



# Publicação de dados de pesquisa no contexto da comunicação científica: cenário terminológico-conceitual

**Fabiana de Melo Amaral Gonçalves Pinto**

Doutora em Ciência da Informação, Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT) / Universidade Federal do Rio de Janeiro (UFRJ), Rio de Janeiro, Rio de Janeiro, Brasil.

Bibliotecária, Universidade Federal Fluminense (UFF), Niterói, Rio de Janeiro, Brasil.

<http://lattes.cnpq.br/3747190237682027>

Email: [fabianamelo@id.uff.br](mailto:fabianamelo@id.uff.br)



Submetido em: 10/09/2023. Aprovado em: 07/05/2024. Publicado em: 18/07/2024.

## RESUMO

Apresenta definições, à luz do movimento da Ciência Aberta, sobre a prática da publicação de dados de pesquisa. Entender o universo terminológico e conceitual, bem como sua aplicação no contexto da Ciência Aberta. Revisão de literatura em base de dados sobre a prática da publicação de dados de pesquisa. Resultados: A comunicação científica via publicação dos dados de pesquisa tende a aumentar e se apresentará sob diversos formatos, podendo ser terminologicamente compreendida sob diversas formas, contudo, estará sempre em sintonia com o comportamento e características de cada área do conhecimento. Tal comportamento demonstra a jovialidade da prática e de sua absorção nas áreas e por atores que compõem o cenário da comunicação científica, oportunizando estudos de áreas de acordo com seus padrões de informação e dados.

**Palavras-chave:** dados de pesquisa; ciência aberta; comunicação científica.

## INTRODUÇÃO

A produção de dados de pesquisa no contexto da *e-science* cresce exponencialmente em um ritmo bastante acelerado e vem sendo tema de debates em vários campos do conhecimento científico. A publicação desses dados é concebida pela comunidade científica ainda de forma bem reticente, embora algumas áreas do conhecimento já estejam bem alavancadas na prática em questão.

O presente estudo buscou apresentar e discutir, a partir da Teoria do Conceito de Dahlberg (1978), as terminologias e definições encontradas na literatura sobre a prática da publicação de dados de pesquisa, à luz do movimento da Ciência Aberta. As ideias de Dahlberg (1978) sobre o conceito e suas relações fazem-se oportunas e contribuem para a elaboração de conceitos, possibilitando ampliar o escopo da discussão sobre as definições propostas para a terminologia referente à publicação de dados de pesquisa. De acordo com Dalhberg (1978, p. 106), “[...] as definições são pressupostos indispensáveis na argumentação e nas comunicações verbais e se constituem elementos necessários na construção de sistemas científicos”. Ou seja, estabelecer uma definição homogênea e sistêmica sobre um determinado termo e seu respectivo conceito é fundamental para o amadurecimento e construção do vocabulário de um domínio, prática ou objeto científico.

Dalhberg (1978) reforça o entendimento anteriormente exposto, tendo em vista a internacionalização da ciência, ao afirmar que:

[...] a importância das definições evidencia-se também quando se tem em vista a comunicação internacional do conhecimento. É pelo domínio perfeito das estruturas dos conceitos que será possível obter também perfeita equivalência verbal (Dahlberg, 1978, p. 106).

Ou seja, as definições terminológicas e conceituais que compõem um domínio e/ou campo científico, bem como sua comunicação, são destacadas como parâmetro para compreensão, alargamento do diálogo nacional e internacional entre os pares.

No âmbito da comunicação científica e internacionalização da ciência, Dalhberg (1978) afirma que:

[...] a importância das definições evidencia-se também quando se tem em vista a comunicação internacional do conhecimento. É pelo domínio perfeito das estruturas dos conceitos que será possível obter também perfeita equivalência verbal”. (Dahlberg, 1978, p. 106).

Contudo, tanto a terminologia encontrada na literatura quanto os conceitos atribuídos à prática da publicação de dados de pesquisa revelam a juventude da temática investigada, ainda muito incipiente na Ciência da Informação. Essas circunstâncias levam, algumas vezes, a um entendimento equivocado sobre o que compreendemos por “publicação de dados de

pesquisa”. Tal afirmação deve-se ao fato de que: a ação de disponibilizar um conjunto de dados em um repositório é considerada completamente diferente do fluxo de trabalho da publicação em um periódico científico.

Campos (2001) também colaborou com o estudo ao destacar que, conforme a Teoria Geral da Terminologia, de Eugen Wuester, o trabalho terminológico estabelece a fixação de conceitos, visando à elaboração de definições orgânicas, além de estabelecer princípios para a criação de novos termos e possibilitar a comunicação mais precisa entre especialistas de diversas áreas do conhecimento no âmbito da Ciência e da Tecnologia, ao nível nacional e internacional (Campos, 2001, p. 71).

Autores como Parsons e Fox (2013), Borgman (2015), Li, Greenberg e Dunic (2020) se referiram a essa inconsistência terminológica como uma “metáfora de publicação de dados”. Sendo assim, entende-se que termos e conceitos devem estar logicamente sincronizados, para produzirem um entendimento harmonioso sobre um campo, prática ou objeto informacional, ainda que estes transitem no domínio da ciência na totalidade, como o movimento da Ciência Aberta, e seu objeto se aplique a diversos domínios, como os dados de pesquisa. A harmonia conceitual pode favorecer o desenvolvimento de estudos interdisciplinares e a colaboração entre pares.

## **METODOLOGIA**

A metodologia proposta consistiu em uma exaustiva revisão de literatura sobre o tema: publicação de dados de pesquisa, com o propósito de entender o universo terminológico e conceitual do objeto e da prática investigada.

Para tal, foram realizadas buscas nas seguintes bases de dados, bibliotecas digitais e repositórios: *Scopus*, *Web of Science*, *Science Direct*, *Information Science and Technology Abstracts* (ISTA), *Library, Information Science & Technology Abstracts with Full Text* (LISTA), *Library and Information Science Abstracts* (LISA), *Springer Link*, Literatura Latino-Americana e do Caribe em Ciências da Saúde (LILACS), MEDLINE, CINAHL, Scielo, Biblioteca Digital Brasileira de Teses e Dissertações (BDTD) do IBICT e o Repositórios Científicos de Acesso Aberto de Portugal (RCAAP).

