

# A teoria da citação de dados: uma revisão da produção científica na América Latina

## *The theory of data citation: a review of scientific production in Latin America*

Caliel Cardoso de OLIVEIRA<sup>1</sup>  0000-0002-8974-0579

Maurício Coelho da SILVA<sup>1</sup>  0000-0002-7923-9457

Caterina Marta Groposo PAVÃO<sup>1</sup>  0000-0003-3712-7200

Fabiano Couto Corrêa da SILVA<sup>1</sup>  0000-0001-5014-8853

Ana Maria Mielniczuk de MOURA<sup>1</sup>  0000-0002-7811-4660

Thiago Henrique Bragato BARROS<sup>1</sup>  0000-0001-7439-5779

### Resumo

Trata-se de uma pesquisa bibliográfica, de caráter qualitativo, que buscou identificar o estado da arte acerca da teoria da citação dos dados na produção científica conduzida na América Latina. Para tanto, foram estabelecidas expressões em português, inglês e espanhol acerca da referida temática, que foram utilizadas para explorar as seguintes bases de dados, repositórios e buscadores: Biblioteca Digital Brasileira de Teses e Dissertações, OasisBR, *La referencia*, *Redalyc*, *Networked Digital Library of Theses and Dissertations*, Portal de Periódicos Capes, *Google Acadêmico*, SciELO e Brapci (Base de Dados Referenciais de Artigos de Periódicos em Ciência da Informação). Após a análise dos trabalhos recuperados, foram considerados somente aqueles que discutiam a temática de citação de dados de pesquisa de maneira aprofundada, com a finalidade de contribuir para a reflexão acerca de uma teoria da citação de dados, totalizando 19 trabalhos. Conclui-se que existe uma ausência significativa de trabalhos na América Latina concernente à teoria da citação de dados, ao mesmo tempo em que foram identificados trabalhos que, embora não se refiram a uma teoria propriamente, oferecem contribuições significativas para a temática de citação de dados de pesquisa e que podem servir de base para o desenvolvimento de trabalhos sobre a teoria da citação de dados. Constatou-se ainda que o Brasil se destacou na produção de trabalhos sobre citação de dados de pesquisa, sendo que dos 19 trabalhos analisados nesta pesquisa, 17 eram produções brasileiras.

**Palavras-chave:** Ciência Aberta. Ciência da Informação. Dados Abertos. Teoria da citação de dados.

### Abstract

*This is a qualitative bibliographic research that sought to identify the state of art regarding the theory of data citation in the scientific production conducted in Latin America. To this end, expressions were established in Portuguese, English, and Spanish about the referred theme which were used to explore the following databases, repositories, and searchers: Biblioteca Digital Brasileira de Teses*

<sup>1</sup> Universidade Federal do Rio Grande do Sul, Faculdade de Biblioteconomia e Comunicação, Programa de Pós-Graduação em Ciência da Informação. R. Ramiro Barcelos, 2705, Santana, 90035-007. Porto Alegre, RS, Brasil. Correspondência para/Correspondence to: C. M. G. PAVÃO. E-mail: <caterina@ufrgs.br>.

Recebido em 14 de setembro de 2021 e aprovado em 18 de novembro de 2021.

Como citar este artigo/How to cite this article

Oliveira, C. C. *et al.* A teoria da citação de dados: uma revisão da produção científica na América Latina. *Transinformação*, v. 34, e210062, 2022. <http://doi.org/10.1590/2318-0889202234e210062>



e Dissertações, Oasisbr, La referencia, Redalyc, Networked Digital Library of Theses and Dissertations, CAPES Journal Portal, Google Scholar, SciELO and Brapci (Reference Database of Journal Articles in Information Science). After the analysis of the retrieved works, only those papers that discussed the topic of citation of research data were considered in depth, in order to contribute to the reflection on a theory of data citation, totaling 19 papers. It is concluded that there is a significant absence of works in Latin America concerning the theory of citation of data; at the same time, works have been identified, although not referring to a theory itself, offering significant contributions to the topic of citation of research data and that can serve as a basis for the development of papers on the theory of data citation. It was also found that Brazil stood out in the production of papers on citation of research data, and of the 19 papers analyzed in this research, 17 were Brazilian productions.

**Keywords:** Open Science. Open Data. Data Citation Theory. Information Science.

## Introdução

A citação tem um papel fundamental na construção do discurso acadêmico-científico. Assim, trata-se de uma prática bastante recorrente na rotina da comunidade científica. Citar trabalhos é uma forma de reconhecer o esforço de pesquisadores que compartilham interesses em comum e, também, de auxiliar na construção e fundamentação teórica das pesquisas científicas. Áreas discursivas, como as ciências sociais e humanas, tendem a recorrer com frequência a artigos e outros textos, buscando solidez na construção argumentativa dos trabalhos, ao mesmo tempo em que também garantem o devido crédito aos estudos que servem como base para novas pesquisas. “*Standing on the shoulders of giants*”, como diz o famoso ditado newtoniano. No contexto da Ciência Aberta, as discussões acerca da citação tomam novas dimensões: passa-se também a discutir e problematizar a citação de dados abertos de pesquisa científica.

