

Tópicos sobre dados abertos para editores científicos

2020

Organizadores:

Milton Shintaku, Luana Farias Sales e Michelli Costa

Autores:

Adeilton Alves Brandão
Adriana Maria Tonini
Benedito Barraviera
Caterina Groposo Pavão
Dalmo Correia
Emanuelle Torino
Enrique Muriel-Torrado
Fabiano Couto Corrêa da Silva
Fábio Castro Gouveia
Gildenir Carolino Santos
Guilherme Ataíde Dias
João Moreira
José Eduardo Santarem Segundo
Juliana Meireles Fortaleza
Luana Farias Sales
Lucas Rocha
Lucélia de Souza

Lúcia da Silveira
Luis Fernando Sayão
Marckson Roberto Ferreira de Sousa
Michelli Costa
Milton Shintaku
Patrícia Henning
Patrícia Rocha Bello Bertin
Paulo Cezar Vieira Guanaes
Paulo Henrique de Assis Santana
Rebeca Fernandes de Moura
Rosilene Paiva Marinho de Sousa
Rui Seabra Ferreira Junior
Sandra Milena Roa-Martínez
Sandro Rautenberg
Silvana Aparecida Borsetti Gregorio Vidotti
Suzana Caetano da Silva Lanne
Thaís Machado Scherrer





TÓPICOS SOBRE DADOS ABERTOS PARA EDITORES CIENTÍFICOS

ASSOCIAÇÃO BRASILEIRA DE EDITORES CIENTÍFICOS

São Paulo

2020

ASSOCIAÇÃO BRASILEIRA DE EDITORES CIENTÍFICOS

Diretoria

Rui Seabra Ferreira Júnior (CEVAP/UNESP, Botucatu, SP)
Presidente

Ricardo Antunes de Azevedo (ESALQ/USP, Piracicaba, SP)
Vice-Presidente

Ana Marlene Freitas de Moraes (IAE/DCTA, São José dos Campos, SP)
Secretário-Geral

Milton Shintaku (Ibict, Brasília, DF)
1º Secretário

Benedito Barraviera (FMB/UNESP, Botucatu, SP)
1º Tesoureiro

Suzana Caetano da Silva Lannes (FCF/USP, São Paulo, SP)
2º Tesoureiro

Conselho Deliberativo – Membros Titulares

Piotr Trzesniak (Universidade Federal de Itajubá, Itajubá, MG)
Elisabete Werlang (UFSC, Florianópolis, SC)

Bruna M. S. Erlandsson (Linceu Editorial, São José dos Campos, SP)

Eloísa da Conceição Príncipe de Oliveira (Ibict, Rio de Janeiro, RJ)

Elizabeth da Costa Mattos (Instituto de Aeronáutica e Espaço, São José dos Campos, SP)

Meryt Tarcila Teixeira Zanini (SBCCV, São Paulo, SP)

Germana Fernandes Barata (Núcleo de Desenvolvimento da Criatividade da Unicamp, Campinas, SP)

Gildenir Carolino Santos (Biblioteca Central da Unicamp, Campinas, SP)

Valtencir Zucolotto (Instituto de Física de São Carlos da USP, São Carlos, SP)

Vera Lúcia Gomes Klein (Instituto de Ciências Biológicas da UFG, Goiânia, GO)

Conselho Fiscal – Membros Titulares

Ilda Fontes (FGV-EAESP, São Paulo, SP)

David Matos Milhomens (Instituto Federal de Educação, Ciência e Tecnologia, Caxias do Sul, RS)

Neusa Maria Bastos Fernandes dos Santos (PUC-SP, São Paulo, SP)



TÓPICOS SOBRE DADOS ABERTOS PARA EDITORES CIENTÍFICOS

Organizadores

Milton Shintaku
Luana Farias Sales
Michelli Costa

ASSOCIAÇÃO BRASILEIRA DE EDITORES CIENTÍFICOS
São Paulo
2020

© 2020 Associação Brasileira de Editores Científicos

Esta obra é licenciada sob uma licença Creative Commons - Atribuição CC BY 4.0, sendo permitida a reprodução parcial ou total, desde que mencionada a fonte.



Impresso no Brasil

Revisão Gramatical

Rafael Teixeira

Design Gráfico, Diagramação

Rafael Fernandez Gomes

Normalização

Ingrid Torres Schiessl

Dados Internacionais de Catalogação-na-Publicação (CIP)

Bibliotecária: Ingrid Schiessl CRB1/ 3084

T674 Tópicos sobre dados abertos para editores científicos / Milton Shintaku; Luana Sales e Michelli Costa (org.). -- Botucatu, SP: ABEC, 2020. --

240 p.

ISBN 978-85-93910-04-3

DOI: 10.21452/978-85-93910-04-3

1. Comunicação científica. 2. Ciência aberta. 3. Dados abertos. 4. Tecnologia da Informação. I. Shintaku, Milton (Org.). II. Sales, Luana (Org.). III. Costa, Michelli (Org.). IV. Título.

CDD 070.5

CDU 655.4



(14) 3815.5095

abec@abecbrasil.org.br

doi@abecbrasil.org.br

Rua Azaleia, 399 - Edifício 3 Office, 7º Andar, Sala 75 - Bairro Chácara Floresta
CEP 18.603-550, Botucatu, SP

Como citar:

SHINTAKU, M.; SALES, L. F; COSTA, M. (org). **Tópicos sobre dados abertos para editores científicos**. Botucatu, SP: ABEC, 2020. 240 p. DOI: 10.21452/978-85-93910-04-3.

Sumário

Agradecimentos Os organizadores	7
Prefácio Rui Seabra Ferreira Junior	9
Apresentação Patrícia Rocha Bello Bertin, Juliana Meireles Fortaleza	11
Ciência Aberta, gestão de dados de pesquisa e novas possibilidades para a editoração científica Luana Farias Sales, Michelli Costa, Milton Shintaku	13
Políticas de Ciência Aberta e a abertura dos dados de pesquisa Michelli Costa	23
Abrindo os dados de acesso e de ações em mídias sociais Fábio Castro Gouveia	31
Ampliando as fronteiras da editoração científica: o papel dos repositórios de dados Luana Faria Sales, Luis Fernando Sayão	39
Um estudo para representação de dados abertos de editoração científica na web de dados Sandro Rautenberg, Lucélia de Souza	47
Web semântica e linked data como aliados na consolidação da Ciência Aberta José Eduardo Santarem Segundo	65
Quais direitos podem ser alegados sobre dados de pesquisa? Paulo Cezar Vieira Guanaes	79

Dados abertos e a Fundação de Amparo à Pesquisa do Estado de São Paulo Milton Shintaku, Suzana Caetano da Silva Lanne	87
Como os indexadores estão vendo a questão de dados abertos de pesquisa para o processo de indexação? Gildenir Carolino Santos	97
Análise sobre dados abertos e regulação autoral no contexto da editoria científica Rosilene Paiva Marinho de Sousa, Guilherme Ataíde Dias, Marckson Roberto Ferreira de Sousa	119
Direitos autorais e os dados de pesquisa abertos Enrique Muriel-Torrado, Rebeca Fernandes de Moura	137
Ciência Aberta, dados abertos e princípios FAIR: uma contribuição dos Países Baixos Patrícia Henning, João Moreira	145
As potencialidades do data paper na ciência atual Lucas Rocha	159
Gestão de dados em periódicos científicos Caterina Groposo Pavão, Fabiano Couto Corrêa da Silva, Lúcia da Silveira	165
Dados de pesquisa: disponibilização ou publicação? Emanuelle Torino, Sandra Milena Roa-Martínez, Silvana Aparecida Borsetti Gregorio Vidotti	183
A Ciência Aberta no CNPq: acesso aberto aos dados e resultados de pesquisas Adriana Maria Tonini, Paulo Henrique de Assis Santana, Thaís Machado Scherrer	203
O Alinhamento das revistas brasileiras sobre Medicina Tropical e Doenças infecciosas e Parasitárias às práticas da Ciência Aberta Benedito Barraviera, Adeilton Alves Brandão, Dalmo Correia, Rui Seabra Ferreira Júnior, Milton Shintaku	215
Política Editorial e os Dados Abertos Milton Shintaku, Rui Seabra Ferreira Junior	231

Agradecimentos

A colaboração entre estudiosos já foi indicada como uma das grandes mudanças na ciência ao final do século passado, com o fim do chamado autor único. Dessa forma, as organizações materializam tal colaboração, em que múltiplos autores constroem uma obra.

Nesse contexto, os autores deste livro agradecem:

À Associação Brasileira de Editores Científico (ABEC Brasil), por meio do seu Presidente, Prof. Rui Seabra Ferreira Junior, pela confiança depositada no desenvolvimento deste projeto;

Aos autores, que atenderam ao convite de forma rápida e eficaz, ofertando os preciosos conhecimentos em forma de capítulos;

À equipe de editoração, que normalizou, revisou e diagramou o livro;

À equipe da Embrapa, que tem atuado como coordenadora do Compromisso de nº 3 do 4º Plano de Ação Nacional em Governo Aberto¹, lançado em outubro de 2018, que visa “Estabelecer mecanismos de governança de dados científicos para o avanço da Ciência Aberta no Brasil”.

À *Open Government Partnership*¹ (OGP), uma iniciativa internacional que busca difundir e incentivar práticas governamentais para a transparência, o acesso à informação pública e a participação social, a qual – por meio dos Planos de Ação Nacionais – nos permitiu operacionalizar o movimento de Ciência Aberta no Brasil;

1 Disponível em: <https://www.opengovpartnership.org/>

Às instituições de todos os participantes do livro, aqui mencionados o Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT) e a Universidade de Brasília (UnB), representando a todos;

Aos que direta ou indiretamente nos ajudaram a organizar esta compilação, o nosso muito obrigado.

Os Organizadores

Prefácio

No ano de 2019, a Associação Brasileira de Editores Científicos (ABEC Brasil), devido à sua representatividade frente a mais de 2.500 periódicos científicos do Brasil foi convidada a participar do 4º Plano de Ação do Brasil para o Governo Aberto, coordenando a Meta 6, denominada: “Articulação com editores científicos para implantação de ações em apoio a Ciência Aberta”. O compromisso deste audacioso plano é o de se estabelecer mecanismos de governança de dados científicos para avanço da Ciência Aberta no país. Nesse sentido, a ABEC tem procurado articular e organizar ações que apoiem os editores científicos a adotarem as premissas da Ciência Aberta, especialmente em relação aos dados abertos de pesquisa em seus respectivos periódicos.

Assim, em junho de 2019, dentro do Curso de Editoração Científica, ocorrido em São Paulo, a palestra “Inovação e Governo Aberto na Ciência: mecanismos de governança de dados científicos para o avanço da Ciência Aberta no Brasil” foi ministrada seguida de uma ampla discussão. Naquele momento, o 4º Plano de Ação do Brasil para o Governo Aberto foi apresentado, sendo essa uma das primeiras ações da ABEC Brasil na realização da Meta 6. Em setembro de 2019, a associação realizou seu evento internacional: ABEC Meeting 2019. O AM2019 foi realizado nas dependências da UniFor na cidade de Fortaleza-CE, tendo como tema “Ciência Aberta”, com a disponibilização de várias palestras e oficinas tratando do tema. No evento, a ABEC Brasil lançou o livro Ciência Aberta para Editores Científicos, com textos que discorrem sobre a Ciência Aberta, voltados às equipes editoriais, atendendo aos compromissos da Associação com o Plano de Ação do Brasil para o Governo Aberto.

Como um dos produtos finais da execução da Meta 6, o presente livro aborda práticas de dados abertos direcionados à editoração científica, oferecendo uma ferramenta de apoio aos editores científicos na adoção das premissas dos dados abertos de pesquisa em suas políticas editoriais. O livro tem ainda por objetivo discutir tópicos que envolvam a adoção dos dados abertos de pesquisa por revistas científicas, o que já é uma realidade em muitos periódicos

internacionais, principalmente aqueles mantidos pelas grandes editoras científicas. Desse modo, a ABEC Brasil atende mais um compromisso assumido na liderança da Meta 6, voltada a articulação dessa temática na comunidade de editores científicos, oferecendo insumos para a melhoria dos periódicos brasileiros e sua adequação à Ciência Aberta.

Boa leitura!

Rui Seabra Ferreira Junior
Presidente da ABEC (2016-2020)



Rui Seabra Ferreira Junior

Lattes: <http://lattes.cnpq.br/5222987608775248>

ORCID: <https://orcid.org/0000-0001-6952-0512>

Presidente da ABEC; Editor Associado do Journal Venomous Animals and Toxins including Tropical Diseases; Pesquisador Associado do CEVAP-UNESP; Prof. Livre-Docente da FMB-UNESP; Pesquisador Pq1C do CNPq; Membro da Academia Brasileira de Ciências Farmacêuticas do Brasil/Academia Nacional de Farmácia.

Apresentação

Patrícia Rocha Bello Bertin
Juliana Meireles Fortaleza

É por meio dos relatos das pesquisas e das trocas que acontecem entre autores e seus pares que o conhecimento científico é validado e legitimado. O papel do editor científico é, portanto, central para o sistema científico global, o qual vem experimentando uma transformação importante na direção de um modo mais aberto de produção e comunicação da ciência.

O paradigma emergente vem sendo designado de “Ciência Aberta”, uma nova dinâmica de produção do conhecimento científico que se apoia sobre a abertura das fronteiras da ciência (Ciência Cidadã), dos métodos e ferramentas científicas, dos cadernos de pesquisa e, não menos importante, das publicações e dos dados de pesquisa. Recai sobre os editores científicos, nesse contexto, a responsabilidade de apoiar as novas formas colaborativas de produção, circulação e apropriação social do conhecimento.

No Brasil, a Parceria para Governo Aberto (do inglês Open Government Partnership) tem se constituído em um importante fórum para que os diversos atores do sistema científico possam debater sobre a gestão de dados de pesquisa no contexto da Ciência Aberta. Entre os 11 (onze) compromissos nacionais que integram o 4º Plano de Ação Nacional em Governo Aberto (http://governoaberto.cgu.gov.br/a-ogp/planos-de-acao/4o-plano-de-acao-brasileiro/4o-plano-de-acao-nacional_portugues.pdf), está o Compromisso nº 3, que visa “estabelecer mecanismos de governança de dados científicos para o avanço da Ciência Aberta no Brasil”.

O presente livro resulta dos esforços empreendidos pela Associação Brasileira de Editores Científicos (Abec Brasil) e parceiros engajados no Compromisso 3 do 4º Plano de Ação Nacional em Governo Aberto, na intenção de melhor compreender a dinâmica da Ciência Aberta e suas implicações sobre o cenário editorial científico nacional. Entre os temas desenvolvidos nesta publicação, estão: data papers e as novas possibilidades de editoração científica; gestão

e abertura de dados em uma variedade de contextos; o papel dos repositórios de dados; web de dados, web semântica e linked data na Ciência Aberta; propriedade intelectual e direito autoral para dados de pesquisa; a interface entre dados abertos e o processo de indexação; princípios FAIR na Ciência Aberta; políticas editoriais e de Ciência Aberta; entre outros tópicos atuais e de interesse da comunidade editorial científica.

Desejamos uma boa leitura e, mais ainda, que a reflexão suscitada pelas páginas a seguir possa produzir um maior engajamento da comunidade editorial nos fóruns de discussão sobre a Ciência Aberta e seus impactos sobre a comunicação científica no País².



Patrícia Rocha Bello Bertin

Lattes: <http://lattes.cnpq.br/4785200171802218>

Orcid: <https://orcid.org/0000-0001-5973-0305>

Biografia: PhD em Gestão da Informação pela Loughborough University, Reino Unido, Patrícia é pesquisadora da Embrapa 2002, atuante nas temáticas da gestão de dados e informações e da Ciência Aberta. Patrícia lidera o Grupo de Interesse em Dados Agrícolas da Research Data Alliance e coordena o compromisso brasileiro pela Ciência Aberta no IV Plano de Ação Nacional em Governo Aberto.



Juliana Meireles Fortaleza

Lattes: <http://lattes.cnpq.br/3704305861823318>

Orcid: <https://orcid.org/0000-0001-6272-7772>

Mestre em Engenharia Agrônoma, pesquisadora da Embrapa desde 2002, atuante nas temáticas da gestão de dados e informações e da Ciência Aberta e faz parte da equipe que coordenada o Compromisso 3 do 4º Plano de Ação Nacional em Governo Aberto.

² Para saber mais e acompanhar as ações do Compromisso pela Ciência Aberta no 4º Plano de Ação Nacional em Ciência Aberta, acesse a wiki: <https://wiki.rnp.br/display/OGPBrasil>.

Ciência Aberta, gestão de dados de pesquisa e novas possibilidades para a editoração científica

Luana Farias Sales
Michelli Costa
Milton Shintaku

1. INTRODUÇÃO

A pesquisa científica é executada conforme uma metodologia, a qual concede às informações resultantes do processo um *status* diferenciado de outras informações. Após a pesquisa ser feita, é preciso comunicar esses resultados em formatos padrão, como livros, artigos ou artigos de evento, enriquecendo o acervo com informações científicas existentes. Em alguns casos, esses resultados podem ser aplicados a fim de melhorar algum produto ou processo.

Todo esse processo pode ser denominado de comunicação científica, sobre o qual Bjork (2007) afirma que a pesquisa científica deve gerar novos conhecimentos e melhorar a qualidade de vida. Para o autor, as pesquisas são fomentadas pela curiosidade dos estudiosos, por problemas científicos ainda não solucionados e por iniciativas econômicas. Evidentemente se requer um ambiente propício para que a pesquisa seja efetuada e obtenham-se resultados apropriados.

A comunicação científica, em sua totalidade, envolve muitos profissionais, pois há etapas que são independentes do estudioso ou equipe que executa a pesquisa. A disseminação dos resultados de pesquisa, por exemplo, envolve

atividades editoriais, nas quais o autor tem papel secundário, tendo o editor como principal ator, visto que ele coordena todo o processo editorial, articulando as atividades dos envolvidos na editoração.

Entretanto, muitos editores no Brasil também são pesquisadores ou possuem grande conhecimento das formas de pesquisa efetuadas nas áreas de abrangência da revista. Independentemente do caso, se é somente editor ou se é também pesquisador, é preciso manter-se atualizado às grandes mudanças do mercado editorial fomentadas principalmente pela evolução tecnológica, que tornou o processo editorial mais dinâmico, e grande parte dos resultados de pesquisa são publicados somente em formato digital. No que tange à editoração científica, é preciso ainda acompanhar o desenvolvimento das novas formas do fazer científico.

E por falar em novos métodos científicos, é preciso jogar a luz sobre a eScience, que é uma forma de fazer científico totalmente pautada no uso e reúso de dados. Propiciada pelo progresso da tecnologia, ela permite que a ciência avance mais rapidamente e de modo mais eficiente rumo a novas descobertas e inovações. Atualmente, muitas áreas têm como insumo científico central dados de pesquisa, os quais são gerados de maneira colaborativa e compartilhados de acordo com regras específicas de cada domínio.

Soma-se à eScience um movimento internacional denominado Ciência Aberta, que defende o acesso aberto não apenas às publicações científicas, mas também aos dados gerados, às metodologias, aos softwares, aos cadernos de laboratórios e até aos resultados negativos – insumos até pouco tempo negligenciados pela ciência, mas que agora se tornam essenciais para garantir, entre outras coisas, a reprodutibilidade da ciência e a economia de recursos.

Assim, a Ciência Aberta se apresenta como um grande guarda-chuva, englobando várias iniciativas que impactam não somente a disseminação dos resultados em artigos, como no caso da editoração científica, mas também iniciativas que dizem respeito às atividades intermediárias de pesquisa, como os cadernos de pesquisa ou resultados negativos. Do mesmo modo, o movimento aponta para outros desdobramentos, como a Ciência Cidadã e recursos educacionais abertos, que tradicionalmente também não estão contemplados nos modelos de comunicação científica tradicionais.

A Ciência Aberta apresenta-se como um desafio a todos os profissionais que estão envolvidos com a pesquisa, na medida em que altera atividades

tradicionalmente estabelecidas, ou mesmo criando novas. Nesse sentido, verifica-se a necessidade de disseminar amplamente todas as iniciativas relacionadas à Ciência Aberta para os profissionais que são direta ou indiretamente afetados pelas mudanças propostas, como no caso das equipes editoriais, que podem não participar da pesquisa, mas se constituem como atores de extrema importância no amplo processo da comunicação científica.

Devido à amplitude do tema Ciência Aberta, torna-se mais eficaz tratar de temas mais específicos dentro de um contexto, como, por exemplo, os dados abertos de pesquisa no processo editorial. Assim, podem-se verificar todas as relações possíveis entre o movimento de dados abertos e a editoração científica, visto que os dados abertos têm sido discutidos no âmbito dos processos editoriais.

2. DADOS NA CIÊNCIA

As mudanças ocorridas na ciência pelo uso das tecnologias digitais afetam a percepção dos seus processos. Sayão e Sales (2014), por exemplo, argumentam que os dados brutos de pesquisa ganham significância própria, transcendendo o papel intermediário e subproduto do processo científico, revelando o seu potencial informacional. Assim, os dados brutos de pesquisa precisam estar disponíveis em plataforma aberta para reúso e preservação.

Dados de pesquisa, como vêm sendo chamado na literatura científica brasileira, podem ser definidos como “todo e qualquer tipo de registro coletado, observado, gerado ou usado pela pesquisa científica, tratado e aceito como necessário para validar os resultados da pesquisa pela comunidade científica” (SALES, SAYÃO, 2019, p. 36). Esses dados de investigação – termo usado em Portugal – podem ser analisados sob diversas perspectivas. Nesta obra, escolhemos trabalhar sob a perspectiva dos dados abertos, atendendo a uma demanda do Compromisso de número 3 do 4º Plano de Ação Nacional em Governo Aberto, lançado em outubro de 2018, o qual visa estabelecer mecanismos de governança de dados científicos para o avanço da Ciência Aberta no Brasil e no qual se encontra um marco específico voltado para editores: o Marco 5, que consiste na articulação com agências de fomento para a implantação de ações de apoio à Ciência Aberta.

Já o Relatório da Organização para Cooperação e Desenvolvimento Econômico (OCDE) apresenta “registros factuais usados como fonte primária para a

pesquisa científica e que são comumente aceitos pelos pesquisadores como necessários para validar os resultados do trabalho científico”. Entretanto, essa definição não explicita todas as possibilidades que os dados abertos podem assumir, visto que depende, em muito, da metodologia utilizada na coleta, geração ou reúso. Assim, podem ser dados de coletas automatizadas ou manuais, resultantes de observações ou simulações, controladas ou não, com uma grande gama de possibilidades, mostrando mais uma vez as diversas facetas que os dados de pesquisa podem assumir e agregando certa complexidade ao tema.

Muitas vezes, dados de pesquisa são associados a coletas quantitativas, por gerarem bases de dados com grande número de registros. Entretanto, como relatam Chauvette, Schick-Makaroff e Molzahn (2019), a abertura dos dados pode ocorrer com dados de pesquisas qualitativas, que possuem características próprias, mas que não as impedem de ser compartilhadas, tendo o amparo de questões epistemológicas, metodológicas e éticas. A ética deve ser o maior guia do compartilhamento de dados de pesquisa, desenvolvendo políticas adequadas ao dados que, em muitos casos, possuem aspectos restritivos. É preciso considerar níveis de compartilhamentos para dados de pesquisa, para que o que precise ser resguardado se mantenha dessa forma. Assim, é preciso atender ao já clássico lema da Ciência Aberta no que tange à gestão de dados de pesquisa, o qual diz: “Tão aberto quanto possível, tão fechado quanto necessário” (EUROPEN COMMISSION, 2016, p. 4).

3. GESTÃO DE DADOS E EDITORAÇÃO CIENTÍFICA

Na perspectiva dos profissionais de informação, a gestão de dados de pesquisa é entendida como um conjunto de atividades gerenciais e tecnológicas, apoiadas por políticas gerais e específicas destinadas a garantir arquivamento, curadoria, preservação e oferta de acesso contínuo aos dados de pesquisa. Na perspectiva dos pesquisadores, a gestão de dados é compreendida como um conjunto de atividades que envolve a geração, o manuseio, o armazenamento e a descrição dos dados. Já para as agências de fomento, a gestão de dados é encarada como um processo de acompanhamento da pesquisa que está sendo desenvolvida com recursos públicos por meio da supervisão dos dados que estão sendo gerados pela pesquisa. As agências encaram os dados ainda como insumos importantes para o desenvolvimento de novas pesquisas, e por isso

vêm solicitando, cada vez mais, planos de gestão de dados que indiquem, entre outras coisas, a destinação dos dados gerados pelas atividades de pesquisa.

Para editores científicos, dados de pesquisa também possuem um papel fundamental, que é garantir a verificação da pesquisa, bem como a reprodutibilidade dos experimentos descritos. Nesse sentido, a gestão de dados de pesquisa pode funcionar como um mecanismo de auxílio aos avaliadores e editores de periódicos.

Somam-se a isso diversas possibilidades de inovar os formatos das publicações científicas que surgem a partir da ligação dos dados de pesquisa aos artigos, entregando aos leitores não apenas o resultado final da pesquisa, mas toda a memória da atividade, dando lugar ao modelo denominado publicação ampliada.

No entanto, para que esses novos formatos sejam possíveis e eficientes, uma gestão de dados cuidadosa deve ser realizada durante todo o ciclo da pesquisa científica. Assim, uma série de princípios vem sendo sugerida pela comunidade de estudiosos sobre a temática Ciência Aberta e gestão de dados, que na sua totalidade é denominada de Princípios FAIR.

Gerenciar dados, tornando-os FAIR ou simplesmente atendendo aos princípios FAIR, significa resumidamente tornar esses dados encontráveis, acessíveis, interoperáveis e reusáveis.

Outro conjunto de princípios interessantes para serem aplicados aos dados de pesquisa no âmbito da Ciência Aberta são os Princípios Panton, a respeito dos quais Maloy (2011) defende que, com apenas poucas exceções, todos os dados gerados em pesquisas devem ser publicados juntamente com os resultados de forma pública. Nesse sentido, destaca-se a renúncia às restrições de acesso e uso, com a disponibilização dos dados em canais de disseminação públicos e com declaração robusta e explícita sobre o uso.

Por fim, apoiando-se em Maloy (2011), destacam-se aqui alguns Princípios Panton que podem servir de apoio aos editores científicos, a saber:

- a. Declaração dos editores:** aos dados ou coleções de dados a serem disponibilizados de forma aberta deve ser agregada uma declaração clara e precisa – apresentando os desejos e expectativas de quem gerou os dados, quanto ao uso ou redestinação – alinhada a questões éticas e jurídicas;

- b. Licenciamento:** cuidado com algumas licenças comumente utilizadas, mas que restringem o reúso amplo dos dados ou coleções de dados, como no caso de Gnu Free Documentation Licence (GFDL), na medida em que mantêm alguns direitos autorais;
- c. Abertura:** utilizar licenças que abram todas as possibilidades, inclusive comerciais ou de obras derivadas, alinhadas à ideia de conhecimento livre;
- d. Licença CC 0 (zero):** encoraja-se o uso da licença *Creative Commons Zero* (CC 0), associando os dados ou coleção ao domínio público, renunciando aos direitos autorais e direitos conexos, assegurados legalmente.

4. CONSIDERAÇÕES FINAIS

A comunicação científica, entendida como o conjunto dos macroprocessos de produção, distribuição e uso da informação, cumpre a importante missão de tornar público o conhecimento resultante das atividades científicas. Os instrumentos para operacionalizar a comunicação científica são objetos de constantes evoluções, resultantes das transformações técnicas e sociais da humanidade. O advento dos periódicos digitais e a modernização dos seus processos refletem as possibilidades tecnológicas do momento e as demandas por mais acesso, alcance e rapidez na publicação dos resultados das atividades científicas. No entanto, ao mesmo tempo que é reflexo de determinadas condições, também é combustível para as novas mudanças.

No bojo das atuais transformações situam-se as iniciativas em prol da Ciência Aberta, que buscam garantir mais transparência, possibilidades de acesso e uso do conhecimento científico. De forma ampla, tais iniciativas movimentam-se para promover a democratização no acesso à ciência e aproximações com a sociedade como um todo, reafirmando e reinventando formas de produzir, divulgar e utilizar o conhecimento científico.

Por maior acesso ao conhecimento, as demandas científicas foram internacionalmente articuladas no início dos anos 2000 a partir de iniciativas que buscaram promover a ampla disponibilização das publicações científicas em acesso aberto. A reverberação desses processos nas comunidades acadêmicas, agências de fomento e editoras científicas ajuda a explicar a ampla adesão aos princípios do acesso aberto as publicações científicas. Como continuação desse movimento de promoção da abertura da ciência,

tem tomado forma e voz às reivindicações para abertura dos dados de pesquisa, bem como os recursos necessários para analisá-los.

A abertura dos dados de pesquisa visa promover melhores condições para avaliação e reuso dos resultados e métodos mais céleres e colaborativos para a produção e distribuição do conhecimento científico. A coerência da abertura dos dados de pesquisa com os princípios basilares da ciência moderna foram anunciados ainda no século XVII, por Oldenburg, em sua proposta de divulgação dos dados e/ou evidências juntamente com as publicações da *Royal Society* (BOULTON, 2013). Os argumentos que fundamentam a proposta são orientados pelo compromisso de produzir conhecimento transparente, refutável e publicamente disponível a quem desejar.

Na contemporaneidade, as editoras científicas são novamente convocadas a contribuir com a democratização da ciência, a partir da organização de soluções que promovam a abertura dos dados de pesquisa. As soluções ora apresentadas são diversas, flexíveis e estão em contínuo processo de construção coletiva.

5. REFERÊNCIAS

BJÖRK, B-C. A model of scientific communication of a global distributed information system. **Information Research**, v. 12, n. 2, 2007. Disponível em: <http://informationr.net/ir/12-2/paper307.html>. Acesso em 21 jan. 2020.

SALES, LF; SAYÃO, L.F. Uma proposta de taxonomia para dados de pesquisa. **Conhecimento em Ação**, Rio de Janeiro, v. 4, n. 1, jan/jun. 2019

CHAUVETTE, A.; SCHICK-MAKAROFF, K.; MOLZAHN, A. E. Open Data in Qualitative Research. **International Journal of Qualitative Methods**, v. 18, p. 1609406918823863, 2019.

EUROPEN COMISSION. **Guidelines on FAIR Data Management in Horizon 2020**. [S.l.], 2016. Disponível em: http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf Acesso em: 26 jun. 2019.

MOLLOY, Jennifer C. The open knowledge foundation: open data means better science. **PLoS biology**, v. 9, n. 12, p. e1001195, 2011.

OECD. **OECD Principles and Guidelines for Access to Research Data from Public Funding**. OECD, 2007. Disponível em: <http://www.oecd.org/sti/inno/38500813.pdf>. Acesso em: 01 jan. 2020.

6. SOBRE OS AUTORES



Luana Farias Sales

Lattes: <http://cnpq.br/9090064478702633>

Orcid: <http://orcid.org/0000-0002-3614-2356>

Doutora em Ciência da Informação pelo Programa de Pós-Graduação do IBICT/UFRJ (2011-2014). Mestre em Ciência da Informação pelo convênio UFF/IBICT (2004-2006), Graduação em Biblioteconomia e Documentação pela Universidade Federal Fluminense (2003). Atualmente exerce o cargo de Coordenadora-Geral de Acesso e Difusão de Acervos do Arquivo Nacional. É docente do Programa de Pós-graduação em Ciência da Informação do convênio IBICT-UFRJ e Coordenadora da Rede de Implementação do GO FAIR Brasil.



Michelli Costa

Lattes: <http://lattes.cnpq.br/5676915146365703>

Orcid: <https://orcid.org/0000-0002-4789-7623>

Doutora e mestre em Ciência da Informação e bacharel em Biblioteconomia pela Universidade de Brasília. Tem experiência na área de Ciência da Informação, atuando principalmente nos seguintes temas: Ciência Aberta, dados de pesquisa, acesso aberto, bibliotecas digitais, bibliotecas populares e bibliotecas prisionais. É coordenadora de extensão e professora da Faculdade de Ciência da Informação da Universidade de Brasília.



Milton Shintaku

Lattes: <http://lattes.cnpq.br/8605833104600600>

Orcid: <https://orcid.org/0000-0002-6476-4953>

Mestre e doutor em Ciência da Informação pela Universidade de Brasília (UnB), coordenador de Articulação, Geração e Aplicação de Tecnologia do Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict), Secretário da Associação Brasileira de Editores Científicos (ABEC) 2017-2019.

7. COMO CITAR

SALES, L. F.; COSTA, M.; SHINTAKU, M. Ciência aberta, gestão de dados de pesquisa e novas possibilidades para a editoração científica. *In*: SHINTAKU, M.; SALES, L. F; COSTA, M. (org). **Tópicos sobre dados abertos para editores científicos**. Botucatu, SP: ABEC, 2020. p. 13-21. DOI: 10.21452/978-85-93910-04-3.cap1.

Políticas de Ciência Aberta e a abertura dos dados de pesquisa

Michelli Costa

A Ciência Aberta é uma forma renovada de produzir, divulgar e usar conhecimento científico. Devido a sua ambição e complexidade, é comum categorizar suas diferentes iniciativas em vertentes. Um elemento que perpassa todas as dimensões dessa ciência são as políticas. Elas são estratégias e consistem em iniciativas para a promoção dos objetivos da Ciência Aberta. Além disso, são declaradas em documentos normativos de países e instituições que fomentam as discussões e as práticas da Ciência Aberta. Os documentos apresentam-se com relevância singular no cenário estratégico na medida em que estabelecem o interesse público e/ou institucional e regulamentam ou orientam a destinação de recursos e as formas de atuação.

As políticas não são apenas regulatórias dos processos que as envolvem, elas são representativas dos paradigmas e valores dos atores que conduzem tais processos em determinados contextos. Albornoz *et al.* (2018) salientam que o contexto no qual se inserem as discussões sobre a Ciência Aberta são marcados por desigualdade histórica e global nos processos de produção e distribuição do conhecimento científico. Os autores ainda ressaltam que as estruturas científicas desiguais “centralizam e privilegiam o conhecimento produzido por pesquisadores do sexo masculino anglo-americanos da América do Norte e Europa e compartilhado por meio de revistas acadêmicas internacionais”³. Com isso, evidencia-se que as políticas para a Ciência Aberta são representativas dos cenários em que estão inseridas e, portanto, não são neutras. Elas precisam ser analisadas e debatidas.

3 Tradução livre.

Para problematizar os elementos relacionados às políticas de promoção da Ciência Aberta, Albornoz *et al.* (2018) analisaram os elementos discursivos de 49 documentos publicados nos cinco continentes por diferentes tipos de instituições, tais como governo, organismos internacionais, associações científicas, universidades e editoras científicas. Os grupos editoriais presentes na análise foram a Elsevier e a *International Association of STM Publishing*. As políticas de ambas foram publicadas em 2017 e incluem orientações para a abertura dos dados vinculados às suas publicações.

Apesar da aparente receptividade de grandes grupos editoriais para os objetivos Ciência Aberta, Albornoz *et al.* (2018) sugerem que as discussões presentes nas políticas enfatizam as iniciativas de Ciência Aberta como forma de controle da qualidade da ciência, sem que os parâmetros da qualidade sejam questionados. A contradição evidenciada revela que os padrões são constituídos nos modelos tradicionais da ciência, que podem privilegiar grupos editoriais comerciais e/ou interesses regionais.

Os interesses regionais podem ser diversos e representar projetos políticos ou econômicos díspares. Ademais, traduzem as demandas locais e os entendimentos acerca dos objetivos e meios para realização da Ciência Aberta. O relatório publicado pela *Digital Curation Centre (DCC)* e *Scholarly Publishing and Academic Resources Coalition (SPARC)* demonstrou a disparidade e as tendências na consolidação de políticas nacionais no continente europeu. O relatório teve como base o levantamento realizado pelo DCC e SPARC em 2016, e foi atualizado em 2018. No documento são indicados 23 países europeus que possuem políticas nacionais para a Ciência Aberta ou que pelo menos promovem iniciativas em âmbito nacional para a promoção da abertura dos dados de pesquisa. Cabe ressaltar que o estudo privilegiou a cobertura dos dados de pesquisa nas políticas analisadas, ou seja, todas as políticas de Ciência Aberta contemplavam os dados de pesquisa. Isso porque, ao tratar do tema, são acionadas outras estruturas científicas, como publicações, softwares e metodologias.

Segundo o relatório, as políticas nacionais para promoção da Ciência Aberta iniciaram-se no continente europeu em 2019 a partir da inclusão da temática no “Código de ética para pesquisa científica na Bélgica”, assinada pelo governo federal do país. Além de ter sido o primeiro documento desse governo na região a orientar sobre a abertura dos dados sua abordagem é singular. Nele é salientado que os dados de pesquisa devem estar livremente disponíveis para qualquer pessoa interessada, com o mínimo de restrições possível. A abertura dos dados é justificada pelos princípios científicos de transparência, colaboração e refutabilidade.

Em 2010 é publicada uma política de agência de fomento de pesquisa nacional que requer abertura das publicações resultantes dos projetos financiados, bem como os dados de pesquisa, os softwares utilizados e a metodologia. A partir de 2014 observa-se um elevado crescimento na elaboração de documentos dessa natureza, especialmente motivados pelas orientações do Programa Horizonte 2020. O programa, que é destinado à promoção da ciência na União Europeia, tem se mostrado forte promovedor das discussões e ações em torno das políticas para a Ciência Aberta. As políticas que surgem a partir daí, já com uma consciência mais amadurecida e homogeneizada dos valores e objetivos dessa ciência, podem ser classificadas em três categorias:

- Lei – firmada pelo governo nacional. Estabelece os direitos e indica as responsabilidades. Comumente, desdobra-se em planos nacionais (que devem incluir os procedimentos e as formas de monitoramento).
- Políticas de agências de fomento – promovidas pelos conselhos nacionais ou regionais de pesquisa, em alguns casos com universidades e agências privadas.
- Plano nacional – promovido pelo governo federal (ministérios). Estabelece condições e estratégias para abertura das informações. Vincula-se ao Programa H2020.

De modo geral, o relatório aponta dois resultados interessantes para considerar o contexto de emergência das políticas de Ciência Aberta no continente europeu. A primeira delas é a classificação acerca do tipo de política, que pode variar em função da sua força. Políticas classificadas como fortes deveriam ter caráter mandatório e apresentar formas de monitoramento, enquanto as classificadas como suaves apresentaram apenas recomendações. O segundo resultado diz respeito à influência do projeto europeu de financiamento da ciência como um dos principais estímulos para as discussões no continente.

Outra força impulsionadora de políticas nacionais para a Ciência Aberta tem se constituído no âmbito da *Open Government Partnership* (OGP) e extrapola os limites da Europa. A OGP é uma articulação internacional, fundada em 2011 por oito países e atualmente conta com a participação de 79. Seu objetivo é promover nas práticas governamentais a democratização no acesso à informação, ampliar a participação social, estimular a governança e responsabilidade pública, além de tecnologias e inovações para a abertura e governança (OGP, 2016). A atuação da OGP nesses países operacionaliza-se por meio de

um plano nacional que é executado a partir de compromissos firmados entre o governo e a sociedade civil.

Dos quase 4 mil compromissos disponíveis na ferramenta OGP Explore, Costa e Braga (2019) analisaram sete que abordaram com prevalência a temática da Ciência Aberta. Todos os compromissos selecionados tratavam de algum nível dessa ciência, entretanto as dimensões contempladas, a abrangência das ações propostas e dos resultados obtidos foram ímpares⁴.

O mais abrangente deles foi o compromisso firmado pelos Estados Unidos e executado entre 2015 e 2017. No período, a agência governamental responsável pela ciência e tecnologia do país se articulou para que todas as agências federais que investissem mais de 100 milhões de dólares por ano em pesquisa implantassem programas para tornar acessíveis e reutilizáveis as publicações e os dados resultantes de pesquisa.

Ainda com uma perspectiva ampla das dimensões da Ciência Aberta, identificaram-se os compromissos firmados pelo Canadá e pela França em 2018, com previsão de encerramento em 2020. De forma geral, ambos se comprometem a maximizar o acesso às informações e aos dados oriundos de pesquisa científica financiada pelo poder público. Adicionalmente, declaram o interesse em criar mecanismos de participação cidadã na ciência.

Os outros países, autores dos documentos localizados, apresentaram uma noção mais restrita da Ciência Aberta. São eles: Albânia, Gana, República Eslovaca e Romênia. A restrição é marcada pela limitação das ações relacionadas às publicações científicas e, em alguns casos, de maneira superficial aos dados. Ademais, salienta-se que não foi possível localizar os resultados das articulações nacionais providas pelos quatro países.

De maneira geral, aponta-se que os sete compromissos identificados propuseram ações nacionais para o desenvolvimento de, pelo menos, uma categoria da Ciência Aberta. As categorias mais representativas desse conjunto foram aquelas relacionadas à promoção do acesso aberto às publicações científicas e abertura dos dados de pesquisa. Outras dimensões, não menos relevantes

4 Os resultados completos do levantamento foram descritos no Relatório do Panorama Preliminar do Desenvolvimento da Ciência Aberta, disponível em: <https://wiki.rnp.br/pages/viewpage.action?pageId=108957259>.

para os objetivos da Ciência Aberta, como a Ciência Cidadã e a avaliação aberta, se mostraram menos expressivas no grupo analisado.

Mais que frentes de atuação diferentes, as categorias representam correntes filosóficas diversas. As diferenças podem ser oriundas das regiões geográficas, dos interesses envolvidos ou do ponto de maturação das discussões colocadas. No entanto, Albornoz *et al.* (2018) chamam a atenção para o fato de que a limitação da discussão em torno das publicações e dados de pesquisa podem ser prejudiciais aos objetivos da Ciência Aberta. Isso porque sua proposta original é modificar as estruturas científicas, o que requer a participação de diversos elementos e atores.

1. APONTAMENTOS PARA REFLEXÃO ACERCA DAS POLÍTICAS DE CIÊNCIA ABERTA

A análise dos diferentes contextos de emergência das discussões e iniciativas para a promoção da Ciência Aberta, bem como a conscientização dos seus elementos discursivos, nos apontam cinco lições para reflexão.

Primeiro, a Ciência Aberta é um projeto de reestruturação da ciência em andamento. Não nos cabe escolher sermos atingidos ou não pelas propostas dessa ciência, pois elas já estão em curso. Tais propostas representam um projeto no campo acadêmico e na relação do setor com a sociedade, setor produtivo e governo. Ainda não está definido se as mudanças serão reformistas ou revolucionárias, mas certamente gerarão impactos nas estruturas científicas. Portanto, nos cabe decidir qual papel assumiremos nesse processo e como influenciaremos nas propostas em construção.

Segundo, a Política Nacional de Ciência Aberta é um marco importante da soberania nacional que se desdobra na ciência, tecnologia e demais interesses sociais. No entanto, para se apresentar como um forte instrumento de promoção da Ciência Aberta, essa política deve ter caráter mandatório, estabelecer responsabilidades e prever formas de monitoramento.

Terceiro, a Política Nacional de Ciência Aberta deve estabelecer garantias acerca da infraestrutura. No entanto, a proposta de infraestrutura deve estar adequada ao contexto em que se insere e não apenas aos padrões internacionais, que por vezes refletem interesses antagônicos ao da ciência em questão. Tais

condições são necessárias para colocar os pesquisadores brasileiros em condições de colaboração internacional e não de dependência, provendo assim condições para superar, ao menos em partes, os desenhos desiguais e injustos da geografia da ciência.

Quarto, o Brasil tem especificidades e potencialidades para influenciar no projeto de Ciência Aberta em curso. Assim como aconteceu – e tem acontecido – com as propostas de acesso aberto às publicações científicas, o Brasil tem potencial para assumir protagonismo e delinear nossas próprias formas de atuação efetivamente coerentes com os objetivos da Ciência Aberta.

Quinto, as discussões acerca da abertura dos dados de pesquisa têm se mostrado como um marco relevante da ampliação e do amadurecimento dos projetos de Ciência Aberta. Sua relevância é estabelecida a partir da sua articulação em uma rede complexa de elementos e atores, especialmente as instituições produtoras de conhecimento científico, as agências de fomento e os editores.

2. REFERÊNCIAS

ALBORNOZ, D. *et al.* Framing Power: tracing key discourses in open science policies. *In: ELPUB 2018. Anais [...]*. Toronto: Elpub, 2018. Disponível em: <https://elpub.episciences.org/4612/pdf>. Acesso em: jan. 2020.

COSTA, M.; BRAGA, T. Compromissos para a promoção da Ciência Aberta no âmbito da Parceria para Governo Aberto. *In: IX ENCONTRO IBÉRICO EDICIC, 9., 2019. Anais [...]*. Barcelona: EDICIC, 2019. Disponível em: <http://eprints.rclis.org/38819/>. Acesso em: 22 jan. 2020.

DIGITAL CURATION CENTER (DCC); SCHOLARLY PUBLISHING AND ACADEMIC RESOURCES COALITION (SPARC). **An analysis of open data and open science policies in Europe**. Reino Unido: DCC; SPARC 2018. [Relatório]. Disponível em: <https://sparceurope.org/new-sparc-europe-report-analyses-open-data-open-science-policies-europe/>. Acesso em: jan. 2020.

OPEN GOVERNMENT PARTNERSHIP (OGP). **3º Plano de Ação Nacional do Brasil**. 2016. Disponível em: https://www.opengovpartnership.org/wp-content/uploads/2001/01/Brazil_Plano-de-Acao-3_2016-2018.pdf. Acesso em: 22 jan. 2020.

3. SOBRE OS AUTORES



Micheli Costa

Lattes: <http://lattes.cnpq.br/5676915146365703>

Orcid: <https://orcid.org/0000-0002-4789-7623>

Doutora e mestre em Ciência da Informação e bacharel em Biblioteconomia pela Universidade de Brasília. Tem experiência na área de Ciência da Informação, atuando principalmente nos seguintes temas: Ciência Aberta, dados de pesquisa, acesso aberto, bibliotecas digitais, bibliotecas populares e bibliotecas prisionais. É coordenadora de extensão e professora da Faculdade de Ciência da Informação da Universidade de Brasília.

4. COMO CITAR

COSTA, M. Políticas de Ciência Aberta e a abertura dos dados de pesquisa. *In*: SHINTAKU, M.; SALES, L. F; COSTA, M. (org). **Tópicos sobre dados abertos para editores científicos**. Botucatu, SP: ABEC, 2020. p. 23-29. DOI: 10.21452/978-85-93910-04-3.cap2.

Abrindo os dados de acesso e de ações em mídias sociais

Fábio Castro Gouveia

1. A CAIXA FECHADA DOS DADOS WEB

A abertura de dados científicos vem se tornando uma demanda de pesquisadores, agências de fomento e governos no mundo todo. Por outro lado, dados de relações entre sítios web, e mais recentemente das mídias sociais, vêm passando por um processo de fechamento e uso comercial. No caso da web, como afirma Gouveia (2012), isto gerou uma crise nos estudos webométricos e conseqüente uma busca por novos caminhos, os quais não se concretizaram como uma continuidade do campo, mas indiretamente alimentaram o interesse em outros focos de pesquisa, como as mídias sociais.

Ao longo dos anos, ocorreram tentativas de correlacionar métricas web às mídias sociais (VAUGHAN; HYSEN, 2002; NEYLON; WU, 2009; PRIEM; HEMMINGER, 2010). De um lado, podemos falar das chamadas métricas ao nível de artigo (*Article Level Metrics – ALM*), que desembocaram no manifesto alt-métrico apresentado por Priem *et al.* (2010), sendo mais à frente retomadas como conceito independente. De outro, informações sobre a repercussão em mídias sociais – a chamada “atenção on-line” – de produções científicas que

ainda são dependentes dos limites impostos pelas plataformas, bem como do trabalho de agregadores de dados, como a Plum Analytics⁵ e a Altmetric⁶.

Informações sobre o acesso (visualização e/ou download) aos artigos nos seus sítios de publicação ou armazenamento são oferecidas tanto por *publishers* de acesso aberto, como a PLOS⁷ e SciELO⁸, quanto em implementações de Repositórios Institucionais. No caso da PLOS, também são fornecidos dados de incorporação ao gerenciador de referências Mendeley⁹ e CiteULike¹⁰ (enquanto o serviço existia), ao número de citações na Scopus¹¹ (uma métrica tradicional), além de dados do Google Scholar¹² e da Crossref¹³, entre outras fontes, e o somatório de “atividades” no Twitter, Facebook, Reddit¹⁴, Wikipedia e comentários na página da revista sobre o artigo. Já o SciELO possui todo um conjunto de dados abertos sobre o acesso aos seus artigos – a SciELO Analytics¹⁵ – apresentado em gráficos e com a possibilidade de análises numa ferramenta ainda considerada em estágio beta.

Apesar dessas iniciativas estruturadas, e as formas mais diversas que se apresentam nos Repositórios Institucionais e em algumas implementações

5 Uma empresa com agregador altmétrico chamado *Plum X* e ligada a *Elsevier* acessível no link: <https://plumanalytics.com/>

6 Um agregador de dados altmétricos da *Digital Science* acessível no link: <http://www.altmetric.com>

7 A PLOS (<https://www.plos.org/>) edita um *mega journal* (*PLOS ONE*) e seis revistas (*Biology*, *Computational Biology*, *Neglected Tropical Diseases*, *Medicine*, *Genetics* e *Pathogens*), dentre outros canais.

8 SciELO – *Scientific Electronic Library Online*, dentre outras funções, é um *publisher* digital de acesso aberto resultado da parceria entre a Fundação de Amparo à Pesquisa de São Paulo (FAPESP) e o Centro Latino-Americano e do Caribe de Informação em Ciências da Saúde (Bireme).

9 Gerenciador de referências online, hoje parte da Elsevier, acessível no link: <http://mendeley.com/>

10 Plataforma de bookmarking social que encerrou suas atividades em março de 2019.

11 Base bibliográfica de resumos e citações da Elsevier. Acessível no link: <http://www.scopus.com>

12 Base indexadora de literatura acadêmica da Google. Acessível no link: <http://scholar.google.com/>

13 Que contém uma base de citações abertas pelo projeto I4OC (*Initiative for Open Citations* – Iniciativa por Citações Abertas)

14 Sítio agregador de mídia e discussão social acessível no link: <https://www.reddit.com/>

15 Disponível em: <https://analytics.scielo.org/>

de gerenciadores de revistas como o Sistema Eletrônico de Editoração de Revistas - SEER (Versão customizada pelo IBICT do OJS¹⁶), os dados de cunho webmétrico nem sempre são obtidos facilmente ou de maneira organizada, tanto por leigos interessados, quanto por autores ou mesmo pesquisadores de estudos métricos. Além desses públicos, editores podem desejar ter acesso a dados de outras revistas para servir de *benchmark* nos seus processos internos avaliativos. Porém, o acesso aos dados é só uma das questões quando se necessita trabalhar com diferentes fontes.

2. WEBMETRIA: DADOS DE ACESSO E DOWNLOADS

Um dos desafios para os dados webmétricos é que eles podem sofrer influências do design e da plataforma onde as páginas e conteúdos são disponibilizados. Em implementações simples do Google Analytics¹⁷, popular ferramenta de *page-tagging*¹⁸ para monitoramento de acesso, não são obtidos dados de acesso direto a documentos PDF (*Portable Data Format*) via Google, bem como o acompanhamento das ações dos usuários nas páginas ao não se configurar os eventos que serão observados. Distorções, principalmente na *bounce rate* (mal traduzida “para o” português “como” “taxa de rejeição”), podem também levar a percepções equivocadas.