A estratégia de busca utilizada nas bases de dados e repositórios considerou os seguintes termos e operadores booleanos: (“*data journal*” OR “*data publish*” OR “*data publication*” AND “*science communication*” OR “*scholarship communication*” AND “*open science*”), no período de 08/11/2021.

A não inclusão do termo “*research data*” na estratégia de busca justifica-se por ter se revelado “dispersivo” para os resultados recuperados em buscas prévias. A opção pelo idioma “inglês” deu-se a partir da representatividade de produção de literatura sobre o assunto (também em buscas prévias).

Os critérios de inclusão adotados na pesquisa consideraram estudos que estabelecessem relação entre a prática da publicação de dados de pesquisa e sua participação

no workflow da comunicação científica. Além disso, buscaram-se documentos que propusessem uma definição terminológica e conceitual para o objeto investigado (a prática da publicação de dados de pesquisa). Os critérios de exclusão referem-se a documentos que não explorassem a prática da publicação de dados de pesquisa no âmbito da Ciência Aberta.

## Panorama terminológico conceitual

No contexto da Ciência Aberta e sua proposta de abertura dos dados de pesquisa, novas maneiras de oferecer transparência à ciência surgem, ampliando o arcabouço terminológico e conceitual do campo científico. Diversos autores contribuem para a pluralidade conceitual e terminológica sobre a prática, mostrando claramente ser essa uma definição em processo de consolidação e amadurecimento. Ao mesmo tempo, revela características interdisciplinares em suas relações ao trazer para a discussão colaborações de múltiplos campos do conhecimento.

Desse modo, para discutir o panorama terminológico e conceitual referente à prática da publicação de dados de pesquisa e suas variações, adotou-se como base a definição proposta por Sales e Sayão (2019) sobre dados de pesquisa. Segundo os autores:

Dado de pesquisa é 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).

A palavra publicar significa, em linhas gerais, dar publicidade; ato de tornar público; registrar; comunicar. No âmbito da ciência, isso significa dar conhecimento e transparência aos pares e à sociedade na totalidade das pesquisas científicas e seus métodos, por meio de um canal formal de comunicação (livros, periódicos, eventos e seus *proceedings*). Meadows (1999) e demais autores chancelam essa afirmação ao apresentarem o periódico científico como instrumento legitimado pela comunidade científica para comunicar e tornar público as pesquisas concluídas ou em andamento, enquanto foram aprovadas e avaliadas pelos pares. É oportuno lembrar que no inglês as duas palavras “*publish*” e “*edit*” têm conotações distintas. Nos Estados Unidos, há clara diferença entre o “*editor*”, responsável intelectualmente pela obra, e o “*publisher*”, responsável pela sua publicação ou editoração, enquanto no Brasil, para ambas as funções, é usada a palavra “*editor*”. Contudo, no contexto da Ciência Aberta e sua proposta de abertura dos dados de pesquisa, novas maneiras de oferecer transparência à ciência surgem, ampliando o arcabouço terminológico e conceitual do campo científico.

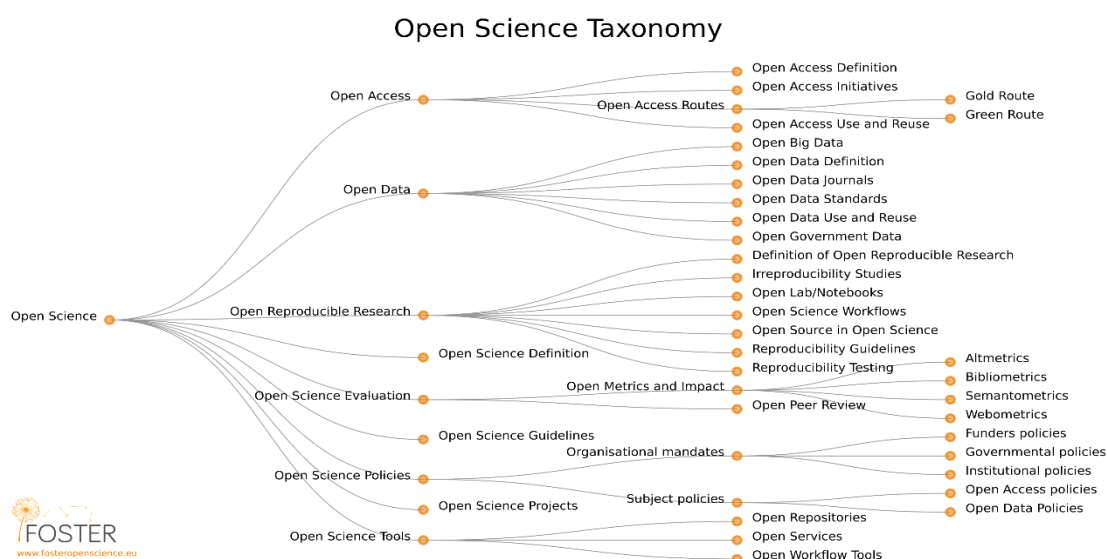
Ao nos referirmos à “publicação de dados de pesquisa”, inúmeras interpretações, aplicações e definições são apresentadas na literatura para o termo, o que ousamos considerar como uma inconsistência terminológica inerente a uma nova prática de comunicação científica.

Apresentamos, a seguir, algumas definições encontradas na literatura, sobretudo internacionais, que se referem à publicação de dados de pesquisa e que possibilitaram

um entendimento mais completo e objetivo sobre a prática investigada. As contribuições aqui apresentadas permitiram-nos identificar características distintas aplicadas aos termos: “periódico de dados de pesquisa”, “publicação de dados de pesquisa”, “documento de dados de pesquisa” e seus correlatos em língua inglesa.

A reflexão tem início a partir da representação ontológica do objeto “dados abertos” no contexto do movimento da Ciência Aberta. A taxonomia desenvolvida pelo projeto *Foster Open Science*<sup>1</sup> (**FIGURA 1**), possibilitou inferir relações no âmbito dos dados de pesquisa aplicados à comunicação científica e ao campo científico de modo geral.

