

**DEBATES:**

Uma introdução à ciência aberta e ao compartilhamento de dados científicos de pesquisa

<https://doi.org/10.47627/gradus.v5i2.165>

Abstract

Open Science is defined as an ecosystem that houses the various movements around open access to scientific information. Scientific data are the basic nuclear unit that supports scientific discoveries. In this sense, the purpose of this paper is to situate scientific research data in the context of Open Science, presenting the definitions of data, sharing and reuse of scientific data. It presents an overview of the challenges regarding the management of scientific data, the type of data and the advantages of sharing and reuse. It presents the computational challenges to the construction of research data repositories. We conclude that diversity, typology and volume of data represent the major challenges in the implementation of repositories and indicate a research directory of disciplinary and multidisciplinary data repositories.

Keywords: open science; research data; data sharing; data reuse; data repository.

Resumo

A Ciência Aberta é definida como um ecossistema que abriga os diversos movimentos em torno do acesso aberto à informação científica. Os dados científicos são a unidade nuclear básica que suporta as descobertas científicas. Nesse sentido, o objetivo deste trabalho é situar os dados científicos de pesquisa no contexto da Ciência Aberta, apresentando as definições de dados, compartilhamento e reuso de dados científicos. O trabalho apresenta um apanhado sobre os desafios quanto à gestão dos dados científicos, à tipologia de dados e às vantagens do compartilhamento e reuso. Apresenta os desafios computacionais para a construção de repositórios de dados de pesquisa. Conclui-se que a diversidade, a tipologia e o volume de dados representam os grandes desafios na implantação dos repositórios e indicam-se diretórios de pesquisa de repositórios de dados disciplinares e multidisciplinares.

Palavras-chave: ciência aberta; dados científicos de pesquisa; compartilhamento de dados; reuso de dados; repositório de dados.

Autores

Karolayne Costa Rodrigues de Lima

<https://orcid.org/0000-0002-6311-8482>

Universidade Federal do Paraná

Marcos Sfair Sunye

<https://orcid.org/0000-0002-2568-5697>

Universidade Federal do Paraná

Histórico do artigo

Recebido 2020-12-01

Aceito 2020-12-07

Publicado 2020-12-23

Introdução

A Ciência Aberta é um ecossistema¹ que reúne as diversas iniciativas, planos e ações em torno da abertura das atividades, processos, métodos, avaliações e produções científicas. O foco da ciência aberta é a abertura de todas as etapas do ciclo de vida da atividade científica e apresenta temas ou macrossistemas que regem as políticas de acesso aberto, metodologias ativas e abertas, avaliação aberta, ferramentas abertas e dados abertos. Cada um desses grandes temas expandem-se em subsistemas vascularizados que particularizam determinada prática da atividade científica, formando, assim, uma interligada rede que tal como um ecossistema biológico interage entre si e com outros.

A Ciência Aberta, de forma ampla, é conceituada como um princípio por meio do qual “o conhecimento científico deve ser livre para as pessoas usarem, reutilizarem e distribuírem sem restrições legais”;² tornando o conhecimento acessível para a sociedade, de modo a fornecer subsídios para o desenvolvimento de uma ciência cidadã onde os indivíduos possam se apropriar do conhecimento gerado pela academia e/ou instituições e centros de pesquisa para modificar sua realidade e retroalimentar a ciência.

As práticas da ciência aberta têm por característica o uso massivo da tecnologia para conectar e inter-relacionar as etapas do ciclo de vida da atividade científica devido à disponibilização de informações e dados em todas as etapas da pesquisa científica, partindo do seu planejamento na etapa inicial com o delineamento de um Plano de Gestão de Dados (PGD) até o reuso de dados científicos de pesquisa.