Assim, obter dados sólidos e bem consolidados dos *downloads* e visualizações dos documentos é o primeiro objetivo a ser alcançado. É importante ainda ter em mente que uma mesma produção pode estar sendo disponibilizada em mais de um local na *web*, ainda que em diferentes versões (*preprints*, por exemplo), e que mais de um formato frequentemente é oferecido dentro do sítio das editoras (html e pdf, por exemplo), além de versões em diferentes idiomas.

16 *Open Journal Systems*, um sistema aberto de gerenciamento de revistas desenvolvido pelo *Public Knowledge Project* - PKP da University of British Columbia. Acessível no link: <http://pkp.sfu.ca/ojs>

17 Acessível no link: <http://analytics.google.com>

18 *Page-tagging* é o método de registro de solicitações aos servidores de conteúdo a partir da marcação das páginas oferecidas e execução de um código para obtenção de informações técnicas sobre quem está solicitando aquele conteúdo.

Como visto anteriormente, estas questões são corretamente contempladas nos dados oriundos da PLOS e da SciELO. Por outro lado, nos casos dos Repositórios e sítios próprios de revistas, podemos ter diferenças no modo como é atribuído a métrica de acesso ao conteúdo.

3. ALTMETRIA: MÍDIAS SOCIAIS E ATENÇÃO ON-LINE

Quando buscamos dados de atenção on-line, encontramos o problema da falta de padronização maior, pois diferentes agregadores têm diferentes formas de obter dados. As diferenças se dão pela definição dos endereços que correspondem ao documento (DOI, URL da revista, PMID, entre outros) e por conta dos contratos de acesso aos dados das plataformas fechadas. Assim, aqui se estabelece uma nova crise de acesso a dados, em linha com o ocorrido à Webometria, porém agora no seu campo derivado, a Altmetria. O Facebook é um exemplo deste processo de paulatina restrição ao acesso de dados, primeiro fechando às postagens públicas de usuários, depois às de grupos e por último às das páginas. Hoje, apenas empresas e softwares certificados podem acessar dados de postagens públicas.

Desde a crise da webometria, não se pode mais ingenuamente considerar que o acesso aos dados consolidados ou sob domínio de empresas comerciais será sempre aberto. Como os campos da cibermetria são dependentes dos agregadores, ou precisam trabalhar com um universo fechado a ser recuperado e estruturado, uma das alternativas que temos é abrir os relatórios que as plataformas oferecem aos geradores de conteúdo.

4. O QUE OS EDITORES PODEM FAZER

Aqueles que fazem a gestão dos sítios das revistas e de suas contas em mídias sociais têm naturalmente acesso às ferramentas de monitoramento. Ter por princípio a abertura dos relatórios e dados gerados pelas ferramentas é certamente um dos primeiros passos. Porém, há todo um conjunto de possibilidades adicionais. Abaixo segue uma lista não exaustiva de recomendações e de ações para a abertura de dados de acesso e de ações em mídias sociais por editores:

- Verificar se o sítio tem uma implementação bem executada de alguma ferramenta de monitoramento de acesso, quer seja de *page-tagging*, como o Google Analytics, quer seja de log de acesso¹⁹;
- Criar uma conta de acesso aos relatórios restrita apenas à visualização para compartilhamento com interessados nos dados (pesquisadores e editores, por exemplo);
- Criar uma página onde seja informada a abertura dos dados de acesso e as formas de solicitá-los ou de acessar e explorar diretamente os relatórios;
- Publicar anualmente um relatório de acesso à revista para fácil referência;
- Disponibilizar planilha com os dados exportados das plataformas de monitoramento para estudos acadêmicos e análises técnicas;
- Ter contas nas mídias sociais que considerar de interesse para seu público, mas tendo em mente que é necessário trabalhar com elas;
- Nas contas em mídias sociais, deixar clara a linha editorial da conta e informar a abertura de dados;
- Publicar anualmente os relatórios obtidos pelas ferramentas de gestão de suas contas em mídias sociais;
- Disponibilizar planilhas com os dados exportados das plataformas de monitoramento de acesso ao sítio e das contas em mídias sociais para estudos acadêmicos e análises técnicas.

Cada uma dessas recomendações e ações pode contribuir não só para os trabalhos dos pesquisadores e para o acompanhamento gerencial de editores, como também permite preservar ao longo do tempo dados que podem subitamente deixar de estar disponíveis. Ter seus dados históricos na forma de relatórios consolidados pode, no futuro, servir para que eles não sejam perdidos, haja vista o fim do Orkut após 11 anos de existência.

19 Todo servidor web tem um arquivo de registro de todas as solicitações feitas a ele. Há softwares que fazem a análise destes dados e os traduzem em gráficos para melhor consulta.

Somente com cada editor abrindo o acesso aos dados de seus periódicos teremos um campo fértil para o desenvolvimento de estudos sobre as ações em mídias sociais das revistas. Subvertendo a visão comercial, onde a estratégia de ação de cada revista seria um segredo de negócio, teremos um ambiente onde os dados abertos de acesso, compartilhamento, debate e uso das produções científicas, tanto na web quanto nas mídias sociais, passam a ser de conhecimento público. Nada mais lógico do que revistas abertas oferecerem abertamente seus dados.

5. REFERÊNCIAS

GOUVEIA, F.C. Novos caminhos e alternativas para a Webometria. **Em Questão**, v.18, p.249-261, 2012. Disponível em: <https://seer.ufrgs.br/EmQuestao/article/view/33270>. Acesso em: 10 jan. 2020.

VAUGHAN, L.; HYSEN, K. Relationship between links to journal web sites and impact factors. **Aslib Proceedings**, v.54, n.6, p.356-361, 2002. DOI: 10.1108/00012530210452555. Disponível em: <https://www.emerald.com/insight/content/doi/10.1108/00012530210452555/full/html>. Acesso em: 19 fev. 2019.

NEYLON, C.; WU, S. Article-level metrics and the evolution of scientific impact. **PLoS Biology**, v.7, n.11, e1000242, 2009. DOI: 10.1371/journal.pbio.1000242. Disponível em: <https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.1000242>. Acesso em: 19 fev. 2019.

PRIEM, J.; HEMMINGER, B.M. Scientometrics 2.0: Toward new metrics of scholarly impact on the social Web. **First Monday**, v.15, n.7, 2010. Disponível em: <https://firstmonday.org/article/view/2874/2570>. Acesso em: 01 mai. 2019.

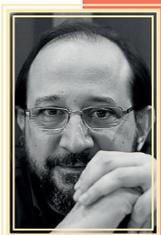
PRIEM, J *et al.* **Altmetrics**: a manifesto. 2010. Disponível em: <http://altmetrics.org/manifesto>. Acesso em: 08 jan. 2019.

6. SOBRE OS AUTORES

Fábio Castro Gouveia

Lattes: <http://lattes.cnpq.br/0733908324235348>

Orcid: <https://orcid.org/0000-0002-0082-2392>



É Tecnologista em Saúde Pública da Fundação Oswaldo Cruz - Brasil, líder do Grupo de Pesquisa Ciência, Dados, Redes e Métricas - (Scimetrics) e participa como pesquisador na Rede Zika Ciências Sociais. Biólogo, mestre em Microbiologia e Imunologia e doutor em Química Biológica (Educação, Gestão e Difusão de Biociências), fez um pós-doutorado curto como Visiting Fellow da Katholieke Universiteit Leuven (Bélgica) selecionado no edital 2009 do Coimbra Group Scholarships Programme for Young Professors and Researchers from Latin American Universities. Desenvolve pesquisas na área Ciência da Informação com ênfase em Estudos Métricos da Informação (Cientometria, Webometria, Altméria e Indicadores de Ciência, Tecnologia e Inovação), Métodos Digitais, STS, Data Science e Tecnologia Blockchain, e na área de Divulgação Científica e Comunicação em Saúde, com ênfase em estudos sobre internet e mídias sociais.

7. COMO CITAR

GOUVEIA, F. C. Abrindo os dados de acesso e de ações em mídias sociais. *In*: SHINTAKU, M.; SALES, L. F.; COSTA, M. (org). **Tópicos sobre dados abertos para editores científicos**. Botucatu, SP: ABEC, 2020. p. 31-37. DOI: 10.21452/978-85-93910-04-3.cap3.

Ampliando as fronteiras da editoração científica: o papel dos repositórios de dados

Luana Farias Sales
Luis Fernando Sayão

1. CONSIDERAÇÕES INICIAIS

A aventura da descoberta científica tem sido guiada pelo método científico que remonta tempos ancestrais e envolve uma abordagem filosófica e prática da ciência. A partir de suas observações sobre o universo, os pesquisadores delineiam conclusões, desenvolvem novas teorias que podem ser testadas, comunicam e compartilham suas descobertas com outros cientistas, permitindo, assim, a revisão crítica e o consenso antes que os novos achados sejam incorporados ao corpo coletivo de conhecimento. Esta é uma parte imprescindível da investigação científica. A comunicação e as discussões abertas estão localizadas no coração dos processos científicos (THE ROYAL SOCIETY, 2012). A convergência entre o método científico e a comunicação dos resultados de pesquisa formam um elo essencial na dinâmica de geração e acumulação do conhecimento científico consensual.

No longo percurso desde da aurora da ciência até os dias de hoje, o passo crucial para os indivíduos que se debruçavam sobre os enigmas da natureza foi sempre tomar consciência das ideias em torno do seu espaço de interesse. Há alguns séculos isto significa obter acesso a livros, periódicos e aos colegas por conversas, cartas e reuniões de sociedades científicas. Romper as barreiras

interpostas ao acesso aos trabalhos publicados e à interlocução com outros cientistas sempre constituiu um fator essencial na superação dos vieses, crenças e limitações de visões individuais do universo.

Nessa direção, o registro das observações da natureza, dos fenômenos sociais e dos ensaios e experimentos que transcorrem nas bancadas dos laboratórios sempre foi parte imprescindível do processo de construção do saber científico. Entretanto, por muito tempo esses registros se manifestam tipicamente nas formas mais tradicionais da comunicação acadêmica, como artigos de periódicos, livros, teses etc. Além disso, esses registros estão presentes também na forma de números, multimídias, modelos, bases de dados, softwares e de outros produtos de pesquisa, como são as anotações nos cadernos de pesquisa e de campo. Essa condição está ainda mais presente na ciência contemporânea, que produz e consome intensamente dados de pesquisa e os torna protagonistas de novas metodologias de pesquisa que desafiam os pressupostos da ciência mais tradicional. Essa nova configuração, determinada pela ciência contemporânea, impõe o desafio de se criar infraestruturas que possibilitem a comunicação e a integração de publicações e de outros produtos de pesquisa, principalmente os dados, formando uma organicidade que permita níveis de interoperabilidade entre esses ativos informacionais. As coleções de dados têm que estar minimamente conectadas às publicações a que eles dão sustentação científica: esse é o patamar básico da organicidade.

2. PUBLICAÇÃO DE DADOS COMO PARTE DA PUBLICAÇÃO DE PESQUISAS

Publicar dados significa, na maioria dos casos, tratá-los, organizá-los e arquivá-los pelo tempo que for necessário em infraestruturas técnicas e gerenciais adequadas. Há algumas décadas os pesquisadores recorriam à criação dos seus próprios *websites* ou servidores de dados. Esses sistemas fragmentados geralmente evoluem para preencher as necessidades dos pesquisadores, entretanto tendem a ser estáticos, isolados e sujeitos a extravios, obsolescência ou evaporação (MILLER, 2016).

Os repositórios de dados foram inicialmente pensados como centros de dados disciplinares, voltados para domínios altamente estruturados e que produzem grandes quantidades de dados, como genética física, astrofísica e ciências ambientais. São infraestruturas complexas de grandes proporções que resultaram

de projetos nacionais e internacionais. Paralelamente observa-se que comunidades científicas específicas começam a evoluir seus sistemas de repositórios de dados com serviços customizados e voltados para suas necessidades e tradições de compartilhamento, interação e reúso de suas áreas. Mais recentemente – por exigências principalmente das agências de fomento e dos periódicos científicos –, a demanda por repositórios de dados está emergindo também na cauda longa da ciência, ou seja, nos domínios científicos nos quais as atividades de pesquisa são desenvolvidas e num grande número de pequenos laboratórios por pequenas equipes, que, coletivamente, produzem a maioria dos resultados científicos. Estes, por sua vez, raramente dispõem de repositórios para a gestão de seus dados.

O que fica claro é que hoje a publicação de dados é parte imprescindível do processo de comunicação científica. A partir desse ponto, configura-se a necessidade de se implantar repositórios multidisciplinares que viabilizem a interconexão de dados de pesquisa e publicações acadêmicas, expandindo as fronteiras da editoração científica na direção de cobrir integralmente a diversidade dos produtos de pesquisa. O Repositório Digital Dryad atua nessa direção.

3. DRYAD – CONECTANDO DADOS E PUBLICAÇÕES

O Dryad²⁰ é uma organização que tem parcerias com várias casas editoriais e periódicos científicos importantes que adotam para suas publicações acadêmicas uma Política Conjunta de Arquivamento de Dados, mais conhecida pela sigla em inglês JDAP, que significa Joint Data Archiving Policy²¹. A JDAP descreve os requisitos necessários para disponibilização pública de publicações que dão suporte a dados de pesquisa, além de ter como princípio norteador o reconhecimento de que infraestruturas administradas de dados pelas comunidades acadêmicas – que sejam abertas e fácil de usar – são necessárias para a implementação dessa Política.

Mesmo considerando que o JDAP e o Dryad são iniciativas distintas, os periódicos que adotam JDAP recomendam Dryad como um repositório apropriado

20 <<https://datadryad.org/stash>>

21 <[http://wiki.datadryad.org/Joint_Data_Archiving_Policy_\(JDAP\)](http://wiki.datadryad.org/Joint_Data_Archiving_Policy_(JDAP))>

para a curadoria dos dados associados aos itens de literatura científica que publicam.

Seguindo os princípios do JDAP, o Dryad operacionaliza um repositório digital que se caracteriza como um repositório multipropósito voltado para uma grande diversidade de tipos de dados de pesquisa. Tem como objetivo contribuir para um mundo onde os dados de pesquisa sejam disponíveis abertamente, integrados com a literatura acadêmica, e possam ser rotineiramente reusados para criar novos conhecimentos. Sua missão é oferecer uma infraestrutura para promover o reúso dos dados que fundamentam a literatura acadêmica e científica. O Dryad se volta tipicamente para as comunidades que não dispõem de repositórios próprios, oferecendo vantagens importantes para os principais atores do ciclo de publicação desse segmento:

- **Para os pesquisadores** – oferece um serviço que apoia o reconhecimento da autoria das coleções de dados geradas/coletadas, tornando-as citáveis; torna as coleções visíveis para outros pesquisadores que podem recuperá-las e reutilizá-las em novas pesquisas.
- **Para as instituições de pesquisa** – mantém íntegra a memória acadêmica por meio do registro e preservação dos dados por longo prazo e torna mais transparente os seus fluxos de pesquisa, na medida em que publica uma parte importante dos seus resultados.
- **Para as editoras científicas** – podem retratar os trabalhos científicos na sua integralidade plena, criando uma organicidade de dados e publicações e oferecendo aos pesquisadores formas múltiplas de acesso a esses ativos.
- **Para a ciência como um todo** – oferece mecanismos mais avançados e confiáveis de revisão por pares; aumenta o grau de reprodutibilidade dos experimentos científicos e o poder de autocorreção da ciência.

Para garantir as possibilidades listadas acima, o Dryad apresenta as seguintes características:

Quadro 4.1 - Características do Dryad

Formato de dados	Flexível em relação aos formatos dos dados, incentivando o uso e o desenvolvimento de padrões da comunidade de pesquisa
Entrada de dados	Encaixa-se no fluxo de trabalho de envio de manuscritos dos periódicos parceiros

Identificadores	Atribui identificadores (DOIs) aos dados para que os pesquisadores possam obter crédito profissional por meio de citação
Visibilidade	Promove a visibilidade dos dados por meio de métricas de uso e de download e permite que o conteúdo seja indexado, pesquisado e recuperado
Qualidade dos dados	Promove a qualidade dos dados por meio de curadores profissionais para garantir a validade dos arquivos e informações descritivas
Licença	O conteúdo é gratuito para download e reuso sob licença Creative Commons Zero (CC0)
Preservação	Os conteúdos são preservados a longo prazo para garantir o acesso indefinidamente
Padronização	Código aberto, tecnologia compatível com os padrões da área

Fonte: Os autores, baseados em informações extraídas da página do Dryad

4. À GUIA DE CONCLUSÃO

A ciência contemporânea e suas interlocuções com os pressupostos da Ciência Aberta demandam novos meios para uma comunicação científica mais completa, que extrapole os limites dos periódicos científicos tradicionais. Nesse cenário, o compartilhamento de dados se torna uma parte essencial da publicação dos resultados de pesquisa. Porém, como desdobramento, coloca o desafio para os editores científicos de oferecer opções de infraestruturas para a publicação de dados que os tornem acessíveis, citáveis e conectados às publicações, além de apoiar a revisão por pares, tendo como referência fluxos de pesquisa mais fidedignos, reproduzíveis e abertos. As infraestruturas de dados são ainda mais importantes para as áreas científicas que não possuem repositórios específicos e que não conseguem conectar seus dados às publicações. Mas elas devem ser construídas em estreita colaboração entre pesquisadores e organizações científicas e apoiadas pelos editores. O Dryad cumpre essa função em escala internacional, porém é recomendável que se possa reproduzir essa ideia no âmbito das casas editoriais brasileiras. Nessa direção, propõe-se o desenvolvimento de repositório multidisciplinar que possa acolher coleções de dados associados a artigos submetidos a periódicos brasileiros, os quais não dispõem de repositórios disciplinares e operam segundo um fluxo de publicação e de revisão por pares adequados à realidade das nossas publicações acadêmicas.

5. REFERÊNCIAS

MILLER, Gary W. Making data accessible: The Dryad experience. **Toxicological Science**, v. 149, n. 1, 2016. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/26728592>>. Acesso em: 20 dez. 2019.

THE ROYAL SOCIETY. **Science as an open enterprise**. London : The Royal Society Science Policy Centre, 2012. Disponível em: <<https://royalsociety.org/~media/policy/projects/sape/2012-06-20-saoe.pdf>>. Acesso em: 23 dez. 2019.

6. SOBRE OS AUTORES



Luana Farias Sales

Lattes: <http://cnpq.br/9090064478702633>

Orcid: <http://orcid.org/0000-0002-3614-2356>

Doutora em Ciência da Informação pelo Programa de Pós-Graduação do IBICT/UFRJ (2011-2014). Mestre em Ciência da Informação pelo convênio UFF/IBICT (2004-2006), Graduação em Biblioteconomia e Documentação pela Universidade Federal Fluminense (2003). Atualmente exerce o cargo de Coordenadora-Geral de Acesso e Difusão de Acervos do Arquivo Nacional. É docente do Programa de Pós-graduação em Ciência da Informação do convênio IBICT-UFRJ e Coordenadora da Rede de Implementação do GO FAIR Brasil.



Luis Fernando Sayão

Lattes: <http://lattes.cnpq.br/3422623122948389>

Orcid: <http://orcid.org/0000-0002-6970-0553>

Possui graduação em Física pela Universidade Federal do Rio de Janeiro (1978), mestrado em Ciência da Informação pela Universidade Federal do Rio de Janeiro/Instituto Brasileiro de Informação em Ciência e Tecnologia (UFRJ/IBICT) e doutorado em Ciência da Informação pela UFRJ/IBICT (1994). Trabalha desde 1980 na Comissão Nacional de Energia Nuclear (CNEN) onde atualmente é líder do grupo de pesquisa em Gestão do Conhecimento Nuclear. Atua ainda como docente do Programa de Pós-Graduação em Biblioteconomia da UNIRIO - Universidade Federal do Estado do Rio de Janeiro e do Programa de Pós-Graduação em Memória e Acervos da Fundação Casa de Rui Barbosa

7. COMO CITAR

SALES, L. F.; SAYÃO, L. F. Ampliando as fronteiras da editoração científica: o papel dos repositórios de dados. *In*: SHINTAKU, M.; SALES, L. F; COSTA, M. (org). **Tópicos sobre dados abertos para editores científicos**. Botucatu, SP: ABEC, 2020. p. 39-45. DOI: 10.21452/ 978-85-93910-04-3.cap4.

Um estudo para representação de dados abertos de editoração científica na web de dados

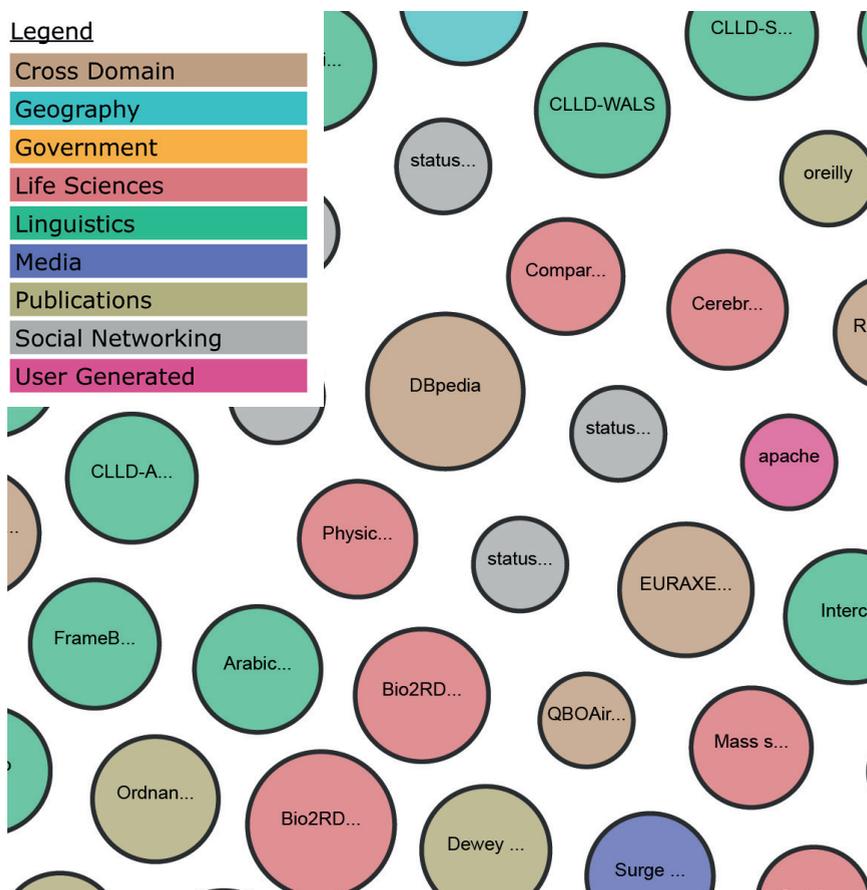
Sandro Rautenberg
Lucélia de Souza

1. INTRODUÇÃO

É notório que os avanços computacionais de processamento, armazenamento e transmissão de dados empoderaram a sociedade nas diversas formas de compartilhamento e uso da informação e do conhecimento. Em outras palavras, mediante os serviços e tecnologias da web, indivíduos têm experimentado o acesso aberto a diversos recursos de dados geograficamente descentralizados, permitindo a recombinação de elementos na busca de informações contextualizadas.

Em face disso, no domínio da Ciência da Informação, as tecnologias de informação e comunicação, principalmente aquelas acessíveis na web, têm contribuído à socialização da comunicação científica (BANDEIRA; FREIRE, 2017). Neste contexto, para os referidos autores, a editoração científica tem se dinamizado ao promover: **i)** a melhor interação entre os indivíduos (editores, leitores e autores) em novos espaços virtuais; e **ii)** a iniciativa do acesso aberto aos insumos de conhecimento. Essa perspectiva também está alinhada à abertura de dados, a qual foi ensejada por Tim Bernes-Lee ao vislumbrar uma nova visão de utilização da web, a Web de Dados (LINKED DATA, 2019).

Figura 5.1 - Representação parcial da Web de Dados



Fonte: adaptada a partir de (ABELE et al., 2019).

Representada na Figura 5.1, a Web de Dados se configura como um espaço virtual formado por inúmeros conjuntos que permitem a publicação e o compartilhamento de dados abertos (AUER, 2014). Em poucas palavras, a Web possibilita a exploração de recursos de dados nos mais variados domínios (W3C, 2019). Segundo Lóscio, Burle e Calegari (2019), essa perspectiva confere aos dados disponibilizados na web os benefícios de **compreensão** – com a publicação de conjuntos de dados e seus respectivos metadados, os seres humanos têm melhor entendimento da natureza, da estrutura e da semântica dos dados compartilhados; **reúso** – aumenta as chances de reutilização dos dados por consumidores, independentemente do domínio de aplicação; **ligação** – ao publicar os dados em consonância com os protocolos da Internet, é possível a criação de relações (*hiperlinks*) entre recursos de dados (conjuntos de dados

e itens de dados); **descoberta** – mediante os *hiperlinks*, autonomamente, os computadores descobrem os conjuntos de dados e navegam entre os recursos digitais, aferindo conhecimento relacional entre os itens compartilhados; **confiança** – com a publicação de metadados de proveniência, os publicadores expressam como os conjuntos de recursos digitais são custodiados ao longo do tempo; **acesso** – utilizando os protocolos de acesso da Internet, promove-se a encontrabilidade e a exploração de recursos digitais pelos agentes (humanos e de software); **interoperabilidade** – permite que os recursos digitais sejam automaticamente convertidos em formatos abertos diversos; e **processabilidade** – adicionalmente à interoperabilidade, a processabilidade permite que as aplicações computacionais manipulem os recursos digitais contidos na Web de Dados de forma automática.

Interdisciplinarmente ao domínio da Ciência da Informação, para socializar os recursos na Web de Dados, pressupõe-se que os metadados constituem-se como os modelos de representação necessários para a publicação de dados. Neste sentido, Schmachtenberg, Bizer e Paulheim (2014) pontuam que os modelos utilizados para representar dados e sua interpretação são imprescindíveis para manter dados semânticos na Web de Dados. Particularmente, a Figura 5.2 melhor ilustra tais pressupostos no domínio da editoração científica. Ou seja, o compartilhamento de dados abertos e semântica inerentes a publicações científicas é mediado pelo uso de modelos de representação (vocabulários ou ontologias) em consonância às melhores práticas de publicação na Web de Dados.

Figura 5.2 - Contexto do trabalho



Fonte: adaptada a partir de (LÓSCIO; BURLE; CALEGARI, 2019).

Assim, considera-se a Web de Dados como um espaço virtual importante no que tange à exploração dos mecanismos para interoperabilidade de dados abertos de periódicos e artigos na web, potencializando a disseminação do conhecimento

científico (PACKER, 2019). Na visão de Bandeira e Freire (2017), isso permeia a apropriação, a produção e o compartilhamento de recursos de conhecimento entre editores, autores e leitores. Disto, pressupõe-se que a reutilização de vocabulários padronizados contribui ao reúso, à processabilidade, à compreensão, à confiança e à interoperabilidade de dados abertos, facilitando a inclusão de recursos informacionais da editoração científica e a expansão da Web de Dados.

Diante do exposto, este capítulo objetiva contribuir no alinhamento dos esforços para organizar, representar e compartilhar recursos de dados do domínio da Ciência da Informação. Especificamente, visa recomendar um modelo de representação já estabelecido e reconhecido na Web de Dados para compreender, reutilizar, interoperar e processar os dados e metadados de recursos inerentes à editoração científica.

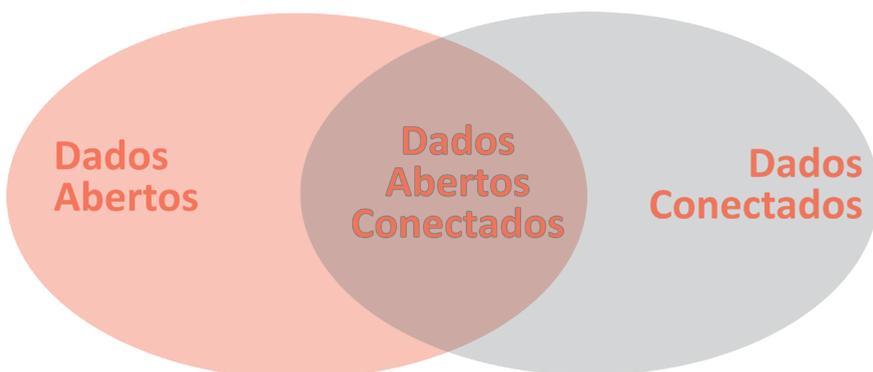
Para tanto, além dessa seção introdutória, este capítulo compreende as discussões sobre: **i) fundamentação teórica** – disserta sobre os conceitos aderentes aos dados abertos conectados como formalismo de representação e publicação de dados abertos de editoração científica na Web de Dados; **ii) procedimento metodológico** – define as diretrizes gerais para a escolha de um modelo de representação de recursos na Web de Dados; **iii) a busca do vocabulário para publicação de dados abertos** – descreve o processo de avaliação realizado na recomendação de um modelo de representação de dados abertos no campo da editoração científica; e **iv) resumo do capítulo** – pontua as assertivas a respeito dos esforços dispendidos e sugere trabalhos futuros.

2. DADOS ABERTOS CONECTADOS E SUA REPRESENTAÇÃO NA WEB DE DADOS

Ao compartilhar os dados abertos e metadados de editoração científica na Web de Dados, dois momentos distintos devem ser observados. Primeiramente, quando um conhecimento a respeito de algum objeto é formalmente representado por meio de um modelo (vocabulários ou ontologias), deve-se relacionar os metadados inerentes do recurso em estudo aos padrões e às Melhores Práticas de Publicação na Web de Dados (LÓSCIO; BURLE; CALEGARI, 2019). No segundo momento, quando os dados de um recurso são materializados em consonância ao modelo de representação escolhido, eles são organizados e estruturados, tornando-se dados conectados. E, se os dados conectados advêm de bases de dados livremente acessíveis e

interoperáveis por agentes (pessoas ou sistemas), eles são classificados como dados abertos conectados. A Figura 5.3 representa o relacionamento entre tais conceitos.

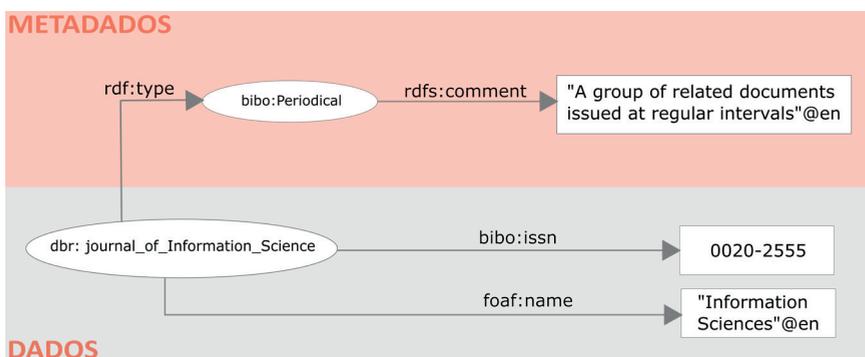
Figura 5.3 - Dados Abertos Conectados e seus benefícios



Fonte: elaborada pelos autores (2019).

Diante disso, o conceito de dados abertos conectados é configurado por duas questões norteadoras: O que são dados abertos? E o que são dados conectados? Resumidamente, os dados abertos são regidos por licenças que definem as regras de distribuição e (re)utilização que as pessoas ou aplicações computacionais devem seguir quando do uso de tais recursos da web (LINKED DATA, 2019). Alguns exemplos de regras podem ser: **a)** a necessidade de citação da fonte original; **b)** o uso não comercial dos dados; **c)** a não derivação ou não adaptação dos dados; **d)** o compartilhamento dos dados em mesmo formato. Por sua vez, os dados conectados referem-se aos dados que são representados mediante os protocolos de acesso da Internet e usam constitutivamente os formalismos da Web de Dados como infraestrutura para promover sua publicidade (BIZER; HEATH; BERNES-LEE, 2009). Para melhor elucidar a contextualização do que é um dado aberto conectado, considere a Figura 5.4.

Figura 5.4 - Exemplo de Dado Aberto Conectado



Fonte: elaborada pelos autores (2019).

O formalismo diante da Web de Dados se constitui pelo uso de um modelo de representação, o qual estrutura a descrição semântica de recursos na web compreensível por agentes humanos e de software. Essa estrutura apresenta a forma de uma tripla (sujeito-predicado-objeto) em que a semântica inerente a um sujeito é intermediada por um predicado, relacionando o sujeito a um objeto. Por exemplo, a Figura 5.4 ilustra a representação parcial do periódico “Information Sciences” como um sujeito (ou recurso), com seus dados e metadados. Em poucas palavras, considerando os protocolos da Internet, um recurso é codificado por **dbr:Journal_of_Information_Sciences**. A partir desse endereço na web, pode-se navegar pelo predicado **rdf:type** e descobrir o objeto que tipifica o recurso em questão como **bibo:Periodical**. Também é possível navegar por meio do predicado **rdfs:comment** e averiguar semanticamente que o referido tipo é conceituado como “A group of related documents issued at regular intervals”, sendo que a tag associada @en indica que o comentário é descrito no idioma inglês. Oportunamente, pode-se descobrir o ISSN do referido periódico por meio do predicado **bibo:issn**, ao qual associa-se um objeto com o valor “0020-0255”. Por fim, pode-se obter o nome associado ao sujeito por meio do predicado **foaf:name**, cujo valor é “Information_Sciences”@en.

Ressalta-se que para tornar tais descrições interoperáveis entre computadores, os preceitos da Web de Dados incentivam o uso de vocabulários e/ou ontologias padronizados e amplamente conhecidos por uma comunidade. No exemplo discutido anteriormente, ao representar o recurso **dbr:Journal_of_Information_Sciences** como dado aberto conectado, fez-se o uso da ontologia BIBO - The Bibliographic Ontology (D’ARCUS; GIASSON, 2019), a

qual reutiliza o vocabulário FOAF – Friend-of-a-Friend (BRICKLEY; MILLER, 2019) em seu modelo de representação. Salienta-se que a adoção de vocabulários e/ou ontologias reconhecidos potencializam o reuso de conceitos, propriedades e relacionamentos comumente aceitos por produtores e consumidores de dados abertos conectados, contribuindo para a expansão representacional e informacional da Web de Dados. Neste sentido, a utilização da BIBO na publicação de dados abertos de editoração científica é discutida em detalhes na Seção 4.

Uma vez definidos os conceitos de dados abertos, dados conectados, dados abertos conectados e vocabulários, a próxima seção aborda o procedimento metodológico que guiou a busca do modelo de representação já estabelecido e reconhecido para publicar dados abertos de editoração científica na Web de Dados.

3. DIRETRIZES DE REUTILIZAÇÃO DE VOCABULÁRIOS DE REPRESENTAÇÃO NA WEB DE DADOS

Para publicar dados abertos conectados, o W3C recomenda a reutilização de vocabulários (ou ontologias) de representação destes recursos, preferencialmente aqueles considerados como um padrão para codificar dados e metadados (LÓSCIO; BURLE; CALEGARI, 2019). Segundo a referida entidade, o uso de vocabulários já estabelecidos: **i)** captura e facilita o consenso em comunidades (publicadores e consumidores de dados); **ii)** aumenta a interoperabilidade; **iii)** reduz redundâncias, incentivando assim a reutilização de dados; **iv)** ajuda na comparação e no processamento automático de dados e metadados; e **v)** evita ambiguidades e conflitos entre elementos ou valores semelhantes. Em suma, a reutilização de vocabulários padronizados contribui para o reuso, processabilidade, compreensão, confiança e interoperabilidade de dados abertos conectados entre os agentes humanos e de software, facilitando a inclusão de recursos informacionais e expandindo o espaço virtual da Web de Dados.

Para reutilizar de uma representação de dados na Web de Dados, Hyland, Atezing e Villazón-Terrazas (2019) sugerem que os publicadores verifiquem:

- a garantia de que os vocabulários sejam publicados por um grupo/organização confiável;

- a garantia de que os vocabulários tenham URIs permanentes; e
- a existência de uma política de versionamento.

Neste sentido, os autores anteriormente referenciados detalham as diretrizes a serem observadas na reutilização vocabulários como modelos de representação. São elas:

- **Vocabulários precisam ser documentados.** Inclui o uso de rótulos e comentários, além de *tags* de idioma apropriadas. O(s) editor(es) deve(m) manter páginas legíveis por humanos que descrevam o vocabulário, juntamente com seus conceitos e propriedades constituintes. Adicionalmente, alguns casos de uso facilmente compreensíveis podem ser definidos e documentados.
- **Vocabulários deveriam ser autodescritivos.** Para cada propriedade ou termo do vocabulário, deve-se fornecer um rótulo, uma definição e um comentário. Os dados autoexplicativos sugerem que as informações sobre cada elemento são fornecidas explicitamente dentro da representação. A capacidade dos dados abertos conectados de se descrever contextualmente contribui para a explicitação da utilidade dos dados. Por exemplo, no vocabulário Dublin Core tem-se um metadado rotulado como *dc:Contributor*, definindo-o como uma entidade responsável por fazer contribuições para o recurso. O comentário presente para este metadado cita que os exemplos de um *Contributor* incluem uma pessoa, uma organização ou serviço.
- **Vocabulários deveriam ser descritos em mais de um idioma.** O multilinguismo deve ser suportado pelo vocabulário. Ou seja, todos os elementos do vocabulário devem ter rótulos, definições e comentários disponíveis em idioma(s) oficial(is) como a língua espanhola e a inglesa. Por exemplo, no vocabulário *Description of a Project (DOAP) Vocabulary*, o termo *doap:Project* é rotulado em diversos idiomas, como as seguintes: “*Project*”@en, “*Projet*”@fr, “*Proyecto*”@es, “*Projekt*”@de, “*Projekt*”@cs, “*Projeto*”@pt.
- **Vocabulários deveriam ser usados em vários conjuntos de dados.** Ressalta-se que a reutilização de vocabulários potencializa os benefícios dos dados abertos conectados. Quando um vocabulário desenvolvido por outrem é usado na representação de outros conjuntos de dados abertos,

evidencia-se um padrão estabelecido na comunidade LOD. E, portanto, o vocabulário configura-se como o melhor candidato para adoção (parcial) e reutilização. Por exemplo, uma análise sobre o uso de vocabulários na Web de Dados revela que o vocabulário Dublin Core é reutilizado em 132 conjuntos de dados.

- **Vocabulários deveriam ser acessíveis de forma perene.** Ao ser selecionado, um vocabulário deveria fornecer alguma garantia de manutenção por um período especificado, idealmente por tempo indeterminado por uma entidade certificadora.
- **Vocabulários deveriam ser publicados por um grupo ou organização confiável.** Embora qualquer pessoa possa criar um vocabulário como modelo de representação para a Web de Dados, sempre é desejável utilizar um vocabulário que atenda aos requisitos de reputação. Neste sentido, pode-se averiguar se os termos do vocabulário são definidos em uma organização de desenvolvimento de padrões (por exemplo, o W3C) ou se são publicados por uma autoridade adequada, no caso, uma agência governamental.
- **Vocabulários deveriam ter URLs persistentes.** É imperativo o acesso persistente ao servidor que hospeda o vocabulário. Por exemplo, o *WGS84 Geo Positioning: an RDF Vocabulary*²² é um dos vocabulários mais usados para representação básica de pontos de georreferenciamento (latitude, longitude e altitude). A versão atual deste vocabulário data de 2009 e encontra-se disponível em <http://www.w3.org/2003/01/geo/wgs84_pos#>.
- **Vocabulários deveriam ter uma política de versionamento.** O editor do vocabulário é responsável por manter a compatibilidade entre as versões do vocabulário ao longo do tempo. As principais alterações nos vocabulários devem ser refletidas na documentação. Tem-se como exemplo os registros dos esforços despendidos na manutenção do WGS84 Geo Positioning ao longo do tempo (Quadro 5.1).

22 http://www.w3.org/2003/01/geo/wgs84_pos#

Quadro 5.1 - Exemplo de comentários do versionamento do WGS84 Geo Positioning: an RDF Vocabulary

01	[...]
02	<rdfs:comment>
03	Recent changes to this namespace:
04	\$Log: wgs84_pos.rdf,v \$
05	
06	Revision 1.22 2009/04/20 15:00:30 timbl
07	Remove the time bits which have been deal with elsewhere eg in iCal.
08	
	[...]
	</rdfs:comment>
	[...]

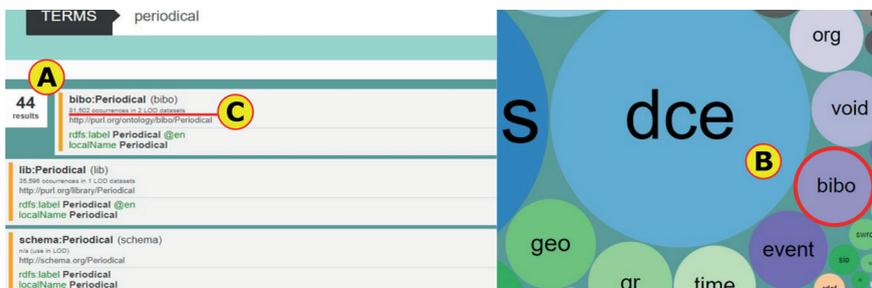
Fonte: adaptada de http://www.w3.org/2003/01/geo/wgs84_pos#.

Considerando as diretrizes anteriormente enumeradas, a próxima seção disserta sobre a aplicação destas recomendações na busca de um modelo de representação adequado à publicação de dados abertos de editoração científica na Web de Dados.

4. EM BUSCA DE UM MODELO DE REPRESENTAÇÃO PARA PUBLICAÇÃO DE DADOS DE EDITORAÇÃO CIENTÍFICA NA WEB DE DADOS

Conforme sugerido por Hyland, Ateazing e Villazón-Terrazas (2019), ao reutilizar um modelo de representação (vocabulários ou ontologias) de dados abertos na Web de Dados, pode-se recorrer ao repositório *Linked Open Vocabulary* (LOV). O referido repositório é compreendido como um observatório do ecossistema de vocabulários presentes na Web de Dados, objetivando promover e facilitar a reutilização de vocabulários bem documentados (VANDENBUSSCHE *et al.*, 2017). Desta forma, a partir de buscas dirigidas no LOV, pode-se avaliar possíveis candidatos de modelos de representação para um conjunto de dados abertos. Em poucas palavras, no referido repositório, a busca por termos já modelados e que são inerentes a algum domínio de interesse configura-se como uma estratégia pertinente para recuperar os modelos de representação bem estabelecidos. Portanto, considerando o domínio da editoração científica, o termo utilizado para a busca de modelos de representação aderentes foi *Periodical*.

Figura 5.5 - Seleção dos vocabulários para o domínio da editoração científica



Fonte: adaptada a partir de (Linked Open Vocabularies, 2019a; Linked Open Vocabularies, 2019b).

Como ilustrado na Figura 5.5a, a partir do repositório LOV são encontradas 44 ocorrências de ontologias ou vocabulários que contêm o termo *Periodical*. E como destacado na 5.5b, a primeira alternativa relevante recuperada é a ontologia *Bibliografic Ontology* (BIBO), devido ao volume de sua utilização na representação de dados abertos. Ressalta-se que em sua essência, a BIBO é usada como uma ontologia de citação ou de classificação, permitindo descrever qualquer tipo de documento bibliográfico na Web de Dados (D'ARCUS; GIAS-SON, 2019). Por exemplo, a referida ontologia já é utilizada para representar

cerca de 81 mil ocorrências de entidades classificadas como `bibo:Periodical` em dois conjuntos de dados abertos (Figura 5.5c). Tais assertivas estão em consonância com as diretrizes listadas na seção anterior, principalmente no que diz respeito ao fato de que os **vocabulários deveriam ser usados em vários conjuntos de dados**.

Em relação à diretriz **vocabulários precisam ser documentados**, os editores da BIBO mantêm um sítio que define as características do vocabulário, juntamente com seus conceitos, propriedades constituintes e exemplos de utilização.

No caso de **vocabulários deveriam ser autodescritivos**, a BIBO fornece rótulos, definições e comentários em linguagem natural, próprios à facilidade de compreensão. Exemplarmente, a Figura 5.6 alude à compreensão para o conceito `bibo:Periodical` (Figura 5.6a). De forma implícita à BIBO (Figura 5.6b), o referido conceito é rotulado como *Periodical*, tendo como definição “um grupo de documentos relacionados emitidos em intervalos regulares” [tradução dos autores].

Figura 5.6 - Representação do conceito *Periodical* na BIBO



Fonte: elaborada pelos autores (2019).

Ainda de acordo a Figura 5.6, pode-se tecer assertivas sobre a diretriz **vocabulários deveriam ter URLs persistentes**. Na Figura 5.6c evidencia-se o URL `<http://purl.org/ontology/bibo/>` como endereço padrão para o acesso aos elementos constituintes da BIBO. Em relação a **vocabulários deveriam ter uma política de versionamento**, a Figura 5.6d aponta que o conceito *Periodical* é representado em uma versão estável. Em suma, isso atesta que a utilização deste elemento como parte de um modelo de representação na Web de Dados é recomendada.

Quanto às diretrizes **vocabulários deveriam ser acessíveis de forma permanente** e **vocabulários deveriam ser publicados por um grupo ou organização confiável**, vale ressaltar que recentemente a renomada organização *Dublin*

Core Metadata Initiative tornou-se uma das entidades responsáveis pela publicação da BIBO (DUBLIN CORE METADATA INITIATIVE, 2019). Dada a reputação e os serviços prestados pela organização, pressupõe-se o contínuo desenvolvimento e o uso regular da BIBO em pesquisas da Ciência da Informação. Adicionalmente, tal fato contribui à diretriz **vocabulários deveriam estabelecer uma política de versionamento**.

Por fim, quanto a **vocabulários deveriam ser descritos em mais de um idioma**, apresentando um grau de multilinguismo, a BIBO ainda não evidencia o atendimento a essa diretriz, limitando a descrição de seus rótulos, definições e comentários somente na língua inglesa.

5. RESUMO DO CAPÍTULO

Neste capítulo é apresentado o processo de escolha de um modelo de representação para recursos digitais da editoração científica como dados abertos conectados. O processo foi guiado em conformidade às Melhores Práticas de Publicação na Web de Dados (LÓSCIO; BURLE; CALEGARI, 2019) e às diretrizes de reutilização de modelos de organização e representação de dados abertos conectados (HYLAND; ATEMEZING; VILLAZÓN-TERRAZAS, 2019). Logo, ao se utilizar o repositório LOV (VANDENBUSSCHE *et al.*, 2017), buscou-se por modelos adequados à representação, à organização e à publicação de registros de editoração científica.

Atentando às diretrizes estabelecidas, definiu-se a BIBO como o modelo mais adequado. A BIBO é constituída por um conjunto de conceitos e propriedades para descrever semanticamente citações e referências bibliográficas em consonância ao modelo RDF e à Web de Dados (D'ARCUS; GIASSON, 2019).

Cabe ressaltar que a BIBO atende a sete das oito diretrizes apresentadas por Hyland; Ateazing e Villazón-Terrazas (2019) para adoção de modelos de representação na Web de Dados. Pontualmente, na BIBO observa-se a preocupação com as diretrizes: **1)** vocabulários precisam ser documentados; **2)** vocabulários deveriam ser autodescritivos; **3)** vocabulários deveriam ser usados em vários conjuntos de dados; **4)** vocabulários deveriam ser acessíveis de forma perene; **5)** vocabulários deveriam ser publicados por um grupo ou organização confiável; **6)** vocabulários deveriam ter URLs persistentes; e **7)** vocabulários deveriam ter uma política de versionamento.

Salienta-se ainda que a partir de 2018 a BIBO é mantida pela Dublin Core Metadata Initiative (DUBLIN CORE METADATA INITIATIVE, 2019). Dada a reputação e os serviços prestados pela referida organização, pressupõe-se o contínuo desenvolvimento e o uso regular da BIBO em pesquisas da Ciência da Informação. Ou seja, diante dessa perspectiva, pressupõe-se que o uso da BIBO como modelo de representação será incentivado em diversos estudos. Como exemplo, conjectura-se que a publicação de dados abertos de editoração científica mediada pela BIBO guiará semanticamente os consumidores de dados na exploração informacional de milhares de recursos disponibilizados na Web de Dados em pesquisas bibliométricas e/ou cientométricas.

Convém destacar que a BIBO apresenta algumas limitações quanto a sua adoção, principalmente no que se refere à diretriz **vocabulários deveriam ser descritos em mais de um idioma**. Neste sentido, não foram observados indícios de multilinguismo da referida ontologia em seus elementos constituintes. Contudo, tal fato aponta para uma lacuna a ser explorada em futuras pesquisas da Ciência da Informação, com a finalidade de traduzir o vocabulário em vários idiomas, promovendo a utilização deste modelo de representação em contextos linguísticos variados.

6. AGRADECIMENTOS

À Fundação Araucária, pelo suporte financeiro ao projeto de pesquisa intitulado “Curadoria Digital e Dados Abertos Conectados: um estudo da preservação de recursos digitais na Web de Dados para estudos cientométricos” (FA - Convênio 49773/19FA).

7. REFERÊNCIAS

ABELE, A. *et al.* **Linking Open Data Cloud**. Disponível em: <http://lod-cloud.net>. Acesso em: 20 nov. 2019.

AUER, S. Introduction to lod2. *In*: AUER, S.; BRYL, V.; TRAMP, C (ed). **Linked Open Data: creating knowledge out of interlinked data**. [S.l.]: Springer, 2014. 215p. DOI: <https://doi.org/10.1007/978-3-319-09846-3>.

BANDEIRA, P. M.; FREIRE, I. M. Movimento de Acesso Aberto no Brasil: contribuição do Instituto Brasileiro de Informação em Ciência e Tecnologia a partir da implementação do Sistema Eletrônico de Editoração de Revistas. **Pesquisa Brasileira em Ciência da Informação e Biblioteconomia**. v. 12, n. 1, p. 057-067, 2017.

BIZER, C.; HEATH, T.; BERNERS-LEE, T. Linked data - the story so far. **International Journal of Semantic Web and Information Systems**, v. 5, n. 1, p. 1–22, 2009.

BRICKLEY, D.; MILLER, L. **FOAF Vocabulary Specification 0.99**. Disponível em: <http://xmlns.com/foaf/spec/>. Acesso em: 20 nov. 2019.

D'ARCUS, B.; GIASSON, F. Bibliographic Ontology Specification. In: GIASSON, F (ed.). **The Bibliographic Ontology**, 04 nov. 2009. Disponível em: <http://bibliontology.com>. Acesso em: 20 nov. 2019.

DUBLIN CORE METADATA INITIATIVE. **DCMI to Maintain the Bibliographic Ontology (BIBO)**. Disponível em: <http://www.dublincore.org/news/2018/11-30-bibo>. Acesso em: 20 de nov. 2019.

HYLAND, B.; ATEMEZING, G.; VILLAZÓN-TERRAZAS, B. **Best Practices for Publishing Linked Data**. Disponível em: <https://www.w3.org/TR/ld-bp>. Acesso em: 20 nov. 2019.

LINKED DATA. Linked data – Wikipédia, a enciclopédia livre. Disponível em: https://pt.wikipedia.org/wiki/Linked_data. Acesso em: 20 nov. 2019.

LINKED OPEN VOCABULARIES. **Linked Data Vocabularies**. Disponível em: <https://lov.linkeddata.es/dataset/lov>. Acesso em: 20 nov. 2019a.

LINKED OPEN VOCABULARIES. **Linked Data Vocabularies**. Disponível em: <https://lov.linkeddata.es/dataset/lov/terms?q=periodical>. Acesso em: 20 nov. 2019b.