**FIGURA 1 – Taxonomia para Ciência**



Fonte: Foster Plus–Foster Open Science (2020)<sup>2</sup>.

Na taxonomia representada pela **FIGURA 1**, os periódicos de dados (*open data journals*) aparecem diretamente ligados aos dados abertos, e estes, por sua vez, representam uma das abas do guarda-chuva da Ciência Aberta. Tais periódicos configuram-se como um braço do movimento, além de ser local para escoamento para publicação de dados. A *FOSTER Plus* é uma iniciativa da União Europeia para a implementação e prática de Ciência Aberta no Horizonte 2020. Seu objetivo é contribuir para a mudança real e duradoura no comportamento dos pesquisadores europeus para garantir que a Ciência Aberta se torne uma realidade normativa.

O Programa *Horizont 2020*<sup>3</sup> é um programa de pesquisa e inovação da União Europeia (UE). Ele promove a transferência de tecnologia e inovação entre Academia e Indústria. Nesta iniciativa, os periódicos de dados, denominados na literatura internacional como “*data*

1 A FOSTER Plus é uma iniciativa da União Europeia para a implementação e prática de Ciência Aberta no Horizont 2020.  
 2 Cf. [https://www.fosteropenscience.eu/themes/fosterstrap/images/taxonomies/os\\_taxonomy.png](https://www.fosteropenscience.eu/themes/fosterstrap/images/taxonomies/os_taxonomy.png)  
 3 Cf. <https://www.fapema.br/wpcontent/uploads/2018/06/Brochure-H2020-Brasil.pdf>.

*journals*” ou “*open data journals*”, são publicações revisadas por pares que reúnem artigos de dados, também conhecidos como documentos de dados ou “*data papers*” (Foster *Open Science*, 2019, *online*).

Os periódicos de dados nascem no âmbito do movimento da Ciência Aberta e sua matéria-prima, os dados de pesquisa, são protagonistas de inúmeras abordagens de pesquisa científica. Representam uma parcela da literatura que concebe a publicação de dados de pesquisa pelo viés tradicional da comunicação científica, ou seja, o periódico científico.

A construção do panorama terminológico iniciou-se com o trabalho de Lawrence *et al.*, (2011), quando entenderam ser

A publicação de dados uma ponte entre humanos e computadores, na medida em que preenche a lacuna entre dados, que são consumidos diretamente por um computador, e informações, que são produzidas a partir de dados, mas consumidas por seres humanos. (Lawrence *et al.*, 2011, p. 15, *tradução nossa*)<sup>4</sup>.

Para estes autores, uma publicação de dados compõe o fluxo de trabalho da comunicação científica, sob os seguintes vieses: 1. Publicação de dados *stand-alone*. 2. Publicação de dados por *proxy*. 3. *Appendix data*. 4. *Journal driven data archival*. 5. *Overlay*.

Na publicação de dados do tipo *stand-alone*, o conjunto de dados é publicado (depositado/disponibilizado) em um repositório sem a necessidade de um documento de dados que o descreva. Na publicação de dados por *proxy*, a publicação dos dados (depósito em repositório) está atrelada à publicação do artigo convencional, porém, apresenta uma desvantagem. A política que rege a preservação a longo prazo (no repositório) do conjunto de dados é independente do documento convencional e limitada pelas políticas e financiamento da instituição que gerencia o repositório.

No tipo de publicação de dados “*Appendix data*”, o conjunto de dados é publicado (arquivado) como material suplementar na própria revista.

Na opção de publicação de dados do tipo “*Journal driven data archival*”, os autores devem reservar uma seção do artigo convencional para fazer a descrição do conjunto de dados produzidos na pesquisa. A descrição dos dados está contida no artigo convencional, portanto, passa pela avaliação dos pares.

O modelo de publicação *overlay* prevê uma publicação de dados específica, ou seja, um artigo de dados em periódico de dados cuja proposta é publicar, via descrição por metadados, o conjunto de dados produzidos por uma pesquisa.

Assim como no modelo *journal driven data archival*, o fluxo de trabalho para a publicação do tipo *overlay* também inclui revisão por pares, tanto para o artigo quanto para o conjunto de dados.

O olhar plural para o termo publicação de dados de pesquisa também foi apresentado por García-García, López-Borrull, Peset (2015), que interpretaram o termo sob três possibilidades:

---

4 Original: “Data publication provides a bridge between human and computer in that it bridgesthe gap between data, which is consumed by a computer directly, and information,which is produced from data, but consumed by humans.” (Lawrence *et al.*, 2011, p. 15).

Publicação como objeto de informação independente em um repositório de dados de pesquisa. 2) Publicação de dados de pesquisa na forma de artigo de dados em um periódico de dados. 3) Publicação dos dados da pesquisa juntamente com o artigo, em um formulário de publicação ampliada. (García-García, López-Borrull, Peset, 2015, p. 846, *tradução nossa*)<sup>5</sup>.

Ao considerarem a publicação de dados um dos principais pilares para uma ciência aberta, sustentável, confiável e moderna, Bloom *et al.*, (2015) contribuíram com o estudo ao entenderem que termo “publicação de dados” se aplica tanto para repositório quanto para periódicos, e, dessa forma, fomenta entendimentos conceituais diversos que envolvem o termo e seus correlatos. Segundo os autores:

Os principais processos da publicação em periódicos de dados consistem na revisão por pares científicos e divulgação dos conjuntos de dados. Naturalmente, para a equipe de revisão, isso geralmente inclui acesso à pré-publicação ao conjunto de dados, que estaria disponível em algum repositório de dados, e também exige soluções explícitas de controle de versão para conjuntos de dados e documentos de dados. (Bloom *et al.*, 2015, p. 11, *tradução nossa*)<sup>6</sup>

[...] periódicos de dados geralmente dependem de repositórios de dados externos para lidar com a gestão de dados. Isso requer uma forte colaboração entre as equipes do periódico e do repositório, além de confiar que o repositório buscará gerenciamento e ingestão de dados de acordo com procedimentos padrão aceitáveis. Revistas e repositórios de dados são incentivados a fazer tais acordos (por exemplo, acordos de nível de serviço) públicos e transparentes para os usuários. (Bloom *et al.*, 2015, p. 13, *tradução nossa*)<sup>7</sup>.

