída adequadamente e contribuir com o avanço da ciência.

Destacamos, por fim, que o tema tem potencial para a realização de outras pesquisas, por ser novo e em razão do persistente descompasso regulatório entre o direito autoral e a ciência aberta ainda não ter sido suficientemente explorado. Como sugestão de temas para futuras pesquisas, indicamos estudos sobre o impacto do direito autoral nas plataformas de publicação criadas por agências de financiamento à pesquisa, formato que repercute diretamente no modelo de negócios de publicações científicas de alto fator de impacto de conglomerados editoriais comerciais; e sobre a questão das restrições impostas por regulamento da OMPI às exceções e limitações ao direito autoral, legalmente reconhecidas em legislações nacionais, inclusive mediante medidas tecnológicas de proteção. Tais restrições configurariam uma privatização do acesso à informação?

REFERÊNCIAS

- ABDO, Alexandre H. Ciência Aberta, da ciência para todos à ciência com todos. **Liinc em Revista**, Rio de Janeiro, v. 10, n. 2, p. 460-471, 2014. Disponível em: <<http://revista.ibict.br/liinc/article/view/3592>. Acesso em: 2 maio 2020.
- ABDO, Alexandre H. **Não à restrição comercial**. 2017. Disponível em: <http://www.em-rede.com/site/ciencia-aberta/n%C3%A3o-%C3%A0-restri%C3%A7%C3%A3o-comercial>. Acesso em 13 ago. 2020.
- ABDO, Alexandre H. O uso da cláusula CC Não-Comercial prejudica a Ciência Aberta. **Blog cienciaaberta.net**. 2018. Disponível em: <<https://www.cienciaaberta.net/o-uso-da-clausula-cc-nao-comercial-prejudica-a-ciencia-aberta/>>. Acesso em: 10 fev. 2020.
- ACESSO ABERTO USP. **Entenda o que é acesso aberto**. 2020. Disponível em: <https://www.aguia.usp.br/apoio-pesquisador/acesso-aberto-usp/entenda-o-que-e-acesso-aberto/>. Acesso em: 21 out. 2020.
- ALBAGLI, Sarita. Ciência aberta em questão. *In*: ALBAGLI, Sarita; MACIEL, Maria Lucia; HANNUD, Alexandre A. (orgs.). **Ciência aberta, questões abertas**. Brasília: IBICT; Rio de Janeiro: UNIRIO, 2015. p. 9-25. Disponível em: <http://livroaberto.ibict.br/bitstream/1/1060/1/Ciencia%20aberta_questoes%20abertas_PORTUGUES_DIGITAL%20%285%29.pdf>. Acesso em: 20 ago. 2019.
- ALBAGLI, Sarita; MACIEL, Maria Lucia. Informação, conhecimento e democracia no capitalismo cognitivo. *In*: COCCO, Giuseppe; ALBAGLI, Sarita (orgs.). **Revolução 2.0 e a crise do capitalismo**. Rio de Janeiro: Garamond, 2012. p. 40-58.
- ALBAGLI, Sarita; CLINIO, Anne; Sabryna, RAYCHTOCK. Ciência Aberta: correntes interpretativas e tipos de ação. **Liinc em Revista**, Rio de Janeiro, v.10, n. 2, p. 434-450, nov. 2014.
- ALEIXANDRE-BENAVENT, Rafael et al. Policies regarding public availability of published research data in pediatrics journals. **Scientometrics**, v. 118, n. 2, p. 439-451, feb. 2019. Disponível em: <<https://springer.proxy.ufrj.br/article/10.1007/s11192-018-2978-1#Tab2>>. Acesso em: 20 mar. 2020.
- AL-KHOURI, Ali M. Data ownership: who owns ‘my data’?. **International Journal of Management & Information Technology**, v. 2, n. 1, 2012. Disponível em: https://www.researchgate.net/publication/331085049_Data_Ownership_Who_Owns_'My_Data'. Acesso em: 22 ago. 2019.
- ANDS (Australian National Data Service). **Research Data Rights Management**. 2018. Disponível em: https://www.ands.org.au/__data/assets/pdf_file/0010/1498744/Research-Data-Rights-Management-Guide.pdf. Acesso em: 16 jan. 2020.
- APPEL, André L.; MACIEL, Maria Lucia; ALBAGLI, Sarita. A e-Science e as novas práticas de produção colaborativa de conhecimento científico. **Revista Internacional de Ciencia Y Sociedad**, v. 3, p. 41-52, 2016. Disponível em: <https://ridi.ibict.br/bitstream/123456789/928/1/470-1274-1-PB.pdf>. Acesso em: 15 ago. 2020.
- ARXIV.ORG. Annual Update, January 2020. 2020. Disponível em: https://arxiv.org/about/reports/2020_update. Acesso em: 13 maio 2020.

- ASSANTE, Massimiliano et al. Science 2.0 Repositories: Time for a Change in Scholarly Communication. **D-Lib Magazine**, v. 21, n.1/2, 2015. Disponível em: <http://www.dlib.org/dlib/january15/assante/01assante.html>. Acesso em: 28 jan. 2020.
- ASSANTE, Massimiliano et al. Are scientific data repositories coping with research data publishing?. **Data Science Journal**, 15, p. 6, 2016. Disponível em: <https://datascience.codata.org/articles/10.5334/dsj-2016-006/>. Acesso em: 28 ago. 2019.
- BALL, A. **How to License Research Data**. 2014. Edinburgh: Digital Curation Centre. Disponível em: <https://www.dcc.ac.uk/guidance/how-guides/license-research-data>. Acesso em: 30 jun. 2020.
- BANZI et al. Evaluation of repositories for sharing individual-participant data from clinical studies. **Trials**, v. 20, article number 169, 2019. Disponível em: <https://trialsjournal.biomedcentral.com/articles/10.1186/s13063-019-3253-3>. Acesso em: 3 fev. 2020.
- BERLIN DECLARATION on Open Access to Knowledge in Science & Humanities. Max-Planck-Gesellschaft. 2003. Disponível em: <https://openaccess.mpg.de/Berliner-Erklaerung>. Acesso em: 30 jun. 2020.
- BETHESDA STATEMENT on Open Access Publishing. Digital Access to Scholarship at Harvard. June, 2003. Disponível em: <https://dash.harvard.edu/handle/1/4725199>>. Acesso em: 30 jun. 2020.
- BIETZ, Matthew J. et al. Opportunities and challenges in the use of personal health data for health research. **Journal of the American Medical Informatics Association**, v. 23, n. e1, p. e42-e48, 2016. Disponível em: <https://academic.oup.com/jamia/article/23/e1/e42/2379811>. Acesso em: 9 ago. 2019.
- BILL & MELINDA GATES FOUNDATION. Global Access. 2003. Disponível em: <http://globalaccess.gatesfoundation.org/>>. Acesso em: 25 out. 2020.
- BILL & MELINDA GATES FOUNDATION. Open Access Policy. 2017a. Disponível em: <https://www.gatesfoundation.org/How-We-Work/General-Information/Open-Access-Policy>>. Acesso em: 3 fev. 2020.
- BILL & MELINDA GATES FOUNDATION. Frequently Asked Questions. 2017b. Disponível em: <https://www.gatesfoundation.org/How-We-Work/General-Information/Open-Access-Policy/Page-2>>. Acesso em: 9 fev. 2020.
- BITTAR, Carlos A. **Direito de autor**. 4. ed. Rio de Janeiro: Forense Universitária, 2008.
- BJÖRK, Bo-Christer; SOLOMON, David. **Open access publishing: developing an effective market for open access article processing charges**. 71 p. 2014. Disponível em: https://web.archive.org/web/20140602195247/http://www.wellcome.ac.uk/stellent/groups/corporatesite/%40policy_communications/documents/web_document/wtp055910.pdf. Acesso em: 30 ago. 2019.
- BJÖRK, Bo-Christer; KORKEAMAKI, Timo. Adoption of the open access business model in scientific journal publishing: A cross-disciplinary study. **arXiv: 2005.01008v1**. 2020. Disponível em: http://search.arxiv.org:8081/paper.jsp?r=2005.01008&qid=1590621285542multi_nCnN_1628488273&qs=open+access. Acesso em: 28 maio 2020.
- BLOOM, Theodora; GANLEY, Emma; WINKER, Margaret. Editorial. Data Access for the Open Access Literature: PLOS's Data Policy. **PLoS Medicine**, v. 11, n. 2, p. e1001607, 2014. Disponível em: <https://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.1001607>. Acesso em: 19 mar. 2020.

BMJ. Author Licence BMJ Publishing Group Limited (BMJ): BMJ Journals – BMJ wholly owned or co-owned journals. 2018. Disponível em: https://s16086.pcdn.co/wp-content/uploads/2018/11/BMJ_Journals_Combined_Author_Licence_2018.pdf. Acesso em: 10 abr. 2020.

BOHLE, Shannon. What is e-science and how should it be managed?. **Nature.com**. 2013. Disponível em: http://www.scilogs.com/scientific_and_medical_libraries/what-is-e-science-and-how-should-it-be-managed. Acesso em: 10 jun. 2019.

BORGMAN, Christine. The conundrum of sharing research data. **Journal of the American Society for Information Science and Technology** (*on line*), v. 63, n. 6. p. 1059-1078, 2012. Disponível em: <http://dx.doi.org/10.2139/ssrn.1869155>. Acesso em: 30 jan. 2020.

BORGMAN, Christine. **Big data, little data, no data**: scholarship in the networked world. Cambridge: MIT Press, 2015.

BORGMAN, Christine L. **Text data mining from the author's perspective**: whose text, whose mining, and to whose benefit?. text data mining national forum statement, final. March 12, 2018. Page 1 of 6. Disponível em: <https://publish.illinois.edu/limitedaccess-tdm/>. Acesso em: 10 maio 2020.

BRANCO, Sérgio; BRITTO, Walter. **O que é Creative Commons?**. Rio de Janeiro: Editora FGV, 2013. 176 p. Disponível em: <https://bibliotecadigital.fgv.br/dspace/bitstream/handle/10438/11461/O%20que%20%C3%A9%20Creative%20Commons.pdf>. Acesso em: 26 jul. 2020.

BRANDÃO, Adeilton. Memórias do Instituto Oswaldo Cruz (*site*). This year, One Hundred and Ten. And counting!. 2019. Disponível em: <https://memorias.ioc.fiocruz.br/110yearstest>. Acesso em: 19 mar. 2020.

BRASIL. Constituição da República Federativa do Brasil (1988). **Diário Oficial da União**, n. 191-A, Brasília, DF, 5 out. 1988, p. 1-32. Disponível em: http://www.planalto.gov.br/ccivil_03/constituicao/constituicao.htm. Acesso em: 20 jul. 2020.

BRASIL. Lei nº 9.610, de 19 de fevereiro de 1998. Lei dos Direitos Autorais. Altera, atualiza e consolida a legislação sobre direitos autorais e dá outras providências. **Diário Oficial da União**, 20 fev. 1998. Disponível em: http://www.planalto.gov.br/ccivil_03/leis/19610.htm. Acesso em: 20 jul. 2020.

BRASIL. Lei nº 13.709, de 14 de agosto de 2018. Lei Geral de Proteção Dados Pessoais (LGPD). **Diário Oficial da União**, 15 ago. 2018. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/L13709.htm. Acesso em: 6 fev. 2020.

BUDAPESTE Open Access Initiative (BOAI). 2002. Disponível em: <https://www.budapestopenaccessinitiative.org/>. Acesso em: 28 ago. 2019.

CABRERA BLÁZQUEZ, F.J. et al. **Exceptions and limitations to copyright**. Estrasburgo: IRIS Plus; European Audiovisual Observatory, 2017. Disponível em: <https://rm.coe.int/iris-plus-2017-1-exceptions-and-limitations-to-copyright/168078348b>. Acesso em: 23 jul. 2020.