LÓSCIO, B. F.; BURLE, C.; CALEGARI, N. Data on the Web Best Practices. **W3C Recommendation**, 31 jan. 2017. Disponível em: <https://www.w3.org/TR/dwbp>. Acesso em: 20 nov. 2019.

PACKER, A. Os desafios da profissionalização. **SciELO em Perspectiva**, 16 jun. 2014. Disponível em: <https://blog.scielo.org/blog/2014/06/16/os-desafios-da-profissionalizacao/>. Acesso em: 20 nov. 2019.

SCHMACHTENBERG M., BIZER C., PAULHEIM H. Adoption of the linked data best practices in different topical domains. *In*: MIKA, P. *et al.* (ed.) **The Semantic Web**: ISWC 2014. [S.I.]: ISWC, 2014, p. 245-260. Lecture Notes in Computer Science, v. 8796.

VANDENBUSSCHE, P.-Y *et al.* Linked Open Vocabularies (LOV): a gateway to reusable semantic vocabularies on the Web. **Semantic Web – Interoperability, Usability, Applicability**, v. 8, n. 3, p. 437-452, 2017.

W3C. **Linked Data**. Disponível em: <https://www.w3.org/standards/semanticweb/data>. Acesso em: 04 set. 2019.

8. SOBRE OS AUTORES



Sandro Rautenberg

Lattes: <http://lattes.cnpq.br/1550342830820869>

ORCID: <https://orcid.org/0000-0002-2375-9365>

Sandro Rautenberg. Professor não-titular – UNICENTRO/PR. Formação. Cientista da Computação e Doutor em Engenharia do Conhecimento pela Universidade Federal de Santa Catarina. Estágio pós-doutoral (Universidade de Leipzig/Alemanha - CAPES - BEX 18228/12-7), tema - ontologias para Linked Open Data. Interesses. Pesquisa nas áreas da Ciência da Computação e da Ciência da Informação, principalmente: Engenharia e Gestão do Conhecimento, Machine Learning, ontologias, Linked Open Data e Cientometria.



Lucélia de Souza

Lattes: <http://lattes.cnpq.br/8989736627689345>

Orcid: <https://orcid.org/0000-0002-9190-7411>

Lucélia de Souza. Professora Adjunta UNICENTRO/PR. Graduação em Informática pela UEPG/PR. Especialização em Administração de Banco de Dados pela UEPG/PR. Mestrado e Doutorado em Informática na UFPR. Pesquisa nas linhas Banco de Dados e Interação Humano-Computador do grupo de pesquisa Tecnologia de Informação e na linha de pesquisa Tecnologias em Web Semântica e Dados Abertos do Laboratório de Aplicações Semânticas e Distribuídas (LASED).

9. COMO CITAR

RAUTENBERG, S.; SOUZA, L. Um estudo para representação de dados abertos de editoração científica na web de dados. *In*: SHINTAKU, M.; SALES, L. F; COSTA, M. (org). **Tópicos sobre dados abertos para editores científicos**. Botucatu, SP: ABEC, 2020. p. 47-63. DOI: 10.21452/ 978-85-93910-04-3.cap5.

Web semântica e linked data como aliados na consolidação da Ciência Aberta

José Eduardo Santarem Segundo

1. INTRODUÇÃO

A necessidade e a proeminência da Ciência Aberta não estão mais em dúvida, já é algo que faz parte do cotidiano do desenvolvimento da ciência. Entretanto, é necessário provar que esse contexto não é apenas uma ideia, uma tendência, mas sim um novo cenário de realidade. Para provar, o quarto paradigma da ciência deve ser a chave definitiva.

O chamado quarto paradigma da ciência, e-Science ou ainda *Data-Driven Science*, que entende os dados como grande aliado e impulsionador para o avanço da ciência moderna, fora previsto por Jim Gray em 2007 (BELL, 2009, p. 13) e tem estado cada dia mais presente nas ações do mundo atual.

Todo esse novo contexto de compartilhamento e uso de dados não pode ser tratado exatamente como uma novidade. Há muitas teorias e pesquisas, algumas nem tão mais recentes, que têm pautado os estudos baseados em dados como um novo paradigma de pesquisa e desenvolvimento. Por isso é notável um grande interesse por vários segmentos de comunidades distintas em também tornar seus dados reutilizáveis. A área científica engrossa esse grupo, tornando públicos os dados de pesquisa.

Entende-se que a partir do momento em que houver um fluxo contínuo de estruturação, publicação e consumo (reúso) de dados científicos, haverá clareza sobre o processo de aceleração de descobertas, além de todas as outras vantagens que a Ciência Aberta proporciona à pesquisa científica. Essas vantagens podem ser facilmente identificadas nos outros capítulos que compõem este livro.

De acordo com Albagli (2015, p. 10), “As abordagens da Ciência Aberta implicam superar a perspectiva de pensar a ciência a partir da sua produtividade intrínseca. Além disso, provocam o abalo de hierarquias, de fontes estabelecidas de autoridade e reputação, dando foco às relações entre ciência e poder e, mais amplamente, entre saber e poder”.

Por hora, neste texto, não há proposição de discutir a Ciência Aberta, mas sim os modos de instrumentalizá-la e torná-la factível, para que se torne recurso fundamental na construção da ciência e para que seja não só aceita, como também requisitada e intrinsecamente inserida em novos paradigmas de evolução da humanidade.

Nos dias de hoje, instrumentalizar, em grande parte das vezes, está caracterizado como atribuir uma tipologia a um contexto, ou uma estrutura organizacional e computacional, dentro de um cenário em que a Internet é tida como o principal instrumento alavancador da Ciência Aberta.

Já em 2001, Berners-Lee, Hendler e Lassila, no primeiro texto sobre Web Semântica publicado na revista Scientific American, diziam que:

[...] se projetada adequadamente, a Web Semântica pode ajudar a evolução do conhecimento humano como um todo. O esforço humano é pego em uma tensão eterna entre a eficácia de pequenos grupos agindo de forma independente e a necessidade de se unir à comunidade em geral. Um pequeno grupo pode inovar de maneira rápida e eficiente, mas isso produz uma subcultura cujos conceitos não são entendidos por outros.

A ideia dos autores está totalmente alinhada com a ideia de Ciência Aberta, de tal modo que:

Coordenar ações em um grande grupo, no entanto, é dolorosamente lento e requer uma enorme quantidade de comunicação. O mundo

trabalha em todo o espectro entre esses extremos, com a tendência de começar pequeno - a partir da ideia pessoal - e avançar para uma compreensão mais ampla ao longo do tempo. Um processo essencial é a união de subculturas quando uma linguagem comum mais ampla é necessária. Frequentemente, dois grupos desenvolvem conceitos muito semelhantes de maneira independente, e descrever a relação entre eles traz grandes benefícios.

Portanto, é possível, a partir dessas afirmações, formular a seguinte pergunta: “As tecnologias da Web Semântica e Linked Data podem instrumentalizar e fortalecer diretamente o desenvolvimento e consolidação da Ciência Aberta?”.

A ideia é responder a essa pergunta a partir do “fluxo organizacional para publicação de dados” proposto por Santarem Segundo (2018, p. 138), quando este afirma que:

Constituir um projeto robusto é garantir uma rotina de publicação de dados, dentro de um contexto que atenda às necessidades da comunidade que tem interesse nos dados e que seja revisto constantemente de forma a melhorar ainda mais o atendimento a essa comunidade, além disso é garantir que os dados atendam a padrões de formalização, tenham licenças apropriadas, dotados de requisitos mínimos de qualidade e que sejam preservados para uso perene.

Portanto, o objetivo deste capítulo é “reapresentar o fluxo organizacional para publicação de dados, modelo baseado em Web Semântica e Linked Data, como modelo norteador para consolidação da publicação de dados científicos e consequentemente da Ciência Aberta”.

Está compreendido no objetivo acima um conjunto de procedimentos que tornam o processo de publicação de dados científicos mais claros e organizados, a fim de instruir editores científicos a avancarem na prática de publicação de dados de pesquisa. Entendem-se aqui não somente os dados brutos, recursos e matéria-prima das pesquisas, mas também os resultados de pesquisa, os próprios *papers* que são o método mais tradicional de apresentar resultados de pesquisa atualmente.

Apesar do fluxo organizacional de Santarem Segundo (2018) não ser dedicado a dados científicos, apresenta-se aqui um contexto particularmente direcionado

a este fim, indo no mesmo sentido do compreendido por Albagli (2015, p. 14), quando afirma que

[...] historicamente, é no compartilhamento e na abertura à produção coletiva e não individual que melhor se desenvolvem a criatividade e a inovatividade. A complexidade dos desafios científicos e a urgência das questões sociais e ambientais que se apresentam às ciências impõem, por sua vez, facilitar a colaboração e o compartilhamento de dados, informações e descobertas.

É fato, portanto, posicionar as tecnologias da Web Semântica e a ligação de dados como fatores fundamentais para consolidar o compartilhamento de dados científicos.

2. WEB SEMÂNTICA E LINKED DATA

A semântica compreende o estudo do significado; é a relação que se estabelece entre dois elementos. Quando transferido para o mundo da Web, a semântica é vista na relação que se estabelece entre recursos, atribuindo à relação um significado. Esses recursos normalmente são continuados como uma URI (um link único na Web) e a ligação entre eles é estabelecida por meio de um vocabulário, que pode ser parte ou não de uma estrutura organizacional como uma ontologia.

A Web Semântica é caracterizada pela busca da compreensão, por agentes de software, do significado do contexto para cada recurso em ambientes Web. É a ideia de que ferramentas e aplicações possam trabalhar em favor dos humanos com objetivo de melhorar semanticamente os processos de recuperação da informação.

Faz parte da ideia e da estrutura da Web Semântica a padronização e organização da informação, a interoperabilidade, a proposição ao reúso, o estabelecimento de regras de inferências e, também, a serendipidade (capacidade de descobrir/encontrar algo ao acaso). Este autor considera a inferência e a serendipidade os grandes propulsores da inteligência e raciocínio automatizado que a Web Semântica pode proporcionar.

É fato dizer então que a Web Semântica está amparada em quatro principais fundamentos. São eles: linguagem estruturada como a XML (ou outra que possa substituí-la como JSON, por exemplo); organização conceitual do RDF; persistência dos dados que o uso de URIs garante; e formalização de um modelo organizacional que o uso de ontologias permite.

O desafio da Web Semântica, portanto, é fornecer uma linguagem que expresse dados e regras de raciocínio sobre os dados e que permita que as regras de qualquer sistema de representação de conhecimento existente sejam exportadas para a Web. A Web Semântica propõe que dados publicados possam ser lidos e compreendidos não apenas por humanos, mas também por máquinas.

Para efetivar a construção de conjuntos de dados interligados utilizando-se dos padrões da Web Semântica, em 2006, Tim Berners-Lee propôs um conjunto de práticas para publicação e interligação de dados na Web, que foi chamado de Linked Data.

Entre as melhores práticas, estão:

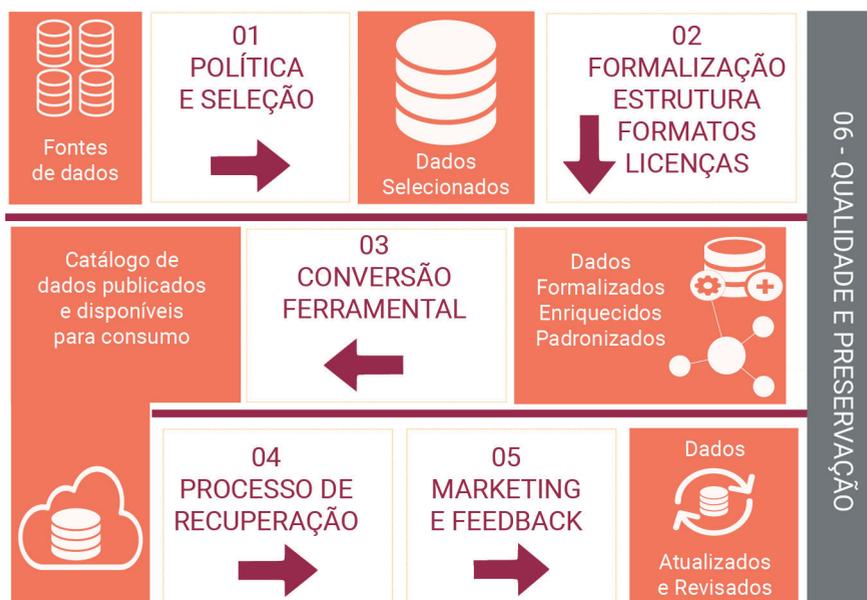
- Usar URIs como nomes para os itens.
- Usar URIs HTTP para que as pessoas possam consultar esses nomes.
- Quando alguém consulta uma URI, deve prover informação RDF útil.
- Incluir sentenças RDF com links para outras URIs, a fim de permitir que itens relacionados possam ser descobertos.

A partir dessas melhores práticas nasceu um projeto chamado Linked Open Data, que inicialmente foi constituído como um grupo de trabalho do W3C e posteriormente tornou-se um projeto independente, baseado em uma grande nuvem de dados interligados, que conecta datasets de várias organizações e instituições. O projeto foi liderado por Heath e Bizer (2011), que compreendem que o LOD tem como princípio a conexão de dados estruturados na Web, permitindo estabelecer links entre itens de diferentes fontes de dados para formar um único espaço de dados global.

3. FLUXO ORGANIZACIONAL PARA PUBLICAÇÃO DE DADOS

Publicado em 2018, o fluxo organizacional de Santarem Segundo, apresentado pela Figura 6.1, tem como premissa organizar o caminho por qual um projeto de publicação de dados deve passar, servindo como subsídio a projetos de publicação de dados em formato aberto, semântico e ligados entre si. É importante ressaltar que grande parte do segmento deste capítulo é baseado no texto “Web Semântica: fluxo para publicação de dados abertos e ligados” de Santarem Segundo (2018).

Figura 6.1 - Fluxo Organizacional para Publicação de Dados.



Fonte: Santarem Segundo (2018)

Como subsídio para consolidação da Ciência Aberta, o fluxo organizacional é apresentado aqui como sugestão para que se disponibilizem dados e resultados de pesquisa com o propósito de que esses dados possam ser reutilizados posteriormente por outrem.

O fluxo organizacional de publicação de dados, que pode ser chamado de projeto de publicação de dados, não é simples, e pode envolver várias pessoas ou divisões de uma organização.

Por meio da Figura 6.1 apresenta-se a proposta de divisão do fluxo organizacional com várias tarefas, que se organizam divididas em seis fases:

Fase 1 – Política e seleção;

Fase 2- Formalização, Estrutura, Formatos e Licenças;

Fase 3 – Conversão Ferramental;

Fase 4 – Processo de Recuperação;

Fase 5 – Marketing e Feedback;

Fase 6 – Qualidade e Preservação.

3.1 Política e seleção

O primeiro passo em projetos de publicação de dados compreende estruturar uma equipe ou definir pessoas que terão que tomar decisões técnicas e administrativas sobre o tema. Cabe a essas pessoas definir uma política que fortaleça a cultura de publicação de dados científicos.

Constituir uma cultura de publicação de dados é levar ao conhecimento de todos os colaboradores de uma instituição o ideal de divulgar dados de forma aberta na internet, além de conscientizar que esse tema deva ser discutido frequentemente e que pode ser rediscutido a qualquer momento. No contexto da publicação de dados científicos deve-se ficar claro para o pesquisador o papel dele no processo.

É importante que o pesquisador colabore com a equipe técnica na identificação do público que pode ter interesse nos dados, na seleção dos dados a serem disponibilizados, na definição da granularidade e na periodicidade de entrega desses dados.

Completa-se essa fase quando houver um pacote de dados que seja um exemplo ou represente uma amostra do que será publicado em definitivo. Essa amostra garantirá à equipe técnica o início do processo de publicação de dados.

3.2 Formalização, Estrutura, Formatos e Licenças

É nessa fase que o dado receberá todo o tratamento técnico, é onde a equipe de TI deverá atuar fortemente para que aconteça uma transformação no dado, incluindo, quando necessário, uma mudança de formato, a ligação com outros dados para que possa ser enriquecido com recursos vindos de outras bases de dados e onde serão atribuídas licenças de uso, para que a comunidade que vai consumir tenha certeza de que pode utilizá-los.

É necessário ficar claro que essa fase é compreendida por muitos projetos como a única de um processo de publicação de dados. Apesar de sua importância técnica, ela é apenas uma das partes do processo. Sua importância fica mais evidente quando se parte da ideia que a maneira como os dados serão disponibilizados pode incentivar ou desestimular o consumo dos mesmos, dependendo de como serão publicados.

Os dados, em geral oriundos de laboratórios ou fontes diferentes, precisam passar por um processo de transformação. É necessário identificar os formatos por meio dos quais eles serão recebidos, associar termos similares, padronizar unidades de medidas, usar vocabulários controlados e torná-los compreensíveis para quem irá consumi-los.

Na sequência, é necessário formalizá-los dentro de uma estrutura organizacional semântica que deve ser baseada em uma ontologia. Usar vocabulários reconhecidos internacionalmente é um dos pontos cruciais para que os dados possam ser coletados e reutilizados para aplicações diversas. O projeto *Linked Open Vocabularies* (LOV)²³ é um ótimo recurso para identificar vocabulários conhecidos e usá-los para dar significado (semântica) à formalização dos dados e construção (ou adaptação) de ontologias.

Após a estruturação, organização e formalização, é importante que eles sejam disponibilizados em um formato de serialização computacional. Disponibilizar dados em formatos como XML, JSON e outras variedades destes atendem ao proposto dessa fase.

É também nessa fase que se atribui uma licença aos dados. É de extrema importância e relevância para quem consumirá os dados que estes expressem como

23 Disponível em: <https://lov.linkeddata.es/dataset/lov/>. Acesso em: 11 dez. 2019

eles podem ser utilizados. É fundamental garantir a segurança e liberdade para o uso dos dados pelos consumidores por meio da clareza da licença atribuída.

3.3 Conversão Ferramental

É nessa fase que o dado passa a ser público, quando ele efetivamente é disponibilizado para consumo.

As maneiras mais simples e básicas remetem à simples disponibilização de arquivos de dados, ou pacotes compactados, que contemplam os arquivos e que mesmo nesse formato podem ser semânticos, mediante o uso de arquivos em serializações adequadas. O que se espera é que tenhamos um conjunto de ferramentas que permitam acesso ao dado das mais variadas formas, tanto para acesso por humanos quanto por máquinas.

Disponibilizar um catálogo com a estrutura dos dados pode favorecer muito o acesso a eles. Grande parte dos *datasets* ao redor do mundo utiliza atualmente a ferramenta CKAN para esta parte do processo, entretanto há outras que podem ser utilizadas. Bancos de dados de grafos disponibilizam acesso via *Sparql EndPoint*, que permitem acesso aos dados por consultas via *Sparql*, tanto por *API's* quanto por *Webservices*. Quando se atinge esse nível de maturidade, de permitir consultas via *Sparql*, há um ganho expressivo na recuperação de dados, principalmente em relação ao contexto semântico de significado dos dados.

A conversão ferramental caracteriza-se principalmente pela constituição de uma cadeia de processos que permite um ciclo constante de publicação de dados, ou seja, a forma de tornar cíclica a tarefa de obtenção e posterior publicação dos dados, agregando as três primeiras fases do fluxo organizacional. Em alguns casos, a produção de dados científicos pode não ter uma periodicidade bem definida, mas é importante que os pesquisadores tenham a certeza de que a informação precisa ser repassada à equipe de publicação dos dados.

Atualmente não há ferramentas disponíveis que possam realizar essa tarefa. Para isso é necessário a construção de scripts de programação que possam automatizar a esteira de procedimentos. Quando isso não for possível, é necessário um cronograma que possa estabelecer regras claras para que o procedimento possa ser cíclico e natural dentro do processo.

3.4 Processo de Recuperação

Essa fase do processo é a primeira que não faz parte da estruturação dos dados, mas sim está posicionada já após os dados serem publicados. Apesar de parecer que o processo de recuperar e acessar os dados seja tarefa de quem tem interesse nos mesmos, pode ser papel da equipe de publicação de dados constituir elementos que possam ajudar no uso deles para uma série de aplicações.

É válido o desenvolvimento de aplicações para que os dados possam ser consumidos pelos pesquisadores da própria organização, o que gera uma retroalimentação institucional. Outro processo enriquecedor é a criação de axiomas que permitam inferência sobre os dados e, principalmente, a aplicação de modelos e técnicas de *machine learning* (aprendizado de máquinas/inteligência artificial) que permitam encontrar padrões e informações até então não percebidos nos dados da própria organização.

Melhorar e enriquecer o processo de consumo dos dados é valorizar todo o trabalho que foi realizado nas três primeiras fases do fluxo organizacional.

3.5 Marketing e Feedback

Muitas vezes há uma grande demanda de trabalho para que se possa publicar dados, porém a maneira como eles ficam disponíveis para a comunidade não é viável o suficiente para que seja usufruído plenamente.

Divulgar nas principais mídias, informar os possíveis interessados, publicar os dados em catálogos de grande acesso, oferecer os dados para serem trabalhados em eventos como Hackatons são algumas das técnicas de marketing que podem ser utilizadas para que a comunidade tenha conhecimento sobre os dados que estão sendo publicados. É de fundamental importância que quem precisa do dado saiba exatamente onde encontrá-los.

A partir do momento em que os dados passam a ser consumidos, há um outro fator que pode ser muito importante para que eles possam ser cada vez mais utilizados pela comunidade que os consome: é o processo de feedback. Em outras palavras, trata-se de criar uma estrutura em que seja possível receber informações da comunidade que está consumindo os dados.

Criar rotinas que analisem o consumo também é bastante importante, visto que esse tipo de atividade permite entender o que realmente tem despertado o interesse da comunidade, quais são os dados de maior circulação, qual a granularidade do dado que mais interessa, qual a forma de acesso mais utilizada etc.

Todo e qualquer *feedback* que possa ser recebido deve ser discutido em grupo e, quando necessário, pode gerar alterações nas três primeiras fases do projeto de publicação de dados. É importante que o pesquisador saiba exatamente qual dos dados gerados tem maior audiência.

Ressalta-se que as solicitações da comunidade podem ser desde parte dos dados que não ficaram disponíveis, passando pela granularidade dos mesmos, o formato de disponibilização, sugestões de alterações na ontologia ou, ainda, nas ferramentas utilizadas para que o dado fique disponível.

É muito importante que se feche o ciclo de publicação de dados com o máximo de atenção às necessidades de quem consome os dados.

3.6 Qualidade e Preservação

A fase seis transcende todas as outras fases anteriores, e está diretamente relacionada a todas elas. Entende-se que dar qualidade e preservar os dados publicados é um item da maior relevância que pode existir em um projeto.

Entende-se que pensar nas questões que envolvem qualidade e preservação está diretamente relacionado à confiança que se pode ter na base. Confiança é um dos elementos mais significativos no processo de uso e reúso de uma base de dados, e tem sido pauta constante de discussão, dada a grande variedade de problemas encontrados em base de dados publicadas em formato aberto e semântico ao redor do mundo, mais especificamente quando tratamos de *datasets* publicados no Linked Open Data.

A preservação dos dados, que acaba sendo intrínseca à qualidade deles e, em alguns casos, compõe parte dos requisitos de qualidade, é um fator determinante para garantir a longevidade dos dados publicados. Sayão e Sales (2012, p. 180) confirmam que “o principal desafio recai na necessidade de se preservar não somente

o conjunto de dados, mas de preservar, sobretudo, a capacidade que ele possui de transmitir conhecimento para uso futuro das comunidades interessadas”.

A qualidade e a preservação garantem a integridade do dado e, além da confiança, estimulam o reuso da base publicada. Em geral, uma base íntegra, que garante preservação e tem qualidade, receberá conexões advindas de outras bases, servindo também como referência para enriquecer dados de bases de outrem.

4. CONSIDERAÇÕES FINAIS

A partir da pergunta apresentada no início deste texto (“As tecnologias da Web Semântica e Linked Data podem instrumentalizar e fortalecer o desenvolvimento e consolidação da Ciência Aberta?”) apontam-se as considerações finais.

É certo que a Ciência Aberta depende de um ciclo de uso e reuso de dados científicos, parte de um processo cíclico de retroalimentação científica. É assim que se obtém a certeza da necessidade de se fazer Ciência Aberta e dos resultados que ela proporciona. Considerando esse contexto, é importante apresentar ferramentas que possam facilitar o processo de uso e reuso dos dados. Entretanto, para isso é necessário que do ponto de vista tecnológico as barreiras de acesso sejam quebradas e, principalmente, que procedimentos de disponibilização de dados sejam constituídos.

Constituir modelos semânticos alinhados às melhores práticas do Linked Data e enriquecer dados de pesquisa ligando seus conceitos a outras bases de dados já publicadas - atribuindo a esses dados qualidade e licenças claras e bem definidas - fortalece a ideia de reuso, além de dar visibilidade, permitindo não apenas que o processo técnico de disponibilização dos dados seja realizado, como também que as pesquisas desenvolvidas em uma organização, assim como seus pesquisadores, sejam reconhecidas universalmente.

Mais do que publicar resultados de pesquisa em repositórios de documentos, publicar dados em formato de *datasets* semânticos é garantir que eles sejam fonte de enriquecimento para outros dados, e conseqüentemente recuperados por meios de outros *datasets*.

A proposição do uso do fluxo organizacional de Santarem Segundo (2018) no contexto da Ciência Aberta pretende propor um modelo a ser seguido por

equipes que tenham interesse em projetos de publicação de dados de pesquisa em formato aberto e semântico, garantindo qualidade e perpetuidade à fonte de dados de uma organização que publica dados de pesquisa.

Formalizar um fluxo de publicação de dados de pesquisa propõe a disponibilização, o consumo, o uso e reúso e, principalmente, uma cultura de consumo de dados de pesquisa que alimenta a ideia da Ciência Aberta e consolida suas práticas de compartilhamento e novas descobertas científicas.

5. REFERÊNCIAS

ALBAGLI, S. Ciência Aberta em questão. *In*: ALBAGLI, S.; MACIEL, M. L.; ABDON, A. H. (Orgs). **Ciência Aberta, questões abertas**. Brasília, DF: IBICT; Rio de Janeiro: UNIRIO, 2015. p. 27-58.

BELL, G. Foreword. *In*: HEY, T.; TANSLEY, S.; TOLLE, K. **The Fourth Paradigm: data-intensive scientific discovery**. redmond, washington: microsoft research, 2009. Disponível em: <https://bit.ly/1iD63DJ>. Acesso em: 11 dez. 2019.

BERNERS-LEE T.; LASSILA, O.; HENDLER, J. The semantic web. **Scientific American**, New York, v. 5, 2001.

BERNERS-LEE, T. **Linked data principles**. 2006. Disponível em: <https://bit.ly/1x6N7XI>. Acesso em: 11 dez. 2019.

HEATH, T.; BIZER, C. **Linked Data: evolving the web into a global data space**. [S.l.:S.n.], 2011. Disponível em: <http://linkeddatatoolkit.com/editions/1.0/>. Acesso em: 11 dez. 2019.

SANTARÉM SEGUNDO, J. E. Web Semântica: fluxo para publicação de dados abertos e ligados. **Informação em Pauta**, v. 3, n. Especial, p. 117-140, 26 nov. 2018. Disponível em: <https://doi.org/10.32810/2525-3468.ip.v3iEspecial.2018.39721.117-140>. Acesso em: 11 dez. 2019.

SAYÃO, L. S. F.; SALES, L. F. Digital curation: a new platform for digital preservation of research data. **Informação & Sociedade: Estudos**, Paraíba, v. 22, n. 3, 2012. Disponível em: <https://bit.ly/2KLvOlq>. Acesso em: 11 dez. 2019.

6. SOBRE OS AUTORES



José Eduardo Santarem Segundo

Lattes: <http://lattes.cnpq.br/5562746387565465>

Orcid: <https://orcid.org/0000-0003-3360-7872>

Docente da Universidade de São Paulo (USP) e do Programa de Pós-Graduação em Ciência da Informação da UNESP Marília. Pós-Doutor pela Western University/Canadá. Doutor e Mestre em Ciência da Informação UNESP-Marília/SP; Bolsista de Produtividade em Pesquisa PQ-2 do CNPq. Coordenador do GT8 - Informação e Tecnologia, da Associação Nacional de Pesquisa e Pós-Graduação em Ciência da Informação (ANCIB). Líder do NEWSDA - Núcleo de Estudos em Web Semântica e Dados Abertos.

7. COMO CITAR

SANTAREM SEGUNDO, J. E. Web semântica e linked data como aliados na consolidação da Ciência Aberta. *In*: SHINTAKU, M.; SALES, L. F; COSTA, M. (org). **Tópicos sobre dados abertos para editores científicos**. Botucatu, SP: ABEC, 2020. p. 65-78. DOI: 10.21452/ 978-85-93910-04-3.cap6.

Quais direitos podem ser alegados sobre dados de pesquisa?

Paulo Cezar Vieira Guanaes

1. INTRODUÇÃO

Há uma tendência a se considerar que a propriedade intelectual ganha protagonismo no movimento de abertura de dados de pesquisa, a nem tão nova agenda do movimento pela Ciência Aberta. Tal abertura comporta o depósito de dados em repositório, o tratamento, o acesso, o compartilhamento e o reuso desses dados, seja em conexão com um artigo publicado, seja pela intenção de tornar os dados acessíveis publicamente. Nesta operação, questões de propriedade intelectual podem se transformar em preocupação para pesquisadores que fazem o depósito, instituições que lhes dão o emprego, financiadores e demais pesquisadores que pretendam reutilizar dados de pesquisa. Em ambos os contextos, indagações são comuns: quais são os direitos legais implícitos nos dados? Como compartilhar e garantir um reuso produtivo de dados? Que licenças usar para permitir acesso a dados?

Antes de quaisquer considerações, a Organização Mundial da Propriedade Intelectual (Ompi) define propriedade intelectual como “a soma dos direitos relativos às obras literárias, artísticas e científicas, às interpretações dos artistas intérpretes e às execuções dos artistas executantes, aos fonogramas e às emissões de radio-difusão, às invenções em todos os domínios da atividade humana, às descobertas científicas, aos desenhos e modelos industriais, às marcas industriais, comerciais e de serviço, bem como às firmas comerciais e denominações comerciais, à proteção contra a concorrência desleal e todos os outros direitos inerentes à atividade intelectual nos domínios industrial, científico, literário e artístico” (OMPI, 2019).

2. QUAIS SÃO OS DIREITOS LEGAIS IMPLÍCITOS NOS DADOS?

Os direitos que entram no radar de proteção da propriedade intelectual e que podem ser reivindicados sobre dados de pesquisa são: segredos industriais, direitos autorais, que incluem direitos sobre bancos de dados e programas de computador, e direitos de patentes. Também pode ser alegado o direito sobre dados relacionados à segurança nacional e ainda há a possibilidade de aqueles que buscam reutilizar dados se depararem com direitos protegidos por contratos, cujas cláusulas poderão conter alguma restrição de uso (CARROL, 2015; GUANAES, 2018).

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, 2018).

No Brasil, em 2018, foi sancionada a Lei Geral de Proteção de Dados (BRASIL, 2018). Dentre os 65 artigos da lei, o artigo 13 é o único que trata de “estudos em saúde pública”. Ele proíbe, nesta hipótese, a revelação de dados pessoais e sua transferência a terceiros. Mas estabelece que dados pessoais de participantes de pesquisa podem ser reutilizados, desde que os participantes deem consentimento e que os dados passem por uma anonimização, processo de remoção de qualquer vestígio que possa levar à identificação do participante da pesquisa. No entanto, mais voltada a modelos de negócio baseados em dados do que à pesquisa científica, a lei só entra em vigor em agosto de 2020.

3. LICENCIAMENTO DE DADOS

Cada vez mais editores ou agências de fomento à pesquisa exigem a abertura de dados de pesquisa (FORTNEY, 2016; BALL, 2014). Para tanto, indicam repositórios públicos – como o Dryad, Zenodo, Figshare – para o depósito de dados, bem como nota-se uma crescente preocupação com a formulação de políticas para compartilhamento de dados, seja por editores, como a PLoS

(2014), ou instituições científicas, como o pioneiro *National Institutes of Health* (NIH), que em 2003 desenvolveu uma de cunho generalista (JORGE; ALBAGLI, 2018). Assim, o licenciamento para (re)utilização de dados de pesquisa ganha importância e precisa oferecer termos de uso claramente definidos, cabendo ao proprietário desses direitos marcar os dados com as permissões associadas. Do contrário, o reuso ficará dificultado quando direitos autorais ou direitos similares estiverem em questão, situação agravada pelas diferentes jurisdições nacionais que aplicam diferentes critérios de criatividade, técnica, trabalho e custos (BALL, 2014).

Neste cenário, as licenças *Creative Commons* (da organização global homônima que as fornece padronizada e gratuitamente), as *Open Data Commons* (da *Open Knowledge Foundation*, que incentiva a abertura de dados de valor à sociedade), entre outras, como a *Open Government License* (criada pelo *National Archives* e usada no Reino Unido), têm sido adotadas por repositórios de dados e vêm cumprindo um papel decisivo na promoção do compartilhamento e reuso de obras criativas, educacionais e científicas (ver Quadro 7.1).

Quadro 7.1 - Algumas licenças para permissão de reuso de dados de pesquisa em bases legais

LICENÇA (CREATIVE COMMONS ²⁴)	SIGLA	CARACTERÍSTICAS
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. 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 copyleft, em que trabalhos derivados terão a mesma licença.
Atribuição Sem Derivações	CC BY-ND	Reuso 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.

24 As seis licenças "Atribuição" da Creative Commons têm em comum a obrigatoriedade do usuário dar o crédito pela criação original

LICENÇA (CREATIVE COMMONS ²⁵)	SIGLA	CARACTERÍSTICAS
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 reúsa 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-NCND	É a mais restritiva das seis licenças "Atribuição". Só permite downloads do trabalho original, que não pode ser alterado ou usado para fins comerciais. Deve-se atribuir o crédito pela criação original.
CC0 1 Universal Dedicção ao Domínio Público ²⁶	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 ²⁶		É um rótulo inserido na obra que "sinaliza" que ela está livre de direitos autorais e em domínio público.
LICENÇA (OPEN DATA COMMONS)	SIGLA	CARACTERÍSTICAS
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.

25 Não são consideradas licenças, mas instrumentos jurídicos para informar a opção do detentor de direitos autorais pelo domínio público

26 Não são consideradas licenças, mas instrumentos jurídicos para informar a opção do detentor de direitos autorais pelo domínio público

LICENÇA (OPEN DATA COMMONS)	SIGLA	CARACTERÍSTICAS
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: Elaborado com base em CREATIVE COMMONS, 2018a; CREATIVE COMMONS1, 2018b; UNIVERSITY LIBRARIE, 2019

4. DOMÍNIO PÚBLICO

Além de suas seis licenças, a *Creative Commons* disponibiliza as ferramentas CC0 1.0 Dedicção ao Domínio Público e Marca de Domínio Público. Ao optar pela primeira, o licenciante estará dedicando a obra que criou ao domínio público. Isto quer dizer que ele renuncia a todos os seus direitos previstos pelas leis de direitos autorais referentes à sua obra, em todo o mundo, na extensão que a legislação permitir. É conhecida como “Sem direitos reservados”. A segunda ferramenta serve para qualquer pessoa informar, por meio de um rótulo, que uma obra está livre de direitos autorais, isto é, que ela está em domínio público. A motivação para uso dessa ferramenta é permitir que as obras em domínio público sejam mais facilmente descobertas e reconhecidas como tal, mas não deve ser usada para a renúncia de direitos (CREATIVE COMMONS, 2018a; 2018b; BALL, 2004).

Segundo FORTNEY (2016), o compartilhamento de dados com a ferramenta CC0 1.0 Dedicção ao Domínio Público é cada vez mais incentivado (ver, p. ex., Dryad e BioMedCentral²⁷) por uma razão muito simples: quanto mais claramente as pessoas souberem que não irão ser processadas por reutilizarem dados, é mais provável que os dados sejam reutilizados.

27 Ver a respeito em: <<https://blog.datadryad.org/2011/10/05/why-does-dryad-use-cc0/>> e <<https://www.biomedcentral.com/about/policies/open-data>>.

5. REFERÊNCIAS

BALL, A. How to License Research Data. **DCC How-to Guides**. Edinburgh: Digital Curation Centre. 2014. Disponível em: <http://www.dcc.ac.uk/resources/how-guides/license-research-data>. Acesso em: 18 abr. 2019.

BRASIL. **Lei n. 13.709, de 14 de agosto de 2018**. Dispõe sobre a proteção de dados pessoais e altera a Lei n. 12.965, de 23 de abril de 2014 (Marco Civil da Internet). Brasília, DF: Presidência da República, 2018. Disponível em: http://www.planalto.gov.br/ccivil_03/_Ato2015-2018/2018/Lei/L13709.htm. Acesso em: 29 maio 2019.

CARROLL, M. 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: 6 jun. 2019.

CREATIVE COMMONS. Sobre as licenças. 2018a. Disponível em: https://creativecommons.org/licenses/?lang=pt_BR. Acesso em: 6 jun 2019.

CREATIVE COMMONS. CC0 1.0 Universal (CC0 1.0) Dedicção ao Domínio Público. 2018b. Disponível em: https://creativecommons.org/publicdomain/zero/1.0/deed.pt_BR. Acesso em: 8 jun. 2018.

FORTNEY, K. Who “owns” your data?. **University of California Office of Scholarly Communication Blog**, 8 set. 2016. Disponível em: <https://osc.universityofcalifornia.edu/2016/09/who-owns-your-data/#policies>. Acesso em: 7 jun. 2019.

GUANAES, P. C.V. (org.). **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: 6 jun. 2019.

JOGE, V. A.; ALBAGLI, S. Compartilhamento de dados de pesquisa em saúde: iniciativas do National Institutes of Health (NIH). **Reciis**, Rev. Eletron.Comun. Inf. Inov. Saúde., Rio de Janeiro, n. 12, v. 4, out./dez. 2018. Disponível em: <https://www.reciis.icict.fiocruz.br/index.php/reciis/article/view/1499/2237>. Acesso em: 7 jun. 2019.

OPEN DATA INSTITUTE (ODI). Reuser's Guide to Open Data Licensing. 2019. Disponível em: <https://theodi.org/article/reusers-guide-to-open-data-licensing/>. Acesso em: 8 jun. 2019.

ORGANIZAÇÃO MUNDIAL DA PROPRIEDADE INTELECTUAL (OMPI). What is Intellectual Property?. 2019. Disponível em: <https://www.wipo.int/about-ip/en/>. Acesso em: 4 jun. 2019.

PLOS MEDICINE. Data availability. 2014. Disponível em: <https://journals.plos.org/plosmedicine/s/data-availability>. Acesso em: 12 abr. 2019.

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: 7 jun. 2019.

6. SOBRE OS AUTORES



Paulo Cezar Vieira Guanaes

Lattes: <http://lattes.cnpq.br/3002783910003215>

Orcid: <https://orcid.org/0000-0003-1248-0462>

Doutorado (em curso) em Ciência da Informação, pelo IBICT/UFRJ, e Mestrado em Ciência - Informação e Comunicação em Saúde, pelo ICICT/Fiocruz. Possui graduação em Direito pela Universidade Estadual do Rio de Janeiro e curso de Formação de Tradutores Inglês-Português, pelo CCE-PUC-Rio. Jornalista e tradutor, é Editor Executivo da revista científica "Trabalho, Educação e Saúde", publicada pela Escola Politécnica de Saúde Joaquim Venâncio, da Fiocruz.

7. COMO CITAR

GUANAES, P. C. V. Quais direitos podem ser alegados sobre dados de pesquisa?. In: SHINTAKU, M.; SALES, L. F.; COSTA, M. (org). **Tópicos sobre dados abertos para editores científicos**. Botucatu, SP: ABEC, 2020. p. 79-85. DOI: 10.21452/978-85-93910-04-3.cap7.

Dados abertos e a Fundação de Amparo à Pesquisa do Estado de São Paulo

Milton Shintaku
Suzana Caetano da Silva Lanne

1. INTRODUÇÃO

A Fundação de Amparo à Pesquisa do Estado de São Paulo (Fapesp) tem uma longa história, prevista desde a Constituição Estadual de 1947 do estado de São Paulo, mas que foi criada somente em 1960 pela Lei Orgânica 5.918, de 18 de outubro de 1960, e regulamentada pelo Decreto 40.132, de 23 de maio de 1962, todas regulamentações estaduais. Assim é a fundação de fomento à pesquisa mais antiga do Brasil e uma das estaduais que mais se destacam pelos resultados obtidos. Monteiro, Lucas e Lecardelli (2019) relatam que a Fapesp é o segundo órgão de fomento que mais apoia a pesquisa no Brasil, a primeira entre as fundações estaduais, logo após o Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), que tem abrangência nacional.

Motoyama (1999) ressalta que, desde a sua criação, a Fapesp tem atuado ativamente na expansão e fomento à pesquisa no estado de São Paulo, aliada a outros órgãos de fomento, e tem seu êxito reconhecido por toda a sociedade. Como um dos fatores de sucesso, o autor ressalta o rigor científico utilizado pelo órgão na seleção de projetos, com vistas a manter certa independência em sua gestão. Nesse sentido, Hamburger (2004) relata que a Fapesp sempre foi pioneira, desbravando as novas fronteiras do conhecimento com apoio a

práticas sistemáticas de pesquisa científica e tecnológica, visando a melhoria da sociedade. Esse ponto foi ressaltado pela independência da fundação, apresentada desde a sua proposta inicial, que se configura como uma personalidade jurídica pública, mas com administração privativa e orçamento mantido por leis e regras estaduais.

A independência e lisura da Fapesp é mantida pela sua estrutura complexa, composta por dois conselhos diretores. Por meio de diretrizes, o Conselho Superior orienta o seu funcionamento, e o Conselho Técnico Administrativo opera as questões de pesquisas e fomentos. Com isso, a Fapesp busca atender à comunidade científica nos seus anseios de modo a subsidiar a sociedade que a mantém por meio de impostos. Em outras palavras, a Fapesp fomenta pesquisas que geram novos conhecimentos, métodos e técnicas com a finalidade de retorná-los à sociedade, enriquecendo-a.

Mesmo a Fapesp sendo um órgão estadual, possui projeção internacional devido a sua atuação, como destaca Domingues e da Costa (2016), para quem a participação da comunidade científica paulista nas decisões e a independência e constância financeira foram fatores primordiais para a visibilidade internacional da fundação. A atuação da Fapesp a coloca em posição de destaque no cenário internacional, com um dos principais representantes brasileiros de fomento à pesquisa.

Muito do destaque explica-se pela atuação da Fapesp nas tendências mundiais de ciências. A participação da comunidade científica paulista na fundação imprime às diretrizes orientações internacionais que a colocam no estado da arte das ciências em muitas áreas. Entre as tendências está a abertura das ciências, que desde o final do século passado tem alterado significativamente os processos científicos nos quais a Fapesp tem atuado desde o início, com ações voltadas ao acesso aberto e aos resultados de pesquisa.

2. A FAPESP E A CIÊNCIA ABERTA

A Ciência Aberta é mais que um movimento, visto que engloba várias ações que impactam nas atividades científicas, em que a tecnologia está muito presente. Historicamente, pode-se dizer que a Ciência Aberta é o resultado de outros movimentos nascidos ainda no final do século passado com os Arquivos Abertos (*Open Archives*) e Acesso Aberto (*Open Access*), resultantes

da evolução tecnológica e crise dos periódicos, respectivamente. A Ciência Aberta vai incorporando preceitos já existentes e propondo novos desafios à comunidade científica mundial. Lima e Lima (2013) apresentaram a sequência que culmina com a Ciência Aberta como um contínuo de evolução, em que cada novo movimento agrega os preceitos anteriores, sendo todos no mesmo contexto.

Fontes Abertas -> Dados Abertos -> Acesso Abertos -> Ciência Aberta

O intuito é tornar a ciência mais aberta, ampliando o compartilhamento de dados de ensaios e experimentos, com maior valorização de atividades científicas e a possibilidade de replicar estudos feitos por outros pesquisadores (MARQUES, 2019). Envolve os fazeres das ciências, tornando-as mais colaborativas e democráticas, na medida em que utiliza a tecnologia como apoio à abertura das suas atividades.

A Fapesp tem atuado em ações ligadas à Ciência Aberta há mais de 30 anos (FAPESP, s. d.), sendo uma das precursoras do movimento no país. Cabe destacar o apoio da fundação à *Scientific Electronic Library Online* (SciELO), pioneira na disseminação livre de resultados de pesquisas, lançado como projeto Fapesp em 1997, inaugurando o *Open Access* no país e se tornando um marco de qualidade de publicações científicas por meio da oferta de infraestrutura de comunicação, que possibilita o acesso aberto de textos completos.

A questão, ainda muito discutida na Ciência Aberta, se refere aos direitos autorais, que ganhou em 2011 fundamentos por meio da publicação da Política para Propriedade Intelectual da Fapesp, por meio da Portaria PR N° 04/2011. Esta política foi fortemente embasada em leis e decretos federais e estaduais, com o objetivo de ter amparo legal em todas as esferas. Mesmo com maior atenção à tecnologia, torna-se uma grande contribuição às questões não totalmente definidas no âmbito da Ciência Aberta.

Com a publicação da “Política para Acesso Abertos às Publicações Resultantes de Auxílio de Bolsas Fapesp”, por meio da portaria CTA N°1/2019, consolida-se uma ação iniciada ainda em 2008 voltada à instituição de Política para Acesso Aberto às Publicações Resultantes de Auxílios e bolsas Fapesp. Com isso, estabelece regras para que resultados de pesquisas fomentadas totalmente ou parcialmente pela Fapesp sejam publicados em periódicos que adotam o Acesso Aberto em suas variações ou que tenham versões depositadas em repositórios que fornecem acesso livre. Assim, garante a democratização do

acesso aos resultados de pesquisas fomentadas pela Fapesp, retornando à população o investimento aplicado.

Conforme o bioquímico britânico Richard Sever (*apud* ANDRADE, 2019), para tornar o conhecimento científico acessível rapidamente, “todos os resultados de pesquisa deveriam ser disponibilizados imediatamente em servidores de preprints, repositórios abertos que reúnem versões preliminares de artigos ainda não avaliados pelos revisores de periódicos científicos, devendo ser obrigatório para a concessão de recursos pelas agências de fomento à pesquisa”.

Com o intuito de tornar a Ciência Aberta mais popular, a Fapesp, em parceria com o jornal Folha de São Paulo, desenvolve um programa para televisão intitulado TV Ciência Aberta, voltado à divulgação do tema. Por isso, os episódios da iniciativa podem ser vistos pelo canal do YouTube, no site da Folha de São Paulo e no próprio site do programa²⁸. Os programas são realizados mensalmente, no auditório da Fapesp, com temas relevantes relacionados à Ciência Aberta.

Salienta-se que a Fapesp articula ações para implementação da Ciência Aberta nas suas variadas modalidades. E a SciELO, por meio da sua rede de revistas, tem um papel importante na disseminação dos resultados de pesquisas, oferecendo acesso livre aos textos completos. Da mesma forma, a política de acesso aberto implementada pela Fapesp torna-se um marco na obrigatoriedade de publicação em acesso aberto pelos projetos fomentados. Em ambos casos revela-se a adesão da Fapesp pelo movimento de Acesso Aberto, incluso na Ciência Aberta.

3. A FAPESP E OS DADOS ABERTOS DE PESQUISA

Dados Abertos, do original em inglês Open Data, não é um conceito novo, tendo estudos de mais de dez anos. Murray-Rust (2008) discute a terminologia em que a abertura na ciências, como no caso do acesso aberto, tenha relação com a queda das barreiras de acesso e das restrições de uso. No caso de dados abertos, é a liberdade de reuso dos dados de pesquisa, sem a necessidade de permissão do seu criador. Com isso, podem-se adicionar novos registros, unir

28 FAPESP (São Paulo). Ciência Aberta. Disponível em: <https://cienciaaberta.fapesp.br/>. Acesso em: 04 nov. 2019.

bases de dados, entre outras ações voltadas à geração de novas análises e novos conhecimentos.

Entretanto, dados de pesquisa possuem características que se apresentam como desafios ao reúso, dadas as diferenças apresentadas. Dados de pesquisa diferem-se conforme a metodologia de coleta utilizada, sendo que o *National Science Board* distingue-os conforme a sua natureza, como observacionais, computacionais e experimentais. Sayão e Sales (2014) defendem a necessidade de mudança nos sistemas de informação diante do cenário promovido pelos dados abertos, em que os resultados da pesquisa não podem ser reduzidos a uma publicação em artigo ou livro.

Em modelos voltados à comunicação científica, como o de Bjork (2007), o depósito dos dados de pesquisa em repositório já era previsto. Assim, dados brutos são resultados da experimentação ou observação, seguindo dois caminhos, sendo um para a análise e outro para depósito em repositórios de dados, para um processo de comunicação dos resultados de pesquisa. No entanto, para que esse processo possa ser efetuado requer-se atenção quanto aos direitos autorais e de propriedade intelectual, assim como as políticas desses repositórios, que orientam que tipo de dados podem compor o seu acervo.

Na “Política para Propriedade Intelectual da Fapesp”, datada de 2011, na seção de fundamentos, dois pontos se destacam em relação aos Dados Abertos resultantes de pesquisa. Na primeira recomendação, a Fapesp espera que os pesquisadores responsáveis compartilhem dados primários de pesquisa, ou mesmo amostras ou coleções resultantes de projetos amparados pela fundação. Na segunda recomendação, a Fapesp compreende as questões legais de direitos dos pesquisadores e suas instituições, mas que esse ponto não reduz a responsabilidade do compartilhamento dos dados. As duas recomendações são importantes para o movimento de dados abertos de pesquisa, pois apresentam a posição da Fapesp no compartilhamento de dados resultantes de pesquisas fomentadas por esse importante órgão.

Em 2014 a Fapesp publicou o livro *Código de Boas Práticas Científicas*²⁹, em que se solicitam ações éticas no processo científico e se discute, entre outras

29 FAPESP. Código de boas práticas científicas. [S.l.]: FAPESP, 2014. 46 p. Disponível em: http://www.fapesp.br/boaspraticas/FAPESP-Codigo_de_Boas_Praticas_Cientificas_2014.pdf. Acesso em: 04 nov. 2019.

coisas, os dados de pesquisa. Nesse sentido, além das orientações relacionadas à precisão e lisura na coleta e registro dos dados, a obra trata do compartilhamento dos dados após o término da pesquisa, possibilitando a colaboração, avaliação, verificação, reúso e continuidade, mantendo as devidas considerações éticas e legais previstas pelas leis.

Outro ponto relevante em relação aos dados abertos na Fapesp foi a determinação de requerer um “Plano de Gestão de Dados³⁰” a ser anexado nas propostas submetidas à Fundação, assinalada por Vidotti *et al.* (2017) como uma das primeiras iniciativas nacionais em relação aos dados abertos. Mesmo que Monteiro, Lucas e Lecardelli (2019) apontem que a Fapesp, com o “Plano de Gestão de Dados”, não atende totalmente aos critérios dos Dados Abertos, ainda se apresenta em destaque, ficando atrás apenas da Fundação Oswaldo Cruz (FioCruz) neste quesito, revelando-se inovadora na relação entre fomento e dados abertos de pesquisa.

Visando à oferta de infraestrutura para depósito de dados de pesquisa, a Fapesp instituiu um Grupo de Trabalho (GT) voltado à criação de uma rede de repositórios abertos de dados de pesquisa. O grupo é formado pelas seis universidades estaduais e federais do estado de São Paulo, incluindo o Instituto Tecnológico da Aeronáutica (ITA), mais a Embrapa Informática Agropecuária, que está localizada no campus da Universidade Estadual de Campinas (Unicamp). Com isso, a Fapesp atua também na oferta de repositórios para que pesquisadores possam depositar os dados resultantes das suas pesquisas.

