

PUBLICAÇÃO DE DADOS DE PESQUISA: PANORAMA TERMINOLÓGICO-CONCEITUAL SOBRE A PRÁTICA NO ÂMBITO DA COMUNICAÇÃO CIENTÍFICA

Publication of research data: terminological-conceptual overview in scientific communication
Publicación de datos de investigación: panorama terminológico-conceptual en comunicación científica



Fabiana de Melo Amaral Gonçalves Pinto
Doutora em Ciência da Informação, Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict/UFRJ), Rio de Janeiro, RJ, Brasil.
Bibliotecária, Universidade Federal Fluminense (UFF), Niterói, RJ, Brasil.
Lattes: <http://lattes.cnpq.br/3747190237682027>
ORCID: <https://orcid.org/0000-0002-8118-3165>

Resumo

Introdução: Apresenta definições, à luz do movimento da Ciência Aberta, sobre a prática da publicação de dados de pesquisa. **Objetivos:** Entender o universo terminológico e conceitual, bem como sua aplicação no contexto da Ciência Aberta. **Metodologia:** 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. **Conclusão:** 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.

Abstract

Introduction: It presents definitions, in the light of the Open Science movement, on the practice of publishing research data. **Objectives:** Understand the terminological and conceptual universe, as well as its application in the context of Open Science. **Methodology:** Literature review of databases on the practice of publishing research data. **Results:** Scientific communication via publication of research data tends to increase and will be presented in different formats, and can be terminologically understood in different ways, however, it will always be in tune with the behavior and characteristics of each area of knowledge. **Conclusion:** Such behavior demonstrates the joviality of the practice and its absorption in the areas and by actors that make up the scenario of scientific communication, providing opportunities for studies of areas according to their standards of information and data.

Keywords: search data; Open Science; scientific communication.

Resumen

Introducción: Presenta definiciones, a la luz del movimiento Open Science, sobre la práctica de publicar datos de investigación. **Objetivos:** Comprender el universo terminológico y conceptual, así como su aplicación en el contexto de la Ciencia Abierta. **Metodología:** Revisión de literatura de bases de datos sobre la práctica de publicar datos de investigación. **Resultados:** La comunicación científica a través de la publicación de datos de investigación tiende a incrementarse y se presentará en diferentes formatos, pudiendo entenderse terminológicamente de diferentes maneras, sin embargo, siempre estará en sintonía con el comportamiento y características de cada área del conocimiento.

Conclusión: Tal comportamiento demuestra la jovialidad de la práctica y su absorción en las áreas y por los actores que componen el escenario de la comunicación científica, brindando oportunidades para los estudios de las áreas según sus estándares de información y datos.

Palabras clave: datos de búsqueda; Ciencia Abierta; comunicación científica.

1. 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 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 é destacada como parâmetro para compreensão, alargamento do diálogo nacional e internacional entre os pares.

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”.

2. 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, buscou-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.

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

O conceito atribuído a “dado de pesquisa” por Sales e Sayão (2018) é abrangente e nos auxiliará na reflexão sobre a prática de publicação bem como sua compreensão na literatura científica. De acordo com 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, 2018 p. 36).

A construção do panorama terminológico teve início com o trabalho de Lawrence *et al.* (2011), quando entendem 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)

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:

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).

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. De acordo com 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 a 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)

Os 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.14, tradução nossa).

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.

3.1 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 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 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).

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 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, 2022, *online*, tradução nossa).

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.

3.2 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).

Hrynaskiewicz e Shintani (2014) compartilharam seu entendimento sobre periódico de dados ao descreverem características e benefícios deste tipo de publicação. De acordo com 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).

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 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.³⁶ (GARCÍA-GARCÍA, LÓPEZ-BORRULL e PESET, 2015, p. 852).

Curty e Aventurier (2017) discutem novos formatos de publicação que estão em ascensão no contexto da Ciência Aberta e que vem 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 exaustivamente 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).