CANDELA, Leonardo et al. Data Journals: A Survey. **Journal of the Association for Information Science and Technology**, v. 66, n. 1, p. 1747-1762, 2015. Disponível em: https://figshare.com/articles/Data_Journals_A_Survey/1358039. Acesso em: 24 set. 2019.

CARALEE, Adams. **SPARC Innovator**: The Bill & Melinda Gates Foundation. 2019. Disponível em: <https://sparcopen.org/our-work/innovator/gates-foundation>. Acesso em: 6 abr. 2020.

- CARBON, Seth et al. An analysis and metric of reusable data licensing practices for biomedical resources. **PLoS One**, v. 14, n. 3, e0213090, 2019. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6436688/>. Acesso em: 25 mar. 2020.
- CARR, David; LITTLER, Katherine. Sharing Research Data to Improve Public Health. **Journal of Empirical Research on Human Research Ethics**, v. 10, n. 3, p. 314-316, 2015. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4547198/>. Acesso em: 22 set. 2019.
- CARROLL, Micahel W. Why Full Open Access Matters. **PLoS Biology**, v. 9, n. 11, e1001210, 2011. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3226455/>. Acesso em: 25 maio 2020.
- CARROLL, Michael W. Sharing Research Data and Intellectual Property Law: A Primer. **PLoS Biology**, v. 13, n. 8, e1002235, 2015. Disponível em: <https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.1002235>. Acesso em: 17 fev. 2020.
- CDL (California Digital Library). University California. **New Dryad is Here**. 2019. Disponível em: <https://cdlib.org/cdlinfo/2019/09/24/new-dryad-is-here>. Acesso em: 20 jan 2020.
- CDL (California Digital Library). University California. **Dryad. 2020**. Disponível em: <https://cdlib.org/services/uc3/dryad/>. Acesso em: 20 jan 2020.
- CHAN, Leslie; OKUNE, Angela; SAMBULI, Nanjira. O que é ciência aberta e colaborativa, e que papéis ela poderia desempenhar no desenvolvimento?. In: ALBAGLI, Sarita; MACIEL, Maria Lucia; HANNUD, Alexandre A. (orgs.). **Ciência aberta, questões abertas**. Brasília: IBICT; Rio de Janeiro: UNIRIO, 2015. p. 91-120. Disponível em: http://livroaberto.ibict.br/bitstream/1/1060/1/Ciencia%20aberta_questoes%20abertas_PORTUGUES_DIGITAL%20%285%29.pdf. Acesso em: 20 mar. 2020.
- CHAN, Leslie. What role for open and collaborative science in development?. University World News, 2014. Disponível em: <https://www.universityworldnews.com/post.php?story=20140905132026576>. Acesso em: 22 mar. 2020.
- COALITION S. Plan S. 2018. Disponível em: <https://www.coalition-s.org/about/>. Acesso em: 10 abr. 2020.
- COALITION S. Rationale for the Revisions Made to the Plan S Principles and Implementation Guidance. 2019. Disponível em: <https://www.coalition-s.org/rationale-for-the-revisions/>. Acesso em: 10 jun. 2020.
- COAR (Confederation of Open Access Repositories). **Next Generation Repositories**. Behaviours and Technical Recommendations of the COAR Next Generation Repositories Working Group. 2017. Disponível em: <https://www.coar-repositories.org/files/NGR-Final-Formatted-Report-cc.pdf>. Acesso em: 30 jul. 2020.
- COMISSÃO EUROPEIA. Parlamento Europeu e Conselho. **Diretiva 96/9/CE, de 11 de março de 1996, relativa à proteção jurídica das bases de dados**. 1996. Disponível em: <https://eur-lex.europa.eu/legal-content/PT/TXT/HTML/?uri=CELEX:31996L0009&from=EN>. Acesso em: 24 ago. 2020.
- COMISSÃO EUROPEIA. **What is horizon 2020?**. 2014. Disponível em: <https://ec.europa.eu/programmes/horizon2020/en/what-horizon-2020>. Acesso em: 7 ago. 2020.

COMISSÃO EUROPEIA. Directorate-General for Research and Innovation. **Open innovation, open science, open to the world: a vision for Europe**. 2016. Disponível em: <http://doi:10.2777/552370>. Acesso em: 3 jun. 2020.

COMISSÃO EUROPEIA. **Guidelines to the rules on open access to scientific publications and open access to research data in horizon 2020**. 2017. Disponível em: https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf. Acesso em: 27 jul. 2020.

COMISSÃO EUROPEIA. Recomendação (UE) de 25 de abril de 2018. Sobre o acesso à informação científica e a sua preservação. **Jornal Oficial da União Europeia**, Bruxelas, L 134/12, 2018. Disponível: <https://eur-lex.europa.eu/legal-content/PT/TXT/HTML/?uri=CELEX:32018H0790&from=EN>. Acesso em: 10 maio 2020.

CONVENÇÃO DE BERNA PARA A PROTEÇÃO DE OBRAS LITERÁRIAS E ARTÍSTICAS. **World Intellectual Property Organization - WIPO**. 1886. Disponível em: https://www.wipo.int/treaties/en/text.jsp?file_id=283698. Acesso em: 15 jul. 2020

COSTA, Michelli P.; LEITE, Fernando C.L. Open access in the world and Latin America: A review since the Budapest Open Access Initiative. **Transinformação**, v. 28, n.1, p. 33-46, 2016. Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0103-37862016000100033&lng=en&nrm=iso. Acesso em: 18 maio 2020.

COSTA, Michelli; LEITE, Fernando C.L. Princípios e recomendações basilares para a comunicação dos dados de pesquisa. **Em Questão**, v. 23, n. 1, 2017. Disponível em: <http://www.redalyc.org/jatsRepo/4656/465649115005/html/index.html#gt9>. Acesso em: 15 set. 2019.

COUSIJN, Helena et al. A data citation roadmap for scientific publishers. **Scientific Data**, 5, article number 180259, 2018. Disponível em: <https://www.nature.com/articles/sdata2018259>. Acesso em: 19 mar. 2020.

CREATIVE COMMONS. PLOS e figshare tornam a publicação científica aberta mais aberta. 2013. Disponível em: <https://creativecommons.org/2013/01/31/plos-and-figshare-make-open-science-publishing-more-open/>. Acesso em: 19 jan. 2010.

CREATIVE COMMONS. **Sobre as licenças**. 2018a. Disponível em: https://creativecommons.org/licenses/?lang=pt_BR. Acesso em: 28 fev. 2019.

CREATIVE COMMONS. CC0 1.0 Universal (CC0 1.0). **Dedicação ao Domínio Público**: sem direito de autor nem direitos conexos. 2018b. Disponível em: https://creativecommons.org/publicdomain/zero/1.0/deed.pt_BR. Acesso em: 13 abr. 2019.

CREATIVE COMMONS. **Integração do Creative Commons, de A a Z**: Seletor de licenças. 2018c. Disponível em: <https://creativecommons.org/platform/toolkit/#license-chooser>. Acesso em: 23 jan. 2010.

CREATIVE COMMONS. **Atribuição 4.0 Internacional (CC BY 4.0)**. 2020. Disponível em: <https://creativecommons.org/licenses/by/4.0/deed.pt>. Acesso em: 23 jan. 2020.

DAVID, Paul A. Can 'Open Science' be Protected from the Evolving Regime of IPR Protections?. **Journal of Institutional and Theoretical Economics (JITE)**, n. 129, p. 1-28, 2004. Disponível em: <http://www-siepr.stanford.edu/workp/swp03011.pdf>. Acesso em: 15 abr. 2020.

DAVID, Paul A. The historical origins of 'open science': an essay on patronage, reputation and common agency contracting in the scientific revolution. **Capitalism and Society**, v. 3, n.2, article 5,

2008. Disponível em: <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2209188>. Acesso em: 3 maio 2020.

DECLARAÇÃO DO MÉXICO. **Declaração conjunta LATINDEX-REDALYC-CLACSO-IBICT recomenda uso não-comercial da produção acadêmica e científica.** 2017. Disponível em: <https://www.aguia.usp.br/noticias/declaracao-conjunta-latindex-redalyc-clacso-ibict-recomenda-uso-nao-comercial-da-producao-academica-e-cientifica/>. Acesso em: 20 out. 2020.

DIGITAL SCIENCE. **Digital Science Report: The State of Open Data 2018.** A selection of analyses and articles about open data, curated by Figshare. 2018. Disponível em: <https://www.digital-science.com/resources/portfolio-reports/state-open-data-2018/>. Acesso em: 11 set. 2020.

DOLDIRINA, C.; Eisenstadt A.; Onsrud, H.; Uhler, P. Legal Approaches for Open Access to Research Data. *LawArXiv Papers*. 2018. Disponível em: <https://osf.io/preprints/lawarxiv/n7gfa>. Acesso em: 15 maio 2020.

DONEDA, Danilo. **Da privacidade à proteção de dados pessoais.** Rio de Janeiro: Revista dos Tribunais, 2019.

DORCH, S.B.F; DRACHEN, T.M; ELLEGAARD, O. The data sharing advantage in Astrophysics. **Proceedings of the International Astronomical Union, General Assembly**, v. 11 (A29A), p. 172-175, 2015. Disponível em: https://www.researchgate.net/publication/283659056_The_data_sharing_advantage_in_astrophysics. Acesso em: 18 maio 2020.

DRYAD. **Why does Dryad use CC0?** *Blog*. 2011. Disponível em: <https://blog.datadryad.org/2011/10/05/why-does-dryad-use-cc0/>. Acesso em: 22 mar. 2020.

DRYAD. Terms of service. 2020. Disponível em: <<https://datadryad.org/stash/terms#submitter-obligations>>. Acesso em: 22 mar. 2020.

DUDZIAK, Elisabeth. Dados de Pesquisa agora devem ser armazenados e citados. 2016. Disponível em: <<http://www.sibi.usp.br/?p=6189>>. Acesso em: 25 fev. 2020.

EISEN, M. B.; BROWN, P.O., VARMUS, H.E. PLoS Medicine: A Medical Journal for the Internet Age. **PLoS Medicine**, v. 1, n. 1, p. e31, 2004. Disponível em: <https://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.0010031>. Acesso em: 19 mar. 2020.

ELLIOT, Roger. Who owns scientific data? The impact of intellectual property rights on the scientific publication chain. **Learned Publishing**, v. 18, p. 91–94, April 2005. Disponível em: <https://www.semanticscholar.org/paper/Who-owns-scientific-data-The-impact-of-intellectual-Elliott/b0a333aeac361ac6036e8fc71c03ac4bba4bc668?p2df>. Acesso em: 10 maio 2020.

EUROPEAN IP HELPDESK. European Commission's Executive Agency for Small and Medium-sized Enterprises. **Protection of a database.** 2020. Disponível em: <http://www.iprhelpdesk.eu/news/protection-database-our-company-has-developed-digital-database-cultural-heritage-how-can-it-be>. Acesso em: 13 ago. 2020.

EXAME. Fiocruz publica a melhor revista científica da América Latina. 2016. Disponível em: <https://exame.abril.com.br/ciencia/fiocruz-publica-a-melhor-revista-cientifica-da-america-latina/>. Acesso em: 19 mar. 2020.