Essa proposição de rede de repositórios atende a uma questão levantada por Malloy (2011), o qual defende que os dados abertos levam a um fazer científico melhor, mas que requer uma nova modalidade de repositórios que atendam ao arquivamento dos dados de modo apropriado, assim como mudanças na editoração científica, para suportar novas formas de trabalho. Com isso, revela a necessidade de mudanças nas políticas editoriais das revistas que vão aderir ao movimento de Dados Abertos, em que pode-se requerer depósito de dados em repositórios para a publicação de artigos.

O arXiv – que foi criado em agosto de 1991 por Paul Ginsparg, da Universidade Cornell (EUA) – é o mais tradicional repositório de *preprints* no mundo. Alguns manuscritos permanecem apenas como *preprints* e outros são publicados. O

30 FAPESP. Plano de Gestão de Dados. [S.l.]: FAPESP,[S.d]. [Online] Disponível em:<http://fapesp.br/gestaodedados/#documentos>. Acesso em: 04 nov. 2019.

SciELO, apoiado pela Fapesp desde sua criação em 1997, recentemente fez parceria com a *Public Knowledge Project* (PKP) para a criação de um servidor de *preprints* para 2020 (ANDRADE, 2019).

Completando alguns aspectos dos dados abertos, a SciELO atua com a publicação de artigos de dados, efetivando um ponto importante no compartilhamento de dados, visto que esse tipo de publicação descreve a procedência, curadoria, cobertura, limitações de formato e uso dos dados (PACKER *et al.*, 2018). Com isso, os dados brutos depositados em repositórios ganham um documento publicado, o qual os descreve e registra a autoria dos pesquisadores que os coletaram.

Diante do proposto, nota-se que a Fapesp apresenta alinhamento com as tendências mundiais da ciência, com ações efetivas e aplicadas, visando atender às necessidades dos pesquisadores quanto aos dados abertos. Além disso, reforça-se a importância do órgão de fomento ante as mudanças impostas pelo avanço da tecnologia e evolução das ciências, aportando recursos para desenvolver estudos em um cenário desafiador.

4. CONSIDERAÇÕES FINAIS

A atividade de pesquisa é primordial para o desenvolvimento de um país. Por isso, muitos países possuem órgãos voltados ao fomento dessas atividades, na medida em que a competição no mundo globalizado é grande e que novos conhecimentos e inovações representam o diferencial. Tanto que, no Brasil, há órgãos de fomento a pesquisas nos âmbitos federal e estadual, os quais atendem a investigações nacionais, de modo que a Fapesp é um dos destaques dessa lista pelo fato de ser um órgão estadual, mas com projeção internacional.

Muito dessa visibilidade deve-se a seu direcionamento científico e à sua independência financeira. Nesse sentido, a Fapesp sempre esteve alinhada às tendências da ciência, fomentando projetos de interesse brasileiro, mas no estado da arte mundial. Na mesma medida em que assume posições de vanguarda, norteia ações, como no caso do Acesso Aberto aos resultados de pesquisa, que teve na SciELO o marco no Brasil e agora com as políticas de acesso aberto aos projetos financiados por ela.

O apoio ao compartilhamento dos dados de pesquisa é outro ponto importante nas ações da Fapesp, que fomenta a criação de infraestrutura por meio de ações coordenadas com entidades de ensino e pesquisa públicas localizadas no estado de São Paulo.

5. REFERÊNCIAS

ANDRADE, R.O. Richard Sever: Para acelerar a comunicação da ciência. **Revista FAPESP**. Ed.285, nov.2019 <https://revistapesquisa.fapesp.br/2019/11/05/richard-sever-para-acelerar-a-comunicacao-da-ciencia/>. Acesso em 08/11/2019

BJÖRK, B.-C. A model of scientific communication of a global distributed information system. 2007. **Informatio Research**, v. 12, n. 2, jan. 2007. Disponível em: <http://informationr.net/ir/12-2/paper307.html>. Acesso em: 04 nov. 2019.

DOMINGUES, A. A.; COSTA, M. C. da. A colaboração internacional da FAPESP: quais contextos para suas transformações?. **Revista Brasileira de História da Ciência**, Rio de Janeiro, v. 9, n. 1, p. 19-35, jan/jun 2016. Disponível em: https://www.sbh.org.br/revistahistoria/view?ID_REVISTA_HISTORIA=56. Acesso em 04 nov. 2019.

FAPESP. **Open Science**. Disponível em: <http://www.fapesp.br/openscience/>. Acesso em: 30 out. 2019.

HAMBURGER, A. I. **Fapesp 40 anos**: abrindo fronteiras. EdUSP, 2004. 552 p.

LIMA, J. A. L.; LIMA, V. A. O. Do *open source* ao *open education*: novos conceitos para educomunicação. In: ENCONTRO DE COMUNICAÇÃO E MÍDIA (Ecom), 2013, Campina Grande-PB. **Anais** [...]. Campina Grande-PB: Ecom, 2013. Disponível em: <https://joaoademar.com.br/ecom.pdf>. Acesso em 04 nov. 2019.

MARQUES, F. Por um novo sistema de recompensas. **Revista FAPESP**. Ed. 281, jul 2019. Disponível em: <https://revistapesquisa.fapesp.br/2019/07/10/por-um-novo-sistema-de-recompensas/> Acesso em 08/11/2019

MOLLOY, J. C. The open knowledge foundation: open data means better science. **PLoS biology**, v. 9, n. 12, p. e1001195, 2011. Disponível em: <https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.1001195>. Acesso em 04 nov. 2019.

MONTEIRO, G.; LUCAS, E. R. de O.; LECARDELLI, J. Financiadores da pesquisa brasileira e os dados científicos abertos. *In: IX ENCONTRO IBÉRICO EDICIC*, 9., 2019, Barcelona. **Anais [...]**. Barcelona: Edicic, 2019. Disponível em: <http://eprints.rclis.org/38555/>. Acesso em 04 nov. 2019.

MOTOYAMA, S.; HAMBURGER, A. I.; NAGAMINI, M. (org.). **Para uma história da FAPESP: marcos documentais**. São Paulo: FAPESP, 1999. 248p. Disponível em: <https://bv.fapesp.br/pt/103/para-uma-historia-da-fapesp-marcos-documentais/>. Acesso em 04 nov. 2019.

MURRAY-RUST, P. Open data in science. **Serials Review**, v. 34, n. 1, p. 52-64, 2008. Disponível em: <http://precedings.nature.com/documents/1526/version/1>. Acesso em 04 nov. 2019.

PACKER, A.L., *et al.* Aos 20 Anos, a Rede SciELO atualiza prioridades e avança para a Ciência Aberta [online]. **SciELO em Perspectiva**, 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 04 nov. 2019.

SAYÃO, L. F.; SALES, L. F. Dados abertos de pesquisa: ampliando o conceito de acesso livre. **Revista Eletrônica de Comunicação, Informação e Inovação em Saúde**, [S.l.], v. 8, n. 2, jun. 2014. ISSN 1981-6278. DOI 10.3395/reciis.v8i2.611. Disponível em: <https://www.reciis.icict.fiocruz.br/index.php/reciis/article/view/611>. Acesso em: 04 nov. 2019.

VIDOTTI, S. A. B. G. *et al.* Repositório de dados de pesquisa para grupo de pesquisa: um projeto piloto. *In: XVIII ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO*, 18., 2017, Marília-SP. **Anais [...]**. Marília-SP: ENANCIB, 2017. Disponível em: <http://enancib.marilia.unesp.br/index.php/xviiienancib/ENANCIB/paper/view/388>. Acesso em 04 nov. 2019.

6. SOBRE OS AUTORES



Milton Shintaku

Lattes: <http://lattes.cnpq.br/8605833104600600>

Orcid: <https://orcid.org/0000-0002-6476-4953>

Mestre e doutor em Ciência da Informação pela Universidade de Brasília (UnB), coordenador de Articulação, Geração e Aplicação de Tecnologia do Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict), Secretário da Associação Brasileira de Editores Científicos (ABEC) 2017-2019.



Suzana Caetano da Silva Lanne

Lattes: <http://lattes.cnpq.br/2988634049584035>

Orcid: <https://orcid.org/0000-0002-4481-7376>

Professora Associada (nível III) da Faculdade de Ciências Farmacêuticas da Universidade de São Paulo. Formação em Engenharia Química pela Escola Superior de Química Oswaldo Cruz (1987), Mestrado (1993) e Doutorado (1997) em Tecnologia Bioquímico-Farmacêutica pela Universidade de São Paulo, Pós-Doutorado no Dipartimento di Ingegneria Chimica e di Processo - Università degli Studi di Genova (UNIGE), Itália (2011). Livre-Docente USP (2008). Editora-Chefe Food Science and Technology (Ciência e Tecnologia de Alimentos) (2005-2010 e de 2013-2019). Diretora da Associação Brasileira dos Editores Científicos-ABEC (2016-2020). IUFOST - Acting President (2019-2020).

7. COMO CITAR

SHINTAKU, M.; LANNE, Z. C. da S. Dados abertos e a Fundação de Amparo à Pesquisa do Estado de São Paulo. *In*: SHINTAKU, M.; SALES, L. F.; COSTA, M. (org). **Tópicos sobre dados abertos para editores científicos**. Botucatu, SP: ABEC, 2020. p. 87-96. DOI: 10.21452/ 978-85-93910-04-3.cap8.

Como os indexadores estão vendo a questão de dados abertos de pesquisa para o processo de indexação?

Gildenir Carolino Santos

1. INTRODUÇÃO

O periódico científico nasceu no século XVII para repertoriar e resumir livros publicados na Europa, onde um deles deu a capacidade de registro dos trabalhos científicos (*The Philosophical Transactions*), enquanto o outro fomentava a divulgação científica (*Le Journal des Sçavans*). Só a partir de 1850 foi que os periódicos científicos começaram a assumir o papel de veículo para contribuições originais (MOSTAFA; TERRA, 2000).

A partir desse feito, do surgimento dos periódicos e da ascensão da internet desde o final da década de 80, a explosão bibliográfica cresceu mais ainda, tornando incontável o surgimento dia após dia de novos títulos e de novos conteúdos relacionados aos periódicos científicos.

Mesmo assim, agora em pleno século XXI, surgem editores de bases de dados e até mesmo grandes *publishers* adotando nas suas diretrizes, e até exigindo, que os editores científicos adotem nas suas políticas editoriais a avaliação aberta por pares. Mais recentemente, avaliou-se como os indexadores estão

acompanhando e enxergando os questionamentos sobre os dados abertos de pesquisa no processo de indexação de periódicos científicos.

O ato da indexação é o processo de transferência e descrição analítica da informação mais importante de Biblioteconomia. É por meio da indexação que sabemos como determinado termos ou assuntos registrados de modo padronizado nos sistemas e recursos informacionais são encontrados. Esses recursos informacionais denominados bases de dados, diretórios, portais e índices e indicam onde uma ou mais publicações (periódicos, revistas, jornais, boletim etc.) encontra-se indexada de maneira manual ou automatizada, destinando-se a eles os responsáveis pelo grande número de informações e dados armazenados de forma segura e padronizada. Saber escolher as fontes de indexação é algo muito importante para a contextualização do periódico e para sua visibilidade por diversas instituições e fornecedores (SANTOS, 2017).

Portanto, é esperado que os sistemas de gerenciamento, que são gerenciados pelos fluxos editoriais periódicos, sejam configurados com plugins de implementação para os artigos, os quais atribuem complementos para os dados de pesquisa, como é o caso do software OJS. Ele integra nas suas configurações o Dataverse³¹, software que agrega dados de pesquisa. Assim, o Dataverse com o OJS possibilita, com grande facilidade, que os indexadores possam captar dados de pesquisa integrantes dos artigos submetidos pelo OJS.

Nesse sentido, ressaltamos que os periódicos indexados são considerados de maior qualidade científica em comparação com periódicos que nunca foram indexados em qualquer tipo de sistema. A indexação passou a ser um processo muito importante nos últimos tempos, tornando-se uma disputa entre os editores, os quais concorrerem entre si para alcançar o patamar de qualidade de suas publicações por meio desse processo junto aos fornecedores de fontes de indexação (bases de dados, diretórios, portas e índices) (SANTOS; XAVIER, 2018).

Os dados de pesquisa que começam a ser vistos como complemento essencial aos documentos convencionais, como os artigos, terão que possuir certa qualidade de dados para que possam embarcar nos indexadores, assim como os artigos são avaliados para tal finalidade de indexação.

31 Dataverse. Disponível em: <https://dataverse.org/about>. Acesso em: 3 dez. 2019.

Para a construção deste ensaio, foi encaminhada por e-mail uma breve questão, de como eles, os indexadores, estão vendo a questão de dados abertos de pesquisa para o processo de indexação, e como isto reflete na pesquisa.

Assim, iremos expor aqui apontamentos sobre a combinação da Ciência Aberta com os Indexadores; da Posição dos Indexadores em Relação aos Dados Abertos de Pesquisa; Novos sistemas de citações abertas e seus Benefícios, finalizando com nossas considerações finais.

2. A CIÊNCIA ABERTA E OS INDEXADORES

De acordo com Abadal (2019):

A Ciência Aberta é uma maneira de conceber a pesquisa científica baseada no trabalho colaborativo, na abertura e transparência de todas as fases da pesquisa (coleta de dados, revisão de especialistas, divulgação, avaliação etc.) e também em a abordagem da ciência à sociedade. A Ciência Aberta é uma transformação radical da maneira de realizar pesquisas científicas e de como transformar seu sistema de avaliação, em síntese, toda uma mudança de paradigma em relação ao sistema atual.

Nas palavras de Oliveira e Silva (2016, p. 6) encontramos a menção para a Ciência Aberta:

é o fio condutor de investigações científicas apoiadas por uma ciberinfraestrutura tecnológica e metodológica que permite o uso, reúso e reprodutibilidade de dados de pesquisa. Destarte, figura o surgimento do novo paradigma da ciência nomeado de quarto paradigma no cenário contemporâneo mundial da comunicação científica.

Já na concepção de Belluzzo (2019, p. 280):

As ações em Ciência Aberta, muito embora tenham o entusiasmo de parcela dos cientistas e de profissionais envolvidos, encontram também muita resistência de outros, não apenas pela dificuldade de aprender a lidar com essas novas práticas, o que demanda a necessidade de absorver novas atividades e habilidades, mas

também pelos deslocamentos de poder que tais mudanças frequentemente envolvem. Há ainda controvérsias sobre a extensão em que cada passo da atividade científica deve estar amplamente disponível para acesso, ou seja, em Acesso Aberto.

Neste caso, sabemos que tudo que está disponível nos indexadores está provido de acesso livre ao conteúdo, como na *Web of Science* e *Scopus*, como as duas maiores bases de dados de assuntos multidisciplinares e de viés estritamente comercial.

Porém, durante a iniciação da proposta deste ensaio, a *Web of Science* se mostrou totalmente a favor e flexível à Ciência Aberta sobre o seu parecer e opinião a respeito do que ele, como indexador, pensa e como está vendo a questão de dados abertos de pesquisa para o processo de indexação.

Quando falamos das bases de dados e sua importância na coleta de dados abertos de pesquisa para qualificar e produzir novos conhecimentos, estamos falando da atuação dos indexadores nesse contexto. Nesse sentido, temos na literatura algumas pesquisas que mostram que esse papel dos indexadores é fundamental, pois, segundo Martins, Anjos e Melo (2015, p. 110-111):

A maior presença da utilização de base de dados em pesquisas científicas pode significar um novo modo de se fazer ciência, utilizando grandes volumes de informações produzidas por projetos de pesquisa. Assim, a cultura do e-Science baseia-se na perspectiva de compartilhar dados entre pesquisadores/cientistas distribuídos por todo o mundo, permitindo que estes possam acessar, manipular, extrair e reusar dados científicos utilizando o acesso remoto e promovendo cooperação, colaboração e interdisciplinaridade. Resumindo, o principal objetivo do *E-Science* (ou *e-Science*) é a concepção e construção de uma rede de colaboração e investigação científica.

Dessa forma, reforçando a origem e a concepção do termo *E-Science* baseado na perspectiva de compartilhar dados entre pesquisadores/cientistas, Albagli, Appel e Maciel (2013) lembram que o termo “*e-Science* foi introduzido por John Taylor em 2001” e é usado para designar a ciência desenvolvida por meio de colaborações globalmente distribuídas viabilizadas pela Internet.

No relato histórico sobre dados de pesquisa, Machado (2015, p. 201) comenta que:

Compartilhar eletronicamente dados de pesquisa não é algo tão novo como parece. Há mais de 40 anos, cientistas da computação já compartilhavam arquivos, de forma anônima, através de FTP1, que era o protocolo padrão de rede usado para transferir arquivos de um computador hospedeiro para outro, em uma rede. Há 23 anos, surgia o arXiv.org, que hoje reúne quase 1 milhão de trabalhos, principalmente das áreas de física, matemática, ciência da computação, estatística. A expressão Acesso Aberto foi lançada com a Declaração de Budapeste da Iniciativa do Acesso Aberto, documento de 2002, de grande importância política. Em 2007, a definição de dados abertos em oito princípios veio dar mais força ao processo de abertura de dados e à ampliação dos usos e reúsos da informação, com impactos também na ciência.

Assim, o discurso atual focado no Acesso Aberto ao conhecimento científico se entrelaça com outro termo usualmente discutido na comunidade científica, ou seja, com dados abertos. Hoje em dia se integram padrões, formatos e sistemas que possibilitam maior comunicação de transparência entre eles, processamento, cruzamento e reúso da informação. O debate sobre o Acesso Aberto falado desde 2002 ficou obsoleto nos dias atuais, “sendo insuficiente para dar conta dos avanços do uso e reúso de uma base de dados aberta” (MACHADO, 2015, p. 202).

Na evolução dos dados científicos *on-line*, podemos considerar como primeiros embriões – justamente ligados a alguns indexadores, ou simplesmente bases de dados – o que Machado (2015, p. 202) apresenta, no Quadro 9.1, a partir do levantamento histórico que o autor fez, e cronologicamente apresentamos em forma de eventos ocorridos para surgimento da evolução dos dados científicos.

Quadro 9.1 - Cronologia dos eventos sobre a evolução dos dados científicos

DATA	ORIGEM	INDEXADOR/ SISTEMA	HISTÓRICO DOS EVENTOS
1966	EUA	ERIC - Educational Resources Information Center ³²	“A primeira iniciativa de criar um banco de dados de bibliografia eletrônica de Acesso Aberto.”

32 <http://www.eric.ed.gov>

DATA	ORIGEM	INDEXADOR/ SISTEMA	HISTÓRICO DOS EVENTOS
1966	EUA	Medline / PubMed ³³	“Banco de dados on-line de livre acesso gerido pela National Library of Medicine e o National Institutes of Health com citações bibliográficas de periódicos da área de biomédica, que mais tarde passaria a ser chamado PubMed, apresentando hoje mais de 14 milhões de artigos completos.”
1971	EUA	Projeto Gutenberg ³⁴	“[...] criado, por Michel Hart, o Projeto Gutenberg com o objetivo de encorajar a produção e a distribuição de e-books (Hart, 2004). O objetivo era disponibilizar publicamente livros que pudessem ser lidos ou impressos a partir do maior número de computadores e programas.”
1974		SLAC - Stanford Linear Accelerator Center ³⁵	Esse sistema começa a [...] “catalogar literatura eletrônica na área de física.”
1974	DE	Deutsches Elektronen Synchrotron ³⁶	Também fomentado para [...] “catalogar literatura eletrônica na área de física.”
1990	Mundial	Internet	[...] “com a expansão mundial da Internet, que começaram a proliferar os bancos de dados de artigos científicos de livre acesso.”
1991	EUA	ArXiv ³⁷	[...] “surgiu o repositório de textos de física, matemática e ciência da computação.”
1992	EUA	Genbank ³⁸	[...] “criado o banco de dados de pesquisa genética.”
1996	EUA	Networked Digital Library of Theses and Dissertations (NDLTD) ³⁹	A Universidade de Virgínia, sob seus cuidados cria um sistema para divulgar as teses e dissertações. [...] “passou a ser o maior banco de teses e dissertações do mundo”.
1997	BR	SciELO (Scientific Electronic Library On-line) ⁴⁰	“Bireme – Centro Latino-Americano e do Caribe de Informação em Ciências da Saúde – com apoio da Fundação de Amparo à Pesquisa do Estado de São Paulo (Fapesp)” [...] cria base de periódicos.

33 <http://pubmed.gov>

34 <http://www.gutenberg.org>

35 <http://www.slac.stanford.edu>

36 <http://www.desy.de>

37 <http://arxiv.org>

38 <http://www.ncbi.nlm.nih.gov/Genbank>

39 <http://www.ndltd.org>

40 <http://www.scielo.org>

DATA	ORIGEM	INDEXADOR/ SISTEMA	HISTÓRICO DOS EVENTOS
2000	UK	BioMed Central ⁴¹	[...] "referência nas áreas de biologia e medicina, juntamente com o PubMed." A BioMed Central é uma editora científica de Acesso Aberto com fins lucrativos, sediada no Reino Unido, que produz mais de 250 periódicos científicos. Todos os seus diários são publicados apenas online. A BioMed Central se descreve como a primeira e maior editora de ciência de Acesso Aberto (Wikipedia).
2000	EUA	PLoS (Public Library of Science) ⁴²	Também se tornou referência nas áreas de biologia e medicina, juntamente com o PubMed. A Public Library of Science, conhecida como PLOS e, anteriormente, PLoS, é um projeto sem fins lucrativos que tem o objetivo de criar uma biblioteca de revistas científicas e publicações afins dentro do modelo de licenciamento de conteúdo aberto, fazendo uso, especificamente, a Creative Commons Atribuição. (Wikipedia).
2002	EUA	California Digital Library, da Universidade da Califórnia ⁴³	Pioneiro na publicação de repositório científico, tendo "um papel fundamental no Acesso Aberto, permitindo a disponibilização de artigos, papers e documentação de pesquisa produzidos na universidade e nos centros de investigação". Lança o sistema eScholarship Repository.
2002	EUA	Papyrus, da Universidade de Montreal ⁴⁴	Outro repositório científico pioneiro na disponibilização de publicações de Acesso Aberto.
2002	UK	E-Prints Soton ⁴⁵	A Universidade de Southampton também se tornou pioneira na área de repositório científico, no campo de preprints e documentos abertos.
2002	MX	Redalyc - Rede de Revistas Científicas da América Latina, Caribe, Espanha e Portugal ⁴⁶	O projeto "Redalyc" é uma base de dados bibliográfica e de uma biblioteca digital de revistas de Acesso Aberto, suportado pela Universidade Autónoma do Estado do México com a ajuda de numerosas outras instituições de ensino superior e dos sistemas de informação. (Wikipedia)

41 <http://www.biomedcentral.com>

42 <https://www.plos.org/>

43 <http://repositories.cdlib.org>

44 <https://papyrus.bib.umontreal.ca>

45 <http://eprints.soton.ac.uk>

46 <https://www.redalyc.org/>

DATA	ORIGEM	INDEXADOR/ SISTEMA	HISTÓRICO DOS EVENTOS
2002	UK	SHERPA – Securing a Hybrid Environment for Research Preservation and Access ⁴⁷	[...] “resultado de um consórcio de 20 bibliotecas universitárias britânicas, cujo objetivo foi estabelecer repositórios nas instituições participantes.”
2003	SWE	DOAJ – Directory of Open Access Journals ⁴⁸	O DOAJ é um website indexador que hospeda uma lista, com curadoria, da comunidade de periódicos de Acesso Aberto, mantida pelo Services for Open Access (IS4OA). Surgiu juntamente no mesmo ano do Movimento do Acesso Aberto (BOAI). (Wikipedia).
2019	BR	Metabuscador de Dados de Pesquisa ⁴⁹	Inaugurado pelo Consórcio CRUESP, o maior repositório de dados de pesquisa do estado de São Paulo.

Fonte: Elaborado pelo autor com a adaptação de Machado (2015).

Lembramos que fizemos contato com o DOAJ e Redalyc e não obtivemos retorno deles sobre o questionamento que solicitamos para a base das respostas do ensaio que apresentaremos mais à frente. No entanto, independentemente de não obtermos o retorno, eles são considerados grandes incentivadores da qualidade, padronização e visibilidade, concentrados na literatura científica da via dourada, os seja, os periódicos de Acesso Aberto. Ambos, aliás, foram iniciados no mesmo ano do Movimento do Acesso Aberto: 2002.

No contexto brasileiro citado no Quadro 1, pudemos contar com a *Scientific Electronic Library On-line* (SciELO), que se tornou a maior base de dados de periódicos de Acesso Aberto de textos completos no mundo, e que nos retornou respondendo sobre o seu posicionamento em relação aos dados abertos de pesquisa que veremos a seguir.

3. POSIÇÃO DOS INDEXADORES EM RELAÇÃO AOS DADOS ABERTOS DE PESQUISA

Ao compor este ensaio, entramos em contato com alguns indexadores. Assim, obtivemos apenas dos cinco mais conhecidos (DOAJ, Scopus, Web of Science,

47 <http://www.sherpa.ac.uk>

48 <https://doaj.org/>

49 <https://metabuscador.uspdigital.usp.br/>

SciELO e Redalyc) o retorno da *Web of Science*, respondido pela própria *Clarivate Analytics* e da *SciELO*. Encaminhado pela responsável do processo de solicitação e avaliação de indexação, conforme apresentado abaixo, os posicionamentos apresentam-se no Quadro 9.2 (*WoS* e *SciELO*):

Quadro 9.2 - Posicionamento dos indexadores em relação aos dados abertos de pesquisa para o processo de indexação

WOS – CLARIVATE ANALYTICS

Esclarecer a posição do Analytics em pesquisa aberta e Acesso Aberto

A pesquisa fornece a base da sociedade moderna. A pesquisa leva a avanços e comunica os resultados, permitindo-nos transformar avanços em vidas melhores – fornecer novos tratamentos para doenças, implementar soluções para desafios como o aquecimento global e construir indústrias inteiras em torno do que antes eram apenas ideias. O Acesso Aberto é a atualização moderna necessária para a comunicação de pesquisa que utiliza totalmente a Internet para o que foi originalmente criada – acelerar a pesquisa.

O Acesso Aberto é a disponibilidade on-line gratuita e imediata de artigos de pesquisa combinada com os direitos de uso total desses artigos no ambiente digital.

Independentemente do modelo de negócios, assinatura ou Acesso Aberto, a comunidade acadêmica deseja acesso à pesquisa; ela quer poder ler o texto completo de um artigo quando clicar nele; quer saber que está obtendo a melhor versão desse artigo. E quer saber que o financiamento é rastreável e transparente.

Quando o artigo é facilmente acessível e compartilhável, a pesquisa é acelerada, as conexões são explícitas e as colaborações são mais naturais e significativas. O processo de descoberta científica é mais eficiente e é mais fácil garantir a integridade e a reprodutibilidade da pesquisa.

A avaliação robusta e a curadoria de nossos dados, independentemente do modelo de publicação, tornam a Web of Science Core Collection o banco de dados global de citações independente de editor mais confiável do mundo. O Web of Science Core Collection é um banco de dados definitivo, confiável e de alta qualidade para revistas, livros e anais de conferências. Somos guiados pelo legado do Dr. Eugene Garfield, inventor do primeiro índice de citação do mundo. Permanecemos fiéis aos seus princípios de objetividade, seletividade e dinâmica de coleta, mas também adaptamos e respondemos aos avanços e mudanças tecnológicas no cenário editorial. Nosso objetivo é refletir o ecossistema de pesquisa em constante mudança, permanecendo absolutamente fiéis aos princípios de seleção do conteúdo da mais alta qualidade.

Nossos Princípios Orientadores

- Somos neutros em relação ao editor e neutros no modelo de publicação.
- Nossas decisões são imparciais e tomadas na qualidade dos dados, não no modelo de publicação.
- Apoiamos os objetivos da pesquisa aberta porque acreditamos que é a melhor maneira de tornar o processo de descoberta científica mais rápido, mais robusto e mais transparente.
- Fomos os primeiros a garantir que nossos clientes pudessem acessar com facilidade o conteúdo de Acesso Aberto revisado por pares, incorporando esses dados na Web of Science.
- Nossos dados são a melhor combinação de dados robustos de OA e curadoria editorial.
- Estamos sempre procurando maneiras de tornar o processo de descoberta científica mais rápido, robusto e transparente.

Nosso apoio à pesquisa aberta e ao Acesso Aberto

- Investimos em Ciência Aberta e na comunidade aberta por meio de uma doação generosa ao Impactstory, sem fins lucrativos. Os resultados desse investimento aprimoraram muito os links para o material Acesso Aberto nos recursos da plataforma Web of Science Core Collection e Web of Science. Como parceiro da comunidade acadêmica, sentimos fortemente que a criação desse algoritmo pertencia ao domínio público; portanto, fornecemos essa tecnologia publicamente para que todos os provedores de informações pudessem tirar proveito dessa melhoria na vinculação ao texto completo.
- Fomos os primeiros a garantir que nossos clientes pudessem acessar com facilidade o conteúdo de Acesso Aberto revisado por pares, incorporando esses dados no Web of Science. Nossos dados são a melhor combinação de dados robustos de OA e curadoria editorial. O Web of Science continua sendo o único banco de dados de citações neutro em editor que abrange largamente o Acesso Aberto controlado por versão e controlado por editoração.
- Investimos em ferramentas abertas – como patrocinador fundador do ORCID em 2010/11, doamos o código do ResearcherID para decolar e continuar trabalhando com o ORCID.
- Habilitamos o acesso com um clique à melhor versão legal de artigos em texto completo para pesquisadores por meio do Kopernio, nossa ferramenta aberta de plug-in para navegador, disponível para todos os usuários para encontrar texto completo no ponto necessário.
- Ajudamos as instituições a rastrear o uso e entender o impacto do conteúdo do Acesso Aberto, incorporando indicadores de OA no InCites Benchmarking and Analytics e Journal Citation Reports.
- Aumentar a transparência ajudando os financiadores a rastrear o uso e o impacto do conteúdo de Acesso Aberto que eles financiam.
- Ajudamos os editores a entender de onde vem o financiamento de Acesso Aberto e como eles podem gerenciar transições de Acesso Aberto.
- Permitimos transparência e abertura na revisão por pares através do Publons e ScholarOne, dando suporte à Open Peer Review trabalhando com os principais editores.
- Usamos nossos dados para ajudar a esclarecer uma paisagem complicada e comovente. Continuamos a oferecer Relatórios Globais de Pesquisa sobre os Principais Tópicos, um dos quais focado na coalizão do Plano S.
- Por meio do Data Citation Index, nosso recurso na plataforma Web of Science que indexa conjuntos de dados e estudos de dados dos principais repositórios internacionais de big data, multidisciplinares, tornamos os dados de pesquisa abertos mais detectáveis e ajudamos pesquisadores e financiadores a rastrear o impacto de seus dados. Esses conjuntos de dados e estudos podem ser pesquisados isoladamente ou em conjunto com os periódicos da Web of Science, anais de conferências, livros e Lietrature de Patentes. Eles também fornecem à comunidade internacional de pesquisa a capacidade de reutilizar esses dados para reduzir efetivamente o processo de pesquisa. Os financiadores determinam que os pesquisadores planejem disponibilizar seus conjuntos de dados e stidies para uso de outras pessoas, o Data Citation Index fornece acesso aos repositórios mais importantes de big data.

SCIELO

- O Programa SciELO está promovendo o alinhamento com as práticas de comunicação de pesquisas da Ciência Aberta das metodologias e tecnologias do Modelo SciELO de Publicação, assim como das políticas editoriais e operação dos periódicos individuais.
- Este alinhamento inclui:
 - (a) a adoção de preprints como versão inicial do manuscrito depositada em servidores de preprints confiáveis;
 - (b) a citação e referenciamento de dados, códigos de programas e outros conteúdos subjacentes aos textos dos artigos.
- Esta implantação deverá ocorrer em três etapas:
 - (i) estimular a citação e referenciamento;
 - (ii) exigir a citação e referenciamento; e
 - (iii) exigir que os conteúdos citados e referenciados estejam em Acesso Aberto exceto em casos especiais.
- O SciELO passará a exigir que até meados de 2020 as políticas editoriais dos periódicos e instruções aos autores já indexados alinhem-se com o descrito acima e que novos periódicos somente serão incluídos quando já tiverem alinhados.

Fonte: SciELO e Clarivate Analytics (respostas encaminhadas por e-mail), 2019.

Por essa razão, comentamos aqui que a *SciELO* mapeia as publicações em suas instruções para alcançar o ingresso na sua base de dados, não devendo deixar de publicar também em língua inglesa. A *SciELO* dá importância para publicações preprint, além de oferecer abertura das avaliações de forma aberta, alcançando os princípios da Ciência Aberta.

Nos demais indexadores (Quadro 9.3), fomos buscar no site deles informações que poderiam responder ao nosso questionamento inicial, e obtivemos os seguintes posicionamentos:

Quadro 9.3 - Posicionamentos dos indexadores não respondentes à pesquisa

INDEXADOR /ORNECEDOR	SERVIÇOS DE INDEXAÇÃO / POSICIONAMENTO SOBRE DADOS ABERTOS DE PESQUISA	FONTE
DOAJ	A equipe do DOAJ adota: Os princípios de transparência (TOP) e boas práticas. Licenças do CCO. Reprodutibilidade. Política de depósito (Sherpa e outros). Identificadores permanentes. Provê de critérios para a publicação contínua e preprint.	Oliveira (2018)
Latindex	O Latindex formou um Grupo de Trabalho para analisar: Implicações da publicação contínua, publicações e preprint, além do catálogo 2.0. Em relação aos dados abertos de pesquisa: Provê de critérios para a publicação contínua.	Oliveira (2018)
Redalyc	A Redalyc em seu site deixa claro que é integrante de registros de: Dados abertos. Interoperabilidade. Reprodutibilidade. Sustentabilidade. Critérios para a publicação contínua.	Oliveira (2018)

INDEXADOR /ORNECEDOR	SERVIÇOS DE INDEXAÇÃO / POSICIONAMENTO SOBRE DADOS ABERTOS DE PESQUISA	FONTE
Elsevier	<p>Pela Elsevier, representando a Scopus, temos o <i>Data Search</i>, lançado em 2016, como uma versão beta da plataforma que é um:</p> <p>“Mecanismo de pesquisa de dados que permite que cientistas e pesquisadores pesquisem muitos tipos e formatos de dados diferentes em uma variedade de repositórios de dados institucionais específicos ao domínio e entre domínios e outras fontes de dados. Os resultados exibem conjuntos de dados de maneira unificada para facilitar a localização de dados de pesquisa relevantes e úteis, permitindo que os usuários visualizem e avaliem rapidamente os dados antes de visualizá-los no repositório de destino. Ao gerar visualizações dos dados reais em linha (por exemplo, planilhas, imagens, mapas interativos etc.), o <i>DataSearch</i> ajuda os usuários a digitalizar vários conjuntos de dados potencialmente interessantes muito mais rapidamente. O <i>DataSearch</i> indexa metadados e dados para facilitar a correspondência de consultas com objetos descritos na pesquisa.”</p> <p>No âmbito de outro produto da Elsevier, temos o <i>Mendeley Data</i>, que é um:</p> <p>“Repositório aberto de dados de pesquisa, onde os pesquisadores podem fazer upload e compartilhar seus dados. O <i>Mendeley Data</i> fornece uma plataforma para mostrar os conjuntos de dados de seus autores e periódicos, tornando-os detectáveis junto com conjuntos de dados de mais de 200 periódicos. Permite que os autores enviem sem esforço seus dados de pesquisa, vinculados ao artigo, via integração personalizada com seu sistema de envio; o rascunho do conjunto de dados estará disponível para os revisores e, assim que o artigo for revisado, aceito e publicado, as páginas do artigo e dos dados serão vinculadas entre si – ou optar por uma exibição incorporada do conjunto de dados na página do artigo. Os dados de pesquisa em uma ampla variedade de formatos podem ser visualizados em linha na página da <i>web</i> do conjunto de dados (e na página do artigo, se você optar pelo recurso incorporado de conjunto de dados), graças ao nosso recurso de visualização de arquivos, permitindo que os leitores visualizem e avaliem facilmente os dados antes de fazer o download. O compartilhamento de dados também é uma maneira para você, como pesquisador, obter exposição para os resultados de sua pesquisa, pois cada conjunto de dados tem um DOI e pode ser citado.”</p> <p>Esses produtos (<i>Data Search</i> e <i>Mendeley Data</i>) da Elsevier apontam que ela está atuando também em prol do movimento e da Ciência Aberta, em relação ao armazenamento e condicionamento dos dados abertos de pesquisa nas pesquisas publicadas em artigos indexadas na base de dados Scopus.</p>	Site ⁵⁰ Elsevier ⁵¹ (2017),

Fonte: Parte adaptada/traduzida de Oliveira (2018).

50 <https://datasearch.elsevier.com>

51 <https://data.mendeley.com/>

Bases indexadoras como a Scopus também estão flexibilizando e aceitando publicações em outros idiomas, o que era antes limitado em maior quantidade na literatura de língua inglesa. Esse conceito está mudando e abarcando também literaturas de língua portuguesa e espanhola com quantidade ascendente. Apesar disso, é nítida no mundo científico a predominância da língua inglesa.

4. NECESSIDADE DE SIMPLIFICAR A INDEXAÇÃO E O COMPARTILHAMENTO DE DADOS

Fazendo uma relação com o questionamento sobre a posição dos indexadores em relação aos dados abertos de pesquisa, Ryvardeen (2019) – gerente digital do periódico *Journal of the Norwegian Medical Association* –, em seu post publicado na *The Scholarly Kitchen*, comenta seu ponto de vista a respeito da necessidade de simplificar a indexação e o compartilhamento de dados:

Indexação e compartilhamento de metadados são a força vital de periódicos acadêmicos. Mesmo com serviços e infraestrutura disponíveis para todos os periódicos, o esforço necessário para participar não é pequeno. Os periódicos que são autopublicados e em suas próprias plataformas precisam de recursos significativos para implementar o compartilhamento e o depósito de metadados.

A complexidade envolvida no entendimento e na criação de serviços como **Crossref**, **Pubmed**, **Emerging Sources Citation Index (ESCI)** e **Google Scholar**, para mencionar apenas alguns, é assustadora para muitas revistas, especialmente fora dos EUA e da Europa Ocidental. Sem usar um parceiro de plataforma, como Silverchair ou Atypon, ou fazer parte de uma grande editora, é muito caro criar e usar esses importantes serviços de troca de dados.

A complexidade da indexação e compartilhamento de dados com repositórios e outros terceiros é uma barreira à mudança para operações menores. Gerenciar contratos de metadados, configurar feeds de dados e adicionar metadados é uma tarefa demorada, mas um custo necessário para fazer negócios. Dito isto, uma vez que você perceba o compromisso de tempo para cada troca ou acordo, poderá, por natureza, tornar-se conservador.

Quanto mais complexidade você assume, maior o risco e o custo. Além disso, ao estabelecer essas conexões técnicas, o pensamento de um dia – eventualmente mover plataformas ou sistema de gerenciamento de conteúdo se torna cada vez mais difícil e caro (grifo nosso).

Ryvarden (2019) comenta que a comunidade de publicações científicas precisa de toda a capacidade possível de mudança, devido a vários fatores:

- A taxa de evolução técnica geral aumentou.
- O setor está enfrentando alguns desafios importantes, como Acesso Aberto e maior foco na privacidade, pressionando os modelos de negócios atuais e a capacidade de financiar mudanças de TI.
- Uma queda na receita de publicidade e uma distribuição física mais cara pressionam os periódicos de impressão a mudarem para a web.

“Para oferecer suporte a esses serviços digitais, a criação e manutenção de sistemas de TI internos está se tornando mais complexa e, portanto, mais cara. Vários fatores criaram uma escassez global de desenvolvedores em todo o mundo ocidental” (RYVARDEN, 2019).

Ryvarden (2019) comenta o seguinte:

Embora periódicos e sociedades bem estabelecidos possam ter recursos internos para entender, manter e experimentar novos parceiros e demandas, publicações menores e mais jovens podem ter dificuldades não apenas com os recursos financeiros, mas também com o conhecimento da equipe para escolher e manter a indústria tecnologia e processos padrão. Provavelmente é fácil esquecer aqueles que trabalham com publicações científicas por 10 a 20 anos, mas pessoas menos experientes geralmente se sentem sobrecarregadas com o tsunami de abreviações e possíveis parceiros.

Ainda assim, Ryvarden (2019) questiona “como essas barreiras para se tornar um membro pleno do mundo das publicações científicas podem ser reduzidas significativamente”. O autor responde ao mesmo tempo que questiona, esclarecendo que “uma simplificação gradual feita por um parceiro na época não dará à comunidade editorial a ajuda significativa necessária.”

Uma razão de quebrar essa barreira é continuar acreditando que “precisamos mudar o sistema atual de envio de dados para facilitar a capacidade de terceiros de coletar facilmente os dados de que precisam. Em vez de tentar simplificar os padrões atuais e dar a cada diário um novo trabalho, é o pequeno grupo de receptores que deve mudar” (RYVARDEN, 2019).

Ryvarden (2019) prossegue argumentando:

Os receptores atuais devem buscar os dados necessários em vez de cada periódico criar e enviar dados. Economicamente, é uma escolha simples: permita que um punhado de indexadores e coletores de dados alterem seus sistemas de entrada, em vez de um setor inteiro tentar concordar e depois mudar milhares de sistemas.

Se todo o conteúdo fosse marcado com um dos formatos de metatag mais comuns, os receptores de dados encontrariam e selecionariam novo conteúdo. Os receptores de dados devem ser permissivos e suportar vários padrões. Ao usar essas tecnologias já amplamente utilizadas, serviços como **PubMed**, vários serviços **Crossref**, Scopus, **Mendeley** e outros podem rastrear e coletar o que desejam. As permissões ainda seriam necessárias, mas poderiam ser simplificadas. Também é possível criar alguns padrões simples para informar sobre permissões - por exemplo, você pode indexar todos os dados, mas não pode servir os artigos completos para os usuários finais (grifo nosso).

Assim, a mudança para o rastreamento levará tempo e uma variação de opinião pelos coletores de dados, mas isso pode e deve ser resolvido. A alternativa é continuar adicionando e alterando os requisitos, um por um, prolongando a situação fragmentada e não democrática em que estamos agora (RYVARDEN, 2019).

5. NOVOS SISTEMAS DE CITAÇÕES ABERTAS

As citações são os elos que unem nosso conhecimento científico e cultural. São dados primários que fornecem a proveniência e uma explicação de como sabemos os fatos (I4OC, 2019)⁵².

52 Disponível em: <https://i4oc.org/>. Acesso em: 3 dez. 2019.

Permitem-nos atribuir e creditar contribuições científicas e permitem avaliar a pesquisa e seus impactos. Em suma, as citações são o veículo mais importante para a descoberta, disseminação e avaliação de todo o conhecimento acadêmico (I4OC, 2019).

De acordo com um post publicado na revista **Nature** em 2014, estima-se que a produção científica global dobra a cada nove anos, no caso das citações, e os sistemas computacionais que as rastreiam permitem que os pesquisadores e o público acompanhem desenvolvimentos significados em qualquer campo. Para que isso seja possível, é essencial ter acesso irrestrito aos dados bibliográficos e de citação em formato legível por máquina.

Enquanto isso, outros sistemas e bases de dados estão se unificando para que seja possível a abertura livremente das citações e dados abertos à pesquisa científica. Nesse viés, temos a iniciativa do projeto *Initiative for Open Citations* (I4OC). O projeto tem atualizado constantemente os dados de citações abertas por meio de grandes publishers e sistema de catálogos, como a Crossref, que oferece em seu catálogo *Crossref Metadata Search* (CMS), mais de 100 milhões de citações, as quais estão devidamente atribuídas com o identificador digital persistente, comumente conhecido como *Digital Object Identifier* (DOI).

O I4OC surgiu em um evento do OASPA em 2016⁵³, como promessa de quebrar a barreira das citações do acesso fechado e disponibilizar o acesso ao público para o aceleração das suas pesquisas. Isso tudo vai ao encontro da Ciência Aberta e do próprio Movimento do Acesso Aberto, iniciado no ano de 2002 em Budapeste.

“Como o número de publicações acadêmicas é dobrado a cada nove anos, as citações – e os sistemas computacionais que as rastreiam – permitem que pesquisadores e o público se mantenham a par de desenvolvimentos significativos em qualquer campo” (I4OC, 2019). Para que isso seja possível, é essencial ter acesso irrestrito aos dados bibliográficos e de citação em formato legível por máquina. A projeção do número de citações abertas pelo sistema já ultrapassa 58%, conforme se observa na Figura 9.1.

53 Open Access Scholarly Publishing (OASPA). Disponível em: <https://oaspa.org/>. Acesso em: 3 dez. 2019.

“Em setembro de 2019, a fração de publicações com referências abertas aumentou de 1% para 59% dos 47,6 milhões de artigos com referências depositadas no Crossref” (I4OC, 2019).

Figura 9.1 - Projeção de citações abertas pela I4OC

Quantas citações estão abertas hoje?



Fonte: I4OC – 03 dez. 2019. Disponível em: <https://i4oc.org/#faq>

“O atual sistema de comunicação acadêmica expõe inadequadamente as redes de conhecimento que já existem em nossa literatura. Os dados de citação geralmente não estão disponíveis gratuitamente para acesso, geralmente estão sujeitos a licenças inconsistentes e difíceis de analisar, e geralmente não são legíveis por máquina” (I4OC, 2019).

Com isso, a **I4OC** (2019) tem os seguintes benefícios de amparo às citações abertas:

- O estabelecimento de uma rede pública global de dados de citações acadêmicas vinculadas para aprimorar a descoberta de conteúdo publicado, tanto de acesso por assinatura quanto por Acesso Aberto. Isso beneficiará particularmente indivíduos que não são membros de instituições acadêmicas com inscrições em bancos de dados de citações comerciais.
- A capacidade de criar novos serviços com base nos dados de citações abertas, para benefício de editores, pesquisadores, agências de fomento, instituições acadêmicas e público em geral, além de aprimorar os serviços existentes.
- A criação de um gráfico de citação pública para explorar conexões entre campos de conhecimento e acompanhar a evolução de ideias e disciplinas acadêmicas.

Vale lembrar que a metade das publicações indexadas na Crossref, uma das consorciadas da I4OC, se encontram para Acesso Aberto, o que resulta em meio bilhão de citações, porém ainda faltam milhões de **referências no Crossref**,

como mostra no estudo de Van Eck *et al.* (2018), que faz um comparativo entre as citações coletadas pela WoS e a Scopus, cuja análise “mostrará que mais de três quartos das referências **na base de dados** WoS, e mais de dois terços das referências de Scopus podem ser encontradas **no catálogo da** Crossref, com cerca de metade dessas referências abertas” (inclusão e grifo nosso).

6. CONSIDERAÇÕES FINAIS

Considera-se que a indexação, em particular a de periódicos, é parte integral do fluxo da comunicação científica, controle bibliográfico, preservação, bibliometria e cienciometria. Pode-se afirmar que a visibilidade e os benefícios em se ter uma publicação indexada são muito grandes, pois o periódico deve pleitear a inclusão em bases de dados nacionais e internacionais, de acordo com a área temática que abrange. Quanto maior o número de bases de dados nacionais e internacionais em que figurar, maior será a valorização de qualidade, produtividade e sua difusão indireta.

Os fornecedores de bases de dados, no caso indexadores, estão atentos e perceptivos em querer adotar como complemento dos artigos os dados abertos de pesquisa, submetidos às publicações no âmbito das diversas áreas do conhecimento.

Sendo assim, a real importância dos periódicos científicos é sua função como canais disseminadores da produção científica nas mais variadas áreas do conhecimento. Seu objetivo primordial é a criação, disseminação e preservação das informações dessas áreas.

A indexação da comunicação científica com certeza é muito relevante e importante, pois sem ela não haveria os rankings de citações e a visibilidade para as publicações, agora alcançando também o leque dos dados abertos de pesquisa.

As grandes bases como Web of Science, que criou em 2015 uma nova base de dados para abarcar a literatura científica de todas as áreas e todos os idiomas, como a ESCI – *Emerging Sources Citation Index*, teve a preocupação e a consciência de que precisaria expandir suas áreas, sem barreiras iniciais no processo de solicitação de indexação no contexto da sua *core collection*. Essa abertura facilita, também, a possibilidade da indexação dos dados abertos de pesquisa.

Assim, a WoS construiu essa base que faria a ponte no processo de filtragem e emergência de novos periódicos futuramente na *core collection* da WoS, isto é, após realizarem um estágio de ganhos bibliométricos e reconhecimentos nas respectivas áreas para alcançarem notoriedade e qualidade científica, possibilitando dessa forma a sua possível inclusão na *core collection*. Todo o processo de solicitação de indexação na WoS antes da criação da ESCI era muito criterioso, e as perspectivas de poder ingressar na WoS eram mínimas.

E como vimos o ponto de vista dessa grande base de dados no Quadro 1, ela está propensa a acolher o recebimento de dados abertos de pesquisa e publicações emergentes.

A Scopus, por meio da Elsevier, mostra que está acompanhando o desenvolvimento e a evolução da Ciência Aberta para com a aplicabilidade dos dados abertos de pesquisa na construção de novos conhecimentos.

A I4OC, como iniciativa fomentada por diversos agentes e instituições, está ciente de que o ponto principal para tornar os dados abertos de pesquisa acessíveis é quebrando o bloqueio ao acesso dos dados brutos de pesquisa, ou seja, as citações, e isso atingirá muito brevemente as grandes editoras para que disponibilizem gratuitamente as citações abertas.

Assim, notamos que as bases de dados de Acesso Aberto e até mesmo as pagas estão se posicionando em favor do acompanhamento da Ciência Aberta, pois o que as move no momento são as pesquisas e artigos que são submetidos dia a dia nestes grandes indexadores. Além disso, a busca frenética dos consumidores, dos usuários e da comunidade científica é consideravelmente oportuna para o aumento da produção científica nas diversas áreas do conhecimento, fazendo com que não percam dados, possibilitando o uso, reúso e reprodutibilidade de uma experiência científica, pois isso mostra que estamos nos movimentando para acompanhar o processo de progresso do conhecimento científico nos diversos lugares do mundo.

7. REFERÊNCIAS

ABADAL, E..Ciência Aberta, um modelo a ser definido com muitos desafios à frente. **Hipertext.Net**, n. 19, p.1-12, 2019. DOI: 10.31009/hipertext.net.2019.i19.01. Disponível em: <http://hdl.handle.net/2445/146558>. Acesso em 17 jan. 2020.

ALBAGLI, S.; APPEL, A. L.; MACIEL, M. L. E-Science e Ciência Aberta: questões debate. *In*: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO (ENANCIB), 14., 2013, Florianópolis. **Anais** [...]. Florianópolis: ANCIB, 2013. Disponível em: <http://ridi.ibict.br/handle/123456789/465>. Acesso em 17 jan. 2020.

BELLUZZO, R. C.B. A competência em informação do profissional bibliotecário em inter-relação com a ciência e o acesso aberto. *In*: SANTOS, G. C.; MARTINS, V. S. G. dos (org.). **Ciência Aberta, sistemas e ambientes de informação**: do acesso às boas práticas de pesquisa. Campinas, SP: Ed.NE-SBU, 2019. (Coleção SBU). ISBN: 978-85-85783-98-3. Disponível em: <https://econtents.bc.unicamp.br/omp/index.php/ebooks/catalog/view/118/978-85-85783-98-3/362-2>. Acesso em: 3 dez. 2019.