A citação de dados de pesquisa tem se desenvolvido como uma das iniciativas da Ciência Aberta, tencionando por uma ciência mais colaborativa, confiável e plural. Os dados de pesquisa se apresentam como uma fonte primária de conhecimento e o seu reuso possibilita que sejam analisados e discutidos sob diferentes perspectivas a partir de pesquisas com objetivos em comum; ou seja, a argumentação construída no parágrafo anterior ganha mais importância e relevância à medida que a ciência busca tornar-se mais transparente. Silvello (2018) aponta que são poucos os esforços da comunidade científica para o desenvolvimento de uma teoria da citação de dados. Essa teoria, no contexto da Ciência Aberta, mostra-se essencial para justificar e compreender de forma mais complexa a prática de citação de dados abertos, dessa forma também incentivando práticas mais colaborativas e confiáveis para uma comunicação científica mais inclusiva.

Nessa conjuntura, e a partir da problematização apresentada por Silvello (2018) acerca da ausência de trabalhos que discutam uma teoria da citação de dados, este estudo propõe-se a compreender como tem sido abordada a temática na produção científica conduzida na América Latina.

Parte-se de um levantamento bibliográfico de trabalhos produzidos na América Latina que discutem a referida temática e da análise das contribuições dessas produções para a discussão da teoria de citação de dados. Constatou-se que, embora exista de fato uma lacuna de trabalhos que discutam propriamente uma teoria, pesquisadores da América Latina têm dado contribuições significativas para as discussões no que concerne à temática de citação de dados. Tipologias de trabalhos consideradas no levantamento incluem artigos de periódicos científicos, trabalhos acadêmicos (TCC, teses e dissertações e projetos de especialização), capítulos de livros, relatórios de pesquisa e trabalhos publicados em eventos científicos. Buscou-se nas bases de dados, repositórios e buscadores BDTD (Biblioteca Digital Brasileira de Teses e Dissertações), OasisBR, La referencia, Redalyc, NDLTD (Networked Digital Library of Theses and Dissertations), Portal de Periódicos Capes, Google Acadêmico, SciELO e Brapci (Base de Dados Referenciais de Artigos de Periódicos em Ciência da Informação) e chegou-se a 19 trabalhos. Observou-se também que, em nível de América Latina, o Brasil se destacou com 17 produções de um total de 19. Essas contribuições, por sua vez, podem servir como base para outros trabalhos como este, que se proponham a fomentar a discussão sobre uma teoria de citação de dados na América Latina.

### Práticas por uma ciência mais colaborativa e confiável: do acesso aberto aos dados abertos

O movimento pela Ciência Aberta é, em sua essência, a contribuição da comunidade científica ao movimento maior pelo acesso e transparência da informação (Silva; Silveira, 2019b), que se expande cada vez mais, de mãos dadas com a expansão do acesso à internet e da cidadania atuante. Para além das iniciativas de Acesso Aberto do começo do século XXI, o movimento pela Ciência Aberta não apenas entende que a produção científica deve ser compartilhada e tornada acessível para a população comum, responsável por grande parte do financiamento das pesquisas, mas também que o próprio compartilhamento dará origem a novos saberes científicos por meio da integração de diferentes atores da Sociedade da Informação (Silva; Silveira, 2019a).

Borges e Sanz Casado (2021) definem a Ciência Aberta como um termo guarda-chuva sob a justificativa de que ela abrange discussões referentes a diversos processos decorrentes da ampla abertura da comunicação e da produção científica. Dos processos que compõem as práticas e discussões teóricas da Ciência Aberta, o acesso aberto e os dados abertos atuam como iniciativas com a finalidade de tornar as pesquisas e seus resultados mais reproduzíveis, confiáveis e também de desenvolver uma perspectiva mais colaborativa e inclusiva do processo de comunicação científica. Os autores situam a Ciência Aberta no contexto da pandemia de covid-19 ao afirmar tratar-se de um “momentum para ciência aberta” onde iniciativas de acesso aberto e dados abertos se mostram fundamentais ao mesmo tempo em que desafiam a comunidade científica a se aliar, colaborando e partilhando dados de pesquisa em busca de soluções para os fenômenos emergentes na pandemia.

No cerne do movimento pela Ciência Aberta encontra-se a problematização quanto à dinâmica de produção e comunicação dos saberes científicos, que frequentemente pautam-se por uma lógica mercadológica da ciência, priorizando a informação científica como bem econômico (Xavier; Costa, 2010) em vez de um fluxo saudável de divulgação de resultados e ideias. Nessa perspectiva, Fernández Pinto (2020) realizou uma crítica não somente ao mercado científico em sua perspectiva privada, mas à própria Ciência Aberta por fortalecer essas dinâmicas de produção e comunicação científica pautadas em uma perspectiva da ciência como bem econômico. Segundo Fernández Pinto (2020), ao disponibilizar resultados e dados de pesquisas realizadas com financiamento público em acesso aberto, pesquisas realizadas no âmbito privado podem se apropriar desses resultados e dados em benefício próprio, sem necessariamente disponibilizar em troca os resultados e dados das pesquisas desenvolvidas no setor privado. Nesse sentido, a crítica do autor aponta que as práticas da Ciência Aberta podem fortalecer a dinâmica de privatização da informação científica que o próprio movimento da Ciência Aberta antagoniza, ao mesmo tempo em que mostra a urgência de uma ciência cada vez mais aberta que também inclua o setor privado.