- FAPESP. Fapesp lança Rede de Repositórios de dados científicos do estado de São Paulo. Agência Fapesp. 2019. Disponível em: <https://agencia.fapesp.br/fapesp-lanca-rede-de-repositorios-de-dados-cientificos-do-estado-de-sao-paulo/32251/>. Acesso em: 13 set. 2020.
- FELL, Michael J. The economic impacts of open science: a rapid evidence assessment. **Publications**, v. 7, n. 3, 2019. Disponível em: <https://www.mdpi.com/2304-6775/7/3/46/htm>. Acesso em: 3 maio 2020.
- FIGSHARE. **About**. Our path so far. 2020a. Disponível em: <https://figshare.com/about>. Acesso em: 25 jan. 2019.
- FIGSHARE. **How to choose the most appropriate license**. 2020b. Disponível em: <https://knowledge.figshare.com/articles/item/what-is-the-most-appropriate-licence-for-my-data>. Acesso em: 23 jan. 2020.
- FIGSHARE. **Copyright and Licence Policy**. 2020c. Disponível em: <https://knowledge.figshare.com/articles/item/copyright-and-licence-policy>. Acesso em: 23 jan. 2020.
- FIGSHARE. **Terms and conditions**. 2020d. Disponível em: <https://figshare.com/terms>. Acesso em: 23 jan. 2020.
- FORTNEY, Katie. **CC BY and data: not always a good fit**. Office of Scholarly Communication. University of California. 2016. Disponível em: <https://osc.universityofcalifornia.edu/2016/09/cc-by-and-data-not-always-a-good-fit/>. Acesso em: 15 set. 2020.
- FOUCAULT, Michel. **O que é um autor?**. 1969. Disponível em: https://edisciplinas.usp.br/pluginfile.php/4956027/mod_resource/content/1/foucault%2C%20michel%20-%20o%20que%20%C3%A9%20um%20autor.pdf. Acesso em: 7 fev. 2020.
- GANDELMAN, Henrique. **De Gutenberg à Internet**. 5 ed. Rio de Janeiro: Record, 2007.
- GILBERT, Natasha. Legal tussle delays launch of huge toxicity database. **Nature**, 19365, 2016. Disponível em: <https://www.nature.com/articles/nature.2016.19365>. Acesso em 14 set. 2020.
- GOUVEIA, Fábio C. Almetria: métricas de produção científica para além das citações. **Liinc em Revista**, Rio de Janeiro, v. 9, n. 1, p. 214-227, 2013. Disponível em: <http://revista.ibict.br/liinc/article/view/3434/3004>. Acesso em: 09 set. 2019.
- GRAND, Ann et al. Mapping the hinterland: Data issues in open science. **Public Understanding of Science**, v. 25, n. 1, p. 88-103, 2016. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4702280/>. Acesso em: 2 jun. 2020.
- GRANT, Rebecca; HRYNASZKIEWICZ, Iain. The impact on authors and editors of introducing Data Availability Statements at Nature journals. **bioRxiv**: The preprints server for Biology, Feb. 2018. Disponível em: <https://www.biorxiv.org/content/10.1101/264929v1.full>. Acesso em: 30 mar. 2020.
- GREENE, Casey S. et al. Celebrating parasites. **Nature Genetic**, v. 49, n. 4, p. 483-484, 2017. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5710834/>. Acesso em: 2 set. 2020.
- GUANAES, Paulo C.V.; GUIMARÃES, Maria Cristina S. Modelos de gestão de revistas científicas: uma discussão necessária. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 17, n. 1, p. 56-73, 2012. Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1413-99362012000100004&lng=pt&nrm=iso. Acessos em: 23 mar. 2020.

GUANAES, Paulo C.V. et al. **Marcos legais nacionais em face da abertura de dados para pesquisa em saúde: dados pessoais, sensíveis ou sigilosos e propriedade intelectual.** Rio de Janeiro: Fiocruz, 2018. 123 p. Disponível em: <<https://www.arca.fiocruz.br/handle/icict/28838>>. Acesso em: 31 out. 2019.

GUANAES, Paulo C.V.; ALBAGLI, Sarita. Compartilhamento e propriedade intelectual de dados de pesquisa: o caso do periódico *thebmj*. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 20., 2019, Florianópolis. **Anais eletrônicos...** Florianópolis: Portal de Conferências da UFSC, 2019. Disponível em: <https://conferencias.ufsc.br/index.php/enancib/2019/paper/view/523/727>. Acesso em: 02 nov. 2019.

GUANAES, Paulo C.V. Quais direitos podem ser alegados sobre dados de pesquisa?. In: SHINTAKU, Milton; SALES, Luana F.; COSTA, Michelli (orgs.). **Tópicos sobre dados abertos para editores científicos.** Botucatu: ABEC, 2020. 240 p. Disponível em: https://www.abecbrasil.org.br/arquivos/Topicos_dados_abertos_editores_cientificos.pdf. Acesso em: 29 jul. 2020. p. 79-85.

GUÉDON, Jean-Claude. The “green” and “gold” roads to open access: the case for mixing and matching. **Serials Review**, v. 30, n. 4, 2004. Disponível em: <http://eprints.rclis.org/5860/1/science.pdf>. Acesso em: 20 maio 2020.

GUIBAULT, Lucie; WIEBE, Andreas (eds.). **Safe to be open: study on the protection of research data and recommendations for access and usage.** Göttingen: Universidade de Göttingen, 2013. Disponível em: <https://univerlag.uni-goettingen.de/bitstream/handle/3/isbn-978-3-86395-147-4/legalstudy.pdf?sequence=1&>. Acesso em: 20 jul. 2020.

GUIBAULT, Lucie; MARGONI, Thomas. Legal aspects of open access to publicly funded research. In: OECD. **Enquiries into Intellectual Property’s Economic Impact**, 2015. cap. VII, p. 373-414. Disponível em: <http://www.oecd.org/sti/ieconomy/KBC2-IP.Final.pdf>. Acesso em: 20 maio 2020.

HAHNEL, Mark. Solving the research data management problem for educational research institutions. Figshare. **Blog**. 2013. Disponível em: https://figshare.com/blog/Solving_the_research_data_management_problem_for_educational_research_institutions./96. Acesso em: 20 jan. 2020.

HAHNEL, Mark. **Global funders who require data archiving as a condition of grants.** Figshare. 2015. Disponível em: https://figshare.com/articles/Global_funders_who_require_data_archiving_as_a_condition_of_grants/1281141. Acesso em 12 abr. 2020.

HAJDUK, Gabriela K. et al. It is not enough that we require data to be shared; we have to make sharing easy, feasible and accessible too!. **BMJ Global Health**, v. 4, n. 4, p. e001550, 2019. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6666804/>. Acesso em: 25 maio 2020.

HARDWICKE, Tom *et al.* Data availability, reusability, and analytic reproducibility: evaluating the impact of a mandatory open data policy at the journal *Cognition*. **Royal Society Open Science**, 5: 1804482018. 2018. Disponível em: <https://royalsocietypublishing.org/doi/10.1098/rsos.180448>. Acesso em: 10 set. 2020.

HARNAD, Stevan. **Electronic scholarly publication: quo vadis?.** Southampton: University of Southampton Institutional Repository, 1995. Disponível em: <https://eprints.soton.ac.uk/253356/>. Acesso em: 10 mar. 2020.

HARNAD, Stevan et al. The access/impact problem and the green and gold roads to open access: an update. **Serials Review**, v. 34, n. 1, p. 36–40, 2004. Disponível em: <https://eprints.soton.ac.uk/265852/1/serev-revised.pdf>. Acesso em: 19 maio 2020.

HARNAD, Stevan. **Finch Group reviews progress in implementing open access transition amid ongoing criticisms**. The London School of Economics and political science. Department of Government. The Impact of Social Sciences. *Blog*. Nov. 18th, 2013. Disponível em: <https://eprints.soton.ac.uk/371583/1/index.html>. Acesso em: 17 maio 2020.

HRYNASZKIEWICZ, Iain; ALTMAN, Douglas G. Towards agreement on best practice for publishing raw clinical trial data. **Trials**, v. 10, art. number 17, 2009. Disponível em: <https://trialsjournal.biomedcentral.com/articles/10.1186/1745-6215-10-17#citeas>.

HRYNASZKIEWICZ, Iain et al. Preparing raw clinical data for publication: guidance for journal editors, authors, and peer reviewers. **BMJ**, 340, c181, 2010. Disponível em: <https://www.bmj.com/content/340/bmj.c181>. Acesso em: 28 ago. 2019.

HRYNASZKIEWICZ, Iain; COCKERILL, Matthew J. Open by default: a proposed copyright license and waiver agreement for open access research and data in peer-reviewed journals. **BMC Research Notes**, v. 5, n. 1, 2012. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3465200/>. Acesso em: 19 set. 2019.

HRYNASZKIEWICZ, Iain; BUSCH, Stefan; COCKERILL, Matthew J. Licensing the future: report on BioMed Central's public consultation on open data in peer-reviewed journals. **BMC Research Notes**, v. 6, n. 1, 2013. Disponível em: <https://bmresnotes.biomedcentral.com/articles/10.1186/1756-0500-6-318#Tab1>. Acesso em: 20 mar. 2020.

HRYNASZKIEWICZ, Iain. Standardising and Harmonising Research Data Policy in Scholarly Publishing. **International Digital Curation Conference**, v. 12, n. 1, p. 65-71, 2017. Disponível em: <http://www.ijdc.net/article/view/12.1.65>. Acesso em: 20 mar. 2020.

HRYNASZKIEWICZ, Iain. Publishers' responsibilities in promoting data quality and reproducibility. In: BARRET, James E. (editor). **Handbook of Experimental Pharmacology**. Springer, 2019. Disponível em: https://link.springer.com/chapter/10.1007%2F164_2019_290. Acesso em 4 fev. 2020. p.1-30. Acesso em: 24 set. 2019.

HRYNASZKIEWICZ, Iain *et al.* Developing a Research Data Policy Framework for All Journals and Publishers. **Data Science Journal**, v. 19, n. 1, p. 5, 2020. Disponível em: <https://datascience.codata.org/articles/10.5334/dsj-2020-005/>. Acesso em: 29 mar. 2020.

HUGENHOLTZ, Bernt. The New Copyright Directive: Text and Data Mining (Articles 3 and 4). Kluwer Copyright Blog. 2019. Disponível em: http://copyrightblog.kluweriplaw.com/2019/07/24/the-new-copyright-directive-text-and-data-mining-articles-3-and-4/?doing_wp_cron=1598062832.4732871055603027343750. Acesso em: 12 ago. 2020.

ICMJE. **Recommendations**. Clinical Trials. Sharing data. 2018. Disponível em: <http://www.icmje.org/recommendations/>. Acesso em: 27 ago. 2019.

INTELLECTUAL PROPERTY OFFICE. **Exceptions to copyright**: Research. 2014. Disponível em: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/375954/Research.pdf. Acesso em: 19 jul. 2020.

IOANNIDIS, John P. A.; BELBASIS, Lazaros; EVANGELOU, Evangelos. Fifty-year fate and impact of general medical journals. **PLoS ONE**, v. 5, n. 9, e12531, 2010. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2931710/>. Acesso em: 10 mar. 2020.

JORGE, Vanessa A. **Abertura e compartilhamento de dados para pesquisa nas situações de emergência em saúde pública: o caso do vírus Zika**. 2018. 263 f. Tese (Doutorado em Ciência da Informação) – Universidade Federal do Rio de Janeiro; Instituto Brasileiro de Informação em Ciência e Tecnologia, Rio de Janeiro, 2018.

JORGE, Vanessa A.; ALBAGLI, Sarita. Compartilhamento de dados de pesquisa em saúde: iniciativas do National Institutes of Health (NIH). **Reciis**, Rio de Janeiro, v. 12, n. 4, out./dez. 2018. Disponível em: <https://www.reciis.icict.fiocruz.br/index.php/reciis/article/view/1499/2237>. Acesso em: 5 jan. 2001.

KINDLING, Maxi *et al.* The landscape of research data repositories in 2015: a re3data analysis. **D-Lib Magazine**, v. 23, n. 3/4, March/April 2017. Disponível em: <http://www.dlib.org/dlib/march17/kindling/03kindling.html>. Acesso em: 2 set. 2019.