A Ciência Aberta não é um tema novo, visto que sua abordagem está intimamente relacionada com a metodologia da pesquisa. Entretanto, o termo “ciência aberta” ou “*Open Science*” tornou-se comum a partir da discussão da mudança de paradigma científico do modelo computacional da última década, focado na simulação de modelos complexos, para a chamada *e-Science*,³ baseada na exploração de dados com uso massivo de tecnologia e internet. Os dados explorados neste novo paradigma são aqueles gerados pela atividade científica, que dão embasamento aos resultados das pesquisas.

Nesse sentido, o objetivo geral desta comunicação é conceituar os dados científicos de pesquisa no âmbito da ciência aberta, detalhando as boas práticas de disponibilização e reuso de dados em diversas áreas do conhecimento.

¹ Albagli et al., *Ciência aberta, questões abertas* (2015).

² <http://www.isitopendata.org/>.

³ Hey et al., *The fourth paradigm: data-intensive scientific discovery* (2009).

Contexto da ciência aberta

No âmbito internacional, as primeiras iniciativas da ciência aberta surgiram em meados de 1980 nos Estados Unidos, com os movimentos do *software* livre e do acesso aberto à informação científica. Desses movimentos, destacam-se o Manifesto GNU em 1985, a licença tipo *Copyleft* e a *General Public Licence* (GPL), com as discussões em torno da abertura de licenciamento para sistemas *open source*.⁴ Em relação ao acesso aberto, destacam-se a Conferência de Santa Fé em 1999, no estabelecimento de padrões para documentos eletrônicos, e os princípios da *Open Archives Initiative* (OAI); a Declaração de Budapeste em 2002, com a definição de estratégias para o acesso aberto e vias para atingi-lo; a Declaração de Bethesda em 2003, expandindo orientações e recomendações sobre acesso aberto para entidades acadêmicas e científicas; e a Declaração de Berlin, também em 2003, reforçando o compromisso com o acesso livre, pela internet, da informação e do conhecimento científico.⁵

Esses movimentos e declarações foram manifestações a favor da ruptura ao modelo tradicional de comunicação científica, que restringia o acesso aos resultados de pesquisas a revistas fechadas (acessíveis mediante assinatura). Essas manifestações, de forma prática, resultaram na criação dos primeiros periódicos eletrônicos e repositórios digitais que, suportados por *softwares open source* e pela internet, representaram o início de um processo de abertura gradual da atividade científica.⁶ Como exemplo de repositório, destaca-se o ArXiv (1990), disponibilizando *preprints* nas áreas de física, matemática e ciência da computação, e englobando os princípios da OAI.

No contexto internacional, em 1998 ocorreu a criação da The Scholarly Publishing and Academic Resources (Sparc), como um acordo universitário em apoio para melhorar o acesso aberto. Em 2001, foi criada a Public Library of Science (PLOS), como um pacto envolvendo diversos países para ofertar artigos científicos em acesso aberto.

No âmbito nacional, o lançamento da *Scientific Eletronic Library* (SciELO), em 1998, destaca-se como um exemplo das primeiras iniciativas em ciência aberta nacional, devido ao oferecimento de uma plataforma de periódicos científicos, revisados por pares, disponível de forma aberta pela internet para toda a sociedade. Em 2005, o Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT) lançou um manifesto em prol do acesso aberto.

O manifesto do IBICT em 2005 ecoou as iniciativas nacionais existentes, como, por exemplo, o lançamento, em 2004, do Repositório Digital Institucional (RDI) da UFPR,⁷ composto por três bibliotecas digitais (teses e dissertações, periódicos eletrônicos

⁴ Stallman, *Software livre para una sociedad libre* (2004).

⁵ Rios et al., “Manifestos do movimento de acesso aberto” (2019).

⁶ Setenareski et al., *A dinâmica competitiva do mercado mundial de publicações científicas* (2020).

⁷ Sunye et al., “A experiência da UFPR na construção de repositórios digitais, a implantação integrada das ferramentas DSpace, Open Journal System” (2009).

institucionais e biblioteca de digital de imagem e som) e suportado pelos *softwares* livre *DSpace* e *Open Journal System* (OJS), esse último elaborado e mantido pela *Public Knowledge Project* (PKP), entidade sem fins lucrativos pioneira em soluções de *software* para o acesso aberto.