Essa não é uma preocupação recente. Já no final da década de 1980, universidades estadunidenses começaram a sofrer de um mal que já afetava instituições de outras partes do mundo, pois tornava-se quase impossível sustentar a assinatura de periódicos científicos de alta qualidade, ainda que os artigos publicados nessas revistas fossem produzidos dentro da própria instituição (Barbosa; Pavan, 2017). Isso ocorreu devido aos elevados preços cobrados pelas editoras para a assinatura de seus periódicos, apesar de grande parte do financiamento para as pesquisas ter origem no orçamento público (Barbosa; Pavan, 2017).

Como solução para esse problema, surgem os primeiros repositórios de *preprints*; ou seja, repositórios que permitem o compartilhamento de artigos antes de sua publicação final em periódicos tradicionais, permitindo uma etapa inicial de acesso e revisão por pares, sem os custos decorrentes da assinatura dos periódicos (Alvarez; Caregnato, 2017). O primeiro desses repositórios de *preprints* foi o Arxiv, ativo até os dias de hoje, que reúne conteúdo das ciências exatas. Muitos outros repositórios foram criados e seguem sendo utilizados.

Outra iniciativa basilar da Ciência Aberta foi o movimento pelo Acesso Aberto (AA). Encabeçado por diversas declarações, assinadas por membros da comunidade acadêmica em diferentes partes do globo, o movimento pelo AA preconizava a necessidade de facilitar o acesso à produção científica sem vincular esse acesso a condições

financeiras insustentáveis (Kuramoto, 2014). A primeira declaração do movimento AA foi a Declaração de Budapeste, divulgada em 2002, que oferecia dois caminhos para facilitar o acesso a artigos científicos pelo público que não possuía assinaturas de periódicos: o auto-arquivamento e os periódicos de acesso aberto. Sobre o primeiro método, a declaração *Budapest Open Access Initiative* (2002) afirma que, inicialmente, a comunidade acadêmica precisa de instrumentos que auxiliem no depósito dos seus trabalhos em repositórios eletrônicos abertos, enquanto os usuários não precisam saber quais desses repositórios existem ou sua localização desde que sejam facilmente identificáveis e seu conteúdo esteja acessível e recuperável.

Esse trecho denota, primeiramente, a importância de um ambiente acadêmico no qual os pesquisadores tenham as condições necessárias para divulgarem sua produção de maneira não-comercial, fomentando práticas de auto-arquivamento e compartilhamento dos saberes. Além disso, reitera a importância da acessibilidade da produção científica para com o público interessado, seja ele composto por pares ou por membros da sociedade em geral, para os quais a produção acadêmica deve estar disponível e alcançável sem necessidade de conhecimentos técnicos prévios.

Sobre os periódicos de acesso aberto, a declaração *Budapest Open Access Initiative* (2002) afirma que a comunidade científica precisa de recursos e incentivos para o desenvolvimento e consolidação de periódicos engajados com as práticas e ideias de acesso aberto, bem como a transição dos já existentes para a modalidade aberta.

A declaração *Budapest Open Access Initiative* (2002) ainda afirma que o *copyright* não mais seria utilizado como uma barreira de acesso aos trabalhos científicos, uma vez que os periódicos usariam os direitos do autor e as ferramentas necessárias para garantir o acesso à informação sem prejudicar os responsáveis pela produção. Aqui, observa-se a importância que se dá à circulação da produção científica em contraponto à rentabilidade dos periódicos tradicionais. Além disso, reforça-se a questão do acesso aberto, livre e permanente àquilo que é produzido pela Academia, especialmente considerando o caráter público dos investimentos que financiam essa produção.

Embora a Declaração de Budapeste tenha sido pioneira, outras declarações surgiram por todo o mundo, como a Declaração de Bethesda e a Declaração de Berlim, ambas em 2003; a Declaração de Salvador e a Carta de São Paulo, ambas de 2005, dentre outras (Barbosa; Pavan, 2017). Além disso, as mudanças propostas pelas declarações passaram a ser consideradas com seriedade dentro dos órgãos governamentais de fomento à pesquisa, como o IBICT, no Brasil, e até mesmo organizações de cooperação internacional, como a *United Nations Educational Scientific and Cultural Organization* (Unesco), que, impulsionada por iniciativas similares em seus países-membros, oferece cursos, ferramentas e diretrizes para a ampliação do acesso aberto à produção científica em todas as suas facetas, desde artigos de periódicos até anais de conferências e conjuntos de dados de pesquisa (United Nations Educational Scientific and Cultural Organization, 2015).

A aceitação do modelo AA pela comunidade científica pode ser observada no número de periódicos indexados pelo *Directory of Open Access Journals* (DOAJ), que reúne mais de 16.500 periódicos de acesso aberto ou híbridos (Directory of Open Access Journals, 2021). Revistas híbridas são aquelas que ainda exigem assinatura por parte de seus leitores, mas cobram taxas de processamento de artigos (as chamadas APCs) para permitirem a disponibilização imediata do estudo após sua publicação, sem um período de exclusividade (Barbosa; Pavan, 2017). É importante notar que as APCs foram a concessão criada pelas editoras de periódicos para permitirem a transição para o modelo AA sem grandes perdas financeiras, e são cobradas mesmo em muitas revistas totalmente abertas, que não exigem assinatura. Dessa forma, as APCs ainda se configuram num problema de caráter financeiro, visto que, por serem taxas pagas pelo autor do artigo para o periódico onde deseja publicar seu trabalho, dificultam a publicação de trabalhos de autores com poucos recursos institucionais ou sem apoio financeiro governamental em periódicos de grande alcance e renome (Abadal, 2021).