DIAS, D. A. (deborah.dias@clarivate.com). Pesquisa para elaboração de paper [mensagem pessoal]. Mensagem recebida por ppec@unicamp.br em 12 nov. 2019.

ELSEVIER. Elsevier Data Search. **Mendeley Blog**, 2017. Disponível em: <https://blog.mendeley.com/2017/09/18/introducing-elsevier-datasearch-beta-two/>. Acesso em: 15 dez. 2019.

INITIATIVE FOR OPEN CITATIONS (I4OC). About. Disponível em: <https://i4oc.org/>. Acesso em: 3 dez. 2019.

MACHADO, J. Dados abertos e Ciência Aberta. *In*: ALBAGLI, S.; MACIEL, M. L.; ABDO, A. H. (org.). **Ciência Aberta, questões abertas**. Brasília: IBICT; Rio de Janeiro: UNIRIO, 2015.

MARTINS, R. D.; ANJOS, C. R.; MELO, E. S. A atuação do bibliotecário no uso e reúso da informação: e-Science, repositórios institucionais e editoração de periódicos. **Tendências da Pesquisa Brasileira em Ciência da Informação**, v. 8, n. 1, 2015. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/119563>. Acesso em: 14 dez. 2019.

MOSTAFA, S. P.; TERRA, M. Das cartas iluministas às listas de discussão. **DataGramZero**, v. 1, n.3, jun. 2000. Disponível em: <http://www.dgz.org.br/jun00/Fart.htm>. Acesso em 17 jan. 2020.

OLIVEIRA, T. The preprint state: an overview about preprint ecosystem. *In: SciELO 20 anos*, 2018. São Paulo, SP. **Painel** [...]. São Paulo, SP: SciELO, 2018. Disponível em: https://www.slideshare.net/scielo/4-thaiane-moreira-de-oliveira?from_action=save. Acesso em: 3 dez. 2019.

PERES, D. S. (denise.peres@scielo.org). Pesquisa para elaboração de paper [mensagem pessoal]. Mensagem recebida por ppec@unicamp.br em 01 nov. 2019.

RYVARDEN, E. The need to simplify indexing and data sharing. **The Scholarly Kitchen**, Minneapolis. Post. 11 Dic. 2019. Disponível em: <https://scholarlykitchen.sspnet.org/2019/12/11/the-need-to-simplify-indexing-and-data-sharing/>. Acesso em: 3 dez. 2019.

SANTOS, G. C.; XAVIER, I. D. C. M. Fontes de indexação importantes para a pesquisa. **Blog PPEC**, Campinas, v. 2, n. 2, fev. 2018. ISSN 2526-9429. Disponível em: <http://periodicos.sbu.unicamp.br/blog/index.php/2018/03/02/indexacao>. Acesso em 17 jan. 2020

SANTOS, G.S. Indexação de publicações acadêmicas universitárias: portais temáticos e suas vantagens para as publicações periódicas. *In: ALMEIDA, M.L.P. (org.). Produção do conhecimento científico e formação do pesquisador na América Latina: as investigações de políticas educacionais em xeque!*. Campinas, SP: Mercado de Letras, 2017. ISBN 978-857591-439-7.

VAN ECK, N. J. *et al.* Crossref as a new source of citation data: A comparison with Web of Science and Scopus. **CWTS Blog**, 17 jan. 2018. Disponível em: <https://www.cwts.nl/blog?article=n-r2s234>. Acesso em 17 jan. 2020.

VAN NOORDEN, R. Global scientific output doubles every nine years. **Blog Nature**, 2014. Disponível em: <http://blogs.nature.com/news/2014/05/global-scientific-output-doubles-every-nine-years.html>. Acesso em; 19 dez. 2019.

8. SOBRE OS AUTORES



Gildeir Carolino Santos

Lattes: <http://lattes.cnpq.br/1221773207784315>

ORCID: <https://orcid.org/0000-0002-4375-6815>

Bibliotecário formado pela PUC-Campinas. Mestre e Doutor em Educação pela Universidade Estadual de Campinas (UNICAMP). Pós-doutor em Divulgação Científica pelo Laboratório de Estudos Avançados em Jornalismo da Universidade Estadual de Campinas (UNICAMP). Atua como editor científico de periódicos das áreas da Biblioteconomia e Ciência da Informação. Atualmente é gerente do Portal de Periódicos Eletrônicos Científicos da UNICAMP.

9. COMO CITAR

SANTOS, G. C. Como os indexadores estão vendo a questão de dados abertos de pesquisa para o processo de indexação?. In: SHINTAKU, M.; SALES, L. F; COSTA, M. (org). **Tópicos sobre dados abertos para editores científicos**. Botucatu, SP: ABEC, 2020. p. 97-118. DOI: 10.21452/ 978-85-93910-04-3.cap9.

Análise sobre dados abertos e regulação autoral no contexto da editoria científica

Rosilene Paiva Marinho de Sousa
Guilherme Ataíde Dias
Marckson Roberto Ferreira de Sousa

1. INTRODUÇÃO

O avanço das tecnologias de informação e comunicação tem contribuído para o fortalecimento a nível global de uma cultura de acesso aberto, seja por meio de repositórios institucionais ou periódicos científicos, inclusive pelo aprimoramento de metodologias e procedimentos que permitem o uso, reúso e compartilhamento de dados de pesquisa.

A importância do acesso aberto a dados científicos proporciona benefícios para toda a comunidade científica pela promoção ao acesso facilitado dos resultados de pesquisas, permitindo que elas possam ser exploradas em diversos contextos. Nessa perspectiva, torna-se relevante refletir sobre a proteção autoral adequada em face do acesso e compartilhamento dos dados frente a repositórios institucionais, periódicos científicos, *data papers*, *data journals*, entre outras modalidades. Assim, pode-se questionar: como se configuram as conexões existentes entre os conjuntos de dados abertos e os periódicos científicos brasileiros considerando a regulação autoral?

Decorrente dessa problemática, o presente trabalho vislumbra a propositura de uma discussão sobre o impacto do acesso aberto em periódicos científicos

e, em particular, de uma gradativa cultura de dados abertos, considerando a importância da regulação autoral no Brasil na perspectiva da editoria científica. Examina-se o acesso aberto ao conhecimento científico sobre o ponto de vista de seu impacto enquanto instrumento de desenvolvimento socioeconômico e tecnológico a partir do acesso e compartilhamento dos dados de pesquisa, bem como a configuração de iniciativas que permitam pesquisadores fazerem uso e reuso desses dados.

Nesse sentido, torna-se relevante considerar, no contexto da legislação brasileira, como está estabelecida a proteção dos direitos autorais, considerando suas principais características a partir da distinção de dois direitos básicos compreendidos no direito autoral, a saber, o direito moral e o direito patrimonial do autor sobre a obra. A partir dessa visão, apresenta-se uma discussão sobre como os editores científicos poderiam pensar uma política autoral numa abordagem sobre acesso aberto a dados públicos.

A relevância desse trabalho está em apresentar uma contribuição que busca expor características da lei de direitos autorais capazes de auxiliar editores científicos na construção de uma política autoral para a Ciência Aberta. Diante disso, torna possível uma visão voltada para a análise teórica sobre a necessidade de considerar as distinções estabelecidas entre a utilização de licenças abertas, como a *Creative Commons*, e as formas de transferência de titularidade de direitos patrimoniais do autor sobre a obra, adotada no âmbito da Lei de Direitos Autorais.

2. CIÊNCIA ABERTA: ACESSO E COMPARTILHAMENTO DE DADOS DE PESQUISA

As tecnologias de informação e comunicação têm possibilitado um crescente volume de conteúdos variados que se tornam disponíveis em ambientes informacionais digitais. Consequentemente, o uso dessas tecnologias tem contribuído gradativamente para o fortalecimento da cultura de acesso, principalmente por meio de repositórios institucionais e periódicos científicos de acesso aberto. Nesse sentido, Guedes (2010, p. 214), ao enfatizar que a produção de conhecimento e sua apropriação social têm se tornado um tema em expansão a nível global, expõe o seguinte:

[...] Observa-se que a produção/distribuição de informação e conhecimento nesse novo contexto vem assumindo uma forte centralidade. Questão que fica mais clara quando nos deparamos com as políticas de propriedade intelectual, que vêm se tornando cada vez mais restritivas, e com os movimentos contrários à apropriação privada do conhecimento e da informação.

Na conjuntura dos debates que envolvem o acesso aberto ao conhecimento científico, evidencia-se a discussão sobre a função da ciência enquanto instrumento de efetivo desenvolvimento socioeconômico e tecnológico, capaz de proporcionar à sociedade o conhecimento de saberes desenvolvidos no âmbito das pesquisas científicas. Essa visão conduz ao pensamento de que, com o auxílio das tecnologias de informação e comunicação, pode-se pensar o cenário científico numa perspectiva de Ciência Aberta, a qual se caracteriza pelo acesso e compartilhamento dos dados das referidas pesquisas.

Segundo Santos, Almeida e Henning (2017, p. 11), o gradativo volume de dados e seu uso no âmbito científico “tem promovido debates sobre sua importância, usos, potencialidades, desafios e impactos numa sociedade conectada e global, apoiada pelo uso de tecnologias digitais cada vez mais potentes e especializadas”. Nesse contexto, ao realizar a promoção da abertura de dados científicos, a Ciência Aberta permite a agregação, verificação e replicação para seu uso e reúso, permitindo a ampliação dos horizontes das pesquisas e seus resultados práticos nos processos de inovação, visando à melhoria da oferta de produtos e serviços para toda sociedade.

Esse avanço tem permitido discussões sobre a gestão de dados científicos por meio de agências governamentais, evidenciando o surgimento de abordagens sobre publicação de dados (*data publishing*) de pesquisa em repositórios institucionais e periódicos científicos, apresentando modelo de publicação científica diverso do tradicional em suas modalidades e formas. Segundo Curty e Avenirier (2017), a finalidade seria a maximização da relação custo-eficácia dos recursos socioeconômicos, bem como a ampliação do tempo da utilidade e aplicação dos dados originais.

Entre as abordagens existentes, pode-se destacar a abordagem referente aos artigos de dados (*data papers*). Conforme exposto por Austin *et al.* (2016, p. 82), pode-se conceituar *data papers* como:

A data article is a ‘data publishing’ product, also known as a ‘data descriptor’, that may appear in a data journal or any other journal. When publishers refer to ‘data publishing’ they usually mean a data article rather than the underlying dataset. Data articles focus on making data discoverable, interpretable and reusable, rather than testing hypotheses or presenting new interpretations (by contrast with traditional journal articles)⁵⁴.

A importância dessa discussão sobre artigos de dados (*data papers*) acontece pelo fato de servirem como instrumento para que dados sejam mais visíveis, contribuindo para o uso, reúso e compartilhamento, bem como para divulgação da modalidade de política de direito autoral adotado pelo mesmo.

Segundo Dias, Anjos e Araújo (2019, p. 6), “No Brasil existe pouca literatura científica publicada acerca de como os pesquisadores fazem uso e reúso dos dados de pesquisa e como se configuram as práticas de gestão relacionadas com este insumo por parte dos membros desta comunidade”. Neste trabalho, considera-se o conceito de dado adotado por Setzer (2015, online), em que o dado constitui “uma seqüência de símbolos quantificados ou quantificáveis”. Sua importância está em fornecer subsídios para seu tratamento, transmissão e uso.

Visando favorecer o acesso aos dados de pesquisa com financiamento público, a Organização para Cooperação e Desenvolvimento Econômico (*Organization for Economic Co-operation and Development*) assumiu a liderança no desenvolvimento de princípios e diretrizes para que o referido acesso possa ocorrer, entre os quais destacam-se Abertura, Flexibilidade, Transparência, Conformidade Legal, Proteção da Propriedade Intelectual (OECD, 2007). Segundo a referida organização, o efetivo acesso a dados de pesquisa tornou-se importante condição para reforçar a investigação científica aberta.

A declaração de Pequim sobre dados de pesquisa, intitulada “*The Beijing Declaration on Research Data*” (CODATA, 2019, p. 1), em seu preâmbulo esclarece que:

54 Um artigo de dados é um produto de “publicação de dados”, também conhecido como “descriptor de dados”, que pode aparecer em um jornal de dados ou em qualquer outro jornal. Quando os editores se referem à “publicação de dados”, eles geralmente se referem a artigos que indicam a descrição dos dados em vez do próprio conjunto de dados. O Foco dos artigos de dados é tornar os dados detectáveis, interpretáveis e reutilizáveis, em vez de testar hipóteses ou apresentar novas interpretações (em contraste com artigos de periódicos tradicionais) (AUSTIN *et al.*, 2016, p. 82, tradução nossa).

[...] The Beijing Declaration is intended as a timely statement of core principles to encourage global cooperation, especially for public research data. It builds on and acknowledges the many national and international efforts that have been undertaken in the policy and technical spheres on a worldwide basis⁵⁵.

Segundo exposto no princípio 3 da Declaração, os princípios de três a sete referem-se a dados “FAIR” (dados que são encontráveis, acessíveis, interoperáveis e reutilizáveis) que foram projetados para serem processados tanto para máquinas quanto por humanos. Wilkinson *et al.* (2016, p. 4) apresentam um quadro, no qual elencam os principais aspectos abordados em cada princípio, conforme especificado no Quadro 1:

55 A Declaração de Pequim pretende ser uma declaração oportuna dos princípios fundamentais para incentivar a cooperação global, especialmente para dados públicos de pesquisa. Baseia-se e reconhece os muitos esforços nacionais e internacionais que foram empreendidos nas esferas política e técnica em nível mundial (CODATA, 2019, p.1, tradução nossa).

Quadro 1 – Os Princípios Orientadores FAIR

The FAIR Guiding Principles

To be Findable:

- F1. (meta)data are assigned a globally unique and persistent identifier
- F2. data are described with rich metadata (defined by R1 below)
- F3. metadata clearly and explicitly include the identifier of the data it describes
- F4. (meta)data are registered or indexed in a searchable resource

To be Accessible:

- A1. (meta)data are retrievable by their identifier using a standardized communications protocol
 - A1.1 the protocol is open, free, and universally implementable
 - A1.2 the protocol allows for an authentication and authorization procedure, where necessary
- A2. metadata are accessible, even when the data are no longer available

To be Interoperable:

- I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- I2. (meta)data use vocabularies that follow FAIR principles
- I3. (meta)data include qualified references to other (meta)data

To be Reusable:

- R1. meta(data) are richly described with a plurality of accurate and relevant attributes
 - R1.1. (meta)data are released with a clear and accessible data usage license
 - R1.2. (meta)data are associated with detailed provenance
 - R1.3. (meta)data meet domain-relevant community

Fonte: Wilkinson *et al.* (2016, p. 4)

De acordo com Wilkinson *et al.* (2016, p. 4), para que os dados se tornem encontráveis, torna-se necessário que os metadados sejam atribuídos a um identificador persistente, único e global, além de incluir de forma clara e explícita um identificador dos dados que descrevem os metadados registrados num recurso pesquisável. Quanto aos dados serem acessíveis, a recuperação dos metadados deve ocorrer pelo seu identificador por meio de um protocolo padrão de comunicação, que por sua vez deve ser aberto, gratuito, universalmente implementável e que possa permitir um procedimento de autenticação

e autorização quando se fizer necessário, mesmo quando os dados deixarem de ser disponibilizados. No caso da interoperabilidade, devem utilizar uma linguagem formal, acessível, compartilhada e amplamente aplicável, devendo utilizar vocabulário que siga os referidos princípios FAIR e que inclua referências qualificadas para outros metadados. Concernente aos dados reutilizáveis, os metadados devem apresentar descrição com pluralidade de características precisas e relevantes, sendo disponibilizados com licenças de uso claras e acessíveis, bem como deve haver uma associação a uma proveniência detalhada e devem estar de acordo com padrões comunitários relevantes para o domínio.

No Brasil, segundo Dias, Anjos e Rodrigues (2019, p. 177), “A adoção de práticas que possibilitem que pesquisadores brasileiros reusem dados científicos tem começado a ser difundida de forma sistemática recentemente, embora de forma ainda tímida”. Entre as iniciativas destaca-se, primeiramente, o Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT) como pioneiro nas iniciativas voltadas para dados abertos no campo da pesquisa científica. Publicou o denominado Manifesto de Acesso Aberto a Dados da Pesquisa Brasileira para Ciência Cidadã, cujo escopo seria a ampliação de uma política de apoio ao acesso aberto e livre à informação científica no Brasil, formalizando a adesão ao movimento europeu, passando a reconhecer “os dados de pesquisa como um recurso imprescindível para as ações de Ciência Aberta, Ciência para todos, Ciência Cidadã” (IBICT, 2016, on-line). O referido manifesto indica algumas diretrizes gerais, entre as quais destacam-se:

formular políticas institucionais mandatórias para assegurar que os dados de pesquisas gerados por seus pesquisadores sejam apropriadamente gerenciados, tendo em vista o acesso aberto, considerando que são produtos de pesquisa desenvolvidos com recursos públicos;[...]estimular pesquisas colaborativas e interdisciplinares e constituir redes de trabalho colaborativo;[...]incorporar dados de pesquisa na memória acadêmica;[...]estimular e apoiar ações de Informática e de informação para desenvolvimento de repositórios de dados de pesquisa e de outras infraestruturas necessárias à curadoria das coleções de dados de valor reconhecido; adotar hardware e software abertos e outras ações que proporcionem a interoperabilidade tecnológica;[...]recomendar aos pesquisadores e profissionais de informação, como os bibliotecários, por exemplo, a elaboração de plano de gestão de dados de pesquisa para os dados gerados, segundo os padrões adotados para a área disciplinar específica; e publicizar e facilitar

o acesso aos dados de pesquisa , a fim de que a sociedade e o cidadão tomem conhecimento do que é produzido na ciência e seus benefícios (IBICT, 2016, on-line).

Além disso, outra iniciativa que se pode destacar diz respeito à participação do IBICT no primeiro encontro anual da rede internacional de implementação do *GO FAIR*, juntamente com a Fiocruz, que ocorreu na universidade de Leiden, na Holanda. O *Global Open FAIR – (GO FAIR)* – trata-se de uma iniciativa que busca fazer com que dados fragmentados e desconectados sejam encontráveis, acessíveis, interoperáveis e reutilizáveis por máquinas e pessoas, em um ambiente global compartilhado para pesquisa (ICICT, 2019, on-line).

Essas iniciativas, de forma abrangente, têm permitido uma ampliação nas abordagens sobre dados abertos diante da interação dos princípios FAIR e sua associação ao humano-computador.

3. REGULAÇÃO AUTORA NO CONTEXTO DA EDITORIA CIENTÍFICA

Para que se possa discutir sobre as licenças abertas no âmbito do acesso aberto ao conhecimento científico, necessário se faz discutir, no âmbito da legislação brasileira, aspectos de proteção sobre direitos autorais, para posteriormente, numa abordagem sobre acesso aberto a dados públicos, verificar como editores poderiam pensar a respeito da política autoral em periódicos científicos.

O direito autoral no Brasil, além de seguir diretrizes estabelecidas por meio de ratificação de convenções internacionais, encontra amparo tanto a nível constitucional como infraconstitucional. Conforme exposto em Sousa e Dias (2017), no âmbito da Constituição Federal de 1988 (BRASIL, 1988), nos direitos e garantias fundamentais, especificamente no art. 5º, incisos XXVII e XXVIII, encontram-se enunciados dispositivos referentes ao direito autoral, cuja regulamentação encontra-se efetivada por meio da Lei nº 9.610/1998, que altera, atualiza e consolida a legislação sobre direitos autorais (BRASIL, 1998).

No domínio da Lei de Direitos Autorais (LDA), em conformidade com seu artigo 1º, compreende-se por Direito Autoral os direitos de autor e os direitos que lhes são conexos. Nessa perspectiva, por definição do artigo 11 da LDA, “Autor é a pessoa física criadora de obra literária, artística ou científica” (BRASIL,

1998, on-line). Os direitos conexos, segundo o que preceitua Fragoso (2009), compreendem “um direito cuja natureza aproxima-o dos direitos atribuídos aos autores, sem que com eles se confundam”, sendo dessa forma aplicáveis, no que couber aos direitos dos artistas intérpretes ou executantes, dos produtores fonográficos e das empresas de radiodifusão.

Os direitos de autor encontram acolhimento em sua teoria dualista por estarem fundados na coexistência de dois direitos basilares, quais sejam, de natureza moral e patrimonial. Nesse sentido, a LDA estabelece, no artigo 22, que “pertencem ao autor os direitos morais e patrimoniais sobre a obra que criou” (BRASIL, 1998, on-line).

Conforme esclarece Fragoso (2009), os direitos de natureza moral, que envolvem a criação de um autor, tratam-se da manifestação de sua vontade individual, representados pela sua forma de expressão. São os direitos da personalidade, pelo seu caráter de inalienabilidade e irrenunciabilidade, previsto no artigo 49, inciso I, da LDA. Os direitos morais do autor são inalienáveis, pois estão representados por faculdades insuscetíveis de negociação e se ancoram na própria personalidade do autor; já a irrenunciabilidade dos direitos morais tem como pressuposto a impossibilidade fática e moral de renunciar-se à própria personalidade.

No campo do direito autoral, compreendem direitos morais do autor aqueles previstos no artigo 24 da LDA, entre os quais se podem citar o direito de reivindicar a autoria da obra, também denominado de direito de paternidade; o direito a nomeação, correspondente a ter seu nome, pseudônimo ou sinal convencional indicado ou anunciado, como sendo o do autor da obra; de conservar a obra inédita; de assegurar a integridade da obra, resguardando-se de modificações ou prática de atos que possam prejudicá-la ou atingi-lo como autor assim em sua reputação ou honra; de modificação da obra; o direito de retirada, resultante de afronta à reputação e à imagem do autor; de acesso a exemplar único e raro de sua obra com finalidade de preservar sua memória; e o direito ao repúdio, em obra que não participou e que seu nome esteja sendo utilizado, podendo desvincular-se da mesma (BRASIL, 1998).

Quanto aos direitos de natureza patrimonial, tratam-se daqueles de natureza econômica sobre suas criações. O artigo 31 se refere às várias formas de utilização da obra que são independentes entre si, tendo como objetivo garantir ao autor o máximo de aproveitamento sob sua criação (FRAGOSO, 2009). Os direitos patrimoniais do autor estão vinculados ao direito exclusivo de utilização da obra, previsto na LDA, ao determinar, no seu artigo 28, que “cabe ao autor o

direito exclusivo de utilizar, fruir e dispor da obra literária, artística ou científica.” (BRASIL, 1998, on-line). Segundo Sousa e Dias (2017), a LDA reconhece inúmeras faculdades, como a comercialização, divulgação, reprodução, execução e tradução da própria obra.

Ao tratar da transmissão dos direitos do autor, evidencia-se que a transmissão se dá por meio da transferência da titularidade do direito de um indivíduo para outro. A previsão da transmissão de direitos encontra-se no Título III, Capítulo V, da LDA. Com base no artigo 49, inciso I, evidencia-se a proibição à transmissão dos direitos morais do autor, por serem considerados intransferíveis, restando limitada a referida transferência, aos direitos patrimoniais, por meio de atos inter vivos, constituído pela reunião de elementos como agente capaz e objeto lícito e possível. Segundo Fragoso (2009), a transferência da titularidade de um indivíduo para outro pode ser parcial ou total, a título universal ou singular, diretamente pelo autor, sucessor ou outros titulares, mediante licenciamento, concessão, cessão ou outros meios legais.

Conforme exposto em Sousa e Garcia (2015, p. 39), a cessão de direito “é caracterizada pelo aspecto de definitividade e de exclusividade patrimonial do autor que ocorre com todos os atributos ínsitos à propriedade, ou seja, o de fruir, utilizar e dispor, desde que observadas limitações de uso previstas na lei”.

O que distingue a licença da cessão de direitos é que a primeira tem caráter de provisoriedade, isto é, o licenciamento caracteriza-se por ser temporário e raramente exclusivo, bem como não transfere a titularidade dos direitos como ocorre na cessão de direitos autorais. Na referida cessão de direitos autorais, além de realizar essa transferência, ela tem caráter de definitividade e exclusividade.

Com vistas ao caráter definitivo das cessões de direito, com base nas suas modalidades, que constituem a cessão total e a título universal, cessão total a título singular, cessão parcial a título universal e cessão parcial a título singular, interessa para este trabalho a cessão total e a título universal, uma vez que essa modalidade seria a que melhor se adequa a uma proposta de dados abertos em periódicos científicos.

Segundo Sousa e Garcia (2015, p. 39), a cessão total e a título universal não dispõem de limitação espacial e temporal. Nesse sentido, Fragoso (2009, p. 352) esclarece que:

Previstas no artigo 49 e 50 da LDA, as cessões em carácter total se dão sobre o conjunto dos direitos patrimoniais de autor e destinadas a todas as modalidades de utilização (a título universal), abrangendo o direito de utilizar, fruir e dispor da obra objeto da cessão, pelo tempo de duração previsto na lei (Artigos 41 a 44) – pois a cessão é sempre definitiva e sem limitação de território de exploração.

Para melhor compreensão, pode-se pensar em uma publicação *on-line*, que pode gerar diferentes modalidades de utilização, podendo inclusive ser alterada, reproduzida, utilizada em qualquer outro suporte, traduzida ou adaptada, não havendo limitação para as formas de utilização. Caso haja a vedação de apenas uma modalidade de utilização, descaracterizaria a cessão a título universal, adotando-se outra modalidade de cessão entre as previstas na lei. No mesmo sentido, Fragoso (2009, p. 352) expõe que:

um conto ou um romance podem ser traduzidos e adaptados para o cinema ou para a televisão ou transformado em texto teatral e tal texto ser representado e tal representação ser radiodifundida; uma escultura reproduzida em cópias para venda ao público, enfim, não há limitação para as formas de utilização da obra, desde que limitadas às formas, meios e processos existentes à época do contrato.

Importa destacar que, na cessão total e a título universal, não se faz necessário novas autorizações para as modalidades de utilização já existentes. Além disso, torna-se importante evidenciar que a transmissão total e definitiva dos direitos só pode ocorrer por meio de contrato escrito, que na sua ausência estipula-se o prazo máximo de cinco anos, assumindo carácter de licença.

Diante desse contexto, as discussões sobre acesso aberto, inclusive no âmbito dos órgãos que o defende, voltam-se cada vez mais para as licenças abertas, a exemplo do *Creative Commons*, que possui várias modalidades:

Figura 10.1 - Modalidades de licenças do Creative Commons

		Can someone use it commercially?	Can someone create new versions of it?
Most Free			
↑			
			Yup, AND they must license the new work under a Share Alike License.
			Yup, AND the work must be non-commercial, but it can be under any non-commercial license.
			Yup, AND they must license the new work under a Non-Commercial Share Alike license.
	Least Free		

Fonte: Adaptado de Commonwealth of Learning (2019, on-line)

Na Figura 10.1, são apresentadas as modalidades de licenças adotadas pelo *Creative Commons*, indicadas da base para o topo, licenças das mais restritivas para as mais abertas. Nesse sentido, para uma análise mais aprofundada sobre os termos dessas licenças, torna-se necessário que os editores reflitam se essas modalidades são equivalentes ou não às características específicas das licenças e/ou cessões de direitos adotadas pela legislação de direitos autorais no Brasil, a fim de que se possa utilizá-las adequadamente no âmbito do acesso aberto, respeitando a legislação pátria.

Evidencia-se que as licenças no regulamento brasileiro, por serem estabelecidas com tempo determinado, tornam-se incompatíveis, pois, na visão de uma Ciência Aberta, não há como estabelecer um controle sobre o acesso a dados ou publicações científicas por período determinado. Entretanto, seria mais apropriado utilizar as cessões de direitos autorais, na modalidade específica de cessão total a título universal, porque além dela ocorrer por tempo indeterminado, permite a transmissão total e definitiva dos direitos.

Caso se deseje contemplar restrições a dados incluídos na categoria de interesse nacional, privacidade, entre outros, conforme ressaltados no âmbito da *The Beijing Declaration on Research Data*, poderia se pensar na cessão parcial a título singular, que se aplica a um ou alguns direitos patrimoniais, destinados a modalidades de utilização existentes, considerando adequações necessárias.

4. CONSIDERAÇÕES FINAIS

Cada vez mais os editores científicos estão se envolvendo com o acesso aberto ao conhecimento científico. Conseqüentemente, esse envolvimento demanda certo domínio sobre a regulação autoral. Nesse contexto, devem-se observar alguns elementos da LDA, ao distinguir a natureza jurídica dos direitos morais e patrimoniais, considerando que a distinção pressupõe uma regulação peculiar entre a garantia de um direito intrínseco a personalidade do autor, e um direito patrimonial passível de permitir o máximo aproveitamento de sua criação.

É conveniente que os editores compreendam a distinção estabelecida entre licença e cessão de direitos autorais, no âmbito da LDA, para que possam refletir sobre a *Licença Creative Commons*, não incorporada pela legislação brasileira. No Brasil, a comunidade científica que trabalha com acesso aberto, dados abertos, Ciência Aberta, utiliza a ideia desta licença. Entretanto, torna-se importante destacar a necessidade de se refletir se há ou não uma total compatibilidade ou equivalência entre cada licença especificada pelo *Creative Commons* e as modalidades de transferência de titularidade no âmbito da Lei de Direitos Autorais, em particular das licenças e cessões de direitos. Nesse sentido, algumas peculiaridades podem definir essa distinção, uma vez que, pela lei de direitos autorais, o licenciamento caracteriza-se por ser temporário e raramente exclusivo, assim como não transfere a titularidade dos direitos. Já a cessão de direitos, além de realizar essa transferência, tem caráter de definitividade e exclusividade na autorização de uso.

Chega-se à conclusão de que a concretização e expansão de uma cultura de dados aberta no âmbito de uma Ciência Aberta, entre outras características relevantes – como da necessidade de uma política específica que possa envolver a Ciência Aberta –, perpassa pela consolidação e fortalecimento de uma política de proteção autoral adequadamente adotada.

Pensar na cessão total e a título universal de direitos e seus requisitos – como adoção de forma contratual escrita que pode ocorrer por meio de termos de cessão de direitos autorais – talvez seja o caminho mais favorável para se trabalhar com Ciência Aberta, acesso aberto ou dados abertos em pesquisa no Brasil, obedecendo assim à legislação pátria ora em vigor, alcançando a possibilidade de utilização de dados de pesquisas de diversas maneiras.

5. REFERÊNCIAS

AUSTIN, C. C. *et al.* Key components of data publishing: using current best practices to develop a reference model for data publishing. **International Journal on Digital Libraries**, v. 18, p.77-92, 2016.

BRASIL. [Constituição(1988)]. **Constituição da República Federativa do Brasil**. Brasília, DF: Presidência da República, [1988]. Disponível em: http://www.planalto.gov.br/ccivil_03/constituicao/constituicao.htm. Acesso em: 17 jan. 2020.

BRASIL. **Lei nº 9.610, de 19 de fevereiro de 1998**. Altera, atualiza e consolida a legislação sobre direitos autorais e dá outras providências. Brasília, DF: Presidência da República, 1998. Disponível em: http://www.planalto.gov.br/ccivil_03/leis/L9610.htm. Acesso em: 12 dez. 2019.

CODATA. Committee on Data of the International Science Council. **The Beijing Declaration on Research Data**. 2019. Disponível em: <https://dx.doi.org/10.5281/zenodo.3552330>. Acesso em: 14 dez. 2019.

COMMONWEALTH OF LEARNING. Which is the most restrictive Creative Commons (CC) license? Why would an author choose a restrictive CC licence? 2019. Disponível em: <http://discourse.col.org/t/which-is-the-most-restrictive-creative-commons-cc-license-why-would-an-author-choose-a-restrictive-cc-licence/114/1>. Acesso em: 14 dez. 2019.

CURTY, R. G.; AVENTURIER, P. O paradigma da publicação de dados e suas diferentes abordagens. *In*: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 17., 2017, Marília. **Anais** [...]. Marília: ANCIB, 2017. Disponível em: <http://enancib.marilia.unesp.br/index.php/xviiienancib/ENANCIB/paper/viewFile/468/820>. Acesso em: 14 dez. 2019.

DIAS, G. A.; ANJOS, R. L. dos; RODRIGUES, A. A. Os Princípios FAIR: viabilizando o reúso de dados científicos. *In*: DIAS, G. A.; OLIVEIRA, B. M. F. de. **Dados Científicos: perspectivas e desafios**. João Pessoa: Editora UFPB, 2019, p. 177-188.

DIAS, G. A.; ANJOS, R. L. dos; ARAÚJO, D. G. de. A gestão dos dados de pesquisa no âmbito da comunidade dos pesquisadores vinculados aos programas de pós-graduação brasileiros na área da Ciência da Informação: desvendando as práticas e percepções associadas ao uso e reúso de dados. **Liinc em Revista**,

Rio de Janeiro, v.15, n.2, p. 5-31, novembro 2019. Disponível em: <http://revista.ibict.br/liinc/article/view/4683/4306>. Acesso em: 13 dez 2019.

FRAGOSO, J. H. da R. **Direito Autoral**: da antiguidade à Internet. São Paulo: Quartier Latin, 2009.

GUEDES, R. D. O Surgimento dos repositórios institucionais e uma breve análise dos instrumentos legais. *In*: CONGRESSO DE DIREITO DE AUTOR E INTERESSE PÚBLICO, 4., 2010, Florianópolis. **Anais** [...]. Florianópolis: UFSC; Fundação Boiteux, 2010. Disponível em: http://www.gedai.com.br/wp-content/uploads/2010/07/Anais_IV-CODAIP_FINAL.pdf. Acesso em: 11 dez 2019.

INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA (IBICT). **Ibict lança manifesto de acesso aberto a dados da pesquisa brasileira para Ciência Cidadã**. 2016. Disponível em: <http://www.ibict.br/sala-de-imprensa/noticias/item/478-ibict-lanca-manifesto-de-acesso-aberto-a-dados-da-pesquisa-brasileira-para-ciencia-cidada>. Acesso em: 14 dez 2019.

ICICT. **Fiocruz e IBICT marcam presença na Rede Internacional de Implementação do GO FAIR**. 2019. Disponível em: <https://www.icict.fiocruz.br/content/fiocruz-e-ibict-marcam-presen%C3%A7a-na-rede-internacional-de-implementa%C3%A7%C3%A3o-do-go-fair>. Acesso em: 14 dez. 2019.

OECD. **OECD Principles and Guidelines for Access to Research Data from Public Funding**. 2007.

SANTOS, P. X. dos; ALMEIDA, B. de A.; HENNING, P. (org.). **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.

SETZER, V. W. **Dado, Informação, Conhecimento e Competência**. 2015. Disponível em: <https://www.ime.usp.br/~vwsetzer/dado-info.html>. Acesso em: 05 dez. 2019.

SOUSA, R. P. M. de; DIAS, G. A. **A Informação e a Proteção da Propriedade Intelectual**. João Pessoa: Editora da UFPB, 2017.

SOUSA, R. P. M. de; GARCIA, J. C. R. Cessão de Direitos Autorais em Periódicos Científicos Brasileiros. In: DIAS, G. A.; OLIVEIRA, B. M. J. F. de. **Informação, Direito Autoral e Plágio**. João Pessoa: Editora da UFPB, 2015, p. 34-59.

WILKINSON, M. D. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. **Scientific Data**. 2016. Disponível em: <https://dash.harvard.edu/bitstream/handle/1/26860037/4792175.pdf?sequence=1&isAllowed=y>. Acesso em: 12 dez 2019.

6. SOBRE OS AUTORES



Rosilene Paiva Marinho de Sousa

Lattes: <http://lattes.cnpq.br/4465533418771961>

Orcid: <https://orcid.org/0000-0002-4699-8692>

Advogada. Doutorado e Mestrado em Ciência da Informação pela Universidade Federal da Paraíba (UFPB). Mestrado em Direito pelo Centro Universitário de João Pessoa (UNIPÊ). Graduação em Ciências Jurídicas e Sociais e Licenciatura em História pela Universidade Federal de Campina Grande (UFCG). Pesquisas nas áreas de Direito e Tecnologia (ênfase em Direito Empresarial, Propriedade Intelectual, Aspectos Jurídicos da Informação e Proteção de Dados).



Guilherme Ataíde Dias

Lattes: <http://lattes.cnpq.br/9553707435669429>

Orcid: <https://orcid.org/0000-0001-6576-0017>

Graduado em Ciência da Computação (UFPB) e Direito (UNIPE), Mestre em Administração (CCSU/USA), Doutor em Ciência da Informação (USP) e Pós-Doutor (UNESP). Professor Associado IV na Universidade Federal da Paraíba. Membro do Programa de Pós-Graduação em Ciência da Informação da UFPB. Atualmente é bolsista de produtividade em pesquisa (PQ-1D) do CNPq.



Marckson Roberto Ferreira de Sousa

Lattes: <http://lattes.cnpq.br/0221265788966967>

Orcid: <https://orcid.org/0000-0003-2001-1631>

Doutor e Mestre em Engenharia Elétrica pela Universidade Federal da Paraíba (UFPB). Graduação em Engenharia Elétrica (UFPB) e em Direito pelo Instituto de Educação Superior da Paraíba (IESP). Professor Associado do Departamento de Ciência da Informação e do Programa de Pós-graduação em Ciência da Informação da UFPB. Pesquisas nas áreas de Competência Informacional, Interação da Informação e Aspectos Jurídicos da Informação.

7. COMO CITAR

SOUSA, R. P. M. de; DIAS, G. A.; SOUSA, M. R. F. de. Análise sobre dados abertos e regulação autoral no contexto da editoria científica. *In*: SHINTAKU, M.; SALES, L. F; COSTA, M. (org). **Tópicos sobre dados abertos para editores científicos**. Botucatu, SP: ABEC, 2020. p. 119-135. DOI: 10.21452/ 978-85-93910-04-3.cap10.

Direitos autorais e os dados de pesquisa abertos

Enrique Muriel-Torrado
Rebeca Fernandes de Moura

1. INTRODUÇÃO

Os direitos autorais estão presentes em muitas das atividades rotineiras realizadas pelos editores de periódicos científicos. Talvez o direito mais claro de perceber na prática é o exercido antes do ato de publicação: para que um periódico possa publicar um artigo é preciso que os autores façam a cessão de alguns dos direitos autorais em conformidade com os direitos patrimoniais. Assim, os autores permitem aos editores reproduzir e comunicar publicamente sua obra, colocando os arquivos no site da revista à disposição dos interessados.

Aqueles periódicos que são de Acesso Aberto se materializam também com a aceitação prévia da submissão de uma licença *Creative Commons* (CC), a mais utilizada no âmbito científico e cultural, cuja função principal é flexibilizar alguns direitos.

A comunicação científica evolui, procurando novos formatos e modelos. Se agora a grande maioria de periódicos são exclusivamente on-line, o meio digital ajuda a apresentar outros tipos de documentos anexos e/ou complementares, como podem ser vídeos explicativos, pareceres ou dados de pesquisa (também conhecidos como dados científicos). Esta última é uma aposta promissória baseada no compartilhamento e reutilização dos dados utilizados pelos autores nas pesquisas próprias, e que parece chegar no momento adequado, com o desenvolvimento e aprimoramento de técnicas de Big Data e *Machine Learning* que, entre outras funções, podem auxiliar para criar novo conhecimento.

2. DIREITOS AUTORAIS

Os direitos autorais podem ser entendidos como “o poder que o autor, o criador, o tradutor, o pesquisador ou o artista tem de controlar o uso que se faz de sua obra” (DUARTE; PEREIRA, 2009, p. 5). O seu objetivo é proteger legalmente as obras intelectuais pela sua originalidade ou criatividade em favor do autor. Para isso, as obras devem ser criadas e fixadas em qualquer suporte físico, seja tangível ou intangível, como no caso dos artigos científicos.

No Brasil, a regulamentação dos direitos autorais é regida pela Lei de nº 9.610, de fevereiro de 1998, e podemos dividi-los em dois: os direitos morais e os patrimoniais.

Os morais são aqueles vinculados à paternidade da obra (BRASIL, 1998). Estão relacionados ao direito que o autor possui de exigir e reivindicar o seu reconhecimento como criadora da obra. Os principais direitos morais para os fins da comunicação científica são a paternidade e a integridade. A paternidade é o direito que o autor possui de ser reconhecido pela sua obra; já a integridade diz respeito ao direito que o autor tem de autorizar ou se opor a modificações ou alterações de sua obra (MURIEL-TORRADO; FERNÁNDEZ-MOLINA, 2014).

Por outro lado, os direitos patrimoniais estão relacionados diretamente à exploração econômica da obra. Regulamentados através dos artigos 28 ao 45 da Lei de Direitos Autorais (LDA), oferecem proteção às obras por determinado período de tempo. Entre os mais relevantes estão: Reprodução, Comunicação Pública, Distribuição e Divulgação.

A reprodução parcial ou integral de uma obra por parte de terceiros depende da autorização do autor; a comunicação ao público é o direito exercido quando o autor publica sua obra sem exemplares físicos (por exemplo, pela internet); por enquanto, a distribuição se produz quando existem esses exemplares tangíveis; e, por último, o direito de divulgação é o direito que apenas o autor, ou terceiros por ele autorizados, possui de decidir se o seu trabalho será conhecido pelo público e o modo como será feito.

Os direitos autorais precisam se adaptar às novas necessidades que vão surgindo nas diferentes áreas do conhecimento. O Brasil ocupa a 13ª posição entre os países com maior número de publicações durante os anos de 2011 a 2016 (CROSS; THOMSON; SIBCLAIR, 2017). Se nos próximos anos essa produção de

artigos começar a ser acompanhada dos seus dados de pesquisa, os editores precisarão saber o que fazer com eles.

3. OPEN SCIENCE E OPEN DATA (CIÊNCIA ABERTA E DADOS ABERTOS)

Uma definição de Ciência Aberta (*Open Science*) é a apresentada por Vicente-Saez e Martinez-Fuentes (2018, tradução nossa), criada a partir de 75 artigos, a qual com simplicidade integra seus principais conceitos numa frase: “Ciência Aberta é um conhecimento transparente e acessível, que é compartilhado e desenvolvido por meio de redes colaborativas”.⁵⁶

Dentro do contexto da Ciência Aberta, ganham protagonismo os dados de pesquisa abertos (Open Data), definidos pelo Relatório da Organização para a Cooperação e Desenvolvimento Econômico (OECD) como “registros factuais usados como fontes primária para a pesquisa científica e que são comumente aceitos pelos pesquisadores como necessários para validar os resultados do trabalho científico” (OECD, 2007).

Silva (2019, p. 21) entende que o conceito é mais amplo, contemplando também “às distintas ferramentas como protocolos, códigos numéricos, gráficos e tabelas que são necessárias para recolher e organizar os dados, tanto em trabalhos de campo quanto em laboratório”. Isto é, “todas as evidências que um investigador necessita para validar suas conclusões após uma pesquisa.”

Alguns desses dados têm um valor “imediate e duradouro” e, por enquanto, outros ganham valor no decorrer do tempo (SAYÃO; SALES, 2015). É a partir do reuso dos dados que aparece o potencial de incrementar valor aos mesmos, em decorrência de novas pesquisas e, conseqüentemente, novas descobertas, adquirindo assim um valor agregado.

Mas para que esses dados de pesquisa sejam abertos e possam ser utilizados, é preciso que apresentem algum tipo de licença.

56 “Open Science is transparent and accessible knowledge that is shared and developed through collaborative networks.”

4. DIREITOS AUTORAIS E DADOS DE PESQUISA

O conceito de autoria aplicado a dados de pesquisa pode levantar alguns questionamentos. A lei brasileira entende que, para que uma obra tenha proteção, deve ser uma “criação do espírito”, isto é, uma criação do intelecto humano (BRASIL, 1998).

Se na recopilação dos dados de pesquisa existe uma mínima atividade criativa na seleção ou organização das variáveis que vão compor o conjunto de dados, poderia se considerar esse tipo de obra como análoga às bases de dados, que, no Inciso XIII do Art. 7 da LDA, são contempladas como merecedora de proteção: “[...] as coletâneas ou compilações, antologias, enciclopédias, dicionários, bases de dados e outras obras, que, por sua seleção, organização ou disposição de seu conteúdo, constituam uma criação intelectual” (BRASIL, 1998). Nesses casos, poderia se utilizar uma licença CC. A recomendação dos autores seria utilizar licenças mais abertas possíveis, isto é, CC by (reconhecimento) ou CC by sa (reconhecimento - compartilhar igual).

Mas existem outras situações que complicam o caso. Imaginemos o seguinte cenário: utilizamos os dispositivos de contagens de pessoas na entrada de todas as lojas de um bairro. Esses dispositivos registram apenas dados de entrada/saída de pessoas, horários e fluxo geral de público nas lojas. A partir dessas informações, obtemos um arquivo que será nossos dados de pesquisa. Seria questionável afirmar que esses dados de pesquisa têm autoria, já que não parece ser uma “criação do espírito”, apenas uma compilação feita por dispositivos eletrônicos automaticamente. Sem participação humana nem contribuição intelectual, é complicado utilizar o conceito de autoria, pelo que poderia significar que esses dados deveriam formar parte do domínio público imediatamente, como acontece no caso das sentenças judiciais ou das leis, por exemplo. Levando-se em conta esses tipos de pressupostos, talvez uma licença CC *Public Domain Mark*⁵⁷ seja a opção mais adequada, já que não pode se reivindicar *copyright* sobre o que não pode ser protegido.

Nos próximos anos, veremos como se resolvem dilemas como esse, e se acaba sendo um óbice para o desenvolvimento de compartilhamento de Dados Abertos ou não.

57 <https://creativecommons.org/share-your-work/public-domain/pdm/>

5. CONSIDERAÇÕES FINAIS

Os dados de pesquisa são uma realidade que se faz presente e é preciso possuir alguns conhecimentos mínimos sobre direitos autorais para saber o que fazer com eles. Contudo, o aumento desses dados na atualidade é incontestável, pois, à medida que cresce o número de artigos publicados, consequentemente cresce o número de dados produzidos e de dados compartilhados. Nesse sentido, faz-se necessário que os autores e editores, assim como a sociedade, compreendam a relevância da Ciência Aberta e dos Dados Abertos.

As licenças escolhidas para acompanhar aos dados de pesquisa são um assunto fundamental. Agora que alguns indexadores do âmbito latino-americano recomendam utilizar a cláusula CC NC (Não Comercial), é importante lembrar que essas licenças podem ser um empecilho para alguns projetos e podem impedir o compartilhamento dessas obras em páginas web que tenham um simples banner publicitário (ROSNAY, 2010), pois tais licenças não permitem obras derivadas e impedem que sejam realizados usos comerciais delas. O *Pantom Principles* para *Open Data* na Ciência, no seu ponto 3, explicita claramente que se o autor desejar que seus dados sejam efetivamente utilizados por terceiros, estes devem ser abertos, conforme a Ciência Aberta, sendo em particular para “uso não-comerciais e outras cláusulas restritivas não devem ser utilizadas” (MOLLOY, 2011, tradução nossa).

Se queremos que a ciência seja realmente aberta, é preciso que os dados de pesquisa possam ser o mais reutilizáveis possíveis e disponibilizá-los com uma licença como *CC by*.

Por último, é relevante lembrar que as licenças dos dados devem apresentar-se com a intenção de que os seres humanos, assim como as máquinas, possam ler o conteúdo das mesmas. Para isso, deveria tornar-se clara a licença dos dados no periódico, no texto do arquivo quando for possível, e disponibilizar também os metadados do arquivo de dados.

6. REFERÊNCIAS

BRASIL. Lei nº 9.610, de 19 de fevereiro de 1998. Altera, atualiza e consolida a legislação sobre direitos autorais e dá outras providências. **Diário Oficial da União**: seção 1, Brasília, DF, ano 136, n. 36, p. 3-9, 20 fev. 1998. Disponível em: http://www.planalto.gov.br/ccivil_03/LEIS/L9610.htm . Acesso em: 19 nov. 2019.

CROSS, D.; THOMSON, S.; SIBCLAIR, A. **Research in Brazil**: A report for CAPES by Clarivate Analytics. Filadélfia: Clarivate Analytics, 2017. Disponível em: <http://www.capes.gov.br/images/stories/download/diversos/17012018-CAPES-InCitesReport-Final.pdf>. Acesso em: 21 nov. 2019.

DUARTE, E. C. de V. G.; PEREIRA, E. C. **Direito autoral**: perguntas e respostas. Curitiba: UFPR, 2009.

MOLLOY, J.C. The Open Knowledge Foundation: open data means better science. **PloS Biology**, San Francisco, v. 9, n. 12, dec. 2011. Disponível em: <https://doi.org/10.1371/journal.pbio.1001195> . Acesso em: 26 nov. 2019.

MURIEL-TORRADO, E.; FERNÁNDEZ-MOLINA, J. C. Enseñanza digital versus derechos de autor: el papel de la biblioteca universitaria en apoyo de profesores y alumnos. **Encontros Bibli**: revista eletrônica de biblioteconomia e ciência da informação, Santa Catarina, v. 19, n. 39, p. 205-223, jan./abr. 2014. Disponível em: <https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2014v19n39p205/26584> . Acesso em: 20 nov. 2019.

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT. **OECD principles and guidelines for access to research data from public funding**. Paris: OECD, 2007. Disponível em: <http://www.oecd.org/sti/sci-tech/38500813.pdf> . Acesso em: 25 nov. 2019.

ROSNAY, M. D. de. **Creative Commons licenses legal pitfalls**: incompatibilities and solutions. Amsterdam: Institute for Information Law, 2010. Disponível em: <https://halshs.archives-ouvertes.fr/halshs-00671622/document>. Acesso em: 2 dez. 2019.

SAYÃO, L. F.; SALES, L. F. **Guia de gestão de dados de pesquisa para bibliotecários e pesquisadores**. Rio de Janeiro: CNEN/ IEN, 2015. Disponível em: http://www.cnen.gov.br/images/CIN/PDFs/GUIA_DE_DADOS_DE_PESQUISA.pdf. Acesso em: 19 nov. 2019.

SILVA, F. C. C. da. **Gestão de dados científicos**. Rio de Janeiro: Interciência, 2019. 146 p.

VICENTE-SAEZ, R.; MARTINEZ-FUENTES, C. Open science now: a systematic literature review for an integrated definition. **Journal of Business Research**, Amsterdam, v. 88, p. 428-436, jul. 2018. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0148296317305441#s0015>. Acesso em: 21 nov. 2019.

7. SOBRE OS AUTORES



Enrique Muriel-Torrado

Lattes: <http://lattes.cnpq.br/6493679032893430>

Orcid: <http://orcid.org/0000-0003-0969-9500>

Professor do Programa de Pós-Graduação em Ciência da Informação da UFSC é doutor em Informação Científica (Universidad de Granada), licenciado e graduado em Biblioteconomia e Documentação (Universidad de Extremadura). Com experiência em Vigilância Tecnológica se interessa nos atuais desafios sociais e tecnológicos, os direitos autorais e a desinformação.



Rebeca Fernandes de Moura

Lattes: <http://lattes.cnpq.br/7987637404192681>

Orcid: <https://orcid.org/0000-0002-3150-9032>

Servidora Técnico-Administrativo da Universidade Federal de Santa Catarina (UFSC). Mestre em Ciência da Informação pela Universidade Federal de Santa Catarina (UFSC) e doutoranda no Programa de Pós-Graduação em Ciência da Informação da UFSC. Graduada em administração e pesquisa e se interessa por temas relacionados ao direito autoral, plágio e ao acesso aberto à informação.