KNORR-CETINA, Karin. A comunicação na ciência. In: **A ciência tal qual se faz**. GIL, Fernando (ed.) Lisboa: Edições João de Sá da Costa, 1999. p. 375-393. Disponível em: http://kops.uni-konstanz.de/bitstream/handle/123456789/11568/Knorr_spanisch2.pdf?sequence=1&isAllowed=y. Acesso em: 25 jul. 2020.

KOBAYASHI, Shinji; KANE, Thomas B.; PATON, Chris. The privacy and security implications of open data in healthcare: a contribution from the IMIA Open Source Working Group. 2018. **Yearbook of Medical Informatics**, v. 27, n.1, p. 41-47, 2018. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6115211/>. Acesso em: 29 jan. 2020.

KOSTOVA, Patty *et al.* Who owns the data?: open data for healthcare. **Frontiers Public Health**, n. 17, 2016. Disponível em: <https://www.frontiersin.org/articles/10.3389/fpubh.2016.00007/full>. Acesso em: 19 set. 2019.

KRATZ, John E.; STRASSER, Carly. Data publication consensus and controversies. **F1000Research (on line)**, v. 3, n. 94, 2014. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4097345/>. Acesso em: 23 jan.2020.

KRATZ, John E.; STRASSER, Carly. Researcher perspectives on publication and peer review of data. **PLoS ONE**, v. 10, n. 2, e0117619, 2015. Disponível em: <https://www.ncbi.nlm.nih.gov/pubmed/25706992>. Acesso em: 1 set. 2019.

LAAKSO, Mikael; BJÖRK, Bo-Christer. Anatomy of open access publishing: a study of longitudinal development and internal structure. **BMC Medicine**, v. 10, article number 124, 2012. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3478161/>. Acesso em: 10 mar. 2020.

LAGOZE, Carl; VAN DE SOMPEL, Herbert. The Open Archives Initiative Protocol for Metadata Harvesting. 2015. Disponível em: <http://www.openarchives.org/OAI/2.0/openarchivesprotocol.htm>. Acesso em: 16 maio 2020.

LARIVIÈRE, Vincent; HAUSTEIN, Stefanie; MONGEON, Philippe. The Oligopoly of Academic Publishers in the Digital Era. **PLoS ONE**, v. 10, n. 6, e0127502, 2015. Disponível em: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0127502#pone.0127502.ref017>. Acesso em: 11 mar. 2020.

LEONELLI, Sabina. Why the current insistence on open access to scientific data?: big data, knowledge production, and the political economy of contemporary biology. **Bulletin of Science, Technology & Society**, v. 33, n. 1-2, p. 6-11, 2013. Disponível em: <https://journals.sagepub.com/doi/full/10.1177/0270467613496768>. Acesso em 2 jun. 2020.

LEVIN, Nadine et. al. How Do Scientists Define Openness? Exploring the Relationship Between Open Science Policies and Research Practice. **Bulletin of Science Technology & Society**, v. 36, n. 2, p. 128-141, 2016. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5066505/>. Acesso em: 3 jun. 2020.

LIN, Jenifer; STRASSER. Recommendations for the Role of Publishers in Access to Data. **PLoS Biology**, v. 12, n. 10, e1001975, 2014. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4211645>. Acesso em: 5 fev. 2020.

LONGO, Dan L.; DRAZEN, Jeffrey M. Sharig data. Editorial. **The New England Journal of Medicine**, v. 374, p. 276-277, 2016. Disponível em: <https://www.nejm.org/doi/10.1056/NEJMe1516564>. Acesso em 2 set. 2020.

LOUREIRO, Isabel. A menos eurocêntrica de todos: Rosa Luxemburgo e a acumulação primitiva permanente. In: SCHÜTRUMPF, Jörn. **Rosa Luxemburgo e o preço da liberdade**. São Paulo: Fundação Rosa Luxemburgo, 2015. p. 97-107. Disponível em: https://rosaluxspba.org/wp-content/uploads/2015/05/Rosa-Luxemburgo_-versaoweb.pdf. Acesso em: 10 ago. 2020.

MALIN, Bradley et al. Between Access and Privacy: Challenges in Sharing Health Data. **Yearb of Medical Informatics**, v. 27, n. 1, p. 55-59, 2018. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6115244>. Acesso em 11 abr. 2020.

MARQUES, Fernanda; REIS, Juliana. Portal de Periódicos Fiocruz. Ensaios clínicos: registro e publicação. 2016. Disponível em: <https://periodicos.fiocruz.br/pt-br/content/ensaios-cl%C3%ADnicos-registro-e-publica%C3%A7%C3%A3o>. Acesso em: 12 mar 2020.

MATTHEWS, Kirstin R.W.; Ho, Vivian. The grand impact of the Gates Foundation. Sixty billion dollars and one famous person can affect the spending and research focus of public agencies. **Embo Reports**, v. 9, n. 5, p. 409-412, 2008. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2373372/>. Acesso em: 5 abr. 2020.

MAYERNIK, Matthew S.; PHILIPS, Jennifer; NIENHOUSE, Eric. Linking Publications and Data: Challenges, Trends, and Opportunities. **D-Lib Magazine**, v. 22, n. 5/6, 2016. Disponível em: <http://www.dlib.org/dlib/may16/mayernik/05mayernik.html>. Acesso em: 20 set. 2019.

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

MEADOWS, Jack. Os periódicos científicos e a transição do meio impresso para o eletrônico. Tradução de Sely M. S. Costa. **Revista de Biblioteconomia de Brasília**, v. 25, n. 1, p. 5-14, 2001. Disponível em: https://www.brapci.inf.br/_repositorio/2010/10/pdf_29f176742d_0012269.pdf. Acesso em: 10 mar. 2020.

MERRIL, Scott. Techcrunch. Science Data Sharing Site figshare Relaunches, Adds Features. 2012. Disponível em: <https://techcrunch.com/2012/01/17/science-data-sharing-site-figshare-relaunches-adds-features/>. Acesso em: 20 jan. 2020.

MERTON, Robert K. **The normative structure of science**. 1942. Disponível em: <https://www.panarchy.org/merton/science.html>. Acesso em: 10 fev. 2020.

MONS, Barend et al. Cloudy, increasingly FAIR; revisiting the FAIR Data guiding principles for the European Open Science Cloud. **Information Services & Use**, v. 37, n. 1, p. 49-56, 2017. Disponível em: <https://content.iospress.com/articles/information-services-and-use/isu824>. Acesso em: 15 out. 2020

MURRAY-RUST, P. Open data in science. **Nature Preceedings**. 2008. Disponível em: <<https://www.nature.com/articles/npre.2008.1526.1>>. Acesso em: 10 fev. 2020.

MURRAY-RUST, Peter; NEYLON, Cameron; POLLOCK, Rufus; WILBANKS, John. **Panton principles**: principles for open data in science. feb. 2010. Disponível em: <https://pantonprinciples.org/>. Acesso em: 04 jan. 2021.

MURRAY-RUST, P.; MOLLOY, J.C.; CABELL, D. Open content mining. *In*: MOORE, S. (ed.). **Issues in Open Research Data**. Londres: Ubiquity Press, 2014. p. 11-30. Disponível em: <https://www.ubiquitypress.com/site/books/e/10.5334/ban/>. Acesso em: 03 ago. 2010.

NATIONAL ACADEMIES OF SCIENCE (EUA). **Open science by design**: realizing a vision for 21st century research. cap. 2, 3. Washington (DC): National Academies Press (EUA), 2018. Disponível em: <https://www.ncbi.nlm.nih.gov/books/NBK525413/>. Acesso em: 6 fev. 2020.

NATIONAL RESEARCH COUNCIL (US) COMMITTEE ON RESPONSIBILITIES of Authorship in the Biological Sciences. **Sharing publication-related data and materials**: responsibilities of authorship in the life sciences. Washington (DC): National Academies Press (US), 2003. cap. 3. Sharing Data and Software. Disponível em: <https://www.ncbi.nlm.nih.gov/books/NBK97155/>. Acesso em: 30 jan. 2020.

NATIONAL SCIENCE BOARD. **Long-lived digital data collections**: enabling research and education in the 21st century. Alexandria: National Science Foundation: 2005.

NATURE. Editorial. Community cleverness required. **Nature**, n. 455, 2008. Disponível em: <https://www.nature.com/articles/455001a#citeas>. Acesso em: 4 fev. 2020.

NATURE. Announcement: Where are the data? Editorial. **Nature**, v. 537, n. 7619, 2016a. Disponível em: <https://www.nature.com/news/announcement-where-are-the-data1.20541>. Acesso em: 23 jan. 2020.

NATURE. Reportagem de Declan Butler. Wellcome Trust launches open-access publishing venture. **Nature**, 2016b. Disponível em: <https://www.nature.com/news/wellcome-trust-launches-open-access-publishing-venture-1.20220>. Acesso em: 6 abr. 2020.

NATURE. Reportagem de Richard Van Noorden. Gates Foundation research can't be published in top journals. 2017b. **Nature**, 541, 270, 2017a. Disponível em: <https://www.nature.com/news/gates-foundation-research-can-t-be-published-in-top-journals-1.21299>. Acesso em: 5 abr. 2020.

NATURE. Reportagem de Richard Van Noorden. Science journals permit open-access publishing for Gates Foundation scholars. **Nature**, Feb. 14, 2017b. Disponível em: <<https://www.nature.com/news/science-journals-permit-open-access-publishing-for-gates-foundation-scholars-1.21486>>. Acesso em: 5 abr. 2020.

NATURE. Reportagem de Declan Butler. Gates Foundation announces open-access publishing venture. **Nature**, n. 543, 599, 2017c. Disponível em: <https://www.nature.com/news/gates-foundation-announces-open-access-publishing-venture-1.21700>. Acesso em: 6 abr. 2020.

NATURE. Reportagem de Richard Van Noorden. Science journals end open-access trial with Gates Foundation. **Nature**, n. 559, p. 311-312, 2018. Disponível em: <https://www.nature.com/articles/d41586-018-05729-2>. Acesso em: 5 abr. 2020.

NATURE. Reportagem de Priyanka Pulla. The plan to mine the world's research papers. **Nature**, n. 571, p. 316-318, 2019. Disponível em: <https://www.nature.com/articles/d41586-019-02142-1>. Acesso em: 11 maio 2020.

NATURE. Reportagem de Holly Else. Nature journals reveal terms of landmark open-access option. **Nature**, 588, 19-20, nov. 2020. Disponível em: <https://www.nature.com/articles/d41586-020-03324-y>. Acesso em: 02/01/2021.

NATURE MEDICINE. Raising the bar for clinical research. Editorial. **Nature Medicine**, v. 25, 1461, 2019. Disponível em: <https://www.nature.com/articles/s41591-019-0619-8>. Acesso em: 30 mar. 2020.

NATURE NANOTECHNOLOGIE. The importance and challenges of data sharing. Editorial. **Nature Nanotechnology**, v. 15, 83, 2020. Disponível em: <https://www.nature.com/articles/s41565-020-0646-0>. Acesso em: 30 mar. 2020.

NAUDET, Florian et al. Data sharing and reanalysis of randomized controlled trials in leading biomedical journals with a full data sharing policy: survey of studies published in The BMJ and PLOS Medicine. **BMJ**, 360: K400, 2018. Disponível em: <https://www.bmj.com/content/360/bmj.k400>. Acesso em: 20 set. 2019.

NEWMAN, Paul. Editorial. Data papers: peer reviewed publication of high quality data sets. **The International Journal of Robotics Research**, v. 28, n. 5, p. 587, 2009. Disponível em: <https://journals.sagepub.com/doi/pdf/10.1177/0278364909104283>. Acesso em: 30 out. 2019.