Os dados de pesquisa também têm sido tema de debate nas iniciativas de Ciência Aberta (Cunha; Costa, 2020). Vision (2010) afirma que os dados de pesquisa são um bem público e não perdem seu valor ao serem disponibilizados em domínio público, uma vez que vão servir de referência para outras pesquisas, permitindo o

refinamento de métodos de pesquisa e a reutilização desses dados de maneira criativa por parte da comunidade científica. Boulton *et al.* (2011) destacam que, em relação aos periódicos científicos, não é suficiente a presença de resumos a respeito dos dados disponíveis na pesquisa, e que somente a disponibilização completa dos dados em formato utilizável permite averiguar sua confiabilidade e sua reprodutibilidade.

Murray-Rust (2008) complementa ao apontar que os dados de pesquisa devem ser um bem comum compartilhado pela comunidade científica e que o direito de uso desses dados não deve ser limitado por periódicos científicos e editoras, mas estendido à comunidade científica. O autor também aponta que, em casos em que os editores impõem barreiras na abertura e uso dos dados, a comunidade científica pode se valer do uso e desenvolvimento de repositórios de dados abertos institucionais, departamentais, nacionais ou internacionais para contornar esse problema. Fecher e Friesike (2013) mencionam estudos que argumentam pela abertura de dados de pesquisa, incluindo o de Murray-Rust (2008), e destacam o pesquisador como principal responsável pela iniciativa de depositar seus dados em domínio público. Ao mesmo tempo, alertam para a necessidade de existência de sistemas de reconhecimento e recompensa que estimulem a prática de disponibilizar dados de pesquisa, tornando-a viral.

Rocha *et al.* (2018) utilizam a expressão “dados abertos e ligados” (em inglês *linked data*) para discutir um movimento de abertura dos dados onde eles são produzidos e interligados na web utilizando a infraestrutura da Web Semântica e tornando-se mais facilmente identificáveis. Já Ngomo *et al.* (2011) definem o termo “dados ligados” como um conjunto de boas práticas de publicação e interligação de dados estruturados na web introduzido por Tim Berners-Lee e que posteriormente se estabeleceu como os princípios de dados ligados. Nessa perspectiva, as discussões e iniciativas concernentes aos dados abertos de pesquisa científica têm avançado e proporcionado perspectivas mais favoráveis de pesquisas mais confiáveis, processos de comunicação e produção científica mais criativos, colaborativos, acessíveis e reprodutíveis.

## Perspectivas acerca de uma teoria da citação de dados

As citações são um dos pilares da comunicação científica. Embora o hábito de citar autores tenha sido uma prática durante toda a história da Ciência, foi na virada do século XIX para o XX que as citações se consolidaram como mecanismo de reconhecimento, debate e promoção das ideias dentro da comunidade acadêmica, graças à formalização das práticas de citação e ao volume cada vez maior de autores contemporâneos que eram citados por seus pares (Leydesdorff, 1998).

No entanto, embora as citações de autores e trabalhos acadêmicos seja uma prática amplamente difundida e estudada, o mesmo não pode ser dito da citação de dados. De fato, embora seja possível afirmar que tenham havido diversas tentativas de construção de uma Teoria da Citação, ao ponto dos “comos” e os “porquês” da citação já serem compreendidos pela comunidade acadêmica (Silvello, 2018), a Academia ainda carece da existência de uma teoria que se proponha a explicar a citação dos dados de pesquisa que, como exposto, são cada vez mais requisitados e utilizados pela comunidade acadêmica.

A ausência de uma teoria acerca das citações não impede uma análise quantitativa das mesmas (elaborando, por exemplo, índices que apontam quais os trabalhos mais citados em dada área, no caso das citações tradicionais), mas sem uma clareza acerca do que se quer explicar com essas análises torna-se difícil justificar e melhorar sua prática (Leydesdorff, 1998), o que levou à diversas investigações sobre as citações acadêmicas tradicionais, seus agentes e suas motivações. No entanto, o mesmo ainda não ocorreu com as citações de dados de pesquisa. Ao deixar de preencher essa lacuna, a comunidade acadêmica corre alguns riscos, como o de trabalhar com a citação de dados abertos sem entender de fato o que essas citações implicam. Segundo Silvello (2018), o uso de indicadores métricos para avaliar e atribuir relevância à produção científica sem uma teoria que sirva como base para os “porquês” dessa avaliação ser adequada pode ser prejudicial à comunidade científica.

Alguns autores e estudos têm realizado esforços para compreender aspectos da prática de citação de dados que podem servir como base para a discussão de teorias acerca da citação de dados - como é o caso do trabalho de Silveira *et al.* (2020), que discutiu as motivações que levam pesquisadores, principalmente da ciência da informação, a citarem dados ou se sentirem inibidos em citar. Segundo Silveira *et al.* (2020), a principal motivação dos pesquisadores ao citar dados científicos é que essa prática gera uma perspectiva de obter mais colaboração, rastreabilidade, visibilidade e reconhecimento entre os pares, enquanto aspectos como a falta de recursos, um volume muito grande de dados que dificultam sua formatação e limpeza e a ausência de um sistema de recompensas pela disponibilização dos dados podem levar pesquisadores a optarem por não fazê-lo.