A diferenciação conceitual estabelecida pelos autores sustenta que a publicação de dados e o depósito de dados configuram-se em práticas distintas, porém, interligadas por um elo fundamental (o conjunto de dados), o que amplia as possibilidades de transparência e abertura da ciência. Desse modo, estabeleceu-se duas abordagens conceituais: abordagem baseada em repositório e abordagem baseada em periódico.

## Abordagem baseada em repositório

Na abordagem baseada em repositório, a publicação de dados refere-se ao depósito/disponibilização do conjunto de dados em um repositório, garantindo sua curadoria, preservação e potencial para compartilhamento e reutilização em novas pesquisas. De

5 Original: “1) Publicación como objeto de información independiente en un repositorio de datos de investigación. 2) Publicación de datos de investigación en forma de data paper en un data jornal. 3) Publicación de datos de investigación junto al artículo, en la forma de enriched / enhanced publication.” (García-García, López-Borrull, Peset, 2015, p. 846).

6 Original: “Data journal publishing’s core processes consist of scientific peer review and dissemination of the datasets. Naturally, for the review team, this usually comprises pre-publication access to the dataset, which would be available in some data repository, and also demands explicit version control solutions for datasets and data papers.” (Bloom *et al.*, 2015, p. 11).

7 Original: “[...] data journals often rely on external data repositories to handle the actual data management. This requires a strong collaboration between the journal and repository staffs and trust that the repository will pursue data management and ingestion according to acceptable standard procedures. Data journals and data repositories are encouraged to make such agreements (e.g., Service Level Agreements) public and transparent to users.” (Bloom *et al.*, 2015, p. 13).

acordo com essa abordagem, é no repositório de dados que ocorre a publicação, e não no periódico. Encontramos na literatura alguns autores que corroboram essa visão tecnicista para a publicação de dados, conforme veremos a seguir:

A publicação formal de dados fornece um serviço além do simples ato de postar um conjunto de dados em um site, na medida em que inclui uma série de verificações técnicas no conjunto de dados (formato, metadados) ou de natureza mais científica. A publicação formal de dados também fornece ao usuário de dados certas garantias sobre a persistência dos dados e fornece um fórum para que o conjunto de dados seja encontrado e avaliado—uma parte essencial do processo científico. (Callaghan *et al.*, 2013, p. 194, tradução nossa)<sup>8</sup>.

Em artigo que explora o desenvolvimento de competências para a publicação de dados de pesquisa, Stoker, Melzack e McLean (2019, p. 175) se referem à publicação de dados “[...] como o processo de disponibilizar um conjunto de dados por meio de um repositório”. A Biblioteca da Universidade de Sidney, na Austrália, considera que “a melhor maneira de disponibilizar seus dados de pesquisa é publicando-os em um repositório de dados”, ou seja, a publicação de dados pelo viés do depósito em repositório, e não pela publicação em revista. No “Guia de publicação de dados” destacam-se os benefícios de uma publicação de dados:

A publicação de seus dados e conjuntos de dados de pesquisa em um repositório de dados pode permitir que seus dados sejam reutilizados e citados em outros trabalhos. Alguns repositórios podem até fornecer ferramentas para rastrear como as pessoas usam seus dados com a altimetria. Os conjuntos de dados podem ser publicados por si mesmos, sem um trabalho de pesquisa associado, o que permite obter reconhecimento por trabalhos que podem não se qualificar para autoria em um periódico de pesquisa tradicional. (University of Sidney, *online*, tradução nossa)<sup>9</sup>.

Observa-se claramente, por meio desta definição, a proposta de depósito ou armazenamento do conjunto de dados, quando o autor se refere à preservação dos dados a longo prazo.

Thanos *et al.* (2023) compreendem a publicação de dados como uma prática que

[...] permite aos pesquisadores descobrir, compreender e fazer afirmações sobre a confiabilidade e adequação de finalidade dos conjuntos de dados/visualizações em um repositório de dados. O objetivo final da publicação de dados é disponibilizar dados científicos para reutilização tanto dentro das disciplinas originais quanto no âmbito mais amplo comunidade. (Thanos *et al.*, 2023, p. 6)<sup>10</sup>.

---

8 Original: “Formal publication of data provides a service over and above the simple act of posting a dataset on a website, in that it includes a series of checks on the dataset of either a technical (format, metadata) or a more scientific (is the data scientifically meaningful?) nature. Formal data publication also provides the data user with certain assurances about data persistence, and provides a forum for the dataset to be found and evaluated – an essential part of the scientific process.” (Callaghan *et al.*, 2013, p. 194)

9 Original: “Publishing your research data and datasets in a data repository can allow for your data to be reused and cited in other works. Some repositories may even provide tools to track how people use your data, such as Altimetrics. Datasets can be published in their own right, without an associated research paper, which can allow you to gain recognition for work that may not qualify for authorship in a traditional research journal.” (University of Sidney, *online*).

10 Original: “[...] allows researchers to discover, understand, and make assertions about the trustworthiness and fitness for purpose of the datasets/views in a data space. The ultimate aim of Data Publication is to make scientific data available for reuse both within the original disciplines and the wider community.” (Thanos *et al.*, 2023, p. 6).



## Abordagem baseada em periódico

Na abordagem baseada em periódico, ocorre, por vezes, uma variação terminológica entre “documento de dados” e “periódico de dados”. O documento de dados pode referir-se tanto a um artigo de dados quanto a um periódico de dados. Desse modo, Costas *et al.*, (2013) definiram o periódico de dados como:

[...] uma publicação que reúne artigos de dados especializados, cuja finalidade principal é descrever dados (fornecendo informações sobre o quê, onde, por que, como e que tipo de dados), em vez de relatar uma investigação de pesquisa. Como tal, contém fatos sobre dados, não hipóteses e argumentos em apoio a essas hipóteses baseadas em dados, como encontrado em um artigo de pesquisa convencional. Seus objetivos são três: fornecer uma publicação de periódicos citáveis que leve crédito acadêmico a editores de dados; descrever os dados de forma estruturada e legível por humanos; e trazer a existência dos dados à atenção da comunidade acadêmica. (Costas *et al.*, 2013, p.13).