8. COMO CITAR

MURIEL-TORRADO, E.; MOURA, R. F. Direitos autorais e os dados de pesquisa abertos. In: SHINTAKU, M.; SALES, L. F.; COSTA, M. (org). **Tópicos sobre dados abertos para editores científicos**. Botucatu, SP: ABEC, 2020. p. 137-143. DOI: 10.21452/978-85-93910-04-3.cap11.

Ciência Aberta, dados abertos e princípios FAIR: uma contribuição dos Países Baixos

Patricia Henning
João Moreira

1. INTRODUÇÃO

No mundo interconectado dos dias de hoje, onde a informação e o conhecimento têm um papel fundamental no desenvolvimento da sociedade, a abertura das práticas científicas contribuem não apenas para expansão do acesso livre à informação e aos dados de pesquisa, mas também para consolidar um movimento denominado Ciência Aberta que busca, sobretudo, a democratização do conhecimento e o envolvimento da sociedade nas suas práticas, mudando o *modus operandi* da produção científica.

No âmbito da Ciência Aberta, a prática dos “Dados Abertos de Pesquisa” visa dar maior transparência à pesquisa com a finalidade de compartilhamento e reuso dos dados em novas investigações. Nesse sentido, a gestão desses dados assumiu novas configurações, elevando-se a um grau de importância tal, que se tornou tarefa indispensável da atividade científica.

Este relato tem o objetivo de trazer contribuições para a ciência brasileira acerca da experiência das práticas da Ciência Aberta, implantada nos Países Baixos. Inicia-se com um encontro recente e pontual, presidido pelo Conselho da União Europeia, em Amsterdam, em 2016, denominado *Open Science – From*

Vision to Action. O evento contou com a presença de diversos especialistas envolvidos com a temática da Ciência Aberta europeia, o que culminou na criação de um documento intitulado *The Amsterdam Call for Action on Open Science*⁵⁸. De forma muito direta e prática foram discutidas no encontro doze ações voltadas para a evolução da Ciência Aberta na Europa, relacionando os problemas, apontando soluções, sugerindo ações e indicando os efeitos positivos esperados. Esse documento foi um divisor de águas que elevou vários países da União Europeia a outro patamar junto às práticas da Ciência Aberta no mundo. Dos países presentes no encontro, os Países Baixos foram aquele que seguiu as orientações discutidas e, em menos de um ano, em 2017, lançou o seu *National Plan Open Science*⁵⁹, possibilitando a expansão e consolidação da Ciência Aberta no país.

Muito embora a experiência holandesa, no que diz respeito à Ciência Aberta, especialmente aos dados de pesquisa, esteja em um estágio mais avançado que a realidade brasileira, não podemos deixar de considerar que o acesso aberto à informação científica, assim como a abertura dos dados governamentais e de pesquisa, não é uma novidade no Brasil. A comunidade científica brasileira vem, há aproximadamente 15 anos, trabalhando em prol do acesso aberto à informação científica, estando o tema consolidado em várias universidades e institutos de pesquisa nacionais. Conforme consta no site do Instituto Brasileiro em Informação em Ciência e Tecnologia (IBICT), existem no Brasil por volta de cem repositórios de informação científica oficialmente cadastrados no Portal Brasileiro de Publicações Científicas em Acesso Aberto (OASISbr)⁶⁰.

No que diz respeito à abertura dos dados Sales *et al.*, (2019) afirmam que “o acesso à informação e a abertura dos dados governamentais foram fortalecidos após a criação da Lei de Acesso à Informação (LAI)⁶¹ nº 12.527, em 18 de novembro de 2011”. Além disso, o Brasil é um dos membros fundadores da Parceria para o Governo Aberto/*Open Government Partnership* (OGP)⁶² desde 20 de setembro de 2011, data da sua criação. A OGP tem o objetivo de difundir

58 <https://www.government.nl/documents/reports/2016/04/04/amsterdam-call-for-action-on-open-science>.

59 https://www.openscience.nl/files/openscience/2019-02/nationalplanopenscience_en.pdf

60 <http://oasisbr.ibict.br/vufind/>

61 http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2011/lei/l12527.htm

62 <https://www.opengovpartnership.org/>

e incentivar as práticas governamentais a nível internacional referentes ao acesso à informação pública, à transparência dos governos e à participação social. Nesse mesmo ano foi criado o Plano de Ação Nacional, em que foram descritos os compromissos, as estratégias e as atividades em Governo Aberto. Em 2012, foi criada a Infraestrutura Nacional de Dados Abertos (INDA)⁶³ e, em 2016, o “Manifesto de Acesso Aberto a Dados da Pesquisa Brasileira para Ciência Cidadã”⁶⁴ foi lançado pelo IBICT, seguindo as tendências mundiais de abertura de dados de pesquisa.

Atualmente há no Brasil uma mobilização de instituições do governo, da sociedade civil e da academia voltada para a condução da Política de Dados Abertos do Poder Executivo Federal, que está sob gestão da Controladoria-Geral da União (CGU)⁶⁵. Não podemos deixar de ressaltar as atividades do 4º. Plano de Ação Nacional em Governo Aberto 2018-2020⁶⁶, que está em andamento e prevê em seus compromissos (marcos de ação)⁶⁷ que estabelecem: a implantação de uma rede interinstitucional; realização de diagnósticos nacionais e internacionais; definição de diretrizes e princípios para as políticas institucionais; promoção e ações de sensibilização; participação e capacitação; articulação com as agências de fomento e editores científicos; proposição de infraestrutura; proposição de padrões de interoperabilidade para repositórios de dados e de conjunto de indicadores para aferição de maturidade em Ciência Aberta.

É indiscutível o enorme esforço da comunidade científica brasileira em prol da Ciência Aberta, sendo essa a razão pela qual buscamos trazer algumas contribuições, inspiradas na experiência holandesa, para agregar às nossas iniciativas. Nesse sentido, apresentamos um resumo do *National Plan Open Science* holandês, com as suas ambições e seus desdobramentos; mencionamos os princípios FAIR de Wilkinson *et al.* (2016) e as principais iniciativas europeias voltadas para a implantação desses princípios que contam com a presença dos Países Baixos. Por fim, descrevemos o engajamento do pesquisador na gestão

63 <http://wiki.dados.gov.br/>.

64 <http://www.ibict.br/sala-de-imprensa/noticias/item/478-ibict-lanca-manifesto-de-acesso-aberto-a-dados-da-pesquisa-brasileira-para-ciencia-cidada>

65 <http://wiki.dados.gov.br/Politica-de-Dados-Abertos.ashx>

66 https://www.opengovpartnership.org/wp-content/uploads/2018/10/Brazil_Action-Plan_2018-2020_POR.pdf

67 <https://wiki.mrp.br/plugins/servlet/mobile#content/view/107315238>

de dados e como as instituições de ensino e pesquisa e agências de fomento holandesas estão conduzindo as práticas da gestão dos dados alinhadas aos princípios FAIR.

2. O PLANO NACIONAL DE CIÊNCIA ABERTA DOS PAÍSES BAIXOS

O *National Plan Open Science* é o plano nacional de Ciência Aberta dos Países Baixos, lançado em fevereiro de 2017, como o documento orientador dos investimentos e das práticas da Ciência Aberta no país. Esse plano reforça a necessidade de ser ambicioso e de investimento em recursos humanos para a sua consolidação. Apresenta o plano holandês para a Ciência Aberta no período de (2017-2020), envolvendo diferentes atores como pesquisadores e universidades, bibliotecas de ensino e pesquisa, instituições governamentais e agências de financiamento, bem como organizações de apoio e plataformas científicas. Esse plano foi estruturado em três metas de atuação:

- a. **Acesso aberto:** tem como principal ambição alcançar 100% do acesso aberto às publicações científicas, em todas as áreas do conhecimento, até 2020. Pretende também focar no desenvolvimento e implementação de políticas; ampliar o compartilhamento do conhecimento no âmbito da União Europeia e em outros os países envolvidos com o acesso aberto; estimular a mudança de cultura e comportamento no âmbito do acesso aberto; incentivar a transparência como condição para todos os contratos de licenciamento e incentivar a reutilização das publicações sem custos adicionais (MINISTRY OF EDUCATION, CULTURE AND SCIENCE, 2017, p. 21).
- b. **Dados abertos:** a promoção do uso e reutilização dos dados de pesquisa tem como objetivo permitir o acesso aos dados de pesquisa alinhados aos princípios FAIR⁶⁸ como pré-condição técnica e política. Tem ainda a ambição de investir na interoperabilidade por meio de padrões; estabelecer regras e acordos, fornecer documentação para fins de verificação e, ainda, estabelecer diretrizes relativas à privacidade, direitos de propriedade, direitos de reutilização e durabilidade do compartilhamento. Quanto ao armazenamento, a iniciativa visa orientar quais dados precisam ser armazenados por um longo período de tempo, levando-se em consideração o custo e as

68 <https://www.nature.com/articles/sdata201618>

tecnologias necessárias para garantir a reutilização dos mesmos. Para isso, é necessário definir as características dos repositórios usando as certificações necessárias para a sua confiabilidade. Por fim, incentivar a adoção de software e ferramentas de código aberto para assegurar o seu uso no futuro (MINISTRY OF EDUCATION, CULTURE AND SCIENCE, 2017, p. 23).

- c. **Sistemas de recompensa:** visa alinhar os sistemas de avaliação e recompensa às tendências da Ciência Aberta. Para alcançar esse objetivo, pretende-se examinar o Protocolo de Avaliação de Pesquisa nos Países Baixos⁶⁹ e propor ajustes de transição para a Ciência Aberta. Por fim, repensar como é possível utilizar os indicadores alternativos da ciência, como o Altmetrics⁷⁰, por exemplo, na avaliação da pesquisa e dos pesquisadores (MINISTRY OF EDUCATION, CULTURE AND SCIENCE, 2017, p. 25).

2.1 O que veio depois?

O *National Plan Open Science*⁷¹ culminou na Plataforma Nacional de Ciência Aberta dos Países Baixos, que contempla, entre várias ações, o **Open Access.nl**⁷² voltado para o acesso aberto à informação científica. Este é um serviço de responsabilidade de um grupo de quatro instituições holandesas que tem a atribuição de atualizar a comunidade de pesquisa nacional e internacional com informações relevantes sobre o crescimento e as novas tendências do Acesso Aberto. Esse *site* divulga o trabalho de monitoramento do Acesso Aberto nos Países Baixos, constatando que, em 2018, 54% das publicações desse país foram em acesso aberto. Oferece também uma lista de aproximadamente 9.000 títulos de periódicos em acesso aberto para facilitar a escolha dos pesquisadores. Esclarece ainda o papel dos atores envolvidos no processo, desde o pesquisador/autor, passando pelo professor, estudante, gestor, editor, além de destacar a importância dos repositórios, das agências financiadoras e das questões de *copyright* dentro desse cenário.

69 <http://vsnu.nl/files/documenten/Domeinen/Onderzoek/SEP2015-2021.pdf>

70 Altmetrics: métricas alternativas às tradicionais

71 <https://www.openscience.nl/en>

72 <https://www.openaccess.nl/>

No que diz respeito aos dados de pesquisa a plataforma remete para o [site **The National Coordination Point Research Data Management \(LCRDM\)**](https://www.lcrdm.nl/en)⁷³ que é a rede holandesa de gestão de dados formada por representantes de institutos de pesquisa, universidades e agências de financiamento, incluindo também formuladores de políticas e especialistas em tecnologia. Essa rede tem o intuito de colocar em prática políticas de gestão de dados e incorporá-las na educação e na pesquisa das universidades e institutos de pesquisa nos Países Baixos. O *site* é rico contendo diversas diretrizes nacionais e internacionais sobre gestão de dados, privacidade, infraestrutura, custos e *data stewardship*⁷⁴. Além disso, divulga as ações e resultados dos grupos de trabalho, apresenta um glossário, um blog, uma Wiki, uma lista de discussão, cursos, eventos, um cadastro de especialistas e uma lista das instituições parceiras.

Essa plataforma, que é suportada pela organização holandesa de tecnologia de informação para pesquisa e educação, remete também para a iniciativa **SURF**⁷⁵, que dá vários esclarecimentos sobre as questões relacionadas à *copyright* desde o *Dutch Copyright Act*, direitos de autor, uso e duração do *copyright*, aponta as exceções e as diferenças entre os países com relação à temática.

No tocante à terceira meta do plano referente ao sistema de avaliação e recompensa, a principal ação é a do **Drive change in recognition and reward of academics**⁷⁶, de responsabilidade da principal agência de fomento dos Países Baixos⁷⁷, a *Dutch Research Council* (NWO). Por outro lado, a Associação das Universidades dos Países Baixos, juntamente com algumas agências de fomento, se juntou para propor novas trajetórias de carreira e renovação das modalidades de avaliação de pesquisa e recompensa da ciência. Segundo o Dutch Research Council (2019), essas mudanças são “fundamentais para o desenvolvimento de talentos, para o avanço da qualidade e o impacto do ensino e da pesquisa, bem como a promoção de instituições de conhecimento e empregadores”.

73 <https://www.lcrdm.nl/en>

74 Data stewardship: significa administrador de dados

75 <https://www.auteursrechten.nl/en>

76 <https://www.nwo.nl/en/news-and-events/news/2018/11/drive-change-in-recognition-and-reward-of-academics.html>

77 https://www.vsnu.nl/en_GB/

3. OS PRINCÍPIOS FAIR E AS INICIATIVAS DE GESTÃO DE DADOS DE PESQUISA

FAIR é um acrônimo para *Findable, Accessible, Interoperable e Reusable*, que nasceu fruto de uma reunião ocorrida, em janeiro de 2014, em Leiden, Países Baixos, denominada “*Jointly designing a data FAIRPORT*”⁷⁸. Esses princípios são considerados o instrumento norteador para descoberta, acesso, compartilhamento e reutilização dos dados de pesquisa, assim como promove a interoperabilidade global. Os princípios propõem estabelecer orientações para a publicação de recursos digitais, conjunto de dados, códigos e objetos de pesquisa de maneira FAIR (WILKINSON *et al.*, 2018). No entanto, só passaram a ser mundialmente considerados como pilares de boas práticas de gestão de dados após a sua publicação no periódico *Nature - Scientific Data* (2016), quando a Comissão Europeia passou a exigir plano de gestão de dados alinhados ao FAIR para quaisquer projetos financiados no âmbito do H2020. Desde então, a adesão a esses princípios passou a ser solicitado por diversas agências de financiamento da Europa e dos Estados Unidos, tornando-se um caminho sem volta.

Embora esses princípios não sejam uma novidade para a comunidade científica, Henning *et al.*, (2019) lembra que “[...] colocá-los em prática ainda é uma tarefa árdua em função do seu alto grau de subjetividade e complexidade” “[...] podendo gerar problemas devido às diferentes interpretações à eles associadas”. Diversas iniciativas e ferramentas vêm sendo criadas para amenizar tais dificuldades, bem como um ecossistema de dados FAIR foi definido por Hodson *et al.*, (2018) como uma proposta de estruturação e adoção dos princípios FAIR na sua totalidade.

Os Países Baixos têm uma presença significativa junto às iniciativas internacionais voltadas para a gestão e disseminação dos dados de pesquisa FAIR, o que os fortalece nas práticas da Ciência Aberta. Algumas dessas iniciativas são de sua própria criação e liderança, enquanto em outras participam como membros. Uma relação de vinte e uma iniciativas voltadas para a promoção, disseminação, educação, treinamento dos produtos e serviços FAIR, no âmbito da gestão de dados, que conta com a atuante participação dos Países Baixos, consta no APÊNDICE.

78 <https://www.lorenzcenter.nl/lc/web/2014/602/info.php?wsid=602>

Sobre os repositórios de dados, a instituição **Data Archiving and Networked Services (DANS)** disponibiliza a Plataforma NARCIS⁷⁹, a qual reúne quinze diferentes repositórios de dados, entre os quais está o *SURF Repository*⁸⁰. O NARCIS contém mais de 235 mil *data sets*. O Repositório *4TU Centre for Research Data*⁸¹, que existe desde 2010, composto por um consórcio de quatro universidades técnicas holandesas (4TU), é um dos mais importantes do país. Desde 2018 o 4TU vem experimentando funcionalidades com o objetivo de, em 2020, oferecer benefícios como *Application Programming Interface (API)* para acesso automatizado por máquinas, diferentes níveis de acesso e código aberto via *GitHub*. Essas funcionalidades “visam o arquivamento de *software*, estatísticas de uso, espaço restrito para compartilhamento privado dos dados antes da sua publicação e customização de metadados” (DUNNING, p. 4 2019). Além disso, o Repositório de dados possui o selo de qualidade *Data Seal of Approval*⁸², que garante a sua qualidade e integridade, fazendo dele um repositório confiável. Oferece ainda aos seus pesquisadores 1 TB de espaço gratuito por ano. Para os pesquisadores de fora do consórcio, oferece 10GB de espaço gratuito e cobra-se 4,50 euros por cada GB adicional. Entre as suas missões para 2020–2023, chama a atenção a criação de redes específicas disciplinares para estimular a adoção e reuso de dados de pesquisa FAIR.

3.1 O engajamento dos pesquisadores dos Países Baixos na gestão de dados FAIR

Partindo do pressuposto que o pesquisador é o responsável pelos seus dados e que as pesquisas baseadas em dados requerem uma cuidadosa gestão durante todo o ciclo de vida deles, o trabalho de gerenciamento passou a ser de tal modo valorizado que, em pouco tempo, se tornou pré-requisito das práticas de pesquisa. Por essa razão, a profissão de administrador de dados, em inglês *Data Stewardship*, surgiu, nos últimos anos, requerendo habilidades específicas, podendo ser assumidas por diferentes profissionais, tais como bibliotecário, arquivista, administrador, profissional de TI, entre outros. Essa nova função vem crescendo no mundo científico por conta do aumento do nível de exigência

79 <https://www.narcis.nl/search/coll/dataset/Language/en>

80 <https://repository.surfsara.nl/node>

81 <https://researchdata.4tu.nl/en/>

82 <https://www.coretrustseal.org/about/history/data-seal-of-approval/>

das agências de financiamento e também pela complexidade cada vez maior dos dados, requisitando competências e responsabilidades que o pesquisador desconhece e não consegue abraçar.

Cada vez mais o pesquisador necessita de ajuda para realizar a gestão correta dos seus dados. Apesar da função de administrador de dados ainda não ser bem definida, existe um consenso geral da necessidade de treinamento básico como pré-requisito para se trabalhar nessa profissão. O *Research Data Netherlands (rdnl)*⁸³, por exemplo, oferece um curso *on-line* com uma introdução sobre a temática, indo até as questões relativas à código de conduta e integridade da pesquisa do país. O *Netherlands Federation of Medical Centre*, responsável pela iniciativa *Data4Lifescience*, entende que uma política de administração de dados é fundamental para o seu compartilhamento e reutilização. Por essa razão, criou um manual *on-line*, denominado *Handbook for Adequate Natural Data Stewardship (HANDS)*⁸⁴ para ajudar os pesquisadores e os administradores de dados da área da saúde.

Todas as agências de fomento à pesquisa dos Países Baixos, sem exceção, solicitam dos seus candidatos a financiamento um plano de gestão de dados alinhados aos princípios FAIR. Essa exigência passou a ser um problema para o pesquisador que, em alguns casos, não sabe nem por onde começar. Para amenizar tais questões as universidades holandesas oferecem treinamentos periódicos presenciais e *on-line* para preenchimento de planos de gestão, além de disponibilizarem em seus sites informações adicionais sobre a temática. Três bons exemplos desse trabalho estão na *Delft University of Technology*⁸⁵, na *Utrecht University*⁸⁶ e na *University of Amsterdam*⁸⁷, que indicam ferramentas e soluções para que a gestão de dados de pesquisa aconteça de forma correta e alinhada aos princípios FAIR.

A *Delft University of Technology*, por exemplo, tem um projeto de *Data Stewardship*⁸⁸ voltado para ajudar os pesquisadores na tarefa de gerenciar, arquivar

83 <https://researchdata.nl/en/>

84 <https://data4lifesciences.nl/hands2/data-stewardship/>

85 <https://www.tudelft.nl/en/library/current-topics/research-data-management/>

86 <https://www.uu.nl/en/research/research-data-management>

87 <https://rdm.uva.nl/en>

88 <https://www.tudelft.nl/en/library/current-topics/research-data-management/r/support/data-stewardship/>

e compartilhar os seus dados de pesquisa. Para isso, contratou administradores de dados especialistas disciplinares para cada uma das suas oito faculdades, com dedicação exclusiva. Esses profissionais têm a função de orientar os pesquisadores desde o início das suas pesquisas, envolvendo-os nas boas práticas de gestão de dados até o momento do compartilhamento.

Outro exemplo é o centro de competência de dados da *Wageningen University & Research (WUR)*⁸⁹, reconhecida universidade rural holandesa e referência na pesquisa em interoperabilidade no domínio de agricultura e pecuária. Esse tipo de centro de competências para administração de dados é uma tendência que vem sendo adotada por diversas instituições de pesquisa, tendo algumas iniciativas correntes na *Leiden University*, e sendo suportada pela recente rede de implementação do GO FAIR, denominada de *Data Stewardship Competence Centers (DSCC-IN)*⁹⁰, que visa alinhar internacionalmente a criação desses centros.

4. CONSIDERAÇÕES FINAIS

Ao refletir sobre o que foi narrado, percebe-se que os Países Baixos estão bem estruturados nas práticas da Ciência Aberta. Com relação à gestão de dados, pode-se afirmar que a Holanda é um dos países mais avançados do mundo, com o maior investimento na implantação dos princípios FAIR nas suas práticas de pesquisa. No entanto, há que se considerar dois fatores que contribuem para esse sucesso. Em primeiro lugar, sendo o país uma **monarquia constitucional** parlamentar democrática, com valores políticos e sociais bastantes liberais, não poderia deixar de aderir ao movimento de abertura da ciência. Em segundo lugar, esse é um país rico, com um grande investimento em pesquisa, estimulado pelas diversas possibilidades de financiamento, tanto das suas próprias agências de fomento quanto da Comissão Europeia.

Muito embora a realidade da Ciência Aberta brasileira ainda esteja distante da realidade holandesa, percebe-se - diante das iniciativas, ações e intenções brasileiras, relatadas na introdução - que estamos no caminho certo. Além disso, a pesquisa brasileira pode evitar erros que a pesquisa europeia cometeu e dar

89 <https://www.wur.nl/en/Value-Creation-Cooperation/WDCC.htm>

90 <https://www.rd-alliance.org/group/rda-netherlands/post/fwd-go-fair-implementation-network-data-stewardship-competence-centers>

um salto rumo à Ciência Aberta, entendendo as lições aprendidas e reutilizando resultados bem-sucedidos.

A partir do que foi narrado neste texto, podemos visualizar algumas possíveis contribuições concretas para o Brasil, baseadas na experiência holandesa. **Primeiro**, a elaboração de um Plano Nacional de Ciência Aberta com ações e prazos bem definidos, detalhando todos os investimentos financeiros e de recursos humanos necessários. **Segundo**, a criação de uma Plataforma de Ciência Aberta com desdobramentos para cada uma das suas práticas envolvendo as instituições responsáveis e interessadas em cada temática, como tem sido praticado nos Países Baixos. **Terceiro**, investir na criação e na qualidade dos repositórios de dados para que sejam confiáveis, assim como em cursos de capacitação sobre Ciência Aberta de modo a estabelecer cultura sobre a temática e dar maior visibilidade ao tema de acesso aberto e gestão de dados. **Quarto**, incentivar as universidades a criarem as suas políticas de abertura dos dados de pesquisa e passar a cobrar um Plano de Gestão de Dados mínimo em todos os projetos de pesquisa da instituição. **Quinto**, tentar se filiar às possíveis iniciativas listadas no Apêndice visando aprender com elas ou tentar acompanhar o trabalho que elas vêm desenvolvendo a nível internacional. **Por fim**, preparar profissionais (administradores de dados) para dar apoio aos pesquisadores na gestão e compartilhamento, assim como orientá-los quanto à publicação dos seus artigos em acesso aberto.

Algumas das ações ora mencionadas já existem no Brasil e outras estão em fase de negociação e desenvolvimento. Esperamos que as informações aqui apresentadas possam, de alguma maneira, contribuir para a melhoria das práticas abertas da ciência no Brasil, proporcionando maior transparência, abrindo maiores possibilidades para o compartilhamento e a reutilização dos dados e da informação em novas pesquisas, como vem acontecendo no cenário internacional.

5. REFERÊNCIAS

DUNNING, A. *et al.* 4TU.ResearchData Strategy 2020-2023. **4TU.ResearchData**, 2019. Disponível em: https://researchdata.4tu.nl/fileadmin/user_upload/Documenten/4TU.ResearchData_Strategy_2020-2023.pdf . Acesso em: 25 nov. 2019.

DUTCH RESEACH COUNCIL (NOW). **Drive change in recognition and reward of academics**, 2019. Disponível em: <https://www.nwo.nl/en/news-and-events/news/2018/11/drive-change-in-recognition-and-reward-of-academics.html> Acesso em: 25 nov. 2019.

HENNING, P. *et al.* Desmistificando os princípios FAIR: conceitos, métricas, tecnologias e aplicações inseridas no ecossistema dos dados FAIR. **Pesquisa brasileira em ciência da informação e biblioteconomia**, v. 14, n. 4, p. 175-192, 2019. Disponível em: <https://periodicos.ufpb.br/index.php/pbcib/article/view/49701>. Acesso em: 25 nov. 2019

HODSON, S. *et al.* **Turning FAIR data into reality**. Interim report of the European Commission Expert Group on FAIR data. 2018. Disponível em: <https://doi.org/10.5281/zenodo.1285272>. Acesso em: 25 nov. 2019.

LCRDM. **Data stewardship on the map**: A study of tasks and roles in Dutch research institutes. Disponível em: https://www.lcrdm.nl/files/lcrdm/201905/LCRDM%20rapport%20datastewardship_EN_online.pdf. Acesso em: 25 nov. 2019.

MINISTRY OF EDUCATION, CULTURE AND SCIENCE. **National Plan Open Science**, 2017. Disponível em: https://www.openscience.nl/files/openscience/2019-02/nationalplanopenscience_en.pdf. Acesso em: 25 nov. 2019.

SALES, L.*et al.* GO FAIR Brazil: A Challenge for Brazilian Data Science. **Data Intelligence**, p. 238-245, 2019. Disponível em: https://www.mitpressjournals.org/doi/abs/10.1162/dint_a_00046. Acesso em: 25 nov. 2019.

WILKINSON, M. D. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. **Scientific Data**, v. 3, 2016. Disponível em: <https://www.nature.com/articles/sdata201618>. Acesso em: 25 nov 2019.

WILKINSON, M. D. *et al.* A design framework and exemplar metrics for *FAIRness*. **Scientific Data**, v. 5, n. 180118, 2018. Disponível em: <https://www.nature.com/articles/sdata2018118>. Acesso em: 25 nov. 2019.

6. APÊNDICE - INSTITUIÇÕES E INICIATIVAS QUE ATUAM COM GDP/FAIR COM A PARTICIPAÇÃO DOS PAÍSES BAIXOS

	NOME	ENDEREÇO DO SITE
1	Science Europe	https://www.scienceeurope.org/
2	European Research Council	https://erc.europa.eu/
3	National Coordination Point Research Data Management-LCRDM	https://www.edugroepen.nl/sites/RDM_platform/SitePages/Home.aspx
4	Data Archiving and Networked Services - DANS	https://dans.knaw.nl/en/about
5	FORCE 11	https://www.force11.org/
6	GO FAIR	https://www.go-fair.org/
7	FAIRHelth	https://www.fair4health.eu/
8	FAIRPlus	https://fairplus-project.eu/
9	Dutch Techcentre for Life Science - DTL	https://www.dtls.nl/
10	FAIRSharing.org	https://fairsharing.org/
11	Committee on Data-CODATA	http://www.codata.org/
12	Research Data Alliance - RDA Netherlands	https://www.rd-alliance.org/groups/rda-netherlands
13	ELIXir	https://elixir-europe.org/
14	OpenAIRE	
15	Consortium of European Social Science Data Archives - CESSDA	https://www.cessda.eu/
16	Digital Preservation Coalition- DPC	https://dpconline.org/
17	FAIRsFAIR	https://www.fairsfair.eu/
18	SURF Foundation	https://www.surf.nl/en
19	LIBER - Europa	https://libereurope.eu/
20	Data4Lifesciences	https://data4lifesciences.nl/hands2/data-stewardship/
21	FAIRDON	https://fair-dom.org/

7. SOBRE OS AUTORES



Patricia Henning

Lattes: <http://lattes.cnpq.br/0970010723997242>

Orcid: <https://orcid.org/0000-0003-0739-6442>

Mestre em Ciência da Informação pelo IBICT, Doutora em Ciência pelo ICICT/Fiocruz e Pós-Doutora pela University of Twente, na Holanda. Professora aposentada do curso de Graduação em Biblioteconomia da Unirio. Atualmente é professora do Programa de Mestrado em Biblioteconomia da Unirio e professora visitante do Programa de Pós-Graduação em Saúde e Tecnologia no Espaço Hospitalar (PPGSTEH) da Unirio.



João Luiz Rebelo Moreira

Lattes: <http://lattes.cnpq.br/3321809257390602>

Orcid: <https://orcid.org/0000-0002-4547-7000>

Dr. Moreira trabalha na academia desde 2008, bacharel em ciência da comunicação e mestre em informática pela UFRJ (DCC/PPGI) com especialização em inteligência de negócios. Sua pesquisa de doutorado na Universidade de Twente (Holanda) foi em interoperabilidade semântica, com ênfase na integração de dados de serviços inteligentes de emergência suportados por tecnologias de Internet-das-Coisas. Dr. Moreira obteve seu pós-doutorado pela Universidade VU Amsterdam com pesquisa em engenharia de ontologias para workflows científicos com princípios FAIR. Dr Moreira é atualmente professor assistente na Universidade de Twente e representante na Europa da rede de implementação GO-FAIR Brasil

8. COMO CITAR

HENNING, P.; MOREIRA, J. Ciência aberta, dados abertos e princípios FAIR: uma contribuição dos Países Baixos. *In*: SHINTAKU, M.; SALES, L. F; COSTA, M. (org). **Tópicos sobre dados abertos para editores científicos**. Botucatu, SP: ABEC, 2020. p. 145-158. DOI: 10.21452/ 978-85-93910-04-3.cap12.

As potencialidades do data paper na ciência atual

Lucas Rocha

A publicação de artigos científicos é tradicionalmente utilizada para disseminar práticas, metodologias, análises e conclusões que auxiliam no avanço das ciências. Nesse sentido, os dados de pesquisa levantados durante a investigação são importantes, já que, por meio deles, é possível analisar as conclusões a que os pesquisadores chegaram em seus estudos.

No entanto, os dados de pesquisa possuem potencial maior do que servir apenas à análise e conclusões dos artigos científicos aos quais estão vinculados. Há, nos dados, uma grande possibilidade de reúso: dados coletados por um grupo de pesquisa podem ser utilizados por outros, em diferentes contextos de aplicação, para solucionar problemas das mais diferentes naturezas. Essa reutilização, que economiza tempo e dinheiro, pode fazer os avanços científicos acontecerem de maneira mais eficaz.

Independentemente dos seus campos de atuação, pesquisadores produzem dados de pesquisa que, segundo Chen (2017), geralmente acabam restritos às paredes de seus próprios laboratórios, acondicionados precariamente – no que diz respeito às práticas de documentação – em arquivos obscuros que podem se perder com facilidade. Os motivos para essa restrição são os mais diversos possíveis: falta de padrões de metadados que sejam adequados para o gerenciamento dos dados de pesquisa, falta de tradição para compartilhar dados de pesquisa (CHEN, 2017), preço para manutenção e atualização dos conjuntos de dados (CHEN, 2017; HUANG; HAWKINS; QIAO, 2013) e a falta de definição de um tamanho apropriado para os conjuntos de dados (HUANG; HAWKINS; QIAO, 2013) são alguns dos fatores que dificultam seu acondicionamento e compartilhamento.

Para que os dados de pesquisa possam ser reaproveitados e reutilizados, é necessário encontrar mecanismos de documentação apropriados. Documentar dados é mais do que apenas disponibilizá-los aos pesquisadores; é descrevê-los de maneira inteligente, para que sua reutilização seja possível.

Para serem reutilizados, é preciso que esses dados possam ser interpretados dentro de seus contextos de utilização e, para isso, necessitam estar bem descritos e apropriadamente armazenados a fim de serem facilmente encontrados. De acordo com Kratz e Strasser (2014), os dados de pesquisa devem ser passíveis de reprodução e uso, com perspectivas de disponibilidade tanto atual quanto futura. A disponibilidade atual diz respeito aos mecanismos de acesso necessários para que os dados possam ser encontrados, e a disponibilidade futura diz respeito às perspectivas adotadas para sua preservação, como opções de armazenamento e migração de formatos.

Como alternativa para tornar a regra do silêncio uma exceção, uma nova modalidade de publicação aparece no cenário das ciências: os *data papers*, ou artigos de dados, uma “publicação em uma revista cujo propósito é descrever dados em vez de informar sobre uma pesquisa ou suas conclusões. Sendo assim, conteria os dados sem as hipóteses ou os argumentos, os resultados e a discussão que sustentaram” (GARCÍA-GARCÍA; LÓPEZ-BORRULL; PESET, 2015, p. 847). Essa nova modalidade de publicação, que utiliza os *data journals* como periódicos para publicação da descrição consistente, visa desengavetar os dados de pesquisa e torná-los visíveis à comunidade acadêmica, tendo como motivador principal as potencialidades que o reúso de dados pode trazer para o desenvolvimento científico, ao mesmo tempo que oferecem aos pesquisadores a possibilidade de citação de uma maneira tão consistente quanto a de um artigo de pesquisa tradicional.

Diferentemente das publicações tradicionais, não é possível avaliar o conteúdo dos dados de pesquisa, uma vez que sua utilização pode acontecer em diversos contextos. No entanto, o trabalho dos *data journals* é verificar a consistência dos dados, isto é, se estão bem descritos e se o potencial de reúso é ou não relevante para a comunidade científica.

Essa nova alternativa de publicação traz consigo outra vantagem: a possibilidade do aumento de citações. É importante para pesquisadores e grupos de pesquisa que seus trabalhos tenham relevância na comunidade acadêmica, e uma maneira de aumentar o impacto de suas pesquisas – que, muitas vezes, demoram anos para serem concluídas justamente por conta do longo tempo

necessário para coleta e análise de dados – é mediante a citação de dados. Esse potencial pode trazer mais visibilidade às pesquisas que ainda estejam em andamento e fazer com que os trabalhos se tornem mais conhecidos ao longo de sua produção.

Nesse sentido, o editor científico é imperativo. É necessário que essa figura importante no trajeto da publicação científica incentive a visibilidade dos dados dentro da comunidade acadêmica. É um processo de ganho tanto para os periódicos, que teriam mais publicações, mais citações e mais possibilidades de acesso, quanto para os pesquisadores, tanto os que utilizam quanto os que disponibilizam seus dados de pesquisa. Os primeiros, por terem a possibilidade de tornar suas pesquisas mais otimizadas, uma vez que não haveria necessidade de todo o processo de coleta de dados; e os últimos, por aumentar o potencial de citação e publicização de seus trabalhos.

Algumas barreiras para que a prática de publicação de dados se torne uma regra e não uma exceção ainda precisam ser superadas. A produção dos dados de pesquisa é um movimento natural de grande parte dos pesquisadores. No entanto, o mesmo esforço de geração dos dados não acontece no momento de sua curadoria. Mesmo com todos os benefícios apontados na publicação e compartilhamento de dados de pesquisa – aumento de reconhecimento dos pesquisadores, geração de maior notoriedade para consultorias e oportunidades de colaboração, acréscimo nos índices de citação por suas produções terem maior visibilidade (COSTELLO, 2009; AKERS, 2013; KRATZ; STRASSER, 2014) e possibilidade de reúso dos dados (GARCÍA-GARCÍA; LÓPEZ-BORRULL; PESET, 2015) –, poucos são os pesquisadores que tornam seus dados públicos para reutilização (CHEN, 2017; HUANG; HAWKINS; QIAO, 2013).

De acordo com Capurro e Hjørland (2007), a natureza digital da informação é o que a torna especialmente significativa. Desde que a internet deixou o cenário de guerra e passou a fazer parte do cotidiano da população, a discussão de suas potencialidades são parte essencial das ciências, sobretudo em um contexto onde as fronteiras se dissolvem e a comunicação entre diferentes partes do mundo pode ser feita com a rapidez de um clique.

E, no contexto da Ciência Aberta, que tem por objetivo tornar mais eficaz a relação entre recursos sociais e econômicos para promover uma ciência mais transparente, é importante que práticas como a publicação de dados de pesquisa – visando sua reutilização – sejam disseminadas para que diferentes pesquisadores possam utilizar e reutilizar dados de pesquisa de maneira consistente

e eficaz, extinguindo a necessidade de coleta de dados para toda nova pesquisa que se inicie. Incentivar que os pesquisadores se dissociem da ideia de que os dados são propriedade privada também é uma maneira importante de atuação, assim como promover um ambiente propício para que os dados possam ser corretamente descritos e depositados para futuras consultas.

1. REFERÊNCIAS

AKERS, K. Looking out for the little guy: Small data curation. **Bulletin of the American Society for Information Science and Technology**, [S. l.], v. 39, 2013.

CAPURRO, R.; HJORLAND, B. O conceito de informação. **Perspect. ciênc. inf.**, Belo Horizonte, v. 12, n. 1, p. 148-207, abr. 2007.

CHEN, Y. An analysis of characteristics and structures embedded in data papers: a preliminary study. **Libellarium: journal for the research of writing, books, and cultural heritage institutions**, [S. l.], v. 9, n. 2, mar. 2017.

COSTELLO, M. J. Motivating Online Publication of Data. **BioScience**, [S. l.], v. 59, p. 418–427, 2009.

GARCÍA-GARCÍA, A.; LÓPEZ-BORRULL, A.; PESET, F. Data journals: eclosión de nuevas revistas especializadas en datos. **El Profesional de la Informacion**, [S. l.], v. 24, p. 845-848, 2015.

HUANG, X.; HAWKINS, B.; QIAO, G. Biodiversity Data Sharing: Will Peer-Reviewed Data Papers Work?. **BioScience**, [S. l.], v. 63, p. 5-6, 2013.

KRATZ, J.; STRASSER, C. Data publication consensus and controversies. **F1000Research**, [S. l.], v. 10, 2014.

2. SOBRE OS AUTORES



Lucas Rocha

Lattes: <http://lattes.cnpq.br/4807740934260257>

Orcid: <https://orcid.org/0000-0002-0373-4881>

Biografia: Lucas de Lima Rocha é formado em Biblioteconomia e Documentação pela Universidade Federal Fluminense (2014) e possui mestrado em Ciência da Informação (2016) pela mesma Universidade. Atualmente, trabalha como bibliotecário em um Centro Unificado Educacional (CEU), na cidade de São Paulo.

3. COMO CITAR

ROCHA, L. As potencialidades do data paper na ciência atual. *In*: SHINTAKU, M.; SALES, L. F; COSTA, M. (org). **Tópicos sobre dados abertos para editores científicos**. Botucatu, SP: ABEC, 2020. p. 159-163. DOI: 10.21452/978-85-93910-04-3.cap13.

Gestão de dados em periódicos científicos

Caterina Groposo Pavão
Fabiano Couto Corrêa da Silva
Lúcia da Silveira

1. INTRODUÇÃO

Os dados ocupam um papel fundamental na Sociedade da Informação. No entanto, embora a quantidade de dados publicados continue a crescer e termos como *dilúvio de dados* e *big data* hoje caracterizem inúmeras iniciativas de pesquisa, ainda é necessário muito trabalho para possibilitar que sua publicação seja efetivamente recuperável, disponível e reutilizável pelos pesquisadores. Várias barreiras impedem a publicação, desde a falta de atribuição e recompensas, prática de citação do conjunto de dados científicos, problemas de qualidade dos dados, até uma ausência generalizada de cultura de compartilhamento de dados nas diferentes áreas do conhecimento. Tendo em vista que a comunicação científica está mudando e as revistas fazem parte desse processo, os periódicos de dados e os repositórios surgiram como soluções para superar algumas dessas dificuldades.

O presente capítulo indica soluções apontadas na literatura como práticas adotadas em revistas quanto a adequação de uma das dimensões do ecossistema da Ciência Aberta: os dados científicos. A questão que se coloca é: o que os editores de periódicos precisam saber para fazer os ajustes? Para isso, se apresentam os repositórios de dados científicos generalistas e específicos, ou seja, repositórios que servem para o depósito de qualquer tipo de dados científicos. As práticas atuais promovidas por esses repositórios são analisadas com relação a oito aspectos principais da publicação de dados, ou seja, formato do conjunto de dados, documentação, licenciamento, custos de publicação, validação, disponibilidade, descoberta e acesso e citação/referência.

A partir dessa análise, verifica-se que os repositórios implementam práticas bem consolidadas para a preservação de conjuntos de dados, como por exemplo, a qualidade dos metadados. Essas práticas e soluções não podem atender totalmente às necessidades de gerenciamento e uso dos recursos dos conjuntos de dados, especialmente em um contexto em que rápidas mudanças tecnológicas abrem continuamente novas perspectivas de exploração.

2. IMPORTÂNCIA DOS DADOS PARA A COMUNICAÇÃO CIENTÍFICA

À medida que a ciência se torna cada vez mais colaborativa, o compartilhamento de dados científicos torna-se mais importante. Este compartilhamento não inclui apenas o depósito e a preservação, mas, principalmente, o acesso para uso e reúso de dados gerados no ciclo de vida da pesquisa. Nesse contexto, o compartilhamento de dados no paradigma da Ciência Aberta tornou-se uma realidade no mundo todo.

Porém, compartilhar dados ainda não é uma atividade entre os pesquisadores, como mostram duas investigações, uma no contexto europeu de Smit e Gruttemeier (2011) e outra brasileira de Vanz *et al.* (2018). Com diferença temporal de oito anos revelam semelhanças quanto à cultura e predisposição dos pesquisadores em utilizar recursos para disponibilizar os dados científicos.

Smit e Gruttemeier (2011) destacam que aproximadamente 60% dos pesquisadores usam os dados de outros pesquisadores e no estudo de Vanz *et al.* (2018, p. 4) 31,71% dos respondentes declararam utilizar algum repositório para acessar dados produzidos por outros pesquisadores, ou seja, os dados de outras pesquisas são pouco usados e os fatores para isso são muitos, desde a credibilidade do autor que produziu tais dados como, também, a insegurança em replicar dados que podem estar equivocados e causar um efeito cascata nas novas pesquisas e na carreira do pesquisador.

Ambas as pesquisas mostram que os pesquisadores têm resistência em publicar os dados. Smit e Gruttemeier (2011) identificaram que em torno de 40% destes possuem dificuldade em compartilhar. Já em Vanz *et al.* (2018) verificamos que 37,02% não compartilham dado algum e apenas 9,19% dos respondentes afirmaram publicar seus dados. Quanto aos motivos para publicar

dados científicos, estão principalmente solicitação dos periódicos científicos, dos eventos e das agências de fomento.

Considerando estes dados sobre as práticas dos pesquisadores e lançando o olhar sobre os periódicos, verificamos o crescente incentivo para que os autores depositem os dados de suas pesquisas em repositórios de dados ou publiquem antecipadamente em revistas e/ou vinculados à publicação do artigo, podendo constituir diferentes materiais suplementares vinculados aos respectivos manuscritos.

Smit e Gruttemeier (2011) apontam que, mesmo quando os periódicos possuem políticas editoriais para o compartilhamento de dados, os pesquisadores não atendem a essa demanda. Isto ocorre por diferentes razões, haja vista que é necessário um trabalho prévio da instituição de filiação do autor para prepará-lo para atender a essa tendência.

3. INCLUSÃO DE POLÍTICAS EDITORIAIS PARA PERIÓDICOS: DADOS CIENTÍFICOS

Globalmente, as agências de fomento e os mandatos nacionais são as principais propulsoras para que os pesquisadores publiquem os dados científicos. Porém, essa questão faz parte de um movimento mais amplo nas comunidades científicas, juntamente com mandatos dos financiadores, declarações de organizações e sociedades acadêmicas, entre outros (FEDERER *et al.*, 2018).

Os editores de periódicos percebem essa necessidade e começam a inserir em suas políticas editoriais os conjuntos de dados subjacentes a resultados de pesquisa em suas publicações.

A presença dos dados científicos permite maximizar a reprodutibilidade e a replicabilidade das pesquisas já publicadas, bem como criar um corpo maior de dados disponíveis para reutilização e novas análises. Tendo em vista que boa parte do que é produzido cientificamente não tem possibilidade de reuso, replicação e reprodução nos encontramos na chamada crise da reprodutibilidade ou replicação (BACKER, 2016).

Portanto, se faz necessário os editores incorporarem políticas para autores a respeito dos dados científicos. Deve ser claro para os autores quais caminhos

devem seguir para a validação, replicação, reutilização e reprodução dos resultados da pesquisa, e como devem citar os dados científicos oriundos de outras pesquisas para garantir que os autores recebam os devidos créditos, tanto do ponto de vista de quem criou os dados, quanto daqueles que estão reusando ou replicando os mesmos.

Embora as diversas políticas tenham objetivos semelhantes, na prática, os editores de periódicos devem estar sintonizados com a infraestrutura que possuem, atingindo graus de abertura conforme as condições de cada periódico/instituição, afetando claramente as políticas. Sendo assim, alguns periódicos encorajam o compartilhamento dos dados, enquanto outros o exigem (FEDERER *et al.*, 2018).

Federer *et al.* (2018) acreditam que é improvável que os pesquisadores sintam-se motivados suficientemente para compartilhar dados na ausência de requisitos formais ou incentivos específicos, por isso, argumentam que é necessário encorajar e exigir. Neste contexto, os periódicos podem ter um impacto significativo no comportamento dos pesquisadores e no aumento da abertura dos dados científicos para compartilhamento, reuso e reprodutibilidade das pesquisas, vincular dados científicos aos artigos é um grande passo para as revistas e vem acontecendo há mais de uma década (SMIT; GRUTTMEIER, 2011).

Não é apenas com as políticas e as diretrizes que os editores científicos devem se preocupar. Ainda, devem enfrentar e ter soluções para o momento em que começarem a receber conjuntos de dados em grande volume e de complexidade variadas. Portanto, é necessário preparar-se para acompanhar a revisão por pares e o gerenciamento dos mesmos (SMIT; GRUTTMEIER, 2011). A inclusão dos conjuntos de dados enlaçados com os artigos, seja por meio de repositório ou documento complementar, irá exigir da equipe de pareceristas novos perfis de avaliadores.

Da mesma forma que os periódicos convencionais, os periódicos que receberão os conjuntos de dados, ou os periódicos de dados propriamente ditos, necessitam desenvolver diretrizes, critérios e instruções para os revisores. Ao analisar os critérios de avaliação existentes, identificamos as cinco classes de critérios a seguir:

1. Qualidade do manuscrito: os critérios convencionais para avaliar a redação do manuscrito, clareza, organização, aderência ao modelo;

2. Consistência em relação a importância do conjunto de dados: critérios para avaliar a eficácia do conteúdo disponível nos *datasets*;
3. Qualidade dos dados: critérios para avaliar as metodologias utilizadas na produção dos conjuntos de dados;
4. Reutilização de dados: critérios para avaliar a reutilização real dos conjuntos de dados;
5. Utilidade e contribuição dos dados: critérios para avaliar o potencial dos conjuntos de dados para a comunidade.

Outro aspecto importante é a decisão quanto à seleção do repositório onde serão depositados os dados dos artigos publicados pela revista. Este deve estar amparado nas decisões institucionais, no caso das revistas brasileiras, associadas a portais de periódicos, ou ter vinculação organizacional.

A política editorial deve contemplar orientações a respeito de onde o autor irá depositar seus dados, qual o conjunto mínimo de metadados para descrevê-los, assim como qual a indicação em relação à licença de uso e distribuição. Algumas decisões são de responsabilidade dos editores na seleção dos repositórios e elaboração da política editorial de recebimento dos dados científicos.

Ao elaborar a política sobre dados científicos, o editor deve ter algumas questões em mente:

- Identificar qual ou quais os repositórios mais adequados, tendo em vista a área de abrangência da revista e a qualidade do repositório. Os repositórios devem atender os princípios FAIR⁹¹ para promover a descoberta, acesso, interoperabilidade e reutilização de dados compartilhados.
- Informar os autores se será possível armazenar dados em repositórios institucionais ou apenas nos indicados pela revista. Para auxiliar na tarefa de seleção do repositório recomenda-se a consulta ao Re3data.org – *Registry of Research Data Repositories*⁹².

91 <https://www.go-fair.org/fair-principles/>

92 <https://www.re3data.org/>

- Elaborar as orientações para os autores quanto à necessidade da gestão dos dados em repositório indicado pela revista.
- Indicar o formato de citação e referência para dados científicos, adotar um padrão reconhecido pelos organismos internacionais, alinhados com o *Joint Declaration of Data Citation Principles* de 2014⁹³.
- Orientar o papel dos pareceristas no acesso e conferência dos dados científicos. Fazer parceria com algum instituto de estatística ou especialista quando for o caso de grandes volumes de dados e análises complexas;
- Facilitar o registro de um identificador persistente, como por exemplo, DOI ou Handle⁹⁴ ou orientar os autores sobre a necessidade e procedimentos para sua atribuição antes da submissão para a revista.

Se a revista não oferece nenhuma orientação a respeito da necessidade de publicação dos dados, o autor poderá fazer o *upload* em repositórios abertos como o *figshare*⁹⁵ ou o *Dryad Digital Repository*⁹⁶ durante o processo de submissão.

Um importante conceito a ser abordado é o de *data publishing*, que reflete um modelo de publicação científica. Promovem a publicação de conjuntos de dados, publicações acadêmicas, metadados que descrevem um conjunto de dados, ou um grupo de conjuntos de dados publicados de acordo com as práticas acadêmicas padrão. Seu objetivo final é fornecer informações sobre o quê, onde, por que, como e quem é responsável pelos dados. Portanto, a publicação de dados é um pré-requisito para permitir o compartilhamento e a reutilização de dados. Apesar de sua potencialidade, os *Data Journals* não são a solução definitiva e completa para todos os problemas de compartilhamento e reutilização de dados e em alguns casos, são considerados indutores de falsas expectativas na comunidade científica.

Neste cenário de repositórios abertos os *Data Journals* foram propostos como solução para algumas das necessidades discutidas acima. O conceito de *Data Journals* possui pelo menos dois elementos que precisam ser materializados

93 <https://www.force11.org/datacitationprinciples>

94 <https://www.handle.net/>

95 <https://figshare.com>

96 <https://datadryad.org>

em objetos de informação concretos e identificáveis para implementá-lo completamente: o conjunto de dados, ou seja, o sujeito do papel de dados, e o próprio papel de dados, ou seja, o artefato produzido para descrever o conjunto de dados. Assim, o termo documento de dados será usado para se referir apenas ao artefato. Esse artefato é homólogo a artigos de periódicos tradicionais; espera-se que tenha um identificador e um conteúdo com título, autores, resumo, diversas seções e referências.

O manuscrito e o conjunto de dados estão associados a outros objetos de informação, os metadados, que contêm informações úteis para o gerenciamento do objeto correspondente. O formato e o conteúdo dos metadados são geralmente selecionados pela entidade que gerencia os dois elementos, ou seja, o editor de uma revista e o repositório de dados.