Silveira *et al.* (2020) ainda acrescentam que o ato de citar dados de pesquisa científica ultrapassa questões de ordem técnica, uma vez que lida com questões de colaboratividade, inovação e respeito aos direitos autorais dos achados citados. Nessa mesma perspectiva, Sayão e Sales (2014) apontam que citar dados é uma forma de atribuir valor à pesquisa e prover dinâmicas interdisciplinares, compensando barreiras por teóricas e geográficas. Também têm ocorrido avanços no desenvolvimento de ferramentas de identificação e citação de dados, como a *Data Citation Index* (DCI), da *Web of Science* (WoS); a *Mendeley Data*, da Elsevier em parceria com o DataCite e o *Google Dataset Search* (GDS) (Silveira *et al.*, 2020).

Outro aspecto bastante recorrente na literatura são os critérios e padrões necessários, discutidos por Cousijn *et al.* (2018), para que os dados de pesquisa se tornem acessíveis para citação com os princípios FAIR: Localizáveis (*Findable*); Acessíveis (*Accessible*); Interoperáveis (*Interoperable*) e reutilizáveis (*Re-usable*) (Silveira *et al.*, 2020). Silva e Silveira (2019a) discutem os dados abertos como um dos movimentos dentro da Ciência Aberta e o potencial que eles têm por permitirem sua reutilização por pesquisadores que possuem objetivos de estudo em comum. Os autores ainda apontam que, para atingir o máximo de abertura de dados de pesquisas científicas, é necessário que os pesquisadores os disponibilizem por meio de produtos em formatos não proprietários, citando como exemplo o csv ao invés do Excel, juntamente com licenças que tornem claras as utilizações dos dados e sua citação (Silva; Silveira, 2019a).

Nessa conjuntura, embora discussões acerca de teorias da citação de dados ainda sejam pouco abordadas e aprofundadas pela comunidade científica, conforme argumentou Silvello (2018), a literatura científica discute aspectos da citação de dados que são significativamente relevantes para possíveis teorias da citação de dados.

## Procedimentos Metodológicos

Este trabalho trata-se de uma pesquisa bibliográfica, de caráter qualitativo, que visou a explorar as principais bases de dados e repositórios latino-americanos para observar o estado da arte nas discussões acerca da citação de dados de pesquisa, em especial quanto aos aspectos relativos a uma possível teoria de citações de dados. De acordo com Sousa, Oliveira e Alves (2021), a pesquisa bibliográfica ocorre por meio do levantamento ou revisão de obras cujo conteúdo temático é delimitado a partir de uma teoria principal, sendo que a análise dessas obras é o método científico utilizado nesse tipo de pesquisa para o desenvolvimento da pesquisa e da discussão proposta.

Dessa forma, este trabalho se desenvolveu em consonância com as etapas da pesquisa bibliográfica colocadas por Marconi e Lakatos (2003), pois envolveu a identificação e seleção das bases a serem pesquisadas, a localização, a compilação e subsequente análise dos trabalhos pertinentes, e, por fim, a interpretação dos resultados obtidos e de seu conteúdo.

As bases de dados e repositórios explorados foram principalmente as latino-americanas de acesso aberto, sendo elas a Biblioteca Digital Brasileira de Teses e Dissertações (BDTD), a OasisBR, a *La referencia* e a *Redalyc*. Para uma maior abrangência de busca e para evitar perder trabalhos significativamente interessantes, também foram

utilizados a base de dados *Networked Digital Library of Theses and Dissertations* (NDLTD), o Portal de Periódicos Capes e o *Google Acadêmico*, sempre que possível aplicando filtros para recuperar somente os trabalhos produzidos na América Latina. Posteriormente às primeiras buscas, em um segundo momento, foram adicionadas à pesquisa e realizadas buscas também na base de dados Brapci e na SciELO.

Quanto às tipologias dos trabalhos, foram considerados artigos de periódicos científicos, trabalhos acadêmicos (TCC, teses e dissertações e projetos de especialização), capítulos de livros, relatórios de pesquisa e trabalhos publicados em eventos científicos. Quanto aos TCCs, embora eles tenham sido considerados nos levantamentos, não foram analisados, mas constam nos quadros elaborados. O foco foram trabalhos da área da ciência da informação e da biblioteconomia, mas trabalhos de outras áreas do conhecimento também foram selecionados quando, por meio de análise, eram considerados relevantes para a temática. Também; foi aplicado um filtro temporal, abrangendo o ano de 2018 até julho de 2021, por se tratar de um estudo que partiu da problematização de Silvello (2018) acerca da ausência de trabalhos sobre a teoria da citação de dados. Dessa forma, foram priorizados trabalhos posteriores ao de Silvello (2018).

O levantamento da bibliografia foi realizado a partir de expressões pré-estabelecidas acerca da temática de teoria de citação de dados. O termo de busca preferido foi “teoria da citação de dados”, seguido da variável “citação de dados”. Também foram estabelecidas variantes das expressões em inglês e espanhol, elaboradas a partir da revisão de literatura e considerando o escopo das bases e repositórios pesquisados, conforme o Quadro 1.

**Quadro 1** - Expressões de busca utilizadas nas bases de dados.

Expressão em Português	Variantes em inglês	Variantes em espanhol
Teoria da citação de dados	<i>Data citation theory</i>	<i>Teoria de la cita de datos</i>
Citação de dados	<i>Data citation</i>	<i>Cita de datos</i>

Fonte: Elaborado pelos autores (2021).