Na esteira do ArXiv, surgiram outras iniciativas de publicação em acesso aberto, tanto com periódicos quanto com a disponibilização integral de teses e dissertações, *preprints* e outros tipos de publicações acadêmicas, científicas e culturais. Nesse sentido, o surgimento dos repositórios digitais conferiu agilidade e gratuidade na obtenção do texto completo de teses e dissertações, fazendo com que serviços pagos de solicitação de documentos como o COMUT, por exemplo, perdessem relevância.

Na *e-Science*, o desenvolvimento da prática científica é suportado pelas tecnologias de informação e comunicação dentro de uma infraestrutura computacional que oferece ao pesquisador plataformas, ferramentas e ambiente para o desenvolvimento da pesquisa de forma integrada e colaborativa com outros pesquisadores e atores. À medida que o desenvolvimento computacional avança, as barreiras ao acesso aberto diminuem e a disponibilidade de publicações aumenta, o que leva a um novo patamar dentro do acesso aberto: a abertura e exploração de dados científicos de pesquisa, unidade nuclear, básica e fundamental da ciência.

Dados científicos de pesquisa: disponibilização, compartilhamento e reuso

Os dados científicos de pesquisa podem ser entendidos como um registro pontual e preciso que representa o resultado de uma manifestação, ação, reação e/ou acontecimento. Dito de outro modo, dados científicos de pesquisa são os dados coletados, organizados e analisados a fim de fornecerem uma informação sobre determinado fato.

A literatura da área indica que não há ainda um consenso sobre a conceituação de dados científicos de pesquisa.⁸ A definição de um conceito para dados deve ser precedida de contexto e abordagem, pois o dado pode representar informações distintas em cada área do conhecimento.⁹ Uma visão sobre os dados é que eles são considerados a “força vital da ciência e tecnologia” e devem ser processados por humanos e máquinas.¹⁰

Os dados científicos podem se caracterizar como texto, vídeos, imagens, objetos tridimensionais, objetos digitais e impressos, nu-

⁸ Borgman, “The conundrum of sharing research data” (2012).

⁹ da Silva, *Gestão de dados científicos* (2019).

¹⁰ Frederick, “Data, open science and libraries: the data deluge column” (2016).

méricos, morfológicos, taxonômicos, amostras biológicas, químicas e físicas, de ressonância, códigos, tabelas, gráficos, protocolos e demais variedades que dependem de contexto e abordagem. Logo, “os dados científicos são todas as evidências de que um investigador necessita para validar suas conclusões após uma pesquisa”.¹¹

Em sua tipologia, os dados científicos classificam-se como do tipo observacional, computacional e experimental. Os dados observacionais são provenientes da observação direta, em tempo real e, por isso, únicos, e assim podem ser qualificados como registros históricos, como, por exemplo, dados meteorológicos ou dados de uma observação etnográfica. Os dados computacionais provêm da execução de simulações e modelos computacionais que podem ser reproduzidos se preservadas as condições de *hardware* e *software*. Como exemplo de dados computacionais, têm-se dados advindos de simulações em realidade virtual. Os dados do tipo experimental decorrem da execução de procedimentos controlados para comprovação de fenômenos e estão sujeitos à reprodução se também preservadas as condições de ambiente, padrões e procedimentos. Os dados experimentais podem ser resultados de estudos laboratoriais, ou ainda dados provenientes de estudos comportamentais.

Assim, tem-se então que os dados científicos são dados primários que servem para comprovar ou refutar hipóteses, atestar ou não a validade de dados. Entretanto, quanto ao uso, os dados ou conjuntos de dados (*datasets*) podem ser classificados em abertos ou restritos. A abertura e restrição é designada pelo(s) autor(res) ou pela instituição financiadora da pesquisa, sendo o uso, compartilhamento e reuso regulamentados conforme o tipo de licença de direito autoral e pela documentação atrelada aos *datasets*, e/ou às políticas dos repositórios ou ambientes onde os dados estão armazenados.