Hrynaszkiewicz e Shintani (2014) compartilharam seu entendimento sobre periódico de dados ao descreverem características e benefícios deste tipo de publicação. Segundo os autores, o periódico de dados:

[...] possibilita melhor descoberta de dados, através da indexação de artigos em base de dados (por exemplo, Scopus, PubMed); permite crédito via artigo de revista; e maior confiabilidade dos dados, graças a revisão por pares. O periódico de dados também pode fornecer recomendações sobre formatação e deposição de dados. O artigo de dados permite publicação de informações contextuais e metodológicas detalhadas, permitindo validação independente e reprodução de dados. Os periódicos de dados também podem fornecer uma saída para conjuntos de dados e experiências não publicados ou sem prévia publicação. E, ao fazer parceria com repositórios, os periódicos de dados também podem melhorar a acessibilidade e a integração de dados com artigos de periódicos e aprimorar a experiência dos leitores (Hrynaszkiewicz; Shintani, 2014, p. 4, tradução nossa)<sup>11</sup>.

Apesar de não contribuírem autoralmente com o conceito de periódico de dados de pesquisa em artigo publicado sobre consensos e controvérsias sobre a publicação de dados, Kratz e Strasser (2014) apresentam características que norteiam este tipo de publicação e a conduzem à ascensão à “primeira classe” dos produtos de pesquisa, no contexto da comunicação científica. Segundo os autores:

---

11 Original: “[...] improved discoverability of data, through indexing of articles in journal repositories (e.g. Scopus, PubMed); credit via a journal article; and increased reliability of the data, if the journal peer reviews the data. Data journals can also provide recommendations on data formatting and deposition. The data article allows publication of detailed contextual and methodological information, enabling independent validation and reproduction of data. In this regard use of data journals can significantly add to the record of version, as printed journals in particular, are often subject to limitations on the amount of text or supplementary data they can publish. Data journals can also provide an outlet for unpublished datasets and experiments with no previous publication. And by partnering with repositories data journals can also improve the accessibility and integration of data with journal articles and enhance the experience for readers” (Hrynaszkiewicz; Shintani, 2014, p. 4).

A força de um documento de dados está no fornecimento de documentação rica, o que é especialmente útil para dados de pesquisa únicos e heterogêneos de “cauda longa” da ciência. O comprimento e a estrutura do papel de dados variam entre os periódicos, mas a tendência é para um formato curto e bem estruturado. Todas as revistas requerem um resumo, métodos de coleta e uma descrição do conjunto de dados; alguns incentivam os autores a sugerir possíveis usos para os dados (por exemplo, Internet Archaeology, e Open Health Data) [...] Os documentos de dados são definidos com maior nitidez, não pela presença de qualquer informação em particular, mas pela ausência de análises ou conclusões. (Kratz; Strasser, 2014, p. 5, tradução nossa)<sup>12</sup>.

O entendimento proposto por García-García, López-Borrull e Peset (2015) destacam alguma das vantagens da publicação de dados sob o viés do periódico científico e afirmam que este tipo de publicação torna a pesquisa científica mais confiável. De acordo com os autores

[...] as revistas de dados se concentram em dados abertos de pesquisa, o que que incentiva a reutilização, uma vez que descrevem a metodologia de coleta e análise de dados – com profundidade semelhante à exigida nas patentes. Em muitas revistas convencionais, a descrição da coleção de dados é reduzida a etapa de resultados e sua discussão. Os artigos de dados trazem à luz conjuntos de dados que por várias razões não foram publicadas em artigos pesquisa.<sup>36</sup> (García-García; López-Borrull; Peset, 2015, p. 852).

Reforçam o entendimento do cenário proposto para a publicação de dados de pesquisa Assante *et al.*, (2016) ao destacarem três principais atores necessariamente envolvidos na produção de uma publicação de dados de pesquisa: autores (pesquisadores), editores e gestores de repositório de dados de pesquisa.

Destacam o repositório como uma ferramenta de apoio à publicação de dados ao afirmarem que:

Os repositórios de dados científicos são frequentemente propostos como instrumentos para apoiar a publicação de dados, pois fornecem instalações para todos os diferentes atores envolvidos nesse processo. (Assante *et al.*, 2016, p. 2, tradução nossa)<sup>13</sup>.

Embora compreendam os diferentes papéis assumidos, tanto pelo periódico quanto pelo repositório, os autores utilizam o termo publicação de dados ambigamente ao se referirem a fluxos de trabalho distintos no contexto da comunicação científica. Destacam a importância da descrição e documentação do conjunto de dados por meio de uma publicação de dados ao afirmarem que:

---

12 Original: “The strength of a data paper is in providing rich documentation, which is especially useful for unique and heterogeneous “long-tail” research data. Data paper length and structure varies between journals, but the tendency is toward a short, tightly structured format. All journals require an abstract, collection methods, and a description of the dataset; a few encourage authors to suggest potential uses for the data (e.g., Internet Archaeology, and Open Health Data). Data papers are most sharply defined not by the presence of any particular information, but by the absence of analysis or conclusions.” (Kratz; Strasser, 2014, p. 5).

13 Original: “Scientific data repositories are often proposed as instruments for supporting data publishing as they provide facilities for all the different players involved in this process.” (Assante *et al.*, 2016, p. 2).

O objetivo final da publicação de dados é disponibilizar conjuntos de dados para validação e reutilização, tanto pela comunidade científica que a produziu e, mais amplamente, em outras comunidades. Para atingir esse objetivo, um conjunto de dados produzido por uma comunidade de prática deve receber informações auxiliares e contextuais sobre o conjunto de dados, tais como os metadados descritivos, e também, como foram obtidos, ou seja, informações sobre sua proveniência. Esta documentação, em geral influencia profundamente a descoberta, a compreensibilidade, a verificação, a citação e a reutilização prática do conjunto de dados. (Assante *et al.*, 2016, p. 8, tradução nossa)<sup>14</sup>.