Após o levantamento bibliográfico, foram realizadas duas análises para selecionar os trabalhos considerados relevantes à temática proposta. A primeira análise foi feita a partir da leitura do título, do resumo e, quando necessário, do restante da estrutura do trabalho para identificar aqueles que de fato estavam discutindo aspectos significativamente relevantes acerca da teoria da citação de dados. A segunda análise foi mais detalhada, realizada após a seleção, e englobou a leitura dos trabalhos no todo, buscando compreender de que forma eles contribuem para a discussão acerca da referida temática.

## Resultados

A seguir, serão apresentados os resultados dos levantamentos realizados nas bases de dados, repositórios, Portal de Periódicos Capes e *Google Acadêmico*. Ressalta-se que foram realizados levantamentos que não foram incluídos nesta seção: na base de dados NDLTD, não foram obtidos resultados relevantes em português e tampouco foi possível utilizar os filtros da base para obter um volume manejável de resultados em inglês e espanhol devido a problemas técnicos da própria base. Além dela, posteriormente às primeiras buscas foram incluídas à pesquisa as bases de dados BRAPCI e a SciELO, mas ambas não retornaram resultados relevantes além daqueles já identificados nas outras bases de dados. Por esses motivos, a NDLTD, a BRAPCI e a SciELO não constam descritas nesta seção do artigo.

Em todas as bases de dados, repositórios e no *Google Acadêmico* foram utilizadas expressões pré-estabelecidas e suas versões em inglês e espanhol, além de terem sido aplicados filtros quando estavam disponíveis, variando de acordo com a base de dados ou o repositório. Os filtros utilizados estão devidamente descritos nas seções a seguir. A busca pelas expressões entre aspas não recuperou resultados, com exceção do *Google Acadêmico*, onde a expressão “citação de dados” entre aspas permitiu recuperar trabalhos mais relevantes. Salienta-se ainda que frequentemente aconteceu de um mesmo trabalho ser localizado em mais de uma base de dados ou repositório, motivo pelo qual pode ocorrer a reincidência de resultados em diferentes quadros. Foram recuperados um total de 19<sup>2</sup> trabalhos, devidamente descritos nas subseções a seguir.

## Biblioteca Digital Brasileira de Teses e Dissertações

Durante a pesquisa na Biblioteca Digital Brasileira de Teses e Dissertações (BDTD), inicialmente buscou-se pela expressão “teoria da citação de dados”, obtendo 54 resultados. Observando que o buscador realizava a busca identificando cada termo da expressão em separado, optou-se por buscar também pela expressão entre aspas, visando a recuperação exata da expressão. Essa busca entre aspas, porém, não devolveu resultados. Dessa forma, foram analisados os 54 resultados trazidos pela primeira busca, destacando o que era pertinente a esta pesquisa.

Também foi feita uma busca pela expressão variante “citação de dados” com a aplicação de um filtro de temporalidade, resgatando apenas resultados dos últimos três anos (ou seja, a partir de 2018). O filtro de tempo foi aplicado pois considerou-se que remover o termo “teoria” da expressão acabaria por aumentar o número de resultados recuperados, então o filtro de tempo atuaria como contrapeso, reduzindo o volume de resultados. Foram obtidos 148 resultados nessa busca - um volume considerável, mas possível de ser trabalhado-, e constatou-se a presença de um trabalho relevante entre os resultados recuperados. No Quadro 2 foi listado o trabalho selecionado como relevante para a discussão proposta por essa pesquisa.

Quadro 2. Trabalhos selecionados na BDTD.

Instituição de Ensino Superior	Periódico/Evento	Autor/ano	Título do trabalho	Termo utilizado
Universidade Federal de Minas Gerais	N/A - Tese de doutorado	Almeida (2019)	Suporte à gestão de dados de pesquisa: uma ampliação dos serviços oferecidos pelas bibliotecas	Citação de dados

Fonte: Elaborado pelos autores (2021).

A seguir serão apresentados os resultados das buscas realizadas no portal brasileiro OasisBR, que reúne produção científica nacional em acesso aberto.

## OasisBR

O termo “teoria da citação de dados” recuperou 477 resultados, enquanto o termo “citação de dados” recuperou 1.643 resultados. A busca pelo termo “teoria da citação de dados” entre aspas não recuperou resultados, enquanto “citação de dados” entre aspas recuperou cinco resultados, com dois registros duplicados e dois trabalhos pertinentes referentes à temática de citação de dados de pesquisa. Filtrando pela área do conhecimento da Ciência

<sup>2</sup> Destes 19 trabalhos recuperados, 4 consistem em trabalhos de conclusão de curso, razão pela qual foram inseridos nos quadros, mas não foram incluídos nas análises nem nas referências do artigo.



da Informação, “teoria da citação de dados” recuperou um trabalho, enquanto “citação de dados” recuperou quatro resultados, sendo que dois estavam duplicados.

O filtro por área do conhecimento não se mostrou muito confiável, uma vez que as produções científicas são registradas em diferentes áreas do conhecimento (em Ciências Sociais Aplicadas ou especificamente na Ciência da Informação) ou não possuem registro em áreas do conhecimento. O trabalho de Silveira *et al.* (2020), por exemplo, era excluído dos resultados quando o filtro Ciência da Informação era aplicado, embora se trate de um artigo da área. Dessa forma, foi aplicado o filtro de tempo para trabalhos desenvolvidos a partir de 2018, recuperando 548 resultados. Para tornar viável a análise do grande número de trabalhos recuperados, foi aplicado o filtro por revistas, bases de dados ou repositórios da Ciência da Informação ou interdisciplinares.