NIH (National Institutes of Health). **Final NIH statement on sharing research data**. 2003a. Disponível em: <https://grants.nih.gov/grants/guide/notice-files/NOT-OD-03-032.html>. Acesso em: 4 fev. 2020.

NIH (National Institutes of Health). **NIH Data Sharing Policy and Implementation Guidance**. 2003b. Disponível em: https://grants.nih.gov/grants/policy/data_sharing/data_sharing_guidance.htm#data. Acesso em: 4 abr. 2020.

NIH (National Institutes of Health). **NIH Grants Policy Statement. Rights in Data (Publication and Copyrighting)**. 2018. Disponível em: https://grants.nih.gov/grants/policy/nihgps_2013/nihgps_ch8.htm#_Toc271264947. Acesso em: 5 abr. 2020.

NIH (National Institutes of Health). **NIH Data Sharing Policy**. 2019. Disponível em: https://grants.nih.gov/grants/policy/data_sharing. Acesso em: 4 abr. 2020.

NIH (National Institutes of Health). Trans-NIH BioMedical Informatics Coordinating Committee (BMIC). **Data Sharing Resources**. 2020. Disponível em: https://www.nlm.nih.gov/NIHbmic/nih_data_sharing_repositories.html. Acesso em: 4 fev. 2020.

OCDE (Organização para a Cooperação e Desenvolvimento). **Principles and guidelines for access to research data from public funding**. 2007. Disponível em: <http://www.oecd.org/sti/inno/38500813.pdf>. Acesso em: 28 ago. 2019.

OCDE (Organização para a Cooperação e Desenvolvimento). OECD Science, Technology and Industry Policy Papers, n. 25. **Making Open Science a Reality**. OECD Publishing: Paris, 2015a. Disponível em: https://www.oecd-ilibrary.org/science-and-technology/making-open-science-a-reality_5jrs2f963zs1-en. Acesso em: 19 set. 2019.

OCDE (Organização para a Cooperação e Desenvolvimento). Legal aspects of open access to publicly funded research. *In*: OCDE. **Enquiries Into Intellectual Property's Economic Impact**. 2015b. cap. VII. p. 373-401. Disponível em: <https://www.oecd.org/sti/ieconomy/Chapter7-KBC2-IP.pdf>. Acesso em: 02 set. 2019.

OCDE (Organização para a Cooperação e Desenvolvimento). OECDLibrary. **Business models for sustainable research data repositories**. 2017. Disponível em: https://www.oecd-ilibrary.org/science-and-technology/business-models-for-sustainable-research-data-repositories_302b12bb-en. Acesso em: 29 abr. 2020.

ODI (Open Data Institute). **Reuser's Guide to Open Data Licensing**. 2019. Disponível em: <https://theodi.org/article/reusers-guide-to-open-data-licensing/>. Acesso em: 19 abr. 2020.

ODLYZKO, Andrew M. Tragic Loss or Good Riddance? The impending demise of traditional scholarly journals. **Jornal of Universal Computer Science**, Oct 1994. Disponível em: http://www.jucs.org/jucs_0_0/tragic_loss_or_good/odlyzko_a.html. Acesso em: 11 mar. 2020.

ODLYZKO, Andrew M. Economic costs of toll access. Open Access: Key Strategic, Technical and Economic Aspects. **Online Information Review**, v. 31, n. 4, p. 39-43, 2006. Disponível em: http://www.dtc.umn.edu/~Odlyzko/doc/OpenAccess_book-odlyzko-chapter.pdf. Acesso em: 10 maio 2020.

OFFICE OF SCIENCE AND TECHNOLOGIE POLICY. **Expanding public access to the results of federally funded research. the White House**. 2013. Disponível em: <https://obamawhitehouse.archives.gov/blog/2013/02/22/expanding-public-access-results-federally-funded-research>. Acesso em: 17 maio 2020.

OKF (Open Knowledge Foundation). **Open definition**. 2020. Disponível em: <http://opendefinition.org/od/2.1/en/>. Acesso em: 10 ago. 2020.

OHMANN, Christian *et al.* Sharing and reuse of individual participant data from clinical trials: principles and recommendations. **BMJ Open**, 7, e018647, 2017. Disponível em: <https://bmjopen.bmj.com/content/7/12/e018647.citation-tools>. Acesso em: 10 abr. 2020.

OMPI (Organização Mundial da Propriedade Intelectual/World Intellectual Property Organization - WIPO). **WIPO Copyright Treaty (WCT)**. 1996. Disponível em: https://www.wipo.int/edocs/pubdocs/en/wipo_pub_226.pdf. Acesso em: 13 ago. 2020.

OMPI (Organização Mundial da Propriedade Intelectual/World Intellectual Property Organization - WIPO). **What is copyright?**. 2020. Disponível em: <https://www.wipo.int/copyright/en/>. Acesso em: 29 jan. 2019.

OMS (Organização Mundial da Saúde). **Global Strategy and Plan of Action on Public Health, Innovation and Intellectual Property**. Genebra: WHO, 2011. Disponível em: https://www.who.int/phi/implementation/phi_globstat_action/en/. Acesso em: 10 mar. 2020.

OMS (Organização Mundial da Saúde). **International Standards for Clinical Trial Registries**. 2018. Disponível em: https://www.who.int/health-topics/clinical-trials/#tab=tab_1. Acesso em: 12 mar. 2020.

OMS (Organização Mundial da Saúde). **Clinical trials**. 2020. Disponível em: https://www.who.int/health-topics/clinical-trials/#tab=tab_1. Acesso em: 12 mar 2020.

OPAS (Organização Pan-Americana de Saúde, Brasil). **Folha informativa – Covid-19**. 2020.

Disponível em:

https://www.paho.org/bra/index.php?option=com_content&view=article&id=6101:covid19&Itemid=875. Acesso em: 16 maio 2020.

OPEN AIRE. **Repositories should recommend the best OS licences but it should be the uploader who chooses which one**. 2019. Disponível em: <https://www.openaire.eu/2018-09-24-10-44-10?>

Acesso em: 22 jan. 2020.

OPEN AIRE. **How do I license my research data?**. 2020. Disponível em:

<https://www.openaire.eu/how-do-i-license-my-research-data>. Acesso em: 22 jan. 2020.

OPEN RESEARCH DATA TASK FORCE. **Research Data Infrastructures in the UK**. 2017.

Disponível em: <https://www.universitiesuk.ac.uk/policy-and-analysis/research-policy/open-science/Documents/ORDTF%20report%20nr%201%20final%2030%2006%202017.pdf>. Acesso em: 14 set. 2019.

OREGON STATE UNIVERSITY. Research Data Services. **Data papers & data journals**. 2019.

Disponível em: <https://guides.library.oregonstate.edu/research-data-services/data-management-data-papers-journals>>. Acesso em: 22 ago. 2019.

OXENHAM, Simon. Legal confusion threatens to slow data science. **Nature**, v. 536, p. 16-17, 2016.

Disponível em: <https://www.nature.com/articles/536016a>. Acesso em: 14 set. 2020.

PACKER, A.L.; SANTOS, S. Ciência aberta e o novo *modus operandi* de comunicar pesquisa – Parte I [on line]. **SciELO em Perspectiva**, ago. 2019. Disponível em:

<https://blog.scielo.org/blog/2019/08/01/ciencia-aberta-e-o-novo-modus-operandi-de-comunicar-pesquisa-parte-i/>. Acesso em: 22 set. 2019.

PACKER, A.L. et al. SciELO: uma metodologia para publicação eletrônica. **Ciência da Informação**, Brasília, v. 27, n.2, p. 109-121, 1998. Disponível em:

http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-19651998000200002. Acesso em: 4 fev. 2020.

PACKER, A.L. et al. Aos 20 Anos, a Rede SciELO atualiza prioridades e avança para a ciência aberta (*on line*). **SciELO em Perspectiva**, set. 2018. Disponível em:

<https://blog.scielo.org/blog/2018/09/17/aos-20-anos-a-rede-scielo-atualiza-prioridades-e-avanca-para-a-ciencia-aberta/>. Acesso em: 22 set. 2019.

PANHUIS, Willem G. et al. A systematic review of barriers to data sharing in public health. **BMC Public Health**, v. 14, article number: 1144, 2014. Disponível em:

<https://bmcpublihealth.biomedcentral.com/articles/10.1186/1471-2458-14-1144>. Acesso em: 25 mar. 2020.

PERALTA, Patrícia P.; SILVA, Elizabeth F.; TERUYA, Dirceu Y. Busca de consenso entre o direito do autor e o acesso à informação pelo público na rede de computadores: uma ótica dos tratados relativos ao direito autoral. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 16, n. 3, p. 116-130, 2011. Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1413-99362011000300007&lng=en&nrm=iso. Acesso em: 23 jul. 2020.

PETERSEN, Carolyn. Through Patients' Eyes: Regulation, Technology, Privacy, and the Future.

Yearbook of Medical Informatics, v. 27, n. 1, p. 10–15, 2018. Disponível em:

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6115244/>. Acesso em: 22 set. 2019.

- PFEIFFENBERGER, H.; CARLSON, D. *Earth System Science Data (ESSD): A Peer Reviewed Journal for Publication of Data*. **D-Lib Magazine**, v. 17, n. 1/2, 2011. Disponível em: <http://www.dlib.org/dlib/january11/pfeiffenberger/01pfeiffenberger.html>. Acesso em: 30 out. 2019.
- PINHEIRO, Lena V. R. Do acesso livre à ciência aberta: conceitos e implicações na comunicação científica. **Reciis: Rev. Eletron. Comun. Inf. Inov Saúde**, Rio de Janeiro, v. 6, n. 2, p. 153-165, 2014. Disponível em: <https://www.arca.fiocruz.br/bitstream/icict/17107/2/8.pdf>. Acesso em: 23 ago. 2019.
- PINOWAR, Heather; DAY, Roger S.; FRIDSMA, Douglas B. Sharing detailed research data is associated with increased citation rate. **PLoS One**, n. 2, v. 3, 2007. Disponível em: <https://doi.org/10.1371/journal.pone.0000308>. Acesso em: 24 jan. 2010.
- PINOWAR, Heather A.; VISION, Todd J. Data reuse and the open data citation advantage. **PeerJ**, v. 1, e175, 2013. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3792178/>. Acesso em 16 maio 2020.
- PISANI, Elizabeth; ABOUZAHAR, Carla. Sharing health data: good intentions are not enough. **Bulletin World Health Organization**, v. 88, n. 6, p. 462-466, 2010. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2878150/>. Acesso em: 19 set. 2019.
- PLOS. Easier access to PLOS data. **The Official PLOS Blog**. 2013. Disponível em: <https://blogs.plos.org/plos/2013/01/easier-access-to-plos-data/>. Acesso em: 20 jan. 2020.
- PLOS MEDICINE. **Data availability**. 2014. Disponível em: <https://journals.plos.org/plosmedicine/s/data-availability>. Acesso em: 23 jan. 2020
- POPESCU, Iulian G. et al. Correlations on the protection of personal data and intellectual property rights in medical research. **Romanian Journal of Morphology and Embryology**, v. 59, n. 3, p. 1001-1005, 2018. Disponível em: <http://www.rjme.ro/RJME/resources/files/59031810011005.pdf>. Acesso em: 19 set. 2019.
- PRIMEZ, Claude; BRANDÃO, Adeilton A.; MOMEN, Hooman. Doenças infecciosas emergentes e publicação acelerada: quando a saúde pública tem prioridade sobre a formalidade da publicação acadêmica. **Memórias do Instituto Oswaldo Cruz**, Rio de Janeiro, v. 111, n. 5, 2016. Disponível em: <http://dx.doi.org/10.1590/0074-02760160001>. Acesso em: 8 fev. 2020.
- REICHMAN, Jerome H.; OKEDIJE, Ruth L. When Copyright Law and Science Collide: Empowering Digitally Integrated Research Methods on a Global Scale. **Minnesota Law Review**, 2012. Disponível em: https://www.minnesotalawreview.org/wp-content/uploads/2012/08/ReichmanOkediji_MLR1362.pdf. Acesso em: 22 jul. 2020.
- ROCHA, Leandro. **Programa SciELO lança seu repositório de dados, o SciELO Data**. ABEC Brasil. 2020. Disponível em: <https://www.abecbrasil.org.br/novo/2020/08/programa-scielo-lanca-seu-repositorio-de-dados-o-scielo-data/>. Acesso em: 8 set. 2020.
- SANTOS, Paula X.; ALMEIDA, Bethânia A.; HENNING, Patricia. **Livro Verde - Ciência aberta e dados abertos**: mapeamento e análise de políticas, infraestruturas e estratégias em perspectiva nacional e internacional. Rio de Janeiro: Fiocruz, 2017. 141 p. Disponível em: <https://www.arca.fiocruz.br/handle/icict/24117>. Acesso em: 3 jul. 2020.
- SANTOS, Solange; SALES, Denise P.; PACKER, Abel L. SciELO adota CC-BY como atribuição principal de Acesso Aberto (on line). **SciELO em Perspectiva**, 2015. Disponível em: <https://blog.scielo.org/blog/2015/06/19/scielo-adota-cc-by-como-atribuicao-principal-de-acesso-aberto/>. Acesso em: 4 fev. 2020.