A abertura irrestrita dos dados no âmbito da ciência aberta significa a possibilidade de exploração de dados no sentido de uso, reuso, compartilhamento e modificações. Os Princípios FAIR e os Princípios Panton são duas iniciativas que auxiliam os pesquisadores na conceitualização dos dados abertos, recomendando formas de fazê-lo.

Os Princípios FAIR¹² preconizam que um dado aberto é um dado encontrável (*Findable*), acessível (*Accessible*), interoperável (*Interoperable*) e reusável (*Reusable*), sendo que cada princípio possui requisições específicas para cada função. Por exemplo, o princípio de reusabilidade requer que os dados possuam metadados (descrições sobre dados) completos com atributos relevantes e precisos por meio de licenciamento claro e específico, associado a uma procedência detalhada e seguindo padrões relevantes de domínio. Ou seja, para ser FAIR, um dado deve ser recuperável por máquina, estar acessível, poder ser recuperado por meio de

¹¹ da Silva, *Gestão de dados científicos*, p. 21 (2019).

¹² Wilkinson et al., “The FAIR guiding principles for scientific data management and stewardship” (2016).

protocolos de interoperabilidade, como o *Open Archives Initiative Protocol Metadata Harvesting* (OAI-PMH) ou o *Open Archive Initiative Object Reuse and Exchange* (OAI-ORE), e ser reusável por meio da aplicação de licenciamento aberto do tipo *Creative Commons* (CC). O tipo de licença de direito autoral mínima recomendada é a cc-by, que representa, no mínimo, a obrigatoriedade de créditos ao autor original.

Os princípios Panton¹³ representam a abertura irrestrita dos dados, de modo que um dado possa ser usado apropriado (inclusive comercialmente), compartilhado, reusado e modificado sem que exista a necessidade de conferir créditos ao autor original do dado ou *dataset*. O tipo de licença recomendado pela Panton é a licença do tipo CCZero (renúncia aos direitos autorais).

Nessa acepção, observa-se que tornar os dados científicos disponíveis de forma aberta e irrestrita significa colocar à disposição da comunidade científica e da sociedade a possibilidade de obter novas descobertas e análises, complementar ou ressignificar outras pesquisas, tudo a partir do compartilhamento e reuso de dados.

A disponibilização aberta, o compartilhamento e o reuso de dados científicos apresentam vantagens e benefícios para a ciência em si, para os pesquisadores e para a sociedade, tais como: validar os processos científicos e seus resultados; promover a reprodutibilidade de pesquisas; promover a preservação digital de longo prazo, mantendo a integridade dos dados; proporcionar economia de recursos na fase de coleta de dados; promover novas análises e ressignificações de pesquisas; promover a transparência na aplicação de recursos e financiamento; promover visibilidade dos pesquisadores na forma de citações, convites para colaboração, ampliação das redes de pesquisa; e tornar visível e acessível para a sociedade as atividades das universidades, centros e institutos de pesquisa.¹⁴

Em uma abordagem simplista, o compartilhamento de dados representa a liberação dos dados para outros, enquanto o reuso é a ação de utilizar os dados disponibilizados abertamente por outros em uma pesquisa secundária. Entretanto, tanto o compartilhamento quanto o reuso de dados enfrentam barreiras e desafios que residem nas seguintes dimensões: política, organizacional, sócio-cultural, estrutural-tecnológica, econômica e legal.¹⁵

A dimensão política está relacionada com elaboração de políticas públicas de informação pelas entidades governamentais e/ou financiadoras que contemplem planejamento, estratégias, objetivos e recursos, a fim de incentivar e viabilizar a disponibilização, acesso, preservação, compartilhamento e reuso de dados.