Em artigo publicado no *International Journal on Digital Libraries*, Austin *et al.*, (2017) contribuíram para o estudo a partir das seguintes definições:

Um periódico de dados é um diário (invariavelmente acesso aberto) que publica artigos de dados. O periódico de dados geralmente fornece modelos para a descrição dos dados e oferece aos pesquisadores orientações sobre onde depositar e como descrever e apresentar seus dados. Dependendo do periódico, esses modelos podem ser genéricos ou focados na disciplina. Alguns periódicos ou seus editores mantêm seus próprios repositórios. Além de apoiar a vinculação bidirecional entre um artigo de dados e o (s) seu (s) conjunto (s) de dados correspondente (s) e facilitar práticas de identificação persistentes, os periódicos de dados fornecem fluxos de trabalho para garantir a qualidade (ou seja, revisão por pares de dados) [...] (Austin *et al.*, 2017, p. 82, tradução nossa)<sup>15</sup>.

Curty e Aventurier (2017) discutem novos formatos de publicação em ascensão no contexto da Ciência Aberta, e vêm gradativamente conquistando seus espaços entre os membros da comunidade científica. Para os autores,

[...] os artigos de dados elevam os dados científicos à condição de protagonistas, pois se dedicam a descrever exhaustivamente a coleção de dados, acompanhados de descrições do contexto, do percurso metodológico e dos aspectos procedimentais da pesquisa, e das possíveis aplicações dos dados. Essa abordagem pode se materializar por meio da publicação dos artigos de dados em periódicos científicos híbridos, receptivos aos data papers, ou periódicos dedicados à publicação de dados. Os periódicos de dados também trazem à superfície para comunidades de interesse, coleções de dados com maior potencial de reuso, tendo em vista a garantia de maior detalhamento documentação, e melhor endossamento. Por preservar características dos periódicos de dados e de cristalizada aceitação e de ampla aquiescência entre os membros da comunidade científica, como *peer-review*, periodicidade e corpo editorial, esta abordagem tem maior potencial de receptividade no meio científico. (Curty; Aventurier, 2017, p. 18).

---

14 Original: "The ultimate goal of data publishing is to make datasets available for validation and reuse both within the scientific community that has produced it and, more widely, within other communities. In order to fully meet this objective a dataset produced by a community of practice has to be endowed with auxiliary data providing contextual information about the dataset, like what it is about – descriptive metadata – and how it has been obtained – data provenance. This documentation at large, deeply influences dataset discoverability, understandability, verification and practical re-use." (Assante *et al.*, 2016, p.8).

15 Original: "A data journal is a journal (invariably open access) that publishes data articles. The data journal usually provides templates for data description and offers researchers guidance on where to deposit and how to describe and present their data. Depending on the journal, such templates can be generic or discipline focused. Some journals or their publishers maintain their own repositories. As well as supporting bi-directional linking between a data article and its corresponding dataset(s), and facilitating persistent identification practices, data journals provide workflows for quality assurance (i.e. data peer review) [...]" (Curty; Aventurier, 2017, p. 18).

No Brasil, encontramos a contribuição de pesquisadores da Universidade Estadual Paulista Júlio de Mesquita Filho–UNESP sobre o assunto. Em recente pesquisa para a proposição de uma estrutura comum para o artigo de dados de pesquisa, Roa Martínez, Vidotti e Santana (2017) definem um artigo de dados como:

[...] um novo tipo de comunicação científica que faz parte das classes de documentação resultantes do processo Publicação de dados. Esse tipo de artigo descreve um conjunto de dados de pesquisa acessível por outros pesquisadores junto com informações próprio, ou seja, com metadados relacionados ao processo a partir do qual os dados foram fornecidos os métodos e considerações de coleta, formatos, condições específicas e autores entre outras (Roa-Martínez; Vidotti e Santana, 2017, p. 5).

Em trabalho apresentado no *International Conference on Grey Literature*, Farace, Frantzen e Smith (2017, p. 27) discutem o documento de dados como um novo tipo de literatura cinza, além de entendê-lo como um dos meios mais tangíveis de implementação dos princípios de dados FAIR.

Ao proporem uma discussão onde a publicação de dados é vista sob a ótica da organização do conhecimento, Schöpfel *et al.*, (2019) concebem o documento de dados da seguinte maneira:

Ao contrário dos documentos de pesquisa comuns, o principal objetivo dos documentos de dados é descrever conjuntos de dados, incluindo as condições e o contexto de sua aquisição e sua utilidade potencial, em vez de relatar e discutir resultados. Além disso, é geralmente assumido que papéis de dados são papéis curtos com até 4 páginas (Schöpfel *et al.*, 2019, p. 3, tradução nossa)<sup>16</sup>.

Em artigo recente publicado no periódico *Journal of the Association for Information Science and Technology*, Li, Greenberg e Dunic (2020) consideram o documento de dados um gênero científico emergente e o definem como: “publicação acadêmica de metadados pesquisáveis, documento que descreve um conjunto de dados acessível on-line, ou um grupo de conjuntos de dados, publicado conforme a norma práticas acadêmicas”. (Li; Greenberg; Dunic, 2020, p. 172)

Cao (2022), em artigo de dados sobre a importância desta tipologia de publicação científica, considera que:

[...] os documentos de dados, comparados com outros tipos de artigos acadêmicos, são um novo tipo de artigo acadêmico e têm se tornado um importante meio de comunicação acadêmica. Um documento de dados é um artigo sobre a descrição de recursos de dados e os métodos de obtenção de recursos de dados. Os documentos de dados promovem o uso eficaz dos dados recursos por meio de revisão por pares, armazenamento permanente, compartilhamento aberto e outras estruturas, para que os dados desempenhem um valor acadêmico maior (Cao, 2022, p. 1).

---

16 Original: “Unlike usual research papers, the main purpose of data papers is to describe datasets, including the conditions and context of their acquisition and their potential utility, rather than to report and discuss results. Also, it is generally assumed that data papers are short papers with up to 4 pages” (Schöpfel *et al.*, 2019, p. 3).

Com base nas contribuições dos autores que compõem a discussão e que não esgotam a investigação sobre o tema, e, ainda, ancorados na trajetória histórica e questões epistemológicas que envolvem a prática da publicação de dados e sua sustentação terminológica e empírica no campo científico, propomos, a seguir, definições para os termos: Periódico de dados de pesquisa, Artigo de dados, Depósito de dados de pesquisa e Publicação de dados de pesquisa. Tais proposições terminológicas e conceituais buscam auxiliar o entendimento e o papel desses objetos e práticas no contexto da infraestrutura informacional da comunicação científica, gestão e compartilhamento de dados de pesquisa.