A reflexão teórica-conceitual proposta, a partir da teoria do conceito de Dalhberg (1978), busca compreensão desse universo complexo, por vezes novo, e quase sempre ambíguo, que envolve os dados de pesquisa e sua interoperabilidade semântica e tecnológica.

4. Considerações finais

A comunicação científica via publicação dos dados produzidos em pesquisas tende a aumentar e se apresenta sob diversos formatos, podendo ser terminologicamente compreendido 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.

Referências

CALLAGHAN, Sarah; MURPHY, Fiona; TEDDS, Jonathan; ALLAN, Rob; KUNZE, John; LAWRENCE, Rebecca; MAYERNIK, Matthew; WHYTE, Angus. A. Processes and procedures for data publication: a case study in the geosciences. **Int. J. Digital Curation**, United Kingdom, v. 8, n. 1, p. 193-203, 2013. DOI: <https://doi.org/10.2218/ijdc.v8i1.253>. Disponível em: <http://www.ijdc.net/article/view/8.1.193>. Acesso em: 19 fev. 2019.

BLOOM, Theodora; DALLMEIER-TIESEN, Sunje; MURPHY, Fiona; AUSTIN, Claire; WHYTE, Angus; TEDDS, Jonathan; NURNBERGER, Amy; RAYMOND, Lisa; STOCKHAUSE, Martina; VARDIGAN, Mary. Workflows for Research Data Publishing: Models and Key Components. **International Journal of Digital Libraries**, [s. l.], special edition, p. 1 - 28, 2015. DOI: <https://doi.org/10.5281/zenodo.20308>. Disponível em: <https://zenodo.org/record/20308#.Xs2UMjdKjGg>. Acesso em: 1 fev. 2019.

COSTAS, Rodrigo; MEIJER, Ingeborg; ZAHEDI, Zohreh; WOUTERS, Paul. **The Value of Research Data Metrics for datasets from a cultural and technical point of view**. Copenhagen, Denmark: 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_Knowledge_Exchange_Report_April_2013. Acesso em: 1 fev. 2019.

CURTY, Renata Gonçalves; AVENTURIER, Pascal. 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, Ingetraut. 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.

GARCÍA-GARCÍA, Alicia; LÓPEZ-BORRULL, Alexandre; PESET, Fernanda. Data journals: eclosión de nuevas revistas especializadas en datos. **El profesional de la información**, v. 24, n. 6, pp. 845-854, 2015. Disponível em: <https://doi.org/10.3145/epi.2015.nov.17>. Acesso em: 22 maio 2020.

HRYNASZKIEWICZ, Iain; SHINTANI, Yoko. Scientific Data: an open access and open data publication to facilitate reproducible research. **J-Stage**, Japan, v. 57, n. 9, p. 629-640, 2014. Disponível em: https://www.jstage.jst.go.jp/article/johokanri/57/9/57_629/article/-char/ja/. Acesso em: 22 maio 2020

LAWRENCE, Bryan; JONES, Catherine; MATTHEWS, Brian; PEPLER, Sam; CALLAGHAN, Sarah. 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. Disponível em: <http://www.ijdc.net/article/view/181>. Acesso em: 19 fev. 2019.

SALES, Luana Farias; SAYÃO, Luís Fernando. A ciência invisível: revelando os dados da cauda longa da pesquisa. *In*: ENCONTRO NACIONAL DE PESQUISA E PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO - ENANCIB, 19., 2018, Marília. **Anais [...]**. Marília: UNESP, 2018. p. 4180-4199. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/103678>. Acesso em: 17 maio 2022.

STOKER, Ryan; MELZACK, Gene; MCLEAN, Jennifer. Developing Culturally Competent Data Publication Resources. **International Journal of Digital Curation**, United Kingdom, 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. **Data publication**: home. Sydney, 4 Oct. 2022. Disponível em: <https://libguides.library.usyd.edu.au/datapublication>. Acesso em: 26 maio 2020.