A dimensão organizacional está relacionada com a forma pela qual as organizações definem suas estruturas e viabilizam a publicação dos dados de pesquisa, seja tanto no sentido de conscientização dos benefícios e vantagens da disponibilização, com-

¹³ Murray-Rust et al., *Panton principles* (2010).

¹⁴ Tenopir et al., "Data sharing by scientists" (2011).

¹⁵ Bueno de la Fuente, *Challenges and strategies for the success of open science* (2016); de Oliveira e da Silva, "Ciência aberta" (2016).

partilhamento e reuso, quanto no oferecimento de condições estruturais de trabalho aos pesquisadores.

A dimensão sócio-cultural está ligada à forma como o pesquisador observa e compreende suas atividades de pesquisa e ao quanto está disposto a compartilhar sua produção com outros. Esta dimensão apresenta-se com um caráter pessoal, pois é relacionada com o apego do autor em relação aos seus dados de pesquisa. Dito de outro modo, a dimensão pessoal diz respeito a quanto o pesquisador está disposto a conceder o uso e reuso dos dados produzidos em sua pesquisa. Essa dimensão divide espaço com a dimensão organizacional da feita de que é preciso um trabalho organizacional conjunto e coletivo para incentivar o pesquisador a compartilhar seus dados.

A dimensão estrutural-tecnológica representa os desafios computacionais e de infraestrutura tecnológica que oportunizem ao pesquisador um ambiente adequado ao trabalho com dados em todas as fases da pesquisa, oferecendo, também, ferramentas para manipulação dos dados e repositórios de dados para armazenamento, curadoria, preservação a longo prazo, acesso e reuso de dados.

A dimensão econômica está interligada com a parte política e organizacional, pois ainda que a disponibilização, o uso e o reuso de dados sejam abertos e gratuitos, existem custos de armazenamento, manutenção e suporte dos dados.

A dimensão legal representa o tipo de licenciamento atrelado aos *datasets* e às condições pré-estabelecidas pelo autor original ou pelo repositório que abriga o dado. Essa dimensão deve atender às novas requisições da Lei Geral de Proteção de Dados (LGPD)¹⁶ no tocante aos dados pessoais sensíveis, consentimento e anonimização.

¹⁶ Lei nº 13.709/18, Lei geral de proteção de dados pessoais (2018).

As dimensões elencadas acima representam uma parte dos desafios quanto ao uso, compartilhamento e reuso de dados científicos de pesquisa. Outra dimensão relevante consiste na elaboração dos repositórios de dados, que são as plataformas por meio das quais serão disponibilizados os dados para acesso. Nesse tocante, desafios como escalabilidade, heterogeneidade e complexidade e qualidade devem ser considerados na elaboração dos repositórios.

Repositório de Dados Científicos de Pesquisa: desafios

Os repositórios de dados são plataformas digitais tecnicamente estruturadas para armazenar e disponibilizar os dados e seus atri-

butos, como: objeto digital, metadados (os atributos dos dados, como autor, título do objeto, tipo, proveniência, entre outros), texto legal sobre licenciamento, documentação e identificador persistente. Os repositórios de dados então cumprem as funções de armazenar, disponibilizar, coletar, preservar e realizar curadoria dos dados. São estruturas hierarquizadas e modulares que devem atender a padrões internacionais de descrição, preservação e interoperabilidade de forma a proporcionar o uso e o reuso de dados sob condições legais.

A concepção e implantação de repositórios de dados científicos é um dos maiores desafios para a Ciência Aberta. Dados científicos possuem dois estados principais: crus (*raw data*) ou já processados. Alguns autores falam em “*dark data*”, que são dados brutos que o pesquisador sequer utilizou em seus estudos.¹⁷

Dados crus, ou ainda não processados, correspondem exatamente àquilo que foi observado, coletado ou gerado em um experimento científico. Esses dados podem ser fundamentais para a reprodutibilidade de um experimento, ou mesmo para assegurar sua confiabilidade. Os dados processados já passaram por algum processo de transformação ou mesmo simplificação, e também são necessários para assegurar a confiabilidade da pesquisa científica.