Foram analisados os artigos publicados nas revistas *Em Questão*, *Encontros Bibli*, *Informação & Informação*, *Informação em pauta*, *Revista Digital de Biblioteconomia e Ciência da Informação*, *Revista ibero-americana de Ciência da Informação* e *Brazilian Journal of Information Science*. Com relação às revistas, bases de dados ou repositórios interdisciplinares, foram analisados o repositório institucional da Unesp, Repositório Institucional da UFC, ForScience: Revista Científica do IFMG, Repositório Institucional da Unipampa, Biblioteca Digital de Teses e Dissertações da UFG, Repositório Científico de Acesso Aberto de Portugal, Repositório Institucional da UFG, Repositório Institucional da UFMG e as bibliotecas digitais de teses e dissertações da Universidade Federal Rural do Semi-Árido (UFERSA), Universidade Federal do Amazonas (UFAM), Universidade Federal do Rio Grande do Sul (UFRGS), Universidade de São Paulo (USP) e do Instituto Presbiteriano Mackenzie.

Filtrando por essas fontes de informação, o número de trabalhos diminuiu para 264, sendo que todos foram analisados. Entretanto, a maioria deles referiam-se à temática de bibliometria e somente três trabalhos que discutiam a citação de dados de pesquisa foram identificados. As buscas através dos termos em inglês e em espanhol não recuperaram novos resultados relevantes. Os trabalhos identificados como pertinentes à temática de teoria da citação de dados encontram-se no Quadro 3.

**Quadro 3** - Trabalhos selecionados na OasisBR.

Instituição de Ensino Superior	Periódico/Evento	Autor/ano	Título do trabalho	Termo utilizado
Universidade Federal de Santa Catarina	Encontros Bibli	Silveira <i>et al.</i> (2020)	Citação de dados científicos: <i>scoping review</i>	Citação de dados

Fonte: Elaborado pelos autores (2021).

Na próxima seção serão apresentados os trabalhos recuperados na rede latinoamericana de repositórios de acesso aberto, a *La Referencia*.

## La Referencia

Inicialmente, realizou-se a busca pela expressão “teoria da citação de dados”, recuperando 66 trabalhos. Ao analisá-los, constatou-se que nenhum deles era pertinente à temática proposta por este estudo. A busca pela expressão entre aspas não retornou resultados relevantes. Em seguida, foi feita a busca pela expressão variante “citação de dados”, obtendo 1,051 resultados. Aplicando o filtro de tempo 2018-2021, o número de resultados foi reduzido para 425. Devido à ausência de um filtro por área do conhecimento, todos os 425 trabalhos recuperados foram analisados, dentre os quais foram identificados três trabalhos úteis.

A seguir, foram realizadas as buscas pelas variações das expressões em inglês. A expressão “*data citation theory*” retornou 166 resultados. Com a aplicação do filtro de tempo de 2018 até 2021, o número de trabalhos recuperados diminuiu para 58, porém não foram identificados trabalhos relevantes. Buscando pela expressão exata entre aspas, também não se obteve resultados. A expressão “*data citation*” recuperou 1.390 resultados, e com filtro de tempo, esse número diminuiu para 462. Foram identificados dois trabalhos relevantes, porém eles já haviam sido recuperados na busca com expressão em português, “citação de dados”. Nenhum novo resultado foi identificado com as expressões em espanhol. Os trabalhos identificados como relevantes podem ser observados no Quadro 4.

**Quadro 4** - Trabalhos selecionados na La referencia.

Instituição de Ensino Superior	Periódico/Evento	Autor/ano	Título do trabalho	Termo utilizado
Universidade Federal de Santa Catarina	Encontros Bibli	Silveira <i>et al.</i> (2020)	Citação de dados científicos: <i>scoping review</i>	Citação de dados
Universidade Federal de Minas Gerais	N/A - Tese de doutorado	Almeida (2019)	Suporte à gestão de dados de pesquisa: uma ampliação dos serviços oferecidos pelas bibliotecas	Citação de dados
Universidade Federal do Rio Grande do Sul	Transinformação	Silva e Silveira (2019)	O ecossistema da ciência aberta	Citação de dados

Fonte: Elaborado pelos autores (2021).

A seção a seguir procura retratar os resultados obtidos durante a busca na *Redalyc*, um sistema de indexação idealizado pela *Universidad Autónoma del Estado de México* e que integra a produção científica da região de maneira a facilitar o fluxo de comunicação científica.

## Redalyc

Inicialmente foi feita a busca pela expressão principal da pesquisa, “teoria da citação de dados”, mas não foram obtidos resultados, nem mesmo com o termo entre aspas. A seguir, foi feita a busca com a variante em português, *citação de dados*, que trouxe 25 resultados, dos quais somente um foi considerado relevante para a discussão acerca da teoria da citação de dados.

A seguir, tiveram início as buscas com o termo em outros idiomas, começando pela língua inglesa. Buscando pela expressão *data citation theory*, obteve-se 0 resultados. Em seguida, foi buscada a expressão *data citation*, para a qual foram obtidos 30 resultados, os quais foram analisados em busca de trabalhos pertinentes a esta pesquisa. Apenas um trabalho foi considerado útil.