**Periódico de dados de pesquisa** – publicação periódica científica revisada por pares que segue os trâmites tradicionais do *workflow* da comunicação científica. Reúne artigos de dados que contêm a descrição do conjunto de dados de uma pesquisa.

**Artigo de dados** – documento cujo objetivo é descrever de forma detalhada e por meio de metadados as condições técnicas (procedimentos), tecnológicas, metodológicas e contextuais em que o conjunto de dados foi produzido, e, assim, possibilitar uma interpretação adequada para fins de compartilhamento e reuso em novas pesquisas. Ao serem publicados em um periódico científico revisado por pares, os artigos de dados ganham visibilidade, citabilidade e reusabilidade, transparência e credibilidade científica.

**Depósito de dados de pesquisa** – processo de armazenamento ou disponibilização de um conjunto de dados em repositórios de dados, podendo estar ou não em acesso aberto.

**Publicação de dados de pesquisa** – ato de comunicar e conferir publicidade aos dados produzidos por uma pesquisa, quer seja por meio de periódico científico, quer seja por meio de repositório de dados ou outro canal de comunicação científica. Refere-se à representação dos dados por meio de metadados e não necessariamente aos dados propriamente ditos.

## **CONSIDERAÇÕES FINAIS**

A comunicação científica via publicação dos dados produzidos em pesquisas tende a aumentar e se apresenta sob diversos formatos, podendo ser terminologicamente compreendida sob diversas formas, contudo, estará sempre em sintonia com o comportamento e características diversificadas de cada área do conhecimento. Tal comportamento demonstra a jovialidade da prática e de sua absorção por áreas e por atores que compõem o cenário da comunicação científica, oportunizando estudos de domínio de acordo com seus padrões de informação e dados.

Ambas as abordagens (publicação via periódico e repositório) não são antagônicas e deseja-se que ocorram de forma concomitante e complementares. Para tal, reforça-se a necessidade da curadoria dos dados, sua adequada representação, além da criação de repositórios disciplinares para absorver esse rico universo de dados especializados.

Os conceitos propostos pelo estudo buscam auxiliar a compreensão desse universo complexo, por vezes novo, e quase sempre ambíguo, que envolve os dados de pesquisa e sua interoperabilidade semântica e tecnológica.

## **REFERENCIAS**

ASSANTE, M.; CANDELA, L.; CASTELLI, D.; TANI, A. Are scientific data repositories coping with research data publishing? **Data Science Journal**, [s. l.], v. 15, n. 6, p. 1–24, 2016. DOI: <http://dx.doi.org/10.5334/dsj-2016-006>.

AUSTIN, C.; BLOON, T.; DALLMEIER- TIESSSEN, S.; KHODIYAR, V. K.; MURPHY, F.; NURNBERGER, A.; RAYMOND, L.; STOCKHAUSE, M.; TEDDS, J.; VARDIGAN, M.; WHYTE, A. Key components of data publishing: using current best practices to develop a reference model for data publishing. **International Journal on Digital Libraries**, [s. l.], v. 18, n. 2, p. 77–92, 2017. DOI 10.1007/s00799-016-0178-2.

BLOOM, T.; DALLMEIER-TIESSSEN, S.; MURPHY, F.; AUSTIN, C. C.; WHYTE, A.; TEDDS, J.; NURNBERGER, A.; RAYMOND, L.; STOCKHAUSE, M.; VARDIGAN, M.; CLARKE, T.; CASTRO, E.; NEWBOLD, E.; MOORE S.; HOLE, B. Workflows for Research Data Publishing: Models and Key Components. **International Journal of Digital Libraries**, [s. l.], special edition, p. 1-28, 2015. Disponível em: <https://zenodo.org/record/20308#.Xs2UMjdKjGg>. Acesso em: 19 fev. 2019.

BORGMAN, C. L. **Big data, little data, no data**: scholarship in the networked world Cambridge: MIT Press, 2015.

CALLAGHAN, S.; TEDDS, J.; KUNZE, J.; MAYERNIK, M. S.; MURPHY, F.; ALLAN, R.; LAWRENCE, R.; WHYTE, A. Processes and procedures for data publication: a case study in the geosciences. **International journal of digital curation**, [s. l.], v. 8, n. 1, p. 193-203, 2013. Disponível em: <http://www.ijdc.net/article/view/8.1.193>. Acesso em: 19 fev. 2019.

CAMPOS, M. L. A. **A organização de unidades do conhecimento em hiperdocumentos**: o modelo conceitual como um espaço comunicacional para a realização da autoria. 2001. Tese (Programa de Pós-graduação em Ciência da Informação) – Instituto Brasileiro de Informação em ciência e Tecnologia, Escola de Comunicação, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2001. Disponível em: <http://www.conexaorio.com/bit/tesemlcampos/TeseMLCampos.pdf>. Acesso em: 5 mar. 2020.

CAO, X. Data papers: an important type of academic articles. **Resources Data Journal**, [s. l.], v. 1, p. 10-16, June 2022. Disponível em: [https://www.jstage.jst.go.jp/article/rdj/1/0/1\\_10/\\_pdf/-char/ja](https://www.jstage.jst.go.jp/article/rdj/1/0/1_10/_pdf/-char/ja). Acesso em: 11 nov. 2023.

COSTAS, R.; MEIJER, I.; COSTAS, R.; ZAHEDI, Z.; WOUTERS, P. **The value of research data metrics for datasets from a cultural and technical point of view**. Copenhagen: Knowledge Exchange Report, 2013. Disponível em: [https://www.researchgate.net/publication/307547241\\_The\\_Value\\_of\\_Research\\_Data\\_\\_Metrics\\_for\\_datasets\\_from\\_a\\_cultural\\_and\\_technical\\_point\\_of\\_view\\_A\\_Knowledg\\_e\\_Exchange\\_Report\\_April\\_2013](https://www.researchgate.net/publication/307547241_The_Value_of_Research_Data__Metrics_for_datasets_from_a_cultural_and_technical_point_of_view_A_Knowledg_e_Exchange_Report_April_2013). Acesso em: 1 fev. 2019.