Sendo assim, um repositório de dados científicos deve, ao mesmo tempo em que descreve e modela os dados, também descrever seus estados e transformações.

A complexidade da construção de Repositórios de Dados Científicos leva em conta ainda novos desafios, como: a escalabilidade que é a capacidade de tratar o tamanho e a velocidade de crescimento da massa de dados; a heterogeneidade e complexidade dos dados, que variam muito em cada subárea do conhecimento; e a qualidade dos dados produzidos pelo pesquisador.

Escalabilidade

O avanço da precisão e do detalhamento dos equipamentos utilizados em experimentos científicos faz com que o volume de dados gerados a cada experimento cresça exponencialmente, gerando problemas de escala no volume de dados armazenados. Um exemplo do volume de dados que podem ser gerados por equipamentos de pesquisa é o acelerador de partículas *Large Hadron Collider* (LHC), da Organização Europeia para Pesquisa Nuclear (Cern), que captura 25 *petabytes* de dados todos os anos.

Planejar o crescimento de um repositório científico cujo potencial de crescimento cresce a cada nova geração de equipamentos é um desafio permanente de planejamento financeiro e tecnológico.

¹⁷ Schembera e Durán, “Dark data as the new challenge for big data science and the introduction of the scientific data officer” (2020).

Heterogeneidade e complexidade

Os equipamentos de pesquisa, além de evoluírem rapidamente na capacidade de coleta e análise e, portanto, na quantidade de dados gerados, evoluem também no tipo de informação que é gerada. Um equipamento que produz apenas espectros em uma determinada análise pode passar a gerar imagens ou imagens 3d, modificando a estrutura do dado gerado.

A própria definição do que é dado cru ou dado processado pode mudar com a evolução do equipamento que pode passar a processar os dados, ao mesmo tempo em que realiza a coleta da massa de informação. Repositórios de dados científicos, mesmo que sejam segmentados em áreas e subáreas, são heterogêneos, e dois *datasets* de um mesmo equipamento podem ter estruturas completamente diferentes.

Encontrar modelos de descrição dos dados em um repositório de dados científicos que consiga levar em conta a complexidade estrutural dos dados é também um desafio permanente.

Qualidade

Os dados gerados pelos pesquisadores, independentemente de serem dados crus ou já processados, podem ter baixa qualidade. Tanto o dado coletado como capturado em algum experimento está sujeito a erros. Por exemplo, a técnica utilizada no uso do equipamento pode ser inadequada, ou a distribuição amostral de dados coletados pode ser distorcida.

Assim, uma solução possível é a exigência de uma análise por pares dos dados gerados antes de sua submissão a um repositório de dados. O processo de revisão por pares, ainda que dentro de um processo de revisão aberta, como preconiza a Ciência Aberta, iria conferir, neste caso, qualidade, confiabilidade e autenticidade aos dados submetidos.

Os desafios acima descritos visam sanar as lacunas estruturais na construção de repositórios de dados para garantir a reprodutibilidade, que é um dos objetivos da Ciência Aberta. E a reprodutibilidade é compreendida não apenas para o uso, a modificação e a republicação dos dados científicos, mas também para a replicação dos procedimentos, códigos, equipamentos e ferramentas utilizados na prática científica. A reprodutibilidade de dados e procedimentos requer a definição de estratégias computacionais e metodológicas para tratar os processos de análise e experimentação de dados de forma que esses possam ser validados e passíveis de reaproveitamento.

Assim, os repositórios de dados devem ser implementados considerando-se a diversidade da tipologia de dados, o volume de dados, a defini-

ção de estratégias de preservação digital (para além de *backup*), de forma a garantir a conservação e acesso aos dados a longo prazo, preservando-se, assim, a memória científica. Os repositórios devem possuir políticas e diretrizes claras de submissão de dados descrevendo seus padrões, critérios, serviços oferecidos, termos de uso e suporte.