SAVAGE, Caroline J.; VICKERS, Andrew J. Empirical Study of Data Sharing by Authors Publishing in PLoS Journals. 2009. **Plos One**, v. 4, n. 9, e7078.

SAYÃO, Luis F.; SALES, Luana F. Dados abertos de pesquisa: ampliando o conceito de acesso livre. **Reciis: Rev. Eletron. Comun. Inf. Inov Saúde**, Rio de Janeiro, v. 8, n. 2, p.76-92, 2014. Disponível em: <<https://www.reciis.icict.fiocruz.br/index.php/reciis/article/viewFile/611/1252>>. Acesso em: 29 maio 2020.

SAYÃO, Luis F.; SALES, Luana F. Algumas considerações sobre os repositórios digitais de dados de pesquisa. **Informação & Informação**, Londrina, v. 21, n. 2, p. 90-115, maio/ago., 2016a. Disponível em: <<http://www.uel.br/revistas/uel/index.php/informacao/article/view/27939>>. Acesso em: 21 set. 2019.

SAYÃO, Luis F.; SALES, Luana F. Curadoria digital e dados de pesquisa. **A to Z: Novas Práticas em Informação e Conhecimento**. Florianópolis, v. 5, n. 2, p. 67-71, 2016b. Disponível em: <https://revistas.ufpr.br/atoz/article/view/49708/3017>. Acesso em: 19 set. 2019.

SCIELO. Sobre o SciELO. **Declaração de Acesso Aberto**. 2019. Disponível em: <<https://www.scielo.org/pt/sobre-o-scielo/declaracao-de-acesso-aberto/>>. Acesso em: 07 fev. 2020.

SCIELO. SciELO Preprints em operação. **SciELO em Perspectiva**, 2020a. Disponível em: <<https://blog.scielo.org/blog/2020/04/07/scielo-preprints-em-operacao/>>. Acesso em: 2 set. 2020.

SCIELO. SciELO Analytics. **Licenças de uso de periódicos**. 2020b. Disponível em: <<https://analytics.scielo.org/w/publication/journal#journals-use-licenses>>. Acesso em: fe. 2020.

SCIENTIFIC DATA. Expanding our generalist data repository options. Blog. 2016. Disponível em: <<http://blogs.nature.com/scientificdata/2016/11/14/expanding-our-generalist-data-repository-options/>>. Acesso em: 21 set. 2019.

SCIENTIFIC DATA. **Aims and Scope**. 2018. Disponível em: <<https://www.nature.com/sdata/publish/for-authors#aims-scope>>. Acesso em: 22 ago. 2019.

SICILIA, Miguel-Angel; GARCIA-BARRIOCANAL, Elena; SÁNCHEZ-ALONSO, Salvador. **Community curation in open dataset repositories: insights from Zenodo**. 2017. Disponível em: <doi: 10.1016/j.procs.2017.03.009>. Acesso em: 24 jan. 2020.

SMITH, Richard. The Trouble With Medical Journals. **Journal of the Royal Society of Medicine**, v. 99, n. 3, p. 115-119, 2006. Disponível em: <<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1383755/>>. Acesso em: 10 mar. 2020.

SOUZA, Allan R. Os direitos morais do autor. **civilistica.com** [on line], Rio de Janeiro, a. 2., n. 1., p. 1-23, 2013. Disponível em: <<http://civilistica.com/wp-content/uploads/2015/02/Souza-civilistica.com-a.2.n.1.2013.pdf>>. Acesso em: 09 ago. 2019.

SOUZA, Carlos A. P. O domínio público e a função social do direito autoral. **Liinc em Revista**, Rio de Janeiro, v.7, n. 2, p. 664- 680, 2011. Disponível em: <<http://revista.ibict.br/liinc/article/view/3314/2924>>. Acesso em: 09 ago. 2019.

SPARC. Scholarly Publishing and Academic Resources Coalition. **Open Access**. 2019. Disponível em: <<https://sparcopen.org/open-access/>>. Acesso em: 29 out. 2019.

SPINAK, Ernesto; PACKER, A. L. Dados científicos: Nature Publishing Group leva adiante a comunicação de dados científicos com sua nova publicação on line de acesso aberto [on line] **SciELO em Perspectiva**, 2014. Disponível em: <https://blog.scielo.org/en/2014/02/04/scientific-data-nature->

publishing-group-moves-the-communication-of-scientific-data-forward-with-its-new-online-open-access-publication/#.XWXgpOhKi71. Acesso em: 21 ago. 2019.

SPRINGER NATURE. **Funders and institutions requiring a CC BY licence for OA articles**. 2018. Disponível em: <https://www.springernature.com/gp/open-research/funding/funders-requiring-cc-by-for-articles>. Acesso em: 10 abr. 2020.

SUBER, Peter. **Problems and opportunities (blizzards and beauty)**. SPARC Open Access Newsletter, issue #111. 2007. Disponível em: <https://legacy.earlham.edu/~peters/fos/newsletter/07-02-07.htm>. Acesso em: 23 maio 2020.

SUBER, Peter. **Open Access**. Cambridge: MIT Press, 2012. Disponível em: <https://openaccesseks.mitpress.mit.edu/>. Acesso em: 10 mar. 2020.

SWAN, Alma. **Diretrizes para políticas de desenvolvimento e promoção do acesso aberto**. Brasília: UNESCO Brasil; IBICT, 2016. 82 p. (Série diretrizes para o acesso aberto). Disponível em: <https://unesdoc.unesco.org/ark:/48223/pf0000246018>>. Acesso em 28 ago. 2019.

TAICHMAN, Darren B. et al. Editorial. Sharing Clinical Trial Data: a Proposal from the International Committee of Medical Journal. **Annals of Internal Medicine**, n. 164, p. 505-506, 2016. Disponível em: <https://www.acpjournals.org/doi/10.7326/M15-2928>>. Acesso em: 19 mar. 2020.

TAICHMAN, Darren B. et al. Editorial. Data Sharing Statements for Clinical Trials: A Requirement of the International Committee of Medical Journal Editors. Editorial. **Annals of Internal Medicine**, 2017. Disponível em: http://www.icmje.org/news-and-editorials/data_sharing_june_2017.pdf>. Acesso em: 19 mar. 2020.

TENNANT, Jonathan P. et al. The academic, economic and societal impacts of Open Access: an evidence-based review. **F1000Research**, 5:632, 2016. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4837983/>>. Acesso em: 26 jun. 2020.

TENOPIR, Carol et al. Data Sharing by Scientists: Practices and Perceptions. 2011. **PLoS One**, v. 6, n. 6, e21101. 2011. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3126798/>>. Acesso em: 21 jan. 2020.

TENOPIR, Carol et al. Changes in Data Sharing and Data Reuse Practices and Perceptions among Scientists Worldwide. **PLoS One**, v. 10, n. 8, e0134826, 2015. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4550246/>>. Acesso em: 1 set. 2019.

TENOPIR, Carol et al. Data sharing, management, use, and reuse: Practices and perceptions of scientists worldwide. 2020. **PLoS One** v. 15, n. 3, p. e0229003, 2020. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7065823/>>. Acesso em: 1 set. 2020.

THEBMJ. Editorials. The new BMJ policy on sharing data from drug and device trials. **BMJ**, n. 345, e7888, 2012. Disponível em: <https://www.bmj.com/content/345/bmj.e7888>>. Acesso em: 26 mar. 2020.

THEBMJ. Editorials. *The BMJ* requires data sharing on request for all trials. **BMJ**, n. 350, h2373, 2015. Disponível em: <https://www.bmj.com/content/350/bmj.h2373>. Acesso em: 26 mar. 2020.

THEBMJ. Publishing Model. **Acesso livre**. 2016. Disponível em: <https://www.bmj.com/about-bmj/publishing-model>. Acesso em: 26 mar. 2020.

THEBMJ. Editorials. Data sharing statements for clinical trials. **BMJ**, n. 357, j2372, Jun. 2017. Disponível em: <https://www.bmj.com/content/357/bmj.j2372>. Acesso em: 26 mar. 2020.

THEBMJ. **Article types and preparation.** Additional requirements by article type. Research. Data Sharing. 2019. Disponível em: <<https://www.bmj.com/about-bmj/resources-authors/article-types>>. Acesso em: 26 mar. 2020.

THEBMJ. **Copyright, open access, and permission to reuse.** 2020. Disponível em: <<https://www.bmj.com/about-bmj/resources-authors/forms-policies-and-checklists/copyright-open-access-and-permission-reuse>>. Acesso em: 26 mar. 2020.

THE HAGUE DECLARATION. **Knowledge Discovery in the Digital Age.** 2015. Disponível em: <https://wayback.archive-it.org/12503/20190730165740/https://thehaguedeclaration.com/wp-content/uploads/sites/2/2015/04/Liber_DeclarationA4_2015.pdf>. Acesso em: 13 maio 2020.

THE LANCET. **2019-nCoV Resource Centre.** 2020. Disponível em: <<https://www.thelancet.com/coronavirus>>. Acesso em 6 fev. 2020.

THE ROYAL SOCIETY. **Science as an open enterprise.** Londres: The Royal Society, 2012. Disponível em: <https://www.academia.edu/18283830/Science_as_an_Open_Enterprise>. Acesso em: 3 ago. 2020.

TUCKER, Katherine et al. Protecting patient privacy when sharing patient-level data from clinical trials. **BMC Medical Research Methodology**, n. 16, supl. 1, p. 77, 2016. Disponível em: <<https://bmcmmedresmethodol.biomedcentral.com/articles/10.1186/s12874-016-0169-4>>. Acesso em: 22 set. 2019.

UHLIR, Paul. The emerging role of open repositories for scientific literature as a fundamental component of the public research infrastructure. In: SICA, Giandomenico. **Open access: Open problems.** Polimetrica: Milão, 2006. p. 57-103. Disponível em: <https://www.researchgate.net/publication/220032324_Open_Access_Open_Problems_2006>. Acesso em: 17 jan. 2021.

UHLIR, Paul. **The Value of Open Data Sharing: A CODATA Report for the Group on Earth Observations.** 2015. Disponível em: <https://www.earthobservations.org/documents/dsp/20151130_the_value_of_open_data_sharing.pdf>. Acesso em: 28 maio 2020.