A maioria dos periódicos de dados adota o esquema convencional de revisão por pares “fechada”, pré-publicação, revisão por pares anônima e privada. Poucos periódicos adotam uma revisão por pares “aberta” para promover imparcialidade e objetividade e reduzir o tempo de publicação.

No caso da revista *Earth System Science Data*⁹⁷, há uma abordagem em dois estágios. Após a submissão, os manuscritos são publicados na *Web* como “documentos de discussão” com o resultado de uma breve revisão realizada por um editor dedicado, solicitado a avaliar o artigo, exceto por seu conteúdo científico – por exemplo, se ele se encaixa no escopo da revista – e em seguida sugere correções técnicas mínimas, como erros de digitação. O documento permanece nesta fase por um período pré-determinado, enquanto a comunidade faz uma revisão e discussão. Todo o trabalho de discussão deve receber pelo menos dois comentários dos avaliadores⁹⁸. Após a fase de discussão pública, solicita-se aos autores que respondam publicamente aos comentários e produzam um manuscrito revisado que, se aprovado pelo editor, é finalmente publicado na revista.

No caso da *F1000 Research*⁹⁹, depois de passar por rápidas verificações iniciais realizadas pela equipe editorial interna, os manuscritos são publicados com o

97 <https://www.earth-system-science-data.net>

98 The two phases of the interactive public discussion. https://www.earth-system-science-data.net/peer_review/interactive_review_process.html

99 <https://f1000research.com>

status “aguardando revisão por pares”. Em seguida, solicita-se aos autores que sugiram cinco avaliadores em potencial para seus manuscritos, que deverão julgar o trabalho cientificamente, atribuindo-lhe um dos três *status* diferentes: “Aprovado”, “Aprovado com Reservas” e “Não Aprovado”. Todos os *status* e comentários dos árbitros são publicados com o manuscrito, sendo uma versão dele. De fato, esse processo leva à publicação de várias versões do manuscrito, que são “aceitas” apenas quando recebem o *status* de “Aprovado” ou “Aprovado com reservas” e um *status* “Aprovado”. Após a aprovação, o documento é indexado podendo ser localizado por meio de serviços como PubMed, Scopus e Google Scholar.

No caso do *Biodiversity Data Journal*¹⁰⁰, além da revisão por pares da “comunidade” (uma revisão por pares “fechada”), os autores podem optar por disponibilizar seu manuscrito para comentários de todos os usuários registrados no periódico e os revisores podem optar por permanecer anônimos ou divulgar seus nomes.

Nos próximos itens serão apresentados os tipos de repositórios de dados científicos e o papel compartilhado nesse contexto.

4. REPOSITÓRIOS DE DADOS CIENTÍFICOS

Os repositórios de dados científicos exercem um papel fundamental na ciência e o seu compartilhamento para uso por terceiros é um tópico de discussão desde a década de 1980 (SILVA, 2019). Implementam práticas sistemáticas de gestão, incluindo a curadoria, preservação, disponibilidade a longo prazo, disseminação e acesso. São amplamente difundidos nas comunidades que os produzem, como na física (SDSS no FermiLab¹⁰¹), na genética (GenBank na NCBI Data¹⁰²) ou nas ciências ambientais (British Atmospheric Data Center¹⁰³).

Recentemente, a demanda por repositórios de dados está emergindo também dentro da chamada “cauda longa” da ciência, ou seja, no contexto daqueles

100 <https://bdj.pensoft.net>

101 <https://www.sdss.org>

102 <https://www.ncbi.nlm.nih.gov/genbank>

103 www.datacentres.nerc.ac.uk

domínios científicos em que a atividade é realizada em um grande número de laboratórios relativamente pequenos e por pesquisadores individuais que coletivamente produzem a maioria dos resultados científicos. Em tais contextos, a pesquisa conduzida é específica, diversificada, de ponta e tradicionalmente carece de repositórios abertos. As necessidades reais das ciências que dependem de menores coletas de dados de “nível de pesquisa” são, no momento, menos compreendidas e não podem ser completamente baseadas em experiências e modelos feitos em repositórios disciplinares padronizados.

Os repositórios de dados científicos são frequentemente propostos como instrumentos para dar suporte à publicação de dados, pois fornecem facilidades aos diferentes atores envolvidos no processo.

Apesar da necessidade premente de práticas apropriadas para a publicação de dados científicos, não há consenso sobre o que os repositórios devem oferecer para apoiar a publicação de dados. Ao analisar algumas soluções existentes, apresentaremos, a seguir, os tipos e as características dos repositórios de dados científicos.

A análise se concentra em repositórios “generalistas” e “específicos”, os primeiros não fazem suposições nem arranjos especiais para aspectos específicos da comunidade ou do tipo de dados e estão abertos para publicar quaisquer conjuntos de dados científicos, como também repositórios específicos de uma área do conhecimento ou junção de áreas afins.

Silva (2019) esclarece que os repositórios generalistas são desenvolvidos especialmente para apoiar a publicação de conjuntos de dados produzidos nos contextos científicos de “cauda longa” devido à heterogeneidade dos possíveis resultados. Por outro lado, os repositórios específicos servem como fonte de referência e parâmetro do desenvolvimento científico em determinadas áreas do conhecimento. Para entender as possibilidades que estão sendo adotadas para o armazenamento de dados, revisamos plataformas de repositórios e apresentamos as diferentes abordagens adotadas que permitem superar as barreiras de compartilhamento de dados. A seguir, apresenta-se os tipos de repositórios e suas características de acordo com a categorização adotada por Silva (2019):

4.1 Institucionais

Repositórios Institucionais (RIs) podem ser definidos como sistemas de informação capazes de capturar, preservar e fornecer acesso à produção intelectual produzida pelos membros de uma instituição. Dentro do contexto específico da academia, um Repositório Institucional pode contribuir para o aumento do prestígio da instituição e nutrir a ideia de uma “memória coletiva de um centro de estudos”, pois gerencia e preserva itens informativos relevantes que, de outra forma, permaneceriam dispersos, desacompanhados ou inacessíveis. Facilitam a via verde¹⁰⁴ ao *Open Access* (OA) ao proporcionar um meio para que os pesquisadores disponibilizem todas as suas publicações, independentemente da abertura da revista original que publicou o artigo.

Com o tempo, essas plataformas também permitiram o armazenamento de dados, facilitando descrições básicas ou complexas de dados, geralmente utilizando identificadores que podem ser empregados para citar e encontrar os dados. Alguns RIs oferecem armazenamento ilimitado de dados, estão respaldados por uma universidade e normalmente são administrados pelos bibliotecários.

Mesmo que os RIs ofereçam confiabilidade, flexibilidade e controle, muitos deles possuem requisitos estritos para aceitar o arquivo dos dados científicos por meio de formatos muito genéricos. Faltam *Application Programming Interface* (API) ou, em português, Interface de Programação de Aplicativos para a interoperabilidade com outros sistemas, alguns somente utilizam um padrão de metadados geral, como o *Dublin Core*¹⁰⁵, e não incorporam os campos de metadados de domínio ou para tipos específicos de dados e vocabulários controlados.

104 A Via Verde consiste na disponibilização realizada pelos próprios autores dos artigos científicos já publicados ou aceitos para publicação, a partir da autorização expressa pelos editores. Disponível em: <https://www.lume.ufrgs.br/handle/10183/194784>

105 <https://dublincore.org/>

4.2 Temáticos

Os repositórios temáticos são aqueles que incluem dados científicos de um campo disciplinar específico. Alguns repositórios temáticos de sucesso são ArXiv¹⁰⁶, Pubmed¹⁰⁷ ou Eprints¹⁰⁸.

Diversas áreas do conhecimento dispõem de repositórios desenhados especificamente para os tipos de dados de seu domínio. Alguns exemplos são o *Protein Data Bank* da *Research Collaboratory for Structural Bioinformatics* (RCSB)¹⁰⁹, para formas 3D das proteínas, ácidos nucléicos e conjuntos complexos; o *GenBank*¹¹⁰, para as sequências de ADN; o *EMDataBank*¹¹¹, para mapas 3D de microscopia eletrônica de densidade, modelos atômicos e metadados associados; o *eCrystals*¹¹², inclui dados cristalográficos de raios X e a *National Oceanographic Data Center* (NODC)¹¹³, contendo dados oceanográficos.

Com o objetivo de fomentar sua reutilização, os repositórios temáticos têm ferramentas analíticas e de descobertas disponíveis junto aos dados. Alguns especialistas sugerem que os dados devem ser protegidos unicamente em repositórios temáticos porque, segundo eles, permitem o uso especializado de metadados e uma maior revisão e validação por especialistas no campo. Entretanto, nem todas as áreas possuem repositórios de dados, o que explica as dificuldades para encontrar o local apropriado para armazenamento nos repositórios existentes, isto se deve a própria especificidade e peculiaridade dos dados.

106 <https://arxiv.org/>

107 <https://www.ncbi.nlm.nih.gov/pubmed/>

108 <https://www.eprints.org/uk/index.php/flavours/eprints-for-research-data-1/>

109 <https://www.rcsb.org/>

110 <https://www.ncbi.nlm.nih.gov/genbank/>

111 <https://www.emdataresource.org/>

112 <http://ecrystals.chem.soton.ac.uk/>

113 <https://www.nodc.noaa.gov/>

4.3 Editoriais

Os repositórios editoriais oferecem características similares a dos repositórios institucionais, porém com alguns atributos especiais que atendem comunidades específicas. Dryad¹¹⁴ é um repositório digital de dados de pesquisa científica e médica, de caráter internacional, que abriga revistas científicas revisadas por pares. Funciona como repositório de diversas disciplinas e também facilita o número do *Digital Object Identifier* (DOI)¹¹⁵, registrado pelo *DataCite*¹¹⁶, conta com revistas tanto de acesso aberto como comerciais. Uma de suas principais diferenças é a capacidade de armazenar qualquer tipo de dados órfão. O Dryad realiza uma leitura automática dos metadados disponibilizados pelas revistas associadas, proporcionando a informação bibliográfica de cada artigo antes de sua publicação.

4.4 De propósito geral

Trata-se de repositórios nos quais qualquer pesquisador pode depositar dados, independentemente de sua filiação institucional. Os dois exemplos mais conhecidos são *Figshare*¹¹⁷ e *Zenodo*¹¹⁸.

Figshare é uma plataforma criada por *Digital Science*¹¹⁹ que permite compartilhar e mostrar os resultados de pesquisas multidisciplinares. Direciona-se a pesquisadores, cientistas, projetos e instituições. É associado ao *F1000 Research*¹²⁰, colabora com a *PLOS – Public Library of Science*¹²¹, a maior revista científica de acesso aberto do mundo e também com *Plum Analytics*¹²², um serviço que quantifica o impacto dos trabalhos de pesquisa publicados. Todo o material publicado no *Figshare* é identificado com um número DOI para facilitar sua

114 <https://datadryad.org/stash>

115 <https://www.doi.org/>

116 <https://datacite.org/does.html>

117 <https://figshare.com>

118 <https://zenodo.org>

119 <https://www.digital-science.com/>

120 <https://f1000research.com/>

121 <https://www.plos.org/>

122 <https://plumanalytics.com/>

localização e citação. Na plataforma podemos buscar apresentações, vídeos, pôsteres, imagens, dados, artigos, entre outros e a preservação dos dados funciona com tecnologia CLOCKSS (*Controlled LOCKSS*)¹²³, uma organização sem fins lucrativos que promove a aliança entre os editores do mundo acadêmico e as bibliotecas para arquivar, de modo sustentável, todo o conteúdo *web* produzido no âmbito científico.

Os usuários podem integrar os dados do repositório com outros *websites* e *blogs* copiando e colando um simples código. Os leitores podem realizar comentários sobre os conjuntos de dados e fazer *download* em arquivos de citação a seus gestores de referência para seu uso posterior. O repositório também oferece a possibilidade de publicar resultados negativos ou sobre experimentos fracassados para que outros pesquisadores poupem o esforço de ter que passar por testes já realizados e como consequência, economizem horas de trabalho e recursos.

O *Zenodo* é uma iniciativa do *OpenAire*¹²⁴, dispõe de uma infraestrutura adequada para a hospedagem de conjunto de dados e outros resultados de pesquisa de projetos europeus. Está construído sobre a plataforma *Invenio* e desenvolvimento no CERN, centro que se ocupa também da gestão da enorme quantidade de dados do *Large Hadron Collider* (LHC)¹²⁵. Como no caso de *Figshare*, o depósito é livre, atribui número DOI e permite conjuntos de dados disponíveis em BibTeX, EndNote e outros formatos bibliográficos. Os usuários podem agregar metadados mais detalhados que os do *Figshare* e os dados são disponibilizados sob licença CC0¹²⁶, ou seja, são de domínio público sem restrições nem solicitações de permissões, exceto para endereços de *e-mail*. Além disso, sempre que permitido, outros usuários do *Zenodo* podem comentar seus arquivos de forma fácil, inscrevendo-se com seu identificador ORCID¹²⁷ ou conta de GitHub¹²⁸.

123 <https://clockss.org/>

124 <https://www.openaire.eu/>

125 <https://home.cern/science/accelerators/large-hadron-collider>

126 https://creativecommons.org/publicdomain/zero/1.0/deed.pt_BR

127 <https://orcid.org/>

128 <https://github.com/>

4.5 Repositórios próprios

Em algumas ocasiões, os pesquisadores arquivam seus dados científicos em um servidor pessoal ou de seus projetos ou na nuvem. Existem variadas opções tecnológicas, como a gratuita de Dropbox, e outras comerciais, como as que oferecem *Amazon Cloud Drive* ou *Microsoft Azure*. A manutenção dos repositórios próprios depende da capacidade do pesquisador para adaptar as necessidades de seus projetos, assim como para levar a cabo ações de preservação como de *backup* e replicação adequadas. Estas são opções menos recomendáveis pela falta de garantias de organização, manutenção e preservação.

5. RESPONSABILIDADE DO COMPARTILHAMENTO DOS DADOS CIENTÍFICOS

Os dados que servem à *Big Science*, bem como aqueles que servem à *Long-tail Science* estão emergindo como instrumentos incentivadores da Ciência Aberta. Beneficiando-se da disponibilidade de dados, os pesquisadores estão trabalhando com uma grande variedade de novos padrões de pesquisa que estão revolucionando a forma como a ciência está sendo conduzida. A realização completa dessa mudança de paradigma, no entanto, requer a abordagem de muitas questões onerosas e desafiadoras.

Embora exista um acordo quase universal sobre os benefícios do “compartilhamento e reutilização de dados” como forma de acelerar o desempenho da ciência, há várias barreiras que dificultam a realização desse objetivo de maneira sistemática e eficaz. Essas barreiras são metodológicas, legais, técnicas e frequentemente relacionadas à falta de incentivos para os pesquisadores compartilharem seus dados. Os efeitos desses obstáculos na ciência são diversos, inclusive quando a disponibilidade dos dados de uma pesquisa é afetada pelo tempo de processamento para publicação do artigo se nenhuma política estiver em vigor. Portanto, práticas e políticas apropriadas de compartilhamento de dados devem ser introduzidas para promover a disponibilidade dos dados. Além disso, devem ser identificados mecanismos para conscientizar a comunidade científica sobre os conjuntos de dados disponíveis, facilitar seu entendimento e promover sua reutilização.

Essa responsabilidade é compartilhada entre todos os atores da comunicação científica, desde as agências de fomento, as universidades, os institutos, os pesquisadores, os avaliadores e a própria sociedade. Os financiadores da ciência de diferentes instâncias podem colaborar estabelecendo critérios nos editais para determinar que a pesquisa só será fomentada com o acesso aberto aos dados, aos artigos e a quaisquer publicação oriunda desse recurso (prática recorrente nas agências internacionais de financiamento de pesquisa).

As universidades devem incorporar o ensino da gestão de dados científicos na formação dos pesquisadores incluindo noções de direitos autorais que permitam entender onde serão publicados os dados, quem detém seus direitos e os embargos de acesso; apontar para a necessidade de estabelecer políticas institucionais de depósito de dados científicos que garantam o crédito da autoria científica. Da mesma forma, são necessárias políticas institucionalizadas de abertura de informação nas quais todo o sistema produtivo da instituição deve estar alinhado com os requisitos internacionais e, principalmente, aos interesses nacionais.

Do ponto de vista do autor, ainda que a revista não ofereça nenhuma orientação a respeito da submissão/inclusão do conjunto dos dados, este poderá autoarquivar os dados, em repositórios abertos como o *figshare*¹²⁹ ou o *Dryad Digital Repository*¹³⁰ antes mesmo do processo de submissão. Além desse processo, o autor precisa planejar e gerenciar o investimento para preservar os dados em longo prazo, prevendo qual de recurso financeiro e o montante que será necessário para realizar a pesquisa e preservá-la em longo prazo.

6. CONSIDERAÇÕES FINAIS

Analisamos as práticas e abordagens atuais dos conjuntos de dados implementadas pelos editores de periódicos a partir da perspectiva de quatro aspectos principais que contribuem para o compartilhamento e reutilização de dados. Verificamos como realizar a descrição dos dados, como promover o acesso, a importância da citação e como fornecer garantias de qualidade.

129 <https://figshare.com>

130 <https://datadryad.org>

De nossas observações, concluímos que os editores de periódicos 1- ainda não têm uma estratégia compartilhada e consolidada para promover uma descrição eficaz do conjunto de dados que favoreça sua reutilização, por exemplo, a única informação do conjunto de dados que é promovida por todos os periódicos é a “disponibilidade”, de maneira que outras informações como “cobertura”, “qualidade” ou “reutilização” são negligenciadas em muitos casos; 2- geralmente confiam nos serviços oferecidos por repositórios e arquivos de dados de terceiros para disponibilizar o conjunto de dados; 3- estão negligenciando muitos dos problemas que afetam a citação de conjuntos de dados, assumindo que as abordagens de citação espelhadas no modelo de publicação científica são suficientes; 4- ainda não foram aprofundadas as questões relacionadas à qualidade dos conjuntos de dados. Em essência, eles continuam a confiar em abordagens consolidadas de revisão por pares, concentrando-se apenas no papel de dados.

Os editores de periódicos estão em uma posição privilegiada para promover práticas de publicação de dados. No entanto, é necessário muito trabalho para alcançar um entendimento comum sobre quais devem ser as práticas para contribuir com o compartilhamento e a reutilização de dados. A promoção da publicação dos dados científicos não é de responsabilidade exclusiva dos editores de periódicos. Em vez disso, os periódicos são parte integrante de todo o ecossistema subjacente às infraestruturas de comunicação científica e são chamados a desenvolver práticas inovadoras, abrangentes e eficazes de publicação de dados em tais ambientes.

7. REFERÊNCIAS

AUSTIN, C. C., *et al.* Key components of data publishing: using current best practices to develop a reference model for data publishing. **International Journal on Digital Libraries**, v. 18, n. 2, jun. 2017. Disponível em: <https://doi.org/10.1007/s00799-016-0178-2>. Acesso em: 16 jan. 2020.

BAKER, M. 1,500 scientists lift the lid on reproducibility. **Nature**, v. 533, n. 7604, p. 452-454, 2016. DOI: 10.1038/533452a. Disponível em: <http://www.nature.com/news/1-500-scientists-lift-the-lid-on-reproducibility-1.19970>. Acesso em: 16 jan. 2020.

FEDERER, L. M., *et al.* Data sharing in PLOS ONE: An analysis of Data Availability Statements. **PLoS ONE**, v. 13, n. 5. Disponível em: <https://doi.org/10.1371/journal.pone.0194768>. Acesso em: 16 jan. 2020.

SILVA, F. C. C. da. **Gestão de dados científicos**. Rio de Janeiro: Interciência. 2019. 146 p.

SMIT, E.; GRUTTEMEIER, H. Are scholarly publications ready for the data era? suggestions for best practice guidelines and common standards for the integration of data and publications. **New Review of Information Networking**, v. 16, n. 1, p. 54-70, 2011. DOI: 10.1080/13614576.2011.574488

VANZ, S. A. de S. *et al.* **Acesso aberto a dados de pesquisa no Brasil: práticas e percepções dos pesquisadores: relatório – 2018**. Porto Alegre: UFRGS, 2018. 91 p.

8. SOBRE OS AUTORES



Caterina Groposo Pavão

Lattes: <http://lattes.cnpq.br/4834791532698069>

Orcid: <http://orcid.org/0000-0003-3712-7200>

Doutora em Informação e Comunicação pela Universidade Federal do Rio Grande do Sul (UFRGS), em 2014. Professora Adjunta do Departamento de Ciência da Informação/FABICO/UFRGS, atuando no curso de graduação de Biblioteconomia e no Programa de Pós-Graduação em Ciência da Informação (PPGCIN), da mesma Universidade. Participa de Grupos de Pesquisa em Comunicação Científica e no NUAWEB - Núcleo de Pesquisa em Arquivamento da Web e Preservação Digital.



Fabiano Couto Corrêa da Silva

Lattes: <http://lattes.cnpq.br/4635807083312321>

Orcid: <https://orcid.org/0000-0001-5014-8853>

Doutor em Información y documentación en la Sociedad del Conocimiento pela Universitat de Barcelona (2017). Professor Adjunto do Departamento de Ciência da Informação/FABICO da Universidade Federal do Rio Grande do Sul (UFRGS), atuando no curso de graduação de Biblioteconomia e o Programa de Pós-Graduação em Ciência da Informação (PPGCIN), ambos da mesma instituição. Atua como vice-líder do Grupo de Pesquisa em Comportamento e Competências InfoComunicacionais (InfoCom).



Lúcia da Silveira

Lattes: <http://lattes.cnpq.br/4927963964595486>

Orcid: <http://orcid.org/0000-0003-1118-2121>

Doutoranda do Programa de Pós-Graduação em Comunicação na UFRGS, possui mestrado em Gestão de Unidades de Informação pela Universidade do Estado de Santa Catarina e graduação em Biblioteconomia pela Universidade Federal de Santa Catarina. Coordenadora do Portal de Periódicos UFSC. Pesquisadora e palestrante nos seguintes temas: acesso aberto, Ciência Aberta, periódico científico, portal de periódicos científico, serviço de editoração para periódicos em bibliotecas universitárias e comunicação científica.

9. COMO CITAR

PAVÃO, C. G.; SILVA, F. B. C. da; SILVEIRA, L. da. Gestão de dados em periódicos científicos. In: SHINTAKU, M.; SALES, L. F; COSTA, M. (org). **Tópicos sobre dados abertos para editores científicos**. Botucatu, SP: ABEC, 2020. p. 165-182. DOI: 10.21452/ 978-85-93910-04-3.cap14.

DADOS DE PESQUISA: DISPONIBILIZAÇÃO OU PUBLICAÇÃO?

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

1. INTRODUÇÃO

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

2. DISPONIBILIZAÇÃO DE DADOS DE PESQUISA

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

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

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

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

3. PUBLICAÇÃO DE DADOS DE PESQUISA

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

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

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

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

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

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

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

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

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

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

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

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

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

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

4. DADOS DE PESQUISA: ALGUMAS DISCUSSÕES

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

5. RECOMENDAÇÕES

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

Para pesquisadores:

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

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

Para gestores de pesquisa:

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

Para gestores de repositórios:

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

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

Para editores de periódicos:

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

6. CONSIDERAÇÕES FINAIS

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

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

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

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

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

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

7. REFERÊNCIAS

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

8. SOBRE OS AUTORES



Emanuelle Torino

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

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

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



Sandra Milena Roa Martínez

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

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

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



Silvana Aparecida Borsetti Gregorio Vidotti

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

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

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

9. COMO CITAR

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

A Ciência Aberta no CNPq: acesso aberto aos dados e resultados de pesquisas

Adriana Maria Tonini
Paulo Henrique de Assis Santana
Thaís Machado Scherrer

1. INTRODUÇÃO

A Parceria para Governo Aberto (*Open Government Partnership* – OGP) foi oficializada em 2011. Atualmente o Brasil está em seu 4º Plano de Ação, que contém 11 compromissos, os quais envolvem mais de 80 instituições entre organizações da sociedade civil e órgãos da Administração Pública (federal, estaduais e municipais), sempre com o intuito de fortalecer os princípios de transparência, participação cidadã, inovação, prestação de contas e responsabilização (*accountability*).

O Compromisso 3 do 4º Plano de Ação consiste em estabelecer mecanismos de governança de dados científicos para o avanço da Ciência Aberta no Brasil, e foram estabelecidos nove marcos para o seu devido cumprimento. O compromisso pretende avançar nos processos relacionados à disponibilização de dados abertos de pesquisa científica por meio do aprimoramento de instrumentos de governança. A conceituação de Ciência Aberta é ampla, ora é definida como um movimento de acesso ao conhecimento científico acessível a todos, mas apesar da variação nos conceitos, é invariável a aceitação de que os dados científicos devem ser localizáveis, interoperáveis, reprodutíveis e reutilizáveis.

Dentro do compromisso 3, o CNPq assumiu a coordenação do Marco 5 - Articulação com agências de fomento para a implantação de ações de apoio à Ciência Aberta – e participa, ainda, dos marcos 7 e 9: implantação de infraestrutura federada piloto de repositórios de dados de pesquisa e proposição de conjunto de indicadores para aferição da maturidade em Ciência Aberta, respectivamente. As ações do compromisso corroboram também com objetivos da Agenda 2030, entre os quais destacamos, dentro do Objetivo 9: “Construir infraestruturas resilientes, promover a industrialização inclusiva e sustentável e fomentar a inovação”, o Item 9.5 - Fortalecer a pesquisa científica, melhorar as capacidades tecnológicas de setores industriais em todos os países, particularmente os países em desenvolvimento, inclusive até 2030, incentivando a inovação e aumentando substancialmente o número de trabalhadores de pesquisa, além do desenvolvimento por milhão de pessoas e do gasto público e privado em pesquisa e desenvolvimento.

Para as ações relativas à participação do CNPq não apenas no Compromisso 3, como também em ações contínuas em prol da Ciência Aberta no Brasil, inicialmente a discussão focou as temáticas: sensibilização, procedimentos operacionais e regulamentação.

Na temática sensibilização, a finalidade foi apresentar, discutir e aprofundar o tema e propor iniciativas de implementação de ações para a Política Nacional de Ciência Aberta com servidores, comunidade científica e membros de comitês de assessoramento. Foi oferecida capacitação aos servidores na temática e há um amplo espectro de possibilidades para avanço nessa frente de trabalho.

Já nos procedimentos operacionais, faz-se necessário o aprofundamento da articulação com parceiros, discutindo questões sobre a disponibilização dos relatórios de pesquisa e dos dados abertos. Já foi viabilizado um novo modelo de relatório contendo estrutura de dados a serem catalogados no repositório de Ciência Aberta, em conjunto com RNP/IBICT, no âmbito do Marco 7 da OGP/CGU.

Nas ações de atualização e ampliação da regulamentação, discussões vêm ocorrendo junto ao Ministério de Ciência, Tecnologia, Inovações e Comunicações (MCTIC) para a elaboração do instrumento que definirá a Política Nacional para a Ciência Aberta. Internamente, já há proposta de alteração no modelo de Chamadas Públicas do CNPq, que passará a conter diretrizes incentivando a inclusão dos dados de pesquisa produzidos no repositório Lattes Data.

O Lattes Data, expansão dentro da Plataforma Lattes, será um repositório para armazenamento de dados de pesquisa, constituindo-se numa das dimensões necessárias para implementação da Ciência Aberta no Brasil, com a finalidade de disponibilizar o acesso ao conhecimento, compartilhamento e reúso de dados pela comunidade científica e acompanhamento pela sociedade.

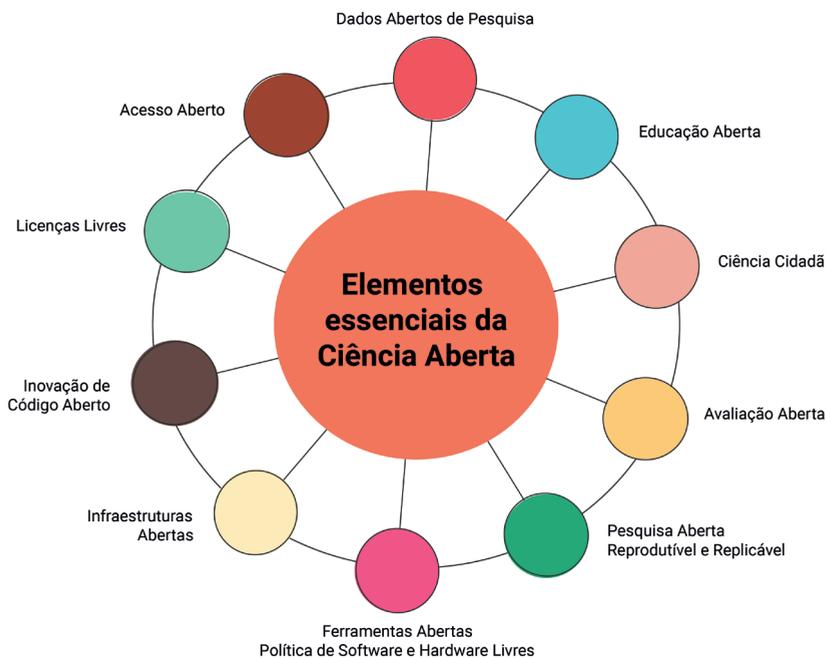
O Lattes Data permitirá armazenar os dados gerados pelas pesquisas científicas financiadas com recursos públicos oriundos do CNPq, de modo a manter sua preservação a longo prazo e, institucionalmente, assentar a governança dos dados científicos.

2. CIÊNCIA ABERTA NO CNPQ: DA INTENÇÃO À PRÁTICA

De acordo com a Declaração do Panamá sobre Ciência Aberta 2018, há dez elementos essenciais, que constam a seguir e são vistos esquematicamente na Figura 16.1:

1. Acesso Aberto;
2. Dados Abertos de Pesquisa;
3. Educação Aberta,
4. Ciência Cidadã,
5. Avaliação Aberta;
6. Pesquisa Aberta Reprodutível e Replicável;
7. Ferramentas Abertas e Política de Software e Hardware Livres;
8. Infraestruturas Abertas;
9. Inovação de Código Aberto;
10. Licenças Livres.

Figura 16.1 - Elementos essenciais da Ciência Aberta



Fonte: Declaración de Panama sobre Ciencia Abierta (2018)

Desde 1999 o CNPq tem estado envolvido com atividades de Acesso Aberto e Dados Abertos de Pesquisa, algumas preparatórias, outras mais efetivas.

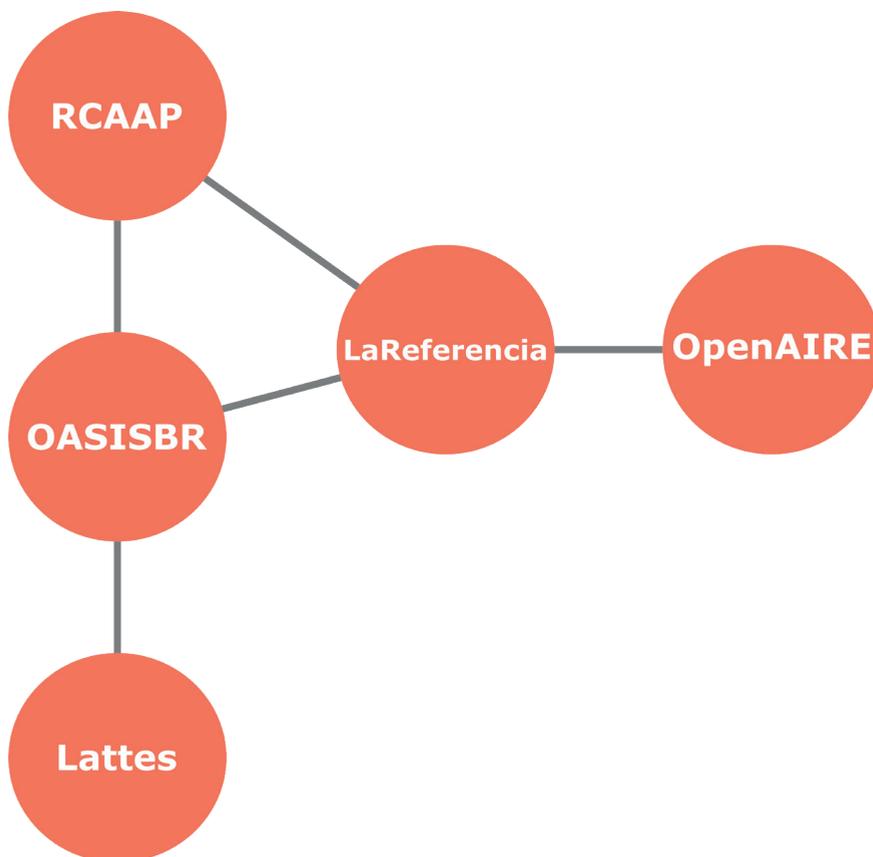
Com o lançamento da Plataforma Lattes, em 1999, o CNPq abraçou uma política de compartilhamento de dados curriculares de pesquisadores com as diferentes agências de fomento nacionais e internacionais. Nesse sentido, seu padrão de registro tem sido adotado por diversas e importantes universidades brasileiras e pela Fundação para a Ciência e Tecnologia (FCT) de Portugal, onde a Plataforma Lattes deu origem à Plataforma DeGóis.

Demonstrando seu interesse em contribuir para a Ciência Aberta, a partir de 2001, a Plataforma Lattes passou a exibir, por meio do currículo Lattes, os textos completos existentes na *Scientific Electronic Library On-line* (SciELO), programa apoiado pela Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), a Organização Pan-Americana de Saúde (OPAS) e o CNPq, contribuindo assim para o acesso aberto aos artigos publicados naquela biblioteca on-line.

Em paralelo ao desenvolvimento da Plataforma Lattes, desenvolveram-se o Programa SciELO, em âmbito internacional, e o Portal Brasileiro de Acesso Aberto à Informação Científica (OASISBR), operado pelo Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT) e que se conecta aos Repositórios Científicos de Acesso Aberto de Portugal (RCAAP) e à Rede Federada de Repositórios Institucionais de Publicações Científicas (La Referencia) que, além do Brasil, congrega os repositórios de mais oito países latino-americanos. Este último repositório, por sua vez, agrega-se ao *Open Access Infrastructure for Research in Europe* (OpenAIRE), repositório de acesso aberto da União Europeia.

Nesse contexto, o CNPq ampliou suas iniciativas em 2019, com a integração da Plataforma Lattes ao OASIS-BR, a fim de que artigos citados nos currículos Lattes passem a integrar aquele repositório nacional, o que aumentará sua visibilidade internacional, pois serão vistos também pelos usuários do RCAAP, La Referencia e OpenAIRE. Da mesma forma, os textos completos existentes no OASIS-BR complementarão as informações dos registros Lattes, ampliando assim as possibilidades de acesso aberto aos artigos científicos, conforme pode ser visto esquematicamente na Figura 16.2.

Figura 16.2 - Cadeia de divulgação e acesso aberto prevista com a integração da Plataforma Lattes ao OASISBR



Fonte: Elaboração própria (2020)

Adicionalmente, no âmbito do Marco 5 do Compromisso 3 do 4º Plano de Ação Nacional em Governo Aberto, o CNPq, em parceria e acordo formalmente estabelecido com o IBICT, iniciou desenvolvimento do que se convencionou denominar Lattes Data: um repositório federado de acesso aberto para dados de pesquisa, ação que situa o CNPq como referência brasileira relativamente ao elemento Dados Abertos de Pesquisa. Considerando que o Lattes Data pode ser desenvolvido de forma complementar ao OASIS-BR e com padrões compatíveis com o OpenAire – que já possui mais de 9.800.000 registros de dados de pesquisa, inclusive com razoável participação de dados brasileiros –, poder-se-á construir repositório integrável àquele instrumento europeu, tornando o Brasil importante ator neste segmento.

3. EVOLUÇÃO DO PROGRAMA EDITORIAL E ESTÍMULOS À CIÊNCIA ABERTA NAS CHAMADAS PÚBLICAS DO CNPQ

No Brasil, a maioria do investimento em CT&I é de origem pública e tais atividades ainda se concentram primariamente nas universidades e institutos de pesquisa. Sob a missão de fomentar a ciência, tecnologia e inovação e atuar na formulação de suas políticas, contribuindo para o avanço das fronteiras do conhecimento, o desenvolvimento sustentável e a soberania nacional, o CNPq tem o compromisso com a sociedade brasileira de garantir financiamento estratégico ao desenvolvimento científico, tecnológico e inovador do Brasil, de modo sustentável, transparente, reprodutível e eficiente, a fim de garantir ciclos de crescimento, promoção da inovação e das atividades econômicas, geração de riqueza, melhoria da priorização de ações com maior resultados dos investimentos. Esses valores estão em consonância com um Estado democrático, que promove cidadania, e com os princípios da Ciência Aberta, a fim de que o fomento público possa gerar dados, pesquisas, impactos e frentes de trabalho que evoluam em divulgação, testes e novos avanços.

Assim, o CNPq entende que a publicação científica, seja por meio de periódicos, livros ou anais de eventos, ainda constitui a principal ferramenta da divulgação e disseminação da produção intelectual, sendo necessário avançar em estruturas e conscientização para uso adequado de mecanismos cada vez mais atuais e de ágil disponibilidade no intuito de que dados e resultados sejam acessíveis, verificáveis e reutilizáveis, proporcionando novas possibilidades para o avanço científico e tecnológico.

Ainda em 2005, o Conselho Deliberativo do CNPq manifestou formalmente seu apoio à publicação em acesso aberto e fazendo recomendações que gradualmente vêm sendo implementadas, como, por exemplo:

- Apoio a bibliotecas eletrônicas, como a SciELO;
- Acordos institucionais de integração da Plataforma Lattes com a SciELO, SCOPUS, entre outros, permitindo acesso integrado de informações de pesquisadores, grupos de pesquisa e publicações;
- A inclusão do critério de acesso aberto nas chamadas do Programa Editorial;

- Repositório de publicações científicas, tecnológicas e de inovação, cuja ideia foi aprovada em 2007, vindo a gerar o OASIS-BR, mesmo que fora do CNPq, de modo que atualmente caminhamos para o início da implementação de um repositório de dados científicos, o Lattes Data.

Não é em vão que o CNPq possui uma longa tradição de fomento sistemático à editoração de periódicos científicos, por meio de chamadas anuais, lançadas de modo praticamente ininterrupto há mais de trinta anos. Nas últimas edições, têm sido apoiados cerca de 200 periódicos por ano.

O objetivo do Programa Editorial do CNPq é apoiar e incentivar a editoração e a publicação de periódicos científicos brasileiros em todas as áreas de conhecimento, priorizando o apoio às revistas divulgadas por meio eletrônico, na *internet*, em modo de acesso aberto, ou de forma impressa/eletrônica simultaneamente.

Em 2018, o CNPq incluiu nos critérios da chamada do Programa Editorial pontuação específica para publicações em formato *Open Access*, que não se tratava de critério eliminatório, entretanto trazia impacto significativo na classificação das propostas. Na edição de 2019, o peso dado a este critério foi aumentado e foram incorporadas as recomendações de que os beneficiários do Programa publiquem seus trabalhos segundo os princípios do modelo de acesso aberto, bem como de que depositem seus trabalhos em repositórios eletrônicos de acesso público, acompanhados da seguinte explicação: “A publicação de acesso aberto é propriedade dos seus autores e não necessariamente das revistas e órgãos de publicação, devendo preencher duas condições: a) os autores e detentores do *copyright* garantem o direito do acesso aberto, irrevogável e perpétuo, sem restrições geográficas, a todos os usuários para copiar, usar, distribuir, transmitir, exibir o trabalho em qualquer meio digital para qualquer objetivo responsável, desde que seja reconhecida e atribuída a sua autoria; b) uma versão completa do trabalho, assim como todo material suplementar relacionado, em formato eletrônico padrão, deve ser depositada imediatamente após a publicação inicial em pelo menos um repositório eletrônico. Tal repositório deve ser garantido por uma instituição acadêmica, uma sociedade científica ou similar, uma agência governamental ou outra organização reconhecida, e garantir o acesso aberto, distribuição irrestrita, interoperabilidade e arquivamento de longa duração”.

Posteriormente, tais recomendações – incluindo a possibilidade de utilização de recursos para pagamento de custos de publicação no modelo de acesso aberto – foram incorporadas ao modelo padrão de chamada do CNPq, devendo

ser obrigatoriamente observadas em todas as ações de fomento lançadas pelo órgão desde então.

Ainda há muito o que avançar no que concerne ao estímulo da Ciência Aberta nas chamadas públicas do CNPq. Em geral, seja por falta de conscientização ou de estímulo, percebe-se que os pesquisadores preocupam-se em publicar suas pesquisas em periódicos de maior impacto, sem qualquer reflexão específica sobre disponibilização em acesso aberto. A falta de regulamentação e de repositórios mais amplos também são fatores negativos.

O papel das agências de fomento é fundamental em reconhecer publicações de acesso aberto, valorizá-las na avaliação da produtividade dos pesquisadores e dar a elas visibilidade, como também a repositórios digitais. A articulação entre as agências deve estar centrada no progressivo estabelecimento, fortalecimento e avanço de políticas públicas em prol do incremento de práticas de Ciência Aberta no Brasil, implicando, principalmente, a definição de agendas, propostas e ações dotadas de flexibilidades que permitam a operação conjunta de atores com diferentes características, conexões, habilidades e capacidades. Assim, este Conselho está comprometido, em especial por meio de sua atuação junto ao Plano de Trabalho da OGP, para o estabelecimento de políticas em Ciência Aberta e atividades de conscientização e estímulo ao acesso aberto.

4. CONCLUSÕES

Práticas consolidadas de Ciência Aberta certamente contribuirão para o avanço científico, tecnológico e de inovação em toda a humanidade, seja no que concerne à rapidez no acesso à informação ou no processo de desenvolvimento de descobertas; na eficiência e replicabilidade científica, reduzindo duplicação de esforços, aumentando a confiabilidade e propiciando intercâmbios entre diferentes grupos de pesquisa; além de possibilitar maiores cooperações multidisciplinares e intersetoriais, o que por fim contribuirá para o aumento da disseminação do conhecimento e maiores retornos em Pesquisa e Desenvolvimento.

Reconhece-se, entretanto, que para o avanço da Ciência Aberta são necessárias mudanças de paradigmas, referenciais e modelos avaliativos, num processo contínuo e consistente. Em todo caso, para garantir mais progressos, são necessários não apenas investimentos em infraestrutura mas, principalmente, em conscientização e em desenvolvimento de políticas institucionais e de modelos

para publicação e avaliação das pesquisas que valorizem a disponibilização de dados e resultados em acesso aberto.

Apesar disso, não se pode menosprezar as conquistas alcançadas até o momento. E a interação entre as agências de fomento, instituições científicas, tecnológicas e de inovação, além de editores científicos, é fundamental para que os avanços se aprofundem e se consolidem.

A Parceria para o Governo Aberto foi, sem dúvida, um importante impulsionador de um novo estágio de maturidade da Ciência Aberta no CNPq, tendo como um de seus principais frutos o incentivo à implantação de um repositório de dados científicos no CNPq. Em especial, destaca-se que a materialização da Ciência Aberta no Lattes Data poderá tanto catalisar a investigação científica, promovendo o aumento da eficiência na produção da ciência, quanto potencializar o valor da informação científica gerada.

5. REFERÊNCIAS

BRASIL. Ministério da Transparência. Controladoria-Geral da União. **4º Plano de Ação Nacional em Governo Aberto**. 31 out. 2018. 52 p. Disponível em: https://www.opengovpartnership.org/wp-content/uploads/2018/10/Brazil_Action-Plan_2018-2020_POR.pdf. Acesso em: 22 jan. 2020.

OPEN GOVERNMENT PARTNERSHIP. **Brazil Mid-Term Report 2016-2018**. 23 jun. 2019. 52 p. Disponível em: <https://www.opengovpartnership.org/documents/brazil-mid-term-report-2016-2018/>. Acesso em: 22 jan. 2020.

FREIRE, J. D. **O CNPq e o acesso aberto à informação científica**. 2011. 273 p. Tese (Doutorado em Ciência da Informação, área: Transferência da Informação). Faculdade de Ciência da Informação. Universidade de Brasília, Brasília /DF.

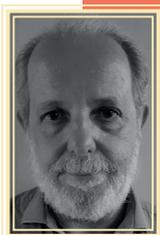
6. SOBRE OS AUTORES



Adriana Maria Tonini

Lattes: <http://lattes.cnpq.br/7318970207481115>

Engenheira Civil pela UFMG (1992), licenciada pela Fundação de Educação para o trabalho de Minas Gerais (1995), Mestre em Tecnologia pelo CEFET-MG (1999) e Doutora em Educação pela UFMG (2007). Atualmente é Diretora de Engenharias, Ciências Exatas, humanas e Sociais do CNPq, além de atuar como Professora do Mestrado em Educação Tecnológica do CEFET-MG e avaliadora institucional Externa do SINAES.



Paulo Henrique de Assis Santana

Lattes: <http://lattes.cnpq.br/9962577608407219>

Engenheiro Mecânico pela Escola de Engenharia de São Carlos (1969), mestre em Ciência Espacial pelo Instituto Nacional de Pesquisas Espaciais (1974), cursos de doutorado, sem desenvolvimento de tese, em Ciência da Computação (1977). Entre 1989 e 2003 foi gestor de tecnologia da informação no CNPq, tendo sido o coordenador do desenvolvimento e implementação da Plataforma Lattes no CNPq e em diversas universidades brasileiras. Atualmente é Coordenador de Informação e Estudos Internacionais do CNPq.



Thaís Machado Scherrer

Lattes: <http://lattes.cnpq.br/8764728978783752>

Bacharel em Meteorologia pela Universidade de São Paulo (2002), Mestre em Meteorologia pela Universidade de São Paulo (2006) e Doutora em Geofísica Aplicada pela Universidade de Brasília (2014). Desde 2004 atua como analista em C&T no Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq, ocupando atualmente a função de Assistente na Diretoria de Engenharias, Ciências Exatas, Humanas e Sociais.

7. COMO CITAR:

TONINI, A. M.; SANTANA, P. H. de A.; SCHERRER, T. M. A Ciência Aberta no CNPq: acesso aberto aos dados e resultados de pesquisas. *In*: SHINTAKU, M.; SALES, L. F; COSTA, M. (org). **Tópicos sobre dados abertos para editores científicos**. Botucatu, SP: ABEC, 2020. p. 203-214. DOI: 10.21452/ 978-85-93910-04-3.cap16.

O Alinhamento das revistas brasileiras sobre Medicina Tropical e Doenças infecciosas e Parasitárias às práticas da Ciência Aberta

Benedito Barraviera
Adeilton Alves Brandão
Dalmo Correia
Rui Seabra Ferreira Junior
Milton Shintaku

1. INTRODUÇÃO

A Editoração Científica, que é um processo vinculado à comunicação, volta-se à disseminação dos resultados de pesquisa por meio de livros, resumos apresentados em eventos ou mesmo revistas científicas. Este processo, na maioria dos casos, é a etapa final da pesquisa. Mesmo sendo uma atividade ligada às investigações científicas, quem o comanda são os editores, por meio de modelos centenários aceitos pela comunidade científica, os quais vêm sendo modificados em decorrência do impacto cada vez maior das novas tecnologias.

Não apenas as mudanças tecnológicas têm impactado a editoração científica, mas a própria evolução ou revolução das ciências, na qual o movimento da Ciência Aberta se apresenta no momento como um protagonista

relevante. Como defende Albagli (2014), a Ciência Aberta é um termo que engloba diversas iniciativas, com perspectivas, pressupostos e implicações distintas. Assim, parte da comunidade científica enxerga esta iniciativa com entusiasmo, na mesma medida em que outra a vê com muita resistência, haja vista as mudanças comportamentais necessárias para a implantação do novo movimento.

A Associação Brasileira de Editores Científicos (ABEC) não ficou fora deste novo desafio; muito pelo contrário, acolheu e participou com intensidade. Assim, o tema principal do ABEC Meeting 2019, ocorrido em Fortaleza, Ceará, foi a Ciência Aberta. Ferreira Jr. (2019) relata que a ABEC disponibiliza frequentemente cursos e textos, divulgando informações relevantes e atualizadas com o melhor do cenário internacional. Para tanto, utiliza os seus canais de comunicação social com o objetivo de divulgar uma produção de conteúdo que estimule e divulgue a importância da implantação da Ciência Aberta no cenário das revistas brasileiras.

Entretanto, a decisão de adotar princípios da Ciência Aberta no processo editorial é de exclusiva iniciativa e responsabilidade dos editores, apoiados no cenário da própria revista e de seus colaboradores e parceiros. A ABEC pode divulgar conceitos, princípios e ideias que visam orientar a adesão. As agências de fomento podem – se desejarem e entenderem os novos paradigmas – direcionar financiamentos principalmente àquelas que aderirem aos novos tempos.

Por outro lado, a biblioteca SciELO (*Scientific Electronic Library Online*) tomou a iniciativa em 2019 de convidar três periódicos nacionais arbitrados, acadêmicos, não comerciais, de acesso aberto e de elevado fator de impacto para participar do novo desafio. A reunião ocorreu em São Paulo sob a coordenação do professor Abel L. Packer e sua equipe; dos editores Benedito Barraviera, Jean-Philippe Chippaux e Rui Seabra Ferreira Junior, do *Journal of Venomous Animals and Toxins including Tropical Diseases*; e dos editores-chefe da Memórias do Instituto Oswaldo Cruz e da Revista da Sociedade Brasileira de Medicina Tropical, respectivamente os professores Adeilton Alves Brandão e Dalmo Correia. Após intensa e profícua discussão, ficou oficialmente comprometido entre os pares o início da implantação progressiva de medidas relacionadas à Ciência Aberta, conforme cronograma proposto pela equipe SciELO.

2. THE JOURNAL OF VENOMOUS ANIMALS AND TOXINS INCLUDING TROPICAL DISEASES

O *Journal of Venomous Animals and Toxins including Tropical Diseases* (JVATiTD) é a publicação oficial do Centro de Estudos de Venenos e Animais Peçonhentos da Universidade Estadual Paulista (CEVAP-UNESP). Publicada ininterruptamente desde 1995, tem desde então inovado e disseminado conhecimentos sobre Toxinologia e Doenças tropicais, infecciosas e parasitárias. Barraviera (2007, 2013, 2015) ressalta que a JVATiTD é uma das três principais revistas do mundo que publicam artigos sobre Toxinologia, tornando-se, portanto, um dos principais canais que contribuem para a disseminação de informações sobre essa temática.

Martins *et al.* (2018), por sua vez, apresentam a trajetória da primeira revista digital brasileira, com distribuição inicialmente em disquetes de 3.5", numa época em que se discutia se o suporte digital iria superar o impresso, além de não haver internet comercial disponível no Brasil. Entre 2000 e 2002, a revista foi distribuída em Compact Disc (CD-Rom) e, a partir de 2003, abriu um site na internet www.jvat.com.br, mudando novamente o suporte que a mantém até os dias de hoje.

Desde que a plataforma *Scientific Electronic Library Online* (SciELO) foi implantada em 1998, os artigos da JVATiTD foram os primeiros digitais a serem selecionados para participar da base. Quando de seu lançamento, o JVATiTD adotou os princípios do movimento de acesso aberto, ratificando a posição de vanguarda e disponibilizando gratuitamente os textos integrais dos artigos publicados.