Os repositórios de dados concentram apenas parte da infraestrutura sobre dados científicos de pesquisa. No entanto, é talvez uma das partes relevantes, visto que representa o ponto de acesso por meio do qual os dados tornam-se visíveis e acessíveis e, por conseguinte, prontos para o reuso. O planejamento de repositório de dados deve considerar a área ou subáreas do conhecimento, respeitando os distintos padrões de descrição entre as áreas, de modo que seja possível a recuperação de informação por máquinas e humanos.

O *Re3data* é um diretório confiável de repositório de dados disciplinares e multidisciplinares por meio do qual é possível pesquisar os repositórios por área do conhecimento, tipo de dados, país, assunto, entre outros. Hoje, o *Re3data* aponta a existência de 11 repositórios de dados no Brasil, sendo a maioria da área multidisciplinar. O *Re3data* indica a existência de quatro repositórios de dados na área de Humanidades/Ciências Sociais Aplicadas e Linguística, por exemplo.¹⁸

¹⁸ <https://www.re3data.org>.

Considerações finais

A temática da Ciência Aberta e a exploração de dados científicos de pesquisa, embora não sejam assuntos novos, ainda são áreas em exploração no Brasil, carecendo de estudos quanto a definições, estruturação de dados, reusabilidade e repositório de dados. O objetivo deste texto foi situar os dados científicos no âmbito da Ciência Aberta, definindo os principais conceitos, bem como sua relação com o ambiente acadêmico, e o uso de ferramentas *open source*.

Os principais desafios com relação ao uso, ao compartilhamento e ao reuso de dados foram explicitados, de forma a assinalar ao pesquisador possibilidades e vantagens quanto à disponibilização de seus dados, bem como na exploração secundária de dados primários.

No contexto internacional, a abertura dos dados tornou-se ação mandatária nas políticas das agências de fomento e financiamento à pesquisa com dois objetivos distintos: tornar disponível e acessível a produção científica (publicações, dados e demais etapas do ciclo de vida da pesquisa) para fora do meio acadêmico, possibilitando a apropriação do conhecimento pela sociedade

de modo geral; e buscar promover transparência na aplicação de recursos, sejam públicos, privados ou mistos.

A transparência quanto à aplicação de recursos públicos, em se tratando da abertura de dados, permite o controle social em relação às ações de entidades governamentais e/ou com aporte público, e ainda a elaboração de indicadores sobre o investimento em ciência.

No Brasil, embora ocorram discussões sobre a abertura de dados em ações coordenadas pelo IBICT, Rede Nacional de Ensino e Pesquisa (RNP) e Ministério de Ciência, Tecnologia, Inovações e Comunicações (MCTIC) e Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), a abertura dos dados é apenas incentivada e não requerida obrigatoriamente. A ampla maioria das agências de fomento nacionais, excetuando a Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), não possuem políticas e/ou indicações mandatórias em seus editais quanto à abertura dos dados e armazenamento em repositórios.¹⁹ Nesse sentido, deve haver um estreitamento da discussão entre as entidades coordenadoras de políticas nacionais de informação científica com as agências de financiamento para que ocorra adequação dos editais de apoio a pesquisa e inclusão de ações de disponibilização e abertura de dados científicos.

A construção de repositório de dados constitui um dos maiores desafios atuais para a Ciência Aberta e é discussão relevante e necessária em prol do desenvolvimento da atividade científica e na transparência científica.

¹⁹ de Lima, “Impacto percebido pelos pesquisadores quanto ao reuso de dados científicos de pesquisa em repositório de dados brasileiros” (2020).