UHLIR, Paul; CLEMENT, Gail (eds). **Legal Interoperability of Research Data: Principles and Implementation Guidelines.** RDA-Codata Legal Interoperability Interest Group. 2016. Disponível em: <http://www.codata.org/uploads/Legal%20Interoperability%20Principles%20and%20Implementation%20Guidelines_Final2.pdf>. Acesso em: 13 ago. 2020.

UNIVERSITY LIBRARIE. King Abdullah University of Science and Technology. **Data Management Planning: Intellectual Property (IP) and Licensing.** 2019. Disponível em: <<https://libguides.kaust.edu.sa/c.php?g=279013&p=3136507>>. Acesso em: 18 abr. 2019.

UNIVERSITY OF CAMBRIDGE. Office of Scholarly Communication. **Blog Unlocking Research.** Is CC-BY really a problem or are we boxing shadows?. 2016. Disponível em: <https://unlockingresearch-blog.lib.cam.ac.uk/?p=555>. Acesso em 3 fev. 2020.

VAN DE SOMPEL, Herbert; LAGOZE, Carl. The Santa Fe Convention of the Open Archives Initiative. **D-Lib Magazine**, v. 6, n. 2, 2000. Disponível em: <http://www.dlib.org/dlib/february00/vandesompel-oai/02vandesompel-oai.html>. Acesso em: 16 maio 2020.

VASILEVSKY, Nicole et al. Reproducible and reusable research: are journal data sharing policies meeting the mark? *PeerJ*, 5, e3208, 2017. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5407277>. Acesso em: 28 ago. 2019.

VIEIRA, Sonia. **Como elaborar questionários**. São Paulo: Atlas, 2009. 159 p.

VISION, Todd J. Open Data and the Social Contract of Scientific Publishing. *BioScience (on line)*, v. 60, n. 5, p. 330-331, 2010. Disponível em: <https://academic.oup.com/bioscience/article/60/5/330/237825>. Acesso em: 03 fev. 2020.

WALLIS, Jillian C.; BORGMAN, Christine L. Who is Responsible for Data? An Exploratory Study of Data Authorship, Ownership, and Responsibility. *ASIS & T Digital Library*, 2012. Disponível em: <https://asistdl.onlinelibrary.wiley.com/doi/full/10.1002/meet.2011.14504801188>. Acesso em: 8 fev. 2020.

WALLIS, Jillian C.; ROLANDO, Elizabeth; BORGMAN, Christine L. If We Share Data, Will Anyone Use Them? Data Sharing and Reuse in the Long Tail of Science and Technology. *PLoS One*, v. 8, n. 7, p. e67332, 2013. Disponível em: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0067332>. Acesso em: 4 jun. 2020.

WEITZEL, Simone R. As novas configurações do Acesso Aberto: desafios e propostas. *Reciis: Rev. Eletron. Comun. Inf. Inov Saúde*, Rio de Janeiro, v. 8, n. 2, 2014. Disponível em: <https://www.reciis.icict.fiocruz.br/index.php/reciis/article/view/447>. Acesso em: 25 mar. 2020.

WELLCOME OPEN RESEARCH. **Data Availability**. 2017. Disponível em: <https://wellcomeopenresearch.org/about/policies#dataavail>. Acesso em 8 abr. 2020

WELLCOME TRUST. **Sharing research data to improve public health: full joint statement by funders of health research**. 2010. Disponível em: <https://wellcome.ac.uk/what-we-do/our-work/sharing-research-data-improve-public-health-full-joint-statement-funders-health>. Acesso em: 4 abr. 2020.

WELLCOME TRUST. **Statement on data sharing in public health emergencies**. 2016. Disponível em: <https://wellcome.ac.uk/press-release/statement-data-sharing-public-health-emergencies>. Acesso em: 6 fev. 2020.

WELLCOME TRUST. **Data, software and materials management and sharing policy**. 2017. Disponível em: <https://wellcome.org/grant-funding/guidance/data-software-materials-management-and-sharing-policy>. Acesso em: 23 jul. 2020.

WELLCOME TRUST. **Wellcome and the Bill & Melinda Gates Foundation join the Open Access coalition**. 2018. Disponível: <https://wellcome.ac.uk/press-release/wellcome-and-bill-melinda-gates-foundation-join-open-access-coalition>. Acesso em: 4 abr. 2020.

WELLCOME TRUST. **Compartilhamento de dados e descobertas relevantes para o novo surto de coronavírus (Covid-19)**. Comunicado à Imprensa. 2020. Disponível em: <https://wellcome.ac.uk/coronavirus-covid-19/open-data>. Acesso em: 20 jul. 2020.

WIKIPEDIA. **Zenodo**. 2019. Disponível em: <https://en.wikipedia.org/wiki/Zenodo>. Acesso em: 23 jan. 2020.

WILKINSON, Mark D. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, v. 3, p. 160018, Mar. 2016. Disponível em: <https://www.nature.com/articles/sdata201618>. Acesso em: 29 ago. 2019.

WILLINSKY, John. **The access principle**: the case for open access to research and scholarship. Cambridge: MIT Press, 2006. Disponível em: <https://pdfs.semanticscholar.org/7131/14a37f06dfda0cfa997bc2cbacb88785a77d.pdf>. Acesso em: 10 mar. 2020.

WOMACK, Ryan P. Research Data in Core Journals in Biology, Chemistry, Mathematics, and Physics. **PLoS One**, v. 10, n. 12, e0143460, 2015. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4670119/#sec008title>. Acesso em: 20 set. 2019.

WOUTERS, Paul; HAAK, Wouter. **Open Data**: The Researcher Perspective. 48p. 2017. Disponível em: https://www.elsevier.com/__data/assets/pdf_file/0004/281920/Open-data-report.pdf. Acesso em: 5 jun. 2020

ZENODO. **Sobre o Zenodo**. 2020a. Disponível em: <https://about.zenodo.org/>. Acesso em: 24 jan. 2020.

ZENODO. **Help**. 2020b. Disponível em: <https://help.zenodo.org/>. Acesso em: 26 jan. 2020.

ZIMAN, John. *Conhecimento Público*. Tradução de Regina Regis Junqueira. v. 8. Belo Horizonte: ed. Itatiaia; São Paulo: ed. da USP, 1979. Coleção o homem e a ciência.

APÊNDICE A - Questionário eletrônico de autoaplicação para editores científicos.

Abertura de dados de pesquisa subjacentes: aspectos do direito autoral na perspectiva das revistas científicas

A abertura de dados de pesquisa subjacentes a artigos publicados em revistas científicas levanta questões tanto a respeito de acesso, compartilhamento e reúso quanto de direitos autorais desses dados a serem disponibilizados publicamente em acesso aberto.

Como você pretende lidar com esse novo desafio?

ABERTURA, ACESSO, COMPARTILHAMENTO E REÚSO DE DADOS DE PESQUISA

1- Você é favorável à disponibilização pública de dados de pesquisa subjacentes aos artigos científicos para fins de compartilhamento e reúso?

Sim Não

Por quê? _____

2- Em relação à abertura de dados de pesquisa subjacentes aos artigos, sua revista:

Já adota.

Adotará.

Não adotará.

Por quê? _____

3- Sua revista exigirá do(s) autor(es) uma declaração de concordância em compartilhar os dados de pesquisa subjacentes ao artigo?

Sim Não Não se aplica [formulário eletrônico tem opção de pular as perguntas que não se aplicam em razão da resposta anterior]

Por quê? _____

4- Em caso positivo, a declaração de concordância assinada pelo autor seria uma pré-condição para que o artigo seja publicado?

Sim Não Não se aplica

Por quê? _____

5- Sua revista exigirá do(s) autor(es) um plano de gestão de dados?

Sim Não Não se aplica

Por quê? _____

6- Sua revista indica(rá) para o(s) autor(es) um ou mais repositórios, específicos de dados, para depósito e disponibilização pública de dados subjacentes ao artigo?

Sim Não Não se aplica

Por quê? _____

7- Sua revista indicaria repositório que aceita embargo que impeça a publicação imediata do conjunto de dados?

Sim Não Não sei

Por quê? _____

8- Dentre os repositórios abaixo, qual a sua revista indicaria? Se quiser, marque mais de uma opção.

- Dryad
- Figshare
- Zenodo
- Dataverse
- Outro. Cite o nome _____
- Não sei.

DIRIEITO AUTORAL APLICÁVEL A DADOS DE PESQUISA

9- Quem deve fazer o envio dos dados de pesquisa subjacentes a um artigo para fins de depósito em repositório?

- O editor.
- A agência financiadora da pesquisa.
- O autor do artigo com capacidade para definir o licenciamento para reuso dos dados.

10- Na sua opinião, a quem cabe decidir sobre qual será a licença pública que descreve os direitos para fins de reuso de dados subjacentes aos artigos?

- À revista
- Ao/À autor/a
- Ao repositório de dados
- Ao detentor dos direitos autorais
- Não sei

11- Conjuntos de dados de pesquisa são passíveis de proteção de direrito autoral. Que licença a revista consideraria adotar para reuso legalizado dos dados de pesquisa subjacentes aos artigos disponibilizados publicamente?

- Sim, respeitamos o direito autoral. Na medida em que a lei permita, entretanto, sugerimos a adoção de licença que permita distribuir, remixar, adaptar, inclusive comercializar o trabalho (CC-BY).
- Concordamos, mas só recebemos dados de pesquisa para depósito sob um instrumento que dedique o conjunto de dados em questão ao domínio público (CC0 Dedicção ao Domínio Público).
- Concordamos, porém indicamos a adoção de licença que permita distribuir, remixar, adaptar um trabalho, mas não comercialmente e desde que os novos trabalhos sejam licenciados sob os mesmos termos (CC-BY-NC-SA).
- Sim, respeitamos os direitos autorais. Na medida em que a lei permita, entretanto, sugerimos a adoção de licença que permita distribuir, remixar, adaptar, exceto comercializar o trabalho (CC-BY-NC).
- Sim, por isso mesmo adotamos uma licença que permite reutilização para qualquer fim, inclusive comercial, mas não se pode compartilhar o trabalho na forma adaptada com outras pessoas (CC-BY-ND).
- Sim, sugerimos a adoção de licença que permita distribuir, remixar, adaptar, incluindo a comercialização, mas o usuário terá de licenciar o novo trabalho sob os mesmos termos da licença que o beneficiou (CC-BY-SA).
- Concordamos e adotamos licença que só permite *download* do trabalho e compartilhamento, mas sem que possa ser alterado de nenhuma forma ou utilizado para fins comerciais (CC-BY-NC-ND).

12- Que repositório de dados de pesquisa a sua revista indicaria levando em conta o tipo de política de direitos autorais que ele adota?

- Indicaria um repositório que cogite obter exclusividade sobre os dados depositados.
- Indicaria um que obtenha um direito não exclusivo para publicar esses dados, mantendo os direitos patrimoniais com o(s) autor(es) ou detentor(es).
- É indiferente.

Por quê? _____

APÊNDICE B - Questionário eletrônico de autoaplicação para gestores de repositórios institucionais e de dados

Abertura de dados de pesquisa subjacentes: aspectos do direito autoral na perspectiva das revistas científicas

A abertura de dados de pesquisa subjacentes a artigos de revistas científicas depositados em repositório levanta questões de acesso, compartilhamento e reuso, mas também de direitos autorais desses dados a serem disponibilizados publicamente em acesso aberto.

Como o seu repositório lida com esse novo papel?

ABERTURA, ACESSO, COMPARTILHAMENTO E REÚSO DE DADOS DE PESQUISA

1- O repositório de dados de pesquisa passou a ter um papel estratégico no sistema de comunicação científica. Indique a ordem de importância dos serviços que ele oferece, enumerando as opções de 1 a 5, sendo 1 a mais importante.