CURTY, R. G.; AVENTURIER, P. O paradigma da publicação de dados e suas diferentes abordagens. *In: ENCONTRO NACIONAL DE PESQUISA E PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO–ENANCIB*, 18., 2017, Marília. **Anais** [...]. Marília: UNESP, 2017. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/105144>. Acesso em: 1 jun. 2020.

DAHLBERG, I. Teoria do conceito. **Ciência da Informação**, Rio de Janeiro, v. 7, n. 2, p. 101-107, 1978. Disponível em: <https://revista.ibict.br/ciinf/article/view/115>. Acesso em: 22 maio 2020.

FARACE, D.; FRANTZEN, J.; SMITH, P. L. Data Papers are Witness to Trusted Resources in Grey Literature: a project use case. *In: NINETEENTH INTERNATIONAL CONFERENCE ON GREY LITERATURE–Public Awareness and Access to Grey Literature*, v. 19, 2017, Rome. **Proceedings** [...]. 2017. p. 27-32, 2017. Disponível em: <https://doi.org/10.26069/GREYNET-2018-000.002-GG>. Acesso em: 22 maio 2020.

FOSTER Open Science. **Open Data Journals**. [S. l.], 2019. Site. Disponível em: <https://www.fosteropenscience.eu/taxonomy/term/114>. Acesso em: 22 maio 2020.

GARCÍA-GARCÍA, A.; LÓPEZ-BORRULL, A.; PESET, F. Data journals: eclosión de nuevas revistas especializadas en datos. **El profesional de la información**, [s. l.], v. 24, n. 6, p. 845-854, 2015. DOI: <http://dx.doi.org/10.3145/epi.2015.nov.17>. Disponível em: <https://revista.profesionaldelainformacion.com/index.php/EPI/article/view/epi.2015.nov.17>. Acesso em: 22 maio 2020.

HRYNASZKIEWICZ, I.; SHINTANI, Y. Scientific Data: an open access and open data publication to facilitate reproducible research. **J-Stage**, Japan, v. 57, n. 9, p. 629-640, 2014. DOI: <https://doi.org/10.1241/johokanri.57.629>. Disponível em: [https://www.jstage.jst.go.jp/article/johokanri/57/9/57\\_629/\\_article/-char/ja/](https://www.jstage.jst.go.jp/article/johokanri/57/9/57_629/_article/-char/ja/). Acesso em: 22 maio 2020.

KRATZ, J.; STRASSER, C. Data publication: consensus and controversies. **F1000Research**, [s. l.], v.3, n. 94, 2014. DOI: <https://doi.org/10.12688/f1000research.3979.3>.

LAWRENCE, B.; JONES, C.; MATTHEWS, B.; PEPLER, S.; CALLAGHAN, S. Citation and peer review of data: moving toward formal data publication. **Int. J. Digital Curation**, United Kingdom, v. 6, n. 2, p. 4-37, 2011. DOI:10.2218/ijdc.v6i2.20r. Disponível em: <http://www.ijdc.net/article/view/181>. Acesso em: 19 fev. 2019.

LI, K.; GREENBERG, J.; DUNIC, J. Data Objects and Documenting Scientific Processes: an analysis of data events in biodiversity data papers. **JASIST – Journal of the Association for Information Science and Technology**, [s. l.], v. 71, n. 2, p.172– 182, 2020. DOI: <https://doi.org/10.1002/asi.24226>.

MEADOWS, A. J. **A comunicação científica**. Brasília: Briquet de Lemos, 1999. 268 p.



PARSONS, M. A.; FOX, P. A. Is data publication the right metaphor? **Data Science Journal**, [s. l.], n. 12, Feb. 2013. Disponível em: [https://www.jstage.jst.go.jp/article/dsj/12/0/12\\_WDS-042/\\_pdf/-char/en](https://www.jstage.jst.go.jp/article/dsj/12/0/12_WDS-042/_pdf/-char/en). Acesso em: 25 maio 2020.

ROA-MARTÍNEZ, S. M.; VIDOTTI, S. A. B.; SANTANA, R. C. Estructura propuesta del artículo de datos como publicación científica. **Revista Española de Documentación Científica**, [s. l.], v. 40, n. 1, enero-marzo 2017. DOI: <http://dx.doi.org/10.3989/redc.2017.1.1375>.

SALES, L. F.; SAYÃO, L. F. Uma proposta de taxonomia para dados de pesquisa. **Conhecimento em Ação**, Rio de Janeiro, v. 4, n. 1, jan./jun. 2019. Disponível em: <https://brapci.inf.br/index.php/res/download/119088>. Acesso em: 11 de set. 2023.

SCHÖPFEL, J.; PROST, H.; ZANE, A. Data papers as a new form of knowledge organization in the field of research data. In: **COLLOQUE INTERNATIONAL ISKO-FRANCE 2019: Données et mégadonnées ouvertes en SHS: de nouveaux enjeux pour l'état et l'organisation des connaissances?** 12.ISKO France, Montpellier, Oct. 2019. DOI: <https://doi.org/10.17026/dans-zk3-jkyb>.

STOKER, R.; MELZACK, G.; MCLEAN, J. Developing Culturally Competent Data Publication Resources. **International Journal of Digital Curation**, Edinburgh, v. 14, n. 1, p. 168-179, 2019. DOI: <https://doi.org/10.2218/ijdc.v14i1.648>. Disponível em: <http://www.ijdc.net/article/view/648>. Acesso em: 26 maio 2020.

UNIVERSITY OF SIDNEY. **Publishing research data**. Sydney, Oct. 2022. Disponível em: <https://libguides.library.usyd.edu.au/datapublication>. Acesso em: 26 maio 2020.

THANOS, C.; MEGHINI, C.; BARTALESI, V.; CORO, G. An exploratory approach to data driven knowledge creation. **Journal of Big Data**, [s. l.], v. 10, n. 29, 2023. Disponível em: <https://doi.org/10.1186/s40537-023-00702-x>. Acesso em: 11 nov. 2023.