Referências

- Albagli, Sarita, Maria Lucia Maciel e Alexandre Hannud Abdo, editores (2015). *Ciência aberta, questões abertas*. Brasília: IBICT. ISBN: 9788570131096. URL: <http://livroaberto.ibict.br/handle/1/1060>.
- Borgman, Christine L. (2012). “The conundrum of sharing research data”. *Journal of the American Society for Information Science and Technology* 63.6, pp. 1059–1078. DOI: 10.1002/asi.22634.
- Bueno de la Fuente, Gema (2016). *Challenges and strategies for the success of open science*. URL: <https://www.fosteropenscience.eu/node/1424>.
- Frederick, Donna Ellen (2016). “Data, open science and libraries: the data deluge column”. *Library Hi Tech News* 33.8, pp. 11–16. DOI: 10.1108/LHTN-09-2016-0040.
- Hey, Tony, Stewart Tansley e Kristin Tolle, editores (2009). *The fourth paradigm: data-intensive scientific discovery*. Redmond: Microsoft Corporation. ISBN: 9780982544204.
- de Lima, Karolayne Costa Rodrigues (2020). “Impacto percebido pelos pesquisadores quanto ao reuso de dados científicos de pesquisa em repositório de dados brasileiros”. Dissertação de mestrado. Curitiba: Universidade Federal do Paraná. URL: <https://www.prppg.ufpr.br/signa/visitante/trabalhoConclusaoWS?idpessoal=64041&idprograma=40001016058P1&anobase=2020&idtc=50>.

- Murray-Rust, Peter, Cameron Neylon, Rufus Pollock e John Wilbanks (2010). *Panton principles: principles for open data in science*. URL: <https://pantonprinciples.org/>.
- de Oliveira, Adriana Carla Silva e Edilene Maria da Silva (2016). “Ciência aberta: dimensões para um novo fazer científico”. *Informação & Informação* 21.2, pp. 5–39. DOI: 10.5433/1981-8920.2016v21n2p5.
- Rios, Fahima, Elaine Oliveira Lucas e Igor Soares Amorim (2019). “Manifestos do movimento de acesso aberto: análise de domínio a partir de periódicos brasileiros”. *Revista Brasileira de Biblioteconomia e Documentação* 15.1, pp. 148–169.
- Schembera, Björn e Juan M. Durán (2020). “Dark data as the new challenge for big data science and the introduction of the scientific data officer”. *Philosophy & Technology* 33.1, pp. 93–115. DOI: 10.1007/s13347-019-00346-x.
- Setenareski, Ligia E., Walter Shima e Marcos S. Sunye (2020). *A dinâmica competitiva do mercado mundial de publicações científicas: tendências e alternativas do acesso aberto*. Curitiba: Appris. ISBN: 9788547334963.
- da Silva, Fabiano Couto Corrêa (2019). *Gestão de dados científicos*. Rio de Janeiro: Interciência. ISBN: 9788571934351.
- Stallman, Richard M. (2004). *Software libre para una sociedad libre*. URL: https://www.gnu.org/philosophy/fsfs/free_software.es.pdf.
- Sunye, Marcos, Ligia E. Setenareski, Fabiano Couto Corrêa da Silva, Edson Ramiro e Lucio Foltran (2009). “A experiência da UFPR na construção de repositórios digitais, a implantação integrada das ferramentas DSpace, Open Journal System”. In: *Implantação e gestão de repositórios institucionais: políticas, memória, livre acesso e preservação*. Editado por Luis Sayão, Lídia Brandão Toutain, Flavia Garcia Rosa e Carlos Henrique Marcondes. Salvador: EDUFBA, pp. 107–122. ISBN: 9788523206550.
- Tenopir, Carol, Suzie Allard, Kimberly Douglass, Arsev Umur Aydinoglu, Lei Wu, Eleanor Read, Mari-beth Manoff e Mike Frame (2011). “Data sharing by scientists: practices and perceptions”. *PloS one* 6.6, e21101. DOI: 10.1371/journal.pone.0021101.
- Wilkinson, Mark D. et al. (2016). “The FAIR guiding principles for scientific data management and stewardship”. *Scientific Data* 3.1, pp. 1–9. DOI: 10.1038/sdata.2016.18.