- Curadoria de dados.
- Adoção dos princípios FAIR.
- Uso de identificadores persistentes, como DOI e ORCID.
- Vinculação dos dados ao artigo.
- Citação de dados.

Outros _____

2- Você é favorável à abertura e à disponibilização pública de conjuntos de dados de pesquisa subjacentes aos artigos científicos para fins de compartilhamento e reuso?

- Sim, se a pesquisa for financiada com recursos públicos, o que funcionaria como o retorno de um bem público à sociedade que pagou pela pesquisa original.
- Não, pois, mesmo financiada com recursos públicos, dados de pesquisa científica não interessam a qualquer público.
- Sim, se a abertura e a disponibilização pública permitirem compartilhamento e reuso sem barreiras legais ou econômicas.

Outros _____

3- Algumas revistas exigem do(s) autor(es) uma declaração de concordância em compartilhar os dados de pesquisa subjacentes ao artigo, chegando até a considerar a não publicação em caso de discordância. Concorda com essa medida?

- Sim Não

Por quê? _____

4- Qual o papel do repositório de dados de pesquisa na comunicação científica?

- Ser um novo modelo de divulgação de dados de pesquisa, viabilizando a avaliação por pares, a citação do autor original e a atribuição de identificadores, como DOI e ORCID, para cada conjunto de dados.
- Oferecer, além dos tradicionais armazenamento e preservação, serviços como curadoria de dados e mineração eletrônica de dados e textos.
- Favorecer o avanço da produção científica por meio do acesso aberto, gratuito e sem barreiras a conjuntos de dados de pesquisa vinculados a artigos científicos.
- Promover um novo tipo de canal formal digital de divulgação de artigos e dados de pesquisa, que represente uma mudança paradigmática na comunicação científica, em que permita também o depósito de figuras, imagens e vídeos sem vínculo com artigos e a formação de comunidades de pesquisa.
- Nenhuma das opções.

5- Seu repositório tem uma diretriz específica para tratar da privacidade, dos dados pessoais, incluindo os dados sensíveis, da pessoa humana que participe de uma pesquisa em saúde pública?

- Sim, somente publicamos dados pessoais e/ou sensíveis se tiverem passado pelo processo de anonimização.
- Não, a privacidade dos dados é uma questão que deve ser tratada entre o pesquisador e o participante da pesquisa.
- Sim, somente publicamos dados pessoais e/ou sensíveis mediante o consentimento expresso do participante da pesquisa e a anonimização desses dados.

Outros _____

DIRIEITO AUTORAL APLICÁVEL A DADOS DE PESQUISA

6- Conjuntos de dados de pesquisa subjacentes a artigos são passíveis de proteção do direito autoral. Marque as opções de licenciamento Creative Commons que julga ideal para esse tipo de dados. Todas as licenças exigem a citação do autor original dos dados.

- Sim, respeitamos o direito autoral. Na medida em que a lei permita, entretanto, sugerimos a adoção de licença que permita distribuir, remixar, adaptar, inclusive comercializar o trabalho (CC-BY).
- Concordamos, mas só recebemos dados de pesquisa para depósito sob um instrumento que dedique o conjunto de dados em questão ao domínio público (CC0 Dedicção ao Domínio Público).
- Concordamos, porém indicamos a adoção de licença que permita distribuir, remixar, adaptar um trabalho, mas não comercialmente e desde que os novos trabalhos sejam licenciados sob os mesmos termos (CC-BY-NC-SA).

- Sim, respeitamos os direitos autorais. Na medida em que a lei permita, entretanto, sugerimos a adoção de licença que permita distribuir, remixar, adaptar, exceto comercializar o trabalho (CC-BY-NC).
- Sim, por isso mesmo adotamos uma licença que permite reutilização para qualquer fim, inclusive comercial, mas não se pode compartilhar o trabalho na forma adaptada com outras pessoas (CC-BY-ND).
- Sim, sugerimos a adoção de licença que permita distribuir, remixar, adaptar, incluindo a comercialização, mas o usuário terá de licenciar o novo trabalho sob os mesmos termos da licença que o beneficiou (CC-BY-SA).
- Concordamos e adotamos licença que só permite *download* do trabalho e compartilhamento, mas sem que possa ser alterado de nenhuma forma ou utilizado para fins comerciais (CC-BY-NC-ND).
- Outros _____

7- Há uma tendência de os repositórios recomendarem aos pesquisadores a escolherem e os editores a apoiarem o instrumento CC0 Dedicção ao Domínio Público, da Creative Commons, para licenciar o reuso de dados. Concorda que há essa tendência?

- Sim Não Não conheço

Por quê? _____

8- É favorável à padronização das normas internacionais de direito autoral para uso por todos os repositórios de dados de pesquisa?

- Sim, não acabaria, mas minimizaria a dificuldade de lidar com diferentes leis de diferentes países.
- Indiferente, pois acho impraticável a chamada “harmonização” das diferentes normas de direito autoral de diferentes países, tendo em vista pesados interesses políticos e econômicos em jogo.
- Não, pois acredito que as licenças alternativas, como as da Creative Commons, praticamente acabaram com as dúvidas quanto à extensão do uso de um direito protegido.

Outros _____

9- A fim de alcançar esse estágio de normas internacionais padrão para todos os repositórios, recomendaria que os legisladores criassem um direito novo, específico para a abertura de dados? Para indicar a sua ordem de prioridade, numere as opções de 1 a 4, sendo 1 a mais prioritária.

- Sim, é imprescindível para o avanço global das pesquisas.
- Não, identifico uma transição no processo de comunicação científica e daqui a pouco tudo se ajusta.
- Sim, sobretudo que levasse em conta casos de emergência em saúde pública, como o Covid-19.
- Sim, mas incluiria nesta cesta todos os movimentos da ciência aberta que porventura se deparem com normas de direito autoral, como o dos recursos educacionais abertos e o da literatura científica.

Outros _____

APÊNDICE C – Roteiro de perguntas para entrevista com editores científicos.

- Na sua opinião, qual é o papel dos periódicos científicos e seus editores na abertura de dados de pesquisa?
- Quais as principais dificuldades e desafios para que os periódicos adotem políticas de abertura de dados subjacentes aos artigos?
- Falta padronização de práticas, como citação, revisão por pares, curadoria e mineração de textos e dados de pesquisa?
- Que motivos levam pesquisadores a compartilhar seus dados e quais os leva a não compartilhar?
- Conhece formas de proteção aos vários tipos de dados existentes? [legal, privacidade, confidencialidade].
- Qual é a sua opinião sobre as políticas dos repositórios e a sua relação com as políticas das revistas no tocante à abertura de dados de pesquisa?
- Qual é a sua opinião sobre o dever legal de observar a propriedade intelectual relacionada a dados de pesquisa, especialmente o licenciamento quanto aos eventuais direitos autorais aplicáveis a esses dados?

APÊNDICE D – Roteiro de perguntas para entrevista com gestores de repositórios digitais e de dados.

- Como avalia o papel dos repositórios de dados no atual sistema de comunicação científica e na disseminação da abertura de dados de pesquisa?
- Quais as principais dificuldades e desafios para que os repositórios adotem políticas de abertura de dados subjacentes aos artigos? Falta padrão nas políticas?
- Qual é a sua opinião sobre a exigência de declaração de concordância assinada pelo pesquisador para o compartilhamento de dados de pesquisa subjacentes a artigos científicos?
- Como percebe o comportamento do pesquisador em abrir seus dados de pesquisa?
- Qual é o seu conhecimento e opinião sobre a adoção de licenças públicas na disponibilização e reúso de dados de pesquisa?
- Já verificou alguma dificuldade do autor de artigo de pesquisa em escolher uma licença que corresponda aos direitos que ele próprio deseja disponibilizar para quem vá reusar os dados da sua pesquisa?

APÊNDICE E – Termo de Consentimento Livre e Esclarecido (enviado aos editores científicos).

Caro Editor(a),

você está sendo convidado a participar do estudo *Abertura de dados de pesquisa subjacentes a artigos: aspectos do direito autoral na perspectiva das revistas científicas*.

A proposta é responder a um Questionário eletrônico de autoaplicação. É um estudo anônimo em que o respondente não é identificado. Basta clicar no link a seguir, responder e enviar de volta:

<https://bit.ly/questionario-paulo-prop1> O estudo é parte de um projeto de pesquisa de doutorado, que conta com a orientação da Prof. Sarita Albagli, do Programa de Pós-Graduação em Ciência da Informação do IBICT-UFRJ.

Os objetivos são:

- a) analisar o regime de direito autoral aplicável a dados de pesquisa, observando o processo de publicação e abertura de dados de pesquisa subjacentes a artigos de periódicos, em particular os da saúde; e
- b) pesquisar políticas de abertura de dados de pesquisa de repositórios de dados, agências de financiamento da pesquisa e revistas científicas, para identificar o quadro político-normativo e inferir eventuais transformações no regime de direito autoral aplicável a dados de pesquisa.

Obrigado por participar. Para informações adicionais:

Pesquisador: Paulo Cezar Vieira Guanaes (PPGCI-IBICT-UFRJ): Tel.: +55 (21) 997363987; e-mail: pcguanaes@gmail.com. **Orientadora:** Sarita Albagli: Tel.: +55 (21) 3873-9453; email: sarita.albagli@gmail.com

Atenciosamente,

Paulo Guanaes / <https://orcid.org/0000-0003-1248-0462>

APÊNDICE F – Termo de Consentimento Livre e Esclarecido (enviado aos gestores de repositórios).

Caro Gestor,

você está sendo convidado a participar do estudo *Abertura de dados de pesquisa subjacentes a artigos: aspectos do direito autoral na perspectiva das revistas científicas*.

A proposta é responder a um Questionário eletrônico de autoaplicação.

É um estudo anônimo em que o respondente não é identificado.

Basta clicar no link a seguir, responder e enviar de volta:

<https://bit.ly/questionario-paulo-gestores-prop1>

O estudo é parte de um projeto de pesquisa de doutorado, que conta com a orientação da Prof. Sarita Albagli, do Programa de Pós-Graduação em Ciência da Informação do IBICT-UFRJ, do Rio de Janeiro.

Os objetivos são:

- a) analisar o regime de direito autoral aplicável a dados de pesquisa, observando o processo de publicação e abertura de dados de pesquisa subjacentes a artigos de periódicos, em particular os da saúde; e
- b) pesquisar políticas de abertura de dados de pesquisa de repositórios de dados, agências de financiamento da pesquisa e revistas científicas, para identificar o quadro político-normativo e inferir eventuais transformações no regime de direito autoral aplicável a dados de pesquisa.

Obrigado por participar.

Para informações adicionais:

Pesquisador: Paulo Cezar Vieira Guanaes (PPGCI-IBICT-UFRJ): Tel.: +55 (21) 997363987; e-mail: pcguanaes@gmail.com. **Orientadora:** Sarita Albagli: Tel.: +55 (21) 3873-9453; email: sarita.albagli@gmail.com

Atenciosamente,

Paulo Guanaes / <https://orcid.org/0000-0003-1248-0462>

APÊNDICE G – Roteiro das perguntas de uma entrevista realizada com umx Product Manager & Make Data Count Principal Investigator, do repositório internacional Dryad.

- How do you evaluate the role of data repositories in the current scientific communication system and in the dissemination of research data opening?
- What are the main difficulties and challenges for the repositories to adopt data openness policies underlying the articles?
- Is there a lack of standard in policies?
- What is your opinion on the requirement for a declaration of agreement signed by the researcher for the sharing of research data underlying scientific articles?
- How do you perceive the researcher's behavior in opening your research data?
- What is your knowledge and opinion about the adoption of public licenses in making research data available and reused?
- Who owns the data?
- Experience at Dryad opening up and sharing research data?