Outro ponto de destaque é a publicação em fluxo contínuo, que foi adotada a partir de 2013. Com isso, ganhou velocidade, flexibilidade e rapidez na publicação. Entre 2013 e 2018, por participar de um *publisher* de renome internacional que cobrava *Article Processing Charge* (APC) para publicar, o JVATiTD pode ter sido considerado uma "revista comercial" (SIMIONATO *et al.*, 2018). Como originalmente não nasceu comercial e nunca visou lucros, em 2019 adotou o modelo de negócio proposto pelo *publisher* SciELO, que se baseia na cobrança de *Article Processing Charge* (APC) necessário e suficiente para manter a sustentabilidade financeira do periódico (CASSARO *et al.*, 2020). Desde a sua criação, a JVATiTD tem sido pioneira em vários paradigmas que foram mudados progressivamente, conforme as necessidades e possibilidades – ou seja: do disquete à Internet sem interrupção (MARTINS *et al.*, 2018). O vanguardismo

da publicação se reflete na posição privilegiada conquistada em 2018 pelo seu Fator de Impacto (FI) de 2.935, publicado pelo *InCites Journal Citation Reports* da *Clarivate Analytics*, pelos Cites/Doc de 2.973 e CiteScore, de 2.630 publicados pelo *Scimago Journal & Country Rank* da Scimago-Scopus e pela participação em indexadores estratégicos e de renome, tais como SciELO, PubMed Central, Europe PubMed Central, Web of Science e Scimago-Scopus (DOMINGUES *et al.*, 2020). Hoje temos orgulho de dizer que o JVATiTD é uma revista acadêmica, não comercial, de acesso aberto, de inserção internacional, de elevado fator de impacto e que estará adotando brevemente, de forma pioneira, “as práticas da Ciência Aberta”.

3. MEMÓRIAS DO INSTITUTO OSWALDO CRUZ

A revista Memórias do Instituto Oswaldo Cruz foi o primeiro dos sete periódicos editados pela Fundação Oswaldo Cruz (Fiocruz) a entrar em circulação. Oficialmente, a sua criação ocorreu em 1907¹³¹, mas a primeira edição da revista foi publicada somente em 1909. A revista MIOC foi uma das primeiras do Brasil a dedicar-se exclusivamente à divulgação dos resultados de pesquisa científica, e em 2019 completou 110 anos de circulação praticamente ininterrupta (entre 1976-1980 não houve circulação). Silva e Lima (2015) relatam que, desde a criação da revista, seus artigos eram publicados em versão bilíngue, apresentando uma preocupação com a internacionalização do conteúdo a fim de dar visibilidade à produção científica desenvolvida no instituto para outros países.

Um grande destaque da MIOC foi a sua segunda edição, a qual publicou um extraordinário artigo do Dr. Carlos Chagas relatando as descobertas relacionadas à doença de Chagas, incluindo o seu agente causador, o protozoário *Trypanosoma cruzi*, e o respectivo inseto vetor, o popular “barbeiro”. Este estudo tem grande importância na medicina tropical do século XX, rendendo ao Dr. Chagas inúmeros prêmios e reconhecimento internacional. A versão completa do artigo está disponível na página web da Memórias do Instituto Oswaldo Cruz, no seu formato original em português e alemão.

A Memórias do Instituto Oswaldo Cruz (MIOC) é uma revista de acesso aberto, que oferece publicação de artigos sem custo para os autores e livre acesso

131 <http://www2.camara.leg.br/legin/fed/decret/1900-1909/decreto-1802-12-dezembro-1907-582504-publicacaooriginal-105302-pl.html>

aos leitores. A revista publica artigos de pesquisa científica original nas grandes áreas da medicina tropical, parasitologia humana e microbiologia médica. Complementarmente, publica resultados de pesquisa em áreas transversais, como bioquímica, imunologia, biologia molecular e celular, genética/genômica, todas relacionadas às grandes áreas do escopo da revista.

O acervo completo, contendo artigos com mais de 100 anos de publicação, está disponível no site da revista. A MIOC é um dos periódicos mais citados na América Latina e é reconhecido internacionalmente como uma das principais revistas nas áreas de Medicina Tropical e Parasitologia.

A MIOC hoje pode ser melhor definida pelo seu subtítulo em inglês: “An international journal of microbes and their vectors causing human infections”.

A revista iniciou o processo de publicação contínua em 2016 e eventualmente divulga “séries de artigos temáticos”, de modo que, em todos os casos, os manuscritos submetidos são analisados por meio da revisão pelos pares. A Memórias optou por este formato de publicação buscando maximizar o uso das suas potencialidades sem a necessidade de aguardar a composição completa dos fascículos ou de edições seriadas. Essa modalidade promove rapidez no processo de comunicação e disponibiliza os resultados das pesquisas de maneira mais ágil, com inúmeras vantagens para se usufruir da informação científica.

Ao longo do último século a Memórias do IOC vem se adaptando aos desafios e transformações do ambiente editorial científico. Um exemplo de resposta aos desafios contemporâneos foi a implementação da publicação rápida como forma de contribuir no enfrentamento de epidemias que demandam agilidade na disseminação de resultados de pesquisa. Assim, foi instituído pelos editores da revista o formato de publicação Fast Trak para trabalhos com os vírus da Zika, Chikungunya, Febre Amarela e Ebola. Após avaliação dos editores, a revista disponibiliza imediatamente em sua página web os artigos científicos sobre os temas, e em seguida os envia para o processo de avaliação pelos pares. Assim, os leitores interessados poderão rapidamente se informar sobre os últimos resultados da pesquisa científica com os agentes e seus vetores. Esse método de pré-publicação se assemelha ao processo de divulgação imediata conhecido como “preprint”, praticado desde 2013 pelo repositório Biorxiv e que aos poucos vem alterando a forma como os resultados da pesquisa biomédica

é divulgada ao público. A MIOC aderiu à publicação de artigos já depositados em repositórios “pre-print” em 2017¹³².

A estrutura editorial atual da revista Memórias é composta de um editor-chefe e dois editores científicos, além de um número variável de editores associados, selecionados entre pesquisadores brasileiros e estrangeiros de áreas correlatas ao escopo da revista. O corpo de editores associados é atualizado a cada dois anos, auxiliando na apropriação de novas práticas e conhecimentos editoriais, além da manutenção da qualidade do processo editorial. A política de publicação está disponível no site da revista, incluindo o escopo, a exigência do uso do inglês como único idioma de publicação, o fluxo de publicação, as diretrizes éticas e de integridade em pesquisa, com alguns links de orientação da Rede Equator. A revista é membro do Comitê de Ética em Publicação (COPE - Committee on Publication Ethics) e busca observar as recomendações do COPE em suas publicações. Ao submeterem seus artigos, os autores devem declarar o conhecimento dos critérios de boas práticas na investigação e publicação científicas recomendados pela Declaração de Singapura.

A taxa de crescimento anual das revistas científicas acelerou nos últimos anos, o que significou uma enorme pressão sobre o processo de publicação científica, sobretudo na revisão pelos pares. A revista MIOC se ajustou ao cenário, aprimorando seu fluxo editorial no intuito de priorizar cada vez mais o seu escopo por meio de ações que a tornem mais alinhada às expectativas de qualidade/relevância de autores e leitores da sua área de atuação. Dessa forma, o processo de análise preliminar dos artigos submetidos se torna cada vez mais sistemático e rigoroso, e como consequência o índice de rejeição/devolução de manuscritos aumentou consideravelmente.

A revista encerrou sua edição impressa em 2015, passando a um formato exclusivamente digital, de acesso aberto e gratuito, adaptando-se às novas rotinas de busca e leitura imediata dos leitores. Lançou-se nas redes sociais e possui contas no Facebook (@MemoriasIOC), Twitter (@MemoriasJournal) e Pinterest (br.pinterest.com/MemoriasJournal). Como parte do corpo editorial, a MIOC dispõe de uma profissional do jornalismo científico como editora associada, a qual é responsável pelo conteúdo das mídias sociais, abordando temas institucionais, infográficos, vídeos (um minuto com o autor) e destacando artigos com maior relevância e inovação. A divulgação da revista também é feita por

132 <https://memorias.ioc.fiocruz.br/recent-posts/item/38-memorias-do-instituto-oswaldo-cruz-welcomes-the-preprints>

meio de e-mail marketing. São elaboradas *newsletters* com conteúdo e temas dos artigos lançados pelas Memórias. Esses dados também são divulgados no Portal de Periódicos da Fiocruz.

A MIOC, em consonância com a Declaration on Research Assessment (DORA), compartilha do compromisso de abrandar o excesso de dependência do Fator de Impacto (FI) e de implementar novos métodos de avaliar a produção científica. Como o FI não reflete exclusivamente a qualidade do conteúdo publicado pela revista, a MIOC enfatiza a prática de outras métricas (alternativas como Altmetrics) que consistem nas menções/referências a artigos MIOC, visitas aos links, cobertura em canais de mídia social, posts em blogs científicos e tweets, medição de compartilhamento, curtidas e downloads dos artigos, atribuindo peso a cada uma dessas ações.

A Memórias do Instituto Oswaldo Cruz está presente no SciELO desde o início e, agora, atende ao desafio de adotar as práticas do movimento da Ciência Aberta, com a tradição de mais de 100 anos e inovação dos tempos atuais.

4. REVISTA DA SOCIEDADE BRASILEIRA DE MEDICINA TROPICAL

A Revista da Sociedade Brasileira de Medicina Tropical é um periódico multidisciplinar, com acesso 100% aberto, com revisão pelos pares, que publica pesquisas originais em todas as áreas da Medicina Tropical (incluindo epidemiologia, estudo clínicos, patologia e imunologia) e doenças infecciosas. É o periódico oficial da Sociedade Brasileira de Medicina Tropical. A revista publica Artigos Originais, Comunicações Breves, Relatos de Caso, Editoriais, Cartas ao Editor, Imagens em Doenças Infecciosas e Parasitárias, Relatórios Técnicos, Obituários e Números Especiais (suplementos). Os artigos de revisão são a convite do editor ou editores associados. A revista publica manuscritos somente em inglês. O seu conteúdo tem acesso livre para os leitores e nenhuma taxa é cobrada dos autores.

A RSBMT faz parte da plataforma *Scientific Electronic Library Online* (SciELO) desde 1999 e toda sua coleção, desde 1967 até janeiro de 2020, está disponível on-line na biblioteca SciELO. A partir de janeiro de 2019, a RSBMT adotou a modalidade de publicação anual contínua, e seu site na internet é <http://www.scielo.br/rsbmt>. Está indexada nos principais indexadores, tais como Journal

Citation Reports-JCR (Thomson Reuters), SciELO, PUBMED, Web of Science e Scimago/Scopus.

Seu fator de impacto vem apresentando elevação consistente desde 2010. Atualmente a RSBMT, em nível internacional e na área de Medicina Tropical com acesso aberto, ocupa o 8º lugar na área de Parasitologia; com acesso aberto ocupa o 9º lugar; e na área de Doenças Infecciosas ocupa o 26º lugar. Em nível nacional, na área de Medicina Tropical com acesso aberto, ocupa o 3º lugar; e na área de Doenças Infecciosas ocupa o 4º lugar, especificamente na classificação dos periódicos (Web of Science/Thomson Reuters, 2018).

Teixeira, Botelho e Petroianu (2013) revelam que as citações da revista têm crescido nos últimos anos, com média maior que outras similares, num modelo a ser seguido por outras, em que o aumento das citações possibilita a indexação por outras bases e o crescimento da visibilidade. Assim, a RSBMT contribui para a transparência da ciência brasileira, visto que Alves e do Rosario (2015) destacam a revista pela produção sobre malária, bem como Sobral (2017) revela o destaque em publicações sobre hanseníase. Outro ponto relevante podia ser visto pela avaliação Qualis da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), na qual a RSBMT aparece também em estratos da geografia, psicologia, ciências ambientais, zootecnia, entre outros.

Como canal oficial da Sociedade Brasileira de Medicina Tropical, a RSBMT adota políticas de acesso aberto, democratizando a medicina tropical e dando-lhe prestígio internacional, tanto que ela é indexada pelo Web of Science e SciELO, entre outros. A visibilidade da revista pode ser verificada pelo seu Fator de Impacto, que a coloca com a sexta posição mundial entre as revistas que tratam de medicina tropical e umas das principais em saúde no país.

Sendo uma revista acadêmica não comercial, de acesso 100% aberto, com visibilidade internacional, 100% gratuita que prioriza a publicação de pesquisas envolvendo doenças negligenciadas, foi convidada a fazer parte do estudo piloto do SciELO para implementação de práticas da “Ciência Aberta”, objetivando aumento da transparência do processo de comunicação, preservação, replicabilidade e reusabilidade dos dados e aceleração da divulgação dos resultados das pesquisas, inovando assim nas práticas editoriais e no conceito de mecanismo de “pos-print” de artigos publicados. Logo incluiu em suas normas, a partir de setembro de 2019, a modalidade de avaliação de artigos no modo “fast-track” e manuscritos depositados em alguns servidores “preprints”.

5. ALINHAMENTO ÀS PRÁTICAS EDITORIAIS DA CIÊNCIA ABERTA

Desde a sua criação em 1998, a biblioteca SciELO tem sido norteadora da editoração científica, inovando sempre e sendo pioneira ao estimular o acesso aberto dos periódicos nacionais. Com o surgimento da Ciência Aberta, novamente como precursora, a SciELO atua para a disseminação e implantação dos novos preceitos. Packer e dos Santos (2019) revelam que a SciELO já projetou estratégias de ação para os próximos cinco anos, visando implementar práticas da comunicação em Ciência Aberta de forma gradual e progressiva.

Neste caminho deu-se a implantação de um projeto piloto, para o qual foram convidados três importantes periódicos nas áreas de Medicina Tropical/Doenças Infecciosas e Parasitárias, de modo que seus editores aceitaram prontamente (BRANDÃO, BARRAVIERA e CORREIA, 2019). Assim, alguns passos iniciais podem ser adotados pela revista, na intenção de atender a premissas da Ciência Aberta, tais como:

- Alteração das políticas editoriais da revista, com explicação sobre a adesão às práticas da Ciência Aberta, a fim de apresentar as premissas nas quais ela repousa;
- Aceitação de manuscritos depositados em servidores de *pre-prints*;
- Mudança progressiva do modelo de avaliação cega pela avaliação aberta, alterando o processo de avaliação por pares;
- Requerimento aos autores do depósito dos dados brutos de pesquisas, assegurado pelas questões legais como parte do processo de submissão dos artigos;
- Criação de mecanismos *post-print* nos artigos publicados pelos periódicos.

Com a adoção dessas ações procura-se atender aos anseios de mudanças ocorridas nas ciências por maior transparência, equidade e rapidez ante a evolução das tecnologias que imprimem às atividades de pesquisa um grande dinamismo. Tornar transparente o processo editorial tem sido uma necessidade, pois, como destaca Vasconcellos (2017), **o editor não é um ditador e a transparência e isonomia deve ser a tônica do processo editorial científico.**

Outro ponto a ser observado é a questão da reprodutibilidade dos resultados de pesquisa. A ciência é realizada por métodos que devem possibilitar a replicabilidade das ações para garantir os mesmos resultados. Assim, disponibilizar os dados brutos de pesquisa assegura a sua preservação em repositórios de dados ou outros sistemas, além de promover o seu reúso para novas análises, cruzamentos ou mesmo agregação de outros dados. Mais que isso, garante a confiabilidade da pesquisa.

Ações como publicação de *pre-print*, publicação em fluxo contínuo, *fast track* e outros aceleram a publicação de resultados de pesquisa, promovendo o dinamismo que a sociedade requer, com o apoio da tecnologia vigente. Portanto, atende às necessidades dos autores e leitores, cada vez mais ávidos por novas informações que respondam as dúvidas de suas pesquisas nos processos de disseminação de resultados e produção de novos conhecimentos.

Nesse contexto, os editores do *Journal of Venomous Animals and Toxins including Tropical Diseases*, *Memórias do Instituto Oswaldo Cruz* e *Revista da Sociedade Brasileira de Medicina Tropical* – em parceria e consonância com as políticas editoriais do SciELO e da Associação Brasileira de Editores Científicos (ABEC) –, com o intuito maior de contribuir para a evolução da editoração científica nacional, ratificaram as ações e se dispuseram a dar início ao novo desafio.

6. CONSIDERAÇÕES FINAIS

O movimento do acesso aberto se baseia principalmente no tripé: acesso aberto e irrestrito aos dados da pesquisa; maior rapidez e visibilidade na comunicação dos artigos; e transparência e abertura do processo de avaliação. O grande desafio na implantação dessas ações está relacionado principalmente à mudança comportamental dos pesquisadores. A abertura do processo de avaliação, publicando os pareceres juntamente com o artigo, dará maior responsabilidade aos pareceristas em aceitar ou não a participação no processo de avaliação, além de moderar e elevar o nível do linguajar na aceitação ou rejeição do artigo científico. Hoje isto é realizado sem tais preocupações, que acabam muitas vezes comprometendo a decisão final dos editores, ou até desconsiderando determinadas avaliações, tendo em vista os evidentes conflitos de interesse. Outra ação importante nesse contexto é a reusabilidade dos dados, muito comum nas revisões sistemáticas com aplicação de metanálise. A reprodutibilidade é uma questão ímpar e fundamental no processo.

BAKER (2016) publicou, na revista *Nature*, um artigo sobre **A crise da reprodutibilidade da ciência mundial**. Foram convidados 1.576 pesquisadores mundo afora e, por meio de um questionário on-line, eles avaliaram a reprodutibilidade da pesquisa científica. Para surpresa do autor, mais de 70% dos pesquisadores tentaram e falharam em reproduzir experimentos de outro cientista, e muito mais da metade não conseguiu reproduzir seus próprios experimentos. Assim, na biologia, 70% da ciência produzida na “Fronteira do Conhecimento” – também denominada “Especulativa” – não é reprodutível; já na medicina, os resultados atingem a marca dos 50%.

LITHGOW (et al., 2018), por sua vez, publicam o seguinte artigo: **Uma longa jornada em direção a resultados reprodutíveis...** Os autores mostram que, após quase um ano de experimentos frustrantes, um colega e seus colaboradores estavam prestes a publicar um artigo descrevendo a incapacidade de sua equipe em reproduzir um artigo de alto nível desenvolvido no próprio laboratório. Seria o primeiro da literatura a mostrar claramente que uma molécula semelhante a uma droga poderia prolongar a vida do ser humano. Os autores concluem: **Se quisermos entender como a vida funciona, descrever detalhadamente a PUBLICAÇÃO é tão importante quanto descrever o que foi observado**. A publicação ratifica a pesquisa de BAKER, realizada em 2016, e abre uma discussão enorme sobre a preferência das revistas científicas em publicar sempre “o que há de novo” e “este artigo será publicado?”. Os editores nunca perguntam: “Esta pesquisa será reprodutível?”.

Dessas observações resta a pergunta: **A Ciência Aberta poderia diminuir tais dificuldades, aumentar o nível de reprodutibilidade da ciência de fronteira e otimizar recursos financeiros investidos na pesquisa?**

Outro paradigma importante a ser vencido na área biológica é o depósito de *pre-prints* em servidores públicos estratégicos. Esse tipo de publicação é uma versão completa do artigo, recebendo inclusive *DOI number* e tornando-se, por consequência, passível de ser citado, com possibilidade posterior de submissão em um periódico para publicação definitiva. Os *pre-prints* contribuiriam para aumentar a velocidade, a visibilidade e preservar rapidamente o ineditismo dos autores e dos resultados da pesquisa científica. A questão é: **os pesquisadores da área biológica aceitarão esse tipo de oportunidade?**

Por fim, a questão da transparência e abertura do processo de avaliação. Esse talvez seja o item mais difícil a ser vencido no âmbito da Ciência Aberta, uma vez que o ser humano não se comporta como uma máquina. Ele tem sentimentos,

capacidades e a complexidade de pensar, de exercer liderança, de ter sensibilidade e generosidade, entre tantos outros atributos. A questão novamente é: **o parecerista conseguiria ser imparcial frente a um conflito de interesse quer seja positivo, quer seja negativo?** É uma resposta difícil de ser obtida. Portanto, a partir da implantação do novo paradigma, acreditamos que será ainda mais difícil encontrar pareceristas que opinem sobre determinado artigo científico conhecendo seus autores, haja vista que até hoje estes profissionais, que são escolhidos pelos editores de acordo com a competência meritocrática, trabalham voluntariamente para os periódicos científicos.

Concluindo: a adoção das práticas da Ciência Aberta às revistas científicas brasileiras certamente contribuirá para alçar um novo patamar de modernidade no processo editorial, atendendo às necessidades atuais da comunidade envolvida nas pesquisas científicas. Assim, a iniciativa de três revistas brasileiras de maior impacto nas áreas da Medicina Tropical e Doenças Infecciosas e Parasitárias adotarem preceitos da Ciência Aberta revela a maturidade desses periódicos em face dos desafios impostos pela comunidade internacional, iniciando-se então uma nova etapa na história do maior parque editorial de revistas acadêmicas, não comerciais e de acesso aberto do mundo, como demonstraram Domingues *et al.* (2020) em recente publicação.

7. REFERÊNCIAS

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

ALVES, A. C.; DO ROSÁRIO, V. E. Bibliografia sobre a malária, com especial relevância para os Países de Língua oficial Portuguesa. **Anais do Instituto de Higiene e Medicina Tropical**, v. 14, p. 125-127, 2015.

BAKER, M. 1,500 scientists lift the lid on reproducibility. **Nature**, v. 533, n. 7604, p. 452-454, 2016. DOI: 10.1038/533452a. Disponível em: <http://www.nature.com/news/1-500-scientists-lift-the-lid-on-reproducibility-1.19970>. Acesso em: 16 jan. 2020.

BARRAVIERA, B. CEVAP journal towards a new phase. **Journal of venomous animals and toxins including tropical diseases**, v. 19, n. 1, 2013. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3707102/>. Acesso em: 16 jan. 2020.

BARRAVIERA, B. CEVAP Journal: the first Brazilian electronic scientific publication turns 20 years old. **Journal of venomous animals and toxins including tropical diseases**, v. 21, n. 1, dez. 2015. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4673739/>. Acesso em: 16 jan. 2020.

BARRAVIERA, B. The Journal of Venomous Animals and Toxins including Tropical Diseases (JVATiTD) from 1995 to 2007. **Journal of venomous animals and toxins including tropical diseases**, v. 13, n. 2, p. 428-429, 2007. Disponível em: http://www.scielo.br/scielo.php?script=sci_issuetoc&pid=1678-919920070002&lng=en&nrm=iso. Acesso em: 16 jan. 2020.

BRANDÃO, A.A., BARRAVIERA, B.; CORREIA, D. Periódicos de Medicina Tropical/Doenças Infecciosas e Parasitárias alinham-se às práticas editoriais da Ciência Aberta [online]. **SciELO em Perspectiva**, 2019. Disponível em: <https://blog.scielo.org/blog/2019/10/23/periodicos-de-medicina-tropical-doencas-infecciosas-e-parasitarias-alinham-se-as-praticas-editoriais-da-ciencia-aberta/#.XjMvmWhKjIU>. Acesso em: 16 jan. 2020.

CASSARO C. V. *et al.* Os desafios enfrentados pela equipe editorial do Journal of Venomous Animals and Toxins including Tropical Diseases na transição entre dois publishers de renome internacional. **Ciência da Informação em Revista**. 2020. [no prelo]

DOMINGUES, R. S. *et al.* Evolução histórica do Fator de Impacto (FI) na base Web of Science (WoS) dos periódicos do Brasil entre 2008 e 2018. **Ciência da Informação em Revista**. 2020. [no prelo].

FERREIRA JR, R. S. Prefácio. *In*: SHINTAKU, M.; SALES, L. F. (org.) **Ciência Aberta para editores científicos**. Botucatu, SP: ABEC, 2020.p. 7-8. DOI: 10.21452/978-85-93910-02-9.

LITHGOW, G. J., DRISCOLL, M.; PHILLIPS P. A long journey to reproducible results. **Nature**, v. 548, n. 7668, p. 687-388, 2017. DOI: <https://dx.doi.org/10.1038/548387a>.

MARTINS, S. de L. P. *et al.* Do disquete às nuvens: a saga da primeira revista eletrônica científica brasileira. **Ciência da Informação em Revista**. v. 5, n. esp., p. 86-100, 2018.

PACKER, A. L.; SANTOS, S. Ciência Aberta e o novo modus operandi de comunicar pesquisa – Parte I [online]. **SciELO em Perspectiva**, 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: 16 jan. 2020.

SILVA, A. P. da; LIMA, M. H. T. de F. Revista "Memórias do Instituto Oswaldo Cruz": uma visão bibliométrica. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 16., 2015, João Pessoa. **Anais** [...]. João Pessoa: UFPB, 2015.

SIMIONATO, J. S. *et al.* Como aumentar o fator de impacto nas bases Web of Science (WoS) e Scopus (Scimago): ações implementadas pelo Journal of Venomous Animals and Toxins including Tropical Diseases. **Ciência da Informação em Revista**. v. 5, n. esp., p. 58-67, 2018.

TEIXEIRA, R. K. C.; BOTELHO, N. M.; PETROIANU, A. Referências de periódicos médicos brasileiros em publicações nacionais. **Revista da Associação Médica Brasileira**, v. 59, n. 6, p. 571-575, 2013.

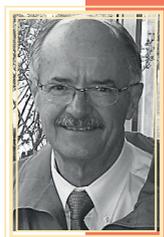
VASCONCELLOS, V. G. de. A função do periódico científico e do editor para a produção do conhecimento no Direito e nas ciências criminais. **Revista Brasileira de Direito Processual Penal**, v. 3, n. 1, p. 9-17, 2017.

8. SOBRE OS AUTORES

Benedito Barraviera

Lattes: <http://lattes.cnpq.br/6840524602748457>

Orcid: <https://orcid.org/0000-0002-9855-5594>



Graduado em Medicina pela Unesp, especialista em Medicina Tropical pela FMUSP, Mestre e Doutor pela Unesp. Professor Livre-Doente e Titular de Infectologia na Unesp. É Pesquisador do Centro de Estudos de Venenos e Animais Peçonhentos da Unesp (Cevap), Editor-chefe do The Journal of Venomous Animals and Toxins including Tropical Diseases (<http://www.jvat.org>) e Professor Titular de Moléstias Infeciosas e Parasitárias da Unesp. Tem experiência em Toxinologia, Medicina Tropical, Pesquisa Clínica e Editoração científica. Foi fundador e diretor do Cevap, Pró-Reitor de Extensão Universitária e Presidente da ABEC. A partir de 2005 tornou-se coordenador do Grupo de Pesquisa em Toxinologia certificado pela Unesp e CNPq. A partir de 2014 aprovou e tornou-se Vice coordenador do National Institute of Science and Technology of Wildlife Science Center (WSC), INCT, CNPq Proc.No. 465429/2014-9. Atualmente coordena dois ensaios clínicos fase I/II. Em 2019 foi indicado e empossado como Membro Titular da Academia de Ciências Farmacêuticas do Brasil. Atualmente coordena dois ensaios clínicos fase I/II sobre o biopolímero de fibrina e o soro antiapifílico.



Adeilton Alves Brandão

Lattes: <http://lattes.cnpq.br/0523061502973878>

Orcid: <https://orcid.org/0000-0001-5877-607X>

Graduado em Farmácia, Universidade Federal de Ouro Preto-MG (1990), Mestre em Biologia Parasitária, Instituto Oswaldo Cruz (1994), Doutor em Ciências, Universidade Federal do Rio de Janeiro (1999) e Certificado de Educação Executiva Strategic Leadership, University of British Columbia, Vancouver, Canada (2013). Atualmente é pesquisador do Instituto Oswaldo Cruz-Fiocruz (biologia molecular de protozoários), e é o Editor-chefe da revista "Memórias do Instituto Oswaldo Cruz".



Dalmo Correia

Lattes: <http://lattes.cnpq.br/0060518181240457>

Orcid: <https://orcid.org/0000-0002-2174-5058>

É professor titular da disciplina de Doenças Infecciosas e Parasitárias da Universidade Federal do Triângulo Mineiro / UFTM. Atua no curso de Pós-Graduação stricto sensu em Medicina Tropical e Infectologia da UFTM, como docente e orientando alunos desde 2000 e tem experiência na área de Medicina Interna, com ênfase em Infectologia (HIV/AIDS e Infecção Hospitalar) e Medicina Tropical, atuando principalmente nos seguintes temas: doença de Chagas, Co-infecção T.cruzi/HIV, Análise da Variabilidade da Frequência Cardíaca na doença de Chagas, Epidemiologia, Clínica, Imunopatogenia e Tratamento das doenças de Chagas e Leishmanioses. Sua experiência na editoração de periódicos científicos inclui a ocupação como Editor Associado da Revista da Sociedade Brasileira de Medicina Tropical de 2000 a 2011 e como Editor Chefe desde agosto de 2012.



Rui Seabra Ferreira Junior

Lattes: <http://lattes.cnpq.br/5222987608775248>

ORCID: <https://orcid.org/0000-0001-6952-0512>

Presidente da ABEC; Editor Associado do Journal Venomous Animals and Toxins including Tropical Diseases; Pesquisador Associado do CEVAP-UNESP; Prof. Livre-Docente da FMB-UNESP; Pesquisador Pq1C do CNPq; Membro da Academia Brasileira de Ciências Farmacêuticas do Brasil/Academia Nacional de Farmácia.



Milton Shintaku

Lattes: <http://lattes.cnpq.br/8605833104600600>

Orcid: <https://orcid.org/0000-0002-6476-4953>

Mestre e doutor em Ciência da Informação pela Universidade de Brasília (UnB), coordenador de Articulação, Geração e Aplicação de Tecnologia do Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict), Secretário da Associação Brasileira de Editores Científicos (ABEC) 2017-2019.

9. COMO CITAR

BARRAVIERA, B. *et al.* O Alinhamento das revistas brasileiras sobre Medicina Tropical e Doenças infecciosas e Parasitárias às práticas da Ciência Aberta. *In*: SHINTAKU, M.; SALES, L. F; COSTA, M. (org). **Tópicos sobre dados abertos para editores científicos**. Botucatu, SP: ABEC, 2020. p. 215-230. DOI: 10.21452/978-85-93910-04-3.cap17.

Política Editorial e os Dados Abertos

Milton Shintaku
Rui Seabra Ferreira Junior

1. INTRODUÇÃO

A *Scientific Library Electronic Online* (SciELO) foi um marco para o movimento de acesso aberto no país, convidando revistas científicas a disponibilizar artigos livremente na internet, ainda na década de 1990. Assim, a SciELO democratiza o conhecimento científico, contribuindo para a visibilidade da ciência nacional, na medida em que oferta artigos sem barreiras de acesso, alinhado às mais novas tendências na publicação científica da época. Com isso, alterou significativamente o processo editorial científico brasileiro, se tornando símbolo de qualidade em publicação científica.

Nesse caminho, a SciELO inova outra vez, apostando nos preceitos da Ciência Aberta, que são alinhados às novas tendências mundiais. Diante disso, em 2016 a SciELO já aconselhava as revistas indexadas por ela a acelerarem a publicação de resultados de pesquisa, como um alinhamento à Ciência Aberta (PARKER *et al.*, 2016). A partir de então, tem trabalhado gradativamente para que as revistas brasileiras adotem políticas editoriais mais abertas, alinhadas às tendências mundiais. Nesse sentido, busca dar maior transparência aos processos editoriais com adoção de práticas como a avaliação aberta em alternativa à duplamente cega, apresentando assim uma grande mudança, trazendo desafios e vantagens. A adoção de práticas que oferecem maior transparência ao processo editorial, como a avaliação aberta e outros, é um desafio a todos envolvidos, pois traz desconfiças dos autores e pareceristas, além de novas atividades aos editores e, assim, requer ações proativas que produzam um ambiente favorável a sua implantação (SHINTAKU *et al.* 2019).

Outro ponto de destaque nas ações da SciELO se refere à promoção da adoção de práticas relacionadas aos dados abertos, na qual os dados de pesquisa devem ser disponibilizados para reuso, com depósito em repositório de dados para a sua preservação. Esse ponto requer grandes ações por parte dos autores e editores, relacionados a direitos autorais, ética, infraestrutura e outros, merecendo cuidados na sua adoção.

O compartilhamento de dados, no entanto, ainda não é unanimidade entre os pesquisadores, dadas as diferenças que existem entre as disciplinas, além de questões éticas e de segurança. Chauvette, Schick-Makaroff e Molzahn (2019) defendem o amparo de questões epistemológicas, metodológicas e éticas para abertura dos dados, em que deve-se atender a todos os seus critérios no intuito de ter segurança na disponibilização dos dados de pesquisa. A epistemologia direciona a pesquisa, influenciando o seu método, e conseqüentemente os dados. Dados qualitativos, por exemplo, possuem singularidades que requerem atenções diferentes dos quantitativos. Wicherts, Bakker e Molenaar (2011), por exemplo, relatam certo receio dos pesquisadores em disponibilizar os dados e outros estudiosos os reanalisarem e encontrarem discrepâncias nos resultados, apresentando conclusões diferentes do trabalho original dos dados, relutando assim na adesão às políticas de dados abertos das revistas.

Assim como a SciELO, a Associação Brasileira de Editores Científicos (ABEC) vem atuando de forma bastante proativa nos últimos dez anos com o objetivo de divulgar as novas tendências internacionais relacionadas à comunicação científica, principalmente no que diz respeito às atividades relacionadas à editoração científica. Com isso, procura promover a Ciência Aberta, fomentando o desenvolvimento dos editores científicos e de seus *staffs* nos seus preceitos, com vistas a facilitar a adesão das revistas não apenas nos cursos e capacitações ofertados nos eventos da associação, mas em publicações, postagens e discussões que visam atender às necessidades dos editores quanto ao enfrentamento dos novos desafios.

2. POLÍTICAS EDITORIAIS COM DADOS ABERTOS

Política editorial, no sentido mais amplo, remete à questão do que e como a informação é posta a público. Restringindo a editoração científica, política editorial pode ser entendida com um conjunto de diretrizes que orientam as atividades

dos envolvidos na disseminação da informação científica, extrapolando as atividades do editor, mesmo que seja o ator principal deste processo. Envolve as estratégias voltadas a responder questões básicas relacionadas ao que publicar, como publicar, em que canal publicar, para quem publicar, entre outros.

Assim, um caminho para ver como as revistas estão adotando os dados abertos é ver as suas políticas editoriais, na medida em que elas refletem a orientação da revista. Políticas editoriais não são eternas e requerem mudanças constantes conforme mudam-se os cenários da área de atuação da revista. Muitas vezes isto passa por fazerem o uso de casos de sucesso implementados em outros periódicos, visto que a ciência é única no mundo, mesmo com as diferenças disciplinares.

Como exemplo, observamos a revista *PLOS One*, que requer a assinatura por parte dos autores na submissão, de uma declaração de estar em conformidade com as suas políticas de dados abertos¹³³, de modo que, se o trabalho for aceito, a declaração é publicada junto ao artigo, no intuito de promover a transparência e o progresso da ciência. Nesse ponto, a *PLOS One* requer o compartilhamento do conjunto mínimo de dados, com os seus metadados, obedecendo a padrões de preparação, registro e deposição de dados, quando aplicáveis. Assim, destacam-se os seguintes pontos:

- **Definição do conjunto mínimo de dados:** A *PLOS One* entende que um conjunto mínimo de dados é capaz de ofertar o necessário para a replicação dos resultados apresentados no artigo, como os dados brutos que geraram médias e outros padrões estatísticos, gráficos ou mesmo dados de análise de imagens, não sendo necessários todos os dados, mas uma parte significativa e relevante. No entanto, não é preciso enviar os dados para a revista, mas apenas depositar os dados em um repositório de dados e indicar a sua localização.
- **Métodos aceitáveis de compartilhamento de dados:** Os conjuntos de dados devem estar disponíveis em repositório generalista ou de especialidade, com metadados descritivos apropriados, a fim de serem acessados e localizados. Da mesma maneira, os conjuntos de dados precisam de identificação, a qual deve ser citada nas referências do artigo, como o DOI ou outro tipo.

133 <https://journals.plos.org/plosone/s/data-availability>

- **Restrições do compartilhamento de dados aceitável:** Em alguns casos, questões éticas ou jurídicas impedem a disponibilização dos dados, mas encoraja-se a compartilhar o máximo de dados permissíveis. Assim, as restrições aceitáveis podem ser:
 - **Dados de terceiros:** Caso o autor utilize dados de terceiros e não possua o direito de compartilhar, deve apresentar o máximo de informações possíveis para que outros pesquisadores possam localizar os detentores dos direitos sobre os conjuntos de dados.
 - **Dados de participantes de pesquisa:** Estudos envolvendo seres humanos requerem considerações éticas e jurídicas maiores, necessitando a passagem por comitê de ética, sendo que em alguns casos os dados dos participantes da pesquisa podem ser descharacterizados. Entretanto, os seguintes casos merecem cuidado:
 - **Dados privativos:** Dados que identifiquem ou comprometam a privacidade de pessoas, instituições, localidades, contatos, datas específicas, entre outros;
 - **Dados qualitativos:** Parte dos dados – como transcrições – pode ser inserida no texto diretamente, com documento aditivo, ou depositado em repositórios, mas deve-se ter cuidado com informações que identifiquem os participantes;
 - **Dados sensíveis:** Não apenas dados de pesquisas em humanos são sensíveis e merecem cuidados. Dados confidenciais sobre espécies ou sítios arqueológicos ameaçados, entre outros, podem requerer restrição de disseminação.
- **Restrições inaceitáveis:** Para as revistas da PLOS, os interesses pessoais não são razão que impeça o compartilhamento dos dados, mesmo envolvendo patentes ou publicações futuras, pois não se justificam.

A Spring Nature¹³⁴, por outro lado, apresenta quatro modelos de políticas diferenciadas voltadas a tipos diferentes de periódicos. Assim, cada revista pode adaptar a política que melhor atenda às suas necessidades de

134 <https://www.springernature.com/gp/authors/research-data-policy/data-policy-types/12327096>

compartilhamento de dados. Esses modelos, de maneira geral, atendem às principais questões a serem seguidas nas políticas de dados abertos, mesmo que de forma superficial, ficando para as revistas os aprofundarem. Os modelos são:

- **O compartilhamento de dados e citação dos dados é encorajado:** Quando possível, a política da revista incentiva os autores a depositarem o conjunto de dados utilizados para as conclusões da pesquisa em repositórios, assim como fazer a citação do conjunto no artigo. Este é o modelo mais brando e pode ser utilizado no início da adoção dos princípios dos dados abertos.
- **O compartilhamento de dados e dados de evidência é encorajado:** A política da revista requer que os dados brutos relevantes e outros materiais utilizados na pesquisa estejam disponíveis para acesso livre e reuso não comercial. Em alguns casos, os dados podem estar no manuscrito ou em arquivo suplementar na submissão. Suas citações podem apresentar apenas dados mínimos sobre os conjuntos. A declaração de disponibilidade de dados deve ser apresentada no artigo, quando possível, por meio de um link.
- **O compartilhamento de dados é incentivado e a declaração de disponibilidade de dados é necessária:** A política da revista requer que os dados brutos relevantes e outros materiais utilizados na pesquisa estejam disponíveis livremente para acesso e para reuso não comercial, desde que não haja problemas de confidencialidade. Para alguns tipos de dados é obrigatório o depósito em repositório específico, da mesma forma que as declarações de disponibilidade dos dados – mesmo quando os dados estão restritos – descrevendo as condições de uso. O artigo deve citar e referenciar os dados utilizados no trabalho, sendo que os revisores são encorajados a verificar a disponibilidade dos dados a fim de validar o cumprimento das políticas.
- **O compartilhamento de dados, dados de evidência e dados de revisão é necessário:** A política da revista exige que os dados utilizados na pesquisa, depositados em repositório ou em arquivo suplementar, estejam disponíveis para acesso público, desde que não tenham empecilho legal ou ético. A declaração de disponibilidade de dados deve apresentá-los, ou, se for restrita, as condições de acesso devem ser expostas. Dados utilizados no trabalho devem ser citados e referenciados no manuscrito. Revisores precisam conferir a declaração de disponibilidade com o objetivo verificar o atendimento à política da revista, assim como considerar se foram feitos

esforços para o compartilhamento dos dados. Este é o modelo mais rígido e pode ser utilizado por revistas com maturidade maior em dados abertos.

A **Elsevier**¹³⁵ apresenta uma política mais simples, mas não menos abrangente no que se refere aos dados abertos. Assim, encoraja o compartilhamento de dados de pesquisa, mantendo o autor no comando do uso e acesso, a fim de valorizar seu trabalho. As diferenças disciplinares devem ser consideradas, bem como a necessidade de curadoria no processo de compartilhamento. Assim, a política da Elsevier visa:

- Encorajar e apoiar o compartilhamento de dados de pesquisa sempre que for possível;
- Padronizar e alinhar as diretrizes no intuito de facilitar a compreensão de onde os dados podem ser depositados e compartilhados para acesso e uso;
- Facilitar o cumprimento das políticas, apoiando declarações de disponibilidade de dados a fim de aumentar a transparência;
- Desenvolver novas ferramentas e serviços voltados ao reuso dos dados;
- Ofertar créditos ao compartilhamento de dados com vistas a incentivar e apoiar práticas de citação de dados;
- Articular, com a comunidade científica, práticas adequadas de revisão de dados no intuito de validar os já publicados;
- Apoiar artigos de dados, revisados pelos pares, fomentando a citação deles por meio dos artigos;
- Apoiar todos os envolvidos na pesquisa, fornecendo infraestrutura de gestão de dados;
- Participar de iniciativas voltadas à descoberta e uso de dados.

135 <https://www.elsevier.com/about/policies/research-data>

Um pouco mais complexa é a política (ou políticas) de compartilhamento de dados apresentada pela **Taylor & Francis**¹³⁶, na qual apresenta cinco níveis a serem utilizados pelos autores no processo de submissão, graduando a abertura dos dados. A Taylor & Francis representa mais de 2.600 revistas por todo o mundo e publica, anualmente, mais de cinco mil livros, sendo uma editora científica respeitada pela mesma comunidade, adotando o acesso aberto por meio de um programa. Assim, as suas políticas de compartilhamento de dados são:

- **Básico (Basic):** Autores são encorajados a compartilhar ou disponibilizar dados e materiais utilizados para realização do trabalho;
- **Compartilhar sob um pedido razoável (Share upon reasonable request):** O autor decide se aceita ou não compartilhar os dados ou outros materiais solicitados;
- **Dados publicamente disponíveis (publicly available data):** O autor publica ou disponibiliza os dados ou materiais utilizados no trabalho, mas decide qual licença de reuso será utilizada;
- **Dados abertos (Open data):** O autor publica ou disponibiliza livremente os dados ou materiais utilizados no trabalho com licença ampla CC BY ou CC0, sem restrição de reuso;
- **Dados abertos e FAIR (Open data and FAIR):** O autor publica, ou disponibiliza livremente, os dados ou materiais utilizados no trabalho, com licença ampla CC BY ou CC0, sem restrição de uso e atendendo aos princípios FAIR (encontrabilidade, acessibilidade, interoperabilidade e reusabilidade).

Em cada uma dessas políticas há um fluxo bem estruturado, apresentando em forma de infográfico as etapas a serem seguidas para uma submissão em revista que usa uma determinada política. Facilita o processo de submissão, na medida em que o autor sabe exatamente o que precisa fazer para atender a uma determinada política de compartilhamento de dados, incluindo detalhes sobre o depósito dos conjuntos de dados, declaração de disponibilidade, citação, revisão pelos pares, entre outros.

Se no mundo as grandes editoras científicas de revistas (*publishers*) já adotam políticas de compartilhamento de dados, no Brasil isto ainda é incipiente. Nesse

136 <https://authorservices.taylorandfrancis.com/data-sharing-policies/>

contexto, conforme Brandão, Barraviera e Correia (2019), a SciELO definiu algumas linhas de ação a serem seguidas para alinhamento às premissas da Ciência Aberta, convidando revistas renomadas a apresentarem projetos-piloto, de modo que três revistas são da área de medicina tropical. Assim, a *Journal of Venomous Animals and Toxins including Tropical Diseases* (JVATiTD), Memórias do Instituto Oswaldo Cruz e Revista da Sociedade Brasileira de Medicina Tropical atenderam ao convite, e gradativamente alteraram suas políticas editoriais para atender, entre outros requisitos, ao compartilhamento de dados e materiais utilizados nos trabalhos. Entretanto, as adequações das políticas editoriais para obedecer às premissas dos dados abertos requerem uma infraestrutura de repositórios de dados de pesquisa que atenda às necessidades das revistas e dos autores.

3. CONSIDERAÇÕES FINAIS

O compartilhamento de dados ou materiais utilizados no trabalho possui questões éticas e jurídicas que merecem cuidados, afetando a política editorial a ser utilizada pela revista. Assim, se algumas pesquisas requerem fortemente a sua análise pelo comitê de ética da instituição em que é executada, outra análise será necessária sobre a disponibilização dos dados resultantes do estudo. Nota-se, então, uma série de mudanças nos procedimentos a serem efetuados durante a pesquisa, com novas etapas a serem executadas. A ética assume o protagonismo na medida em que transcende a execução da pesquisa até o processo de disseminação dos resultados, tornando-se uma preocupação dos editores.

O entendimento, por parte da revista, do que considera como conjuntos de dados ou materiais utilizados no trabalho torna-se primordial para a definição das políticas editoriais de compartilhamento de dados, pois orienta a sua formação. Por exemplo, caso uma pesquisa precise desenvolver um algoritmo para processar os dados, esses programas de computador precisam ser compartilhados? O que pode ser considerado como dado qualitativo? Algumas perguntas precisam ser respondidas pelos editores, dependendo da área de atuação da revista.

Na análise das políticas das grandes editoras científicas, alguns pontos se tornaram evidentes, como a questão do depósito dos dados e materiais e a declaração de disponibilidade, revelando pontos bem apresentados nas políticas. A necessidade de repositórios de dados com identificação unívoca

– como o Digital Object Identifier (DOI), Handle ou outro, disponíveis para o depósito dos dados – é clara. Da mesma maneira, é necessário desenvolver padrões de declaração de disponibilidade com o propósito de facilitar a adoção por editores.

A criação de políticas de dados de pesquisa no Brasil também pode tomar como base as políticas de dados abertos governamentais, que já possui certa maturidade e se diferencia por ter sido implementada por força do Decreto Federal nº 8.777, de 11 de maio de 2016. Com isso, os órgãos federais criaram suas políticas de dados abertos e os depositam em repositórios próprios ou gerais, mantidos pela Controladoria Geral da União (CGU).

Nesse contexto, renova-se a importância da ABEC no processo, visto que os dados abertos não podem mais ser considerados uma tendência, mas uma realidade, cabendo aos editores brasileiros se adaptarem ao novo cenário. Torna-se urgente a discussão com editores científicos sobre a atualização de políticas editoriais, além da necessidade de criação de uma infraestrutura de repositório de dados de pesquisa com todos os requisitos do FAIR.

4. REFERÊNCIAS

BRANDÃO, A.A.; BARRAVIERA, B.; CORREIA, D. Periódicos de Medicina Tropical/Doenças Infecciosas e Parasitárias alinham-se às práticas editoriais da Ciência Aberta [online]. **SciELO em Perspectiva**, 2019 Disponível em: <https://blog.scielo.org/blog/2019/10/23/periodicos-de-medicina-tropical-doencas-infecciosas-e-parasitarias-alinham-se-as-praticas-editoriais-da-ciencia-aberta/>. Acesso em 13 jan. 2020.

CHAUVETTE, A.; SCHICK-MAKAROFF, K.; MOLZAHN, A. E. Open Data in Qualitative Research. **International Journal of Qualitative Methods**, v. 18, p. 1609406918823863, 2019.

PACKER, A. *et al.* Acelerando a comunicação das pesquisas: as ações do SciELO [online]. **SciELO em Perspectiva**, 2016. Disponível em: <https://blog.scielo.org/blog/2016/03/10/acelerando-a-comunicacao-das-pesquisas-as-acoes-do-scielo/>. Acesso em 10 jan. 2020.

SHINTAKU, M. *et al.* Discussão sobre a avaliação aberta, no âmbito da Ciência Aberta. *In: ABEC MEETING, 2019, Fortaleza. Anais [...].* São Paulo: Associação Brasileira de Editores Científicos, 2019. Disponível em: 10.21452/abecmeeting.2019.214. Acesso em 10 jan. 2020.

WICHERTS, J. M.; BAKKER, M.; MOLENAAR, D. Willingness to share research data is related to the strength of the evidence and the quality of reporting of statistical results. *PloS one*, v. 6, n. 11, p. e26828, 2011.

5. SOBRE OS AUTORES



Milton Shintaku

Lattes: <http://lattes.cnpq.br/8605833104600600>

Orcid: <https://orcid.org/0000-0002-6476-4953>

Mestre e doutor em Ciência da Informação pela Universidade de Brasília (UnB), coordenador de Articulação, Geração e Aplicação de Tecnologia do Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict), Secretário da Associação Brasileira de Editores Científicos (ABEC) 2017-2019.



Rui Seabra Ferreira Junior

Lattes: <http://lattes.cnpq.br/5222987608775248>

ORCID: <https://orcid.org/0000-0001-6952-0512>

Presidente da ABEC; Editor Associado do *Journal Venomous Animals and Toxins including Tropical Diseases*; Pesquisador Associado do CEVAP-UNESP; Prof. Livre-Docente da FMB-UNESP; Pesquisador Pq1C do CNPq; Membro da Academia Brasileira de Ciências Farmacêuticas do Brasil/Academia Nacional de Farmácia.

6. COMO CITAR

SHINTAKU, M.; FERREIRA JÚNIOR, R. S. Gestão de dados em periódicos científicos. *In: SHINTAKU, M.; SALES, L. F.; COSTA, M. (org). Tópicos sobre dados abertos para editores científicos.* Botucatu, SP: ABEC, 2020. p. 231-240. DOI: 10.21452/978-85-93910-04-3.cap18.



Em setembro de 2018 o Programa SciELO comemorou seus 20 anos com uma semana de eventos - a Semana SciELO 20 Anos, que culminou um programa de aproximadamente um ano de estudos, análises e debates envolvendo a Rede SciELO, pesquisadores editores, pesquisadores e profissionais relacionados com a comunicação e avaliação da pesquisa e de periódicos. O programa abordou a dinâmica de inovações políticas e operacionais em prol da abertura de todos os conteúdos das pesquisas e dos processos de sua comunicação que caracteriza o modus operandi da ciência aberta, suas vantagens e desafios em especial para os autores e periódicos. O pioneirismo e a experiência dos periódicos SciELO com a publicação em acesso aberto e a familiaridade com a adoção continuada de inovações contribuiu decisivamente para o estabelecimento das linhas prioritárias de ação do Programa SciELO que projeta a adoção da ciência aberta nos próximos anos. O primeiro avanço esperado é a revisão das políticas editoriais dos periódicos, a atualização das instruções aos autores, do formulário de submissão de manuscritos e do formulário de avaliação dos pareceristas. Estas revisões e atualizações os procedimentos relativos a aceitação para avaliação de manuscritos depositados previamente em servidores de preprints confiáveis, a citação e referenciamento dos dados, códigos e algoritmos e outros materiais subjacentes aos artigos e diferentes opções de abertura da revisão por pares. Como a adoção da ciência aberta é uma ação coletiva e cooperativa de todos os atores da pesquisa e sua comunicação, o Programa SciELO se esforça para participar, cooperar e acompanhar as iniciativas nacionais e internacionais. Sendo assim, parabeniza a Associação Brasileira de Editores Científicos (ABEC) por mais essa obra, que aborda assuntos importantes para os editores que desejam adotar as práticas da ciência aberta e particularmente o compartilhamento de dados de pesquisa. A obra contribuirá para a promoção da ciência aberta por meio da discussão e educação.

SciELO e a Ciência Aberta

Abel Laerte Packer, Diretor do Programa SciELO/FAPESP