Finalmente, foram feitas buscas pelas expressões na língua espanhola. A busca por *teoría de la cita de datos* não devolveu resultado. Já a busca por *cita de datos* devolveu 202 resultados, motivo pelo qual foi aplicado o filtro de temporalidade, observando apenas trabalhos publicados a partir de 2018. No entanto, provavelmente devido a algum erro de programação da base, a aplicação desses filtros faz com que a base recupere dezenas de milhares de trabalhos, que supõe-se serem todos aqueles publicados a partir de 2018, sem considerar a expressão inserida para a busca. Problema similar acontece quando é feita a busca com a expressão entre aspas, com milhares de resultados sendo recuperados, reforçando a ideia de que talvez haja alguma instabilidade no buscador da base. Por esses motivos, foram analisados os 202 títulos recuperados, sendo que apenas um artigo foi considerado relevante, e que esse artigo já havia sido localizado em pesquisas anteriores. No Quadro 5, podem ser observados os trabalhos considerados pertinentes à pesquisa.

Na próxima seção apresentam-se os resultados recuperados no Portal de Periódico da Capes, criado pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes) com o objetivo de democratizar o acesso à informação científica em instituições de ensino superior, fomentando o desenvolvimento científico.

**Quadro 5** - Resultados da *Redalyc*.

Instituição de Ensino Superior	Periódico/ Evento	Autor	Título do trabalho	Termo utilizado
Universidade Federal de Santa Catarina	Encontros Bibli	Silveira <i>et al.</i> (2020).	Citação de dados científicos: <i>scoping review</i>	Citação de dados
Universidad del Valle	Ingeniería y competitividad	Machuca-Martínez (2020).	<i>Importance of scientific data and its publication as data paper</i>	<i>Data citation</i>

Fonte: Elaborado pelos autores (2021).

## Portal de Periódicos Capes

No Portal de Periódicos da Capes foi realizada a busca por assunto. A busca utilizando a expressão *teoria da citação de dados* resultou em 772 trabalhos recuperados. Por ser um número muito grande de resultados, o que tornava inviável a análise manual de cada um, foi aplicado o filtro por tópicos disponibilizado pelo Portal. Os tópicos selecionados foram *information science* e *bibliometrics*, juntamente do filtro de tempo de 2018 a 2021. Os trabalhos recuperados diminuíram para 16, mas não foram identificados resultados relevantes entre eles. Quando se buscou pelo termo entre aspas, não foram recuperados resultados.

A expressão *citação de dados* recuperou 1.309 resultados, e o filtro de tempo reduziu para 343 resultados. Foram aplicados os filtros por tópico: *information science*, *library & information science*, *ciência da informação* e *bibliometrics*, diminuindo para 44 o número de trabalhos recuperados. Apenas um trabalho pertinente sobre a temática foi identificado, o qual já havia sido recuperado anteriormente nas outras bases e repositórios.

A expressão em inglês *theory of data citing* não recuperou resultados relevantes. Em seguida, foi utilizada a expressão *data citation* e uma quantidade massiva de trabalhos foram recuperados, totalizando 4.341.641. Aplicando o filtro de tempo, esse número diminuiu para 876.953 resultados. Como nesse caso os resultados foram muitos, optou-se por realizar a pesquisa usando somente o termo entre aspas. O número de resultados diminuiu para 1.262. Por ainda ser um número muito grande de trabalhos a serem analisados, foi aplicado primeiro o filtro pelos tópicos mais gerais ou que mais se aproximavam dos estudos em ciência da informação, excluindo tópicos muito específicos de outras temáticas ou áreas do conhecimento.

Os tópicos selecionados e que permaneceram foram: *sciences (general)*, *research article*, *research and analysis methods*, *social sciences* e *bibliometrics*. O número de trabalhos recuperados diminuiu para 767. Para filtrar ainda mais os resultados, foi utilizado o filtro de tópicos para excluir os trabalhos agrupados nos seguintes tópicos: *physical sciences*, *medicine and health sciences* e *biology and life sciences*. Restou um total de 310 trabalhos, dos quais apenas um foi considerado relevante. As buscas pelas expressões em espanhol não identificaram novos trabalhos relevantes. O Quadro 6 mostra o único resultado relevante identificado no Portal de Periódico da Capes, sendo que o trabalho já havia sido identificado em outras bases de dados.

Na seção seguinte são apresentados os resultados do levantamento bibliográfico no *Google Acadêmico*. O buscador virtual foi o que propiciou um maior número de trabalhos recuperados referentes à temática proposta pela pesquisa.

**Quadro 6** - Resultados do Portal de Periódicos Capes.

Instituição de Ensino Superior	Periódico/Evento	Autor	Título do trabalho	Termo utilizado
Universidade Federal de Santa Catarina	Encontros Bibli	Silveira <i>et al.</i> 2020.	Citação de dados científicos: <i>scoping review</i>	Citação de dados

Fonte: Elaborado pelos autores (2021).

## Google Acadêmico

A busca por *teoria da citação de dados* recuperou 195.000 resultados e, mesmo aplicando o filtro de tempo, ainda foram recuperados 15.500 trabalhos, sendo que a expressão entre aspas não retornou resultados. Para tornar viável uma análise manual, optou-se por realizar a busca no Google Acadêmico utilizando somente a expressão *citação de dados*, que recuperou 27.500 resultados mesmo com o filtro de tempo. Para obter uma precisão maior na busca, decidiu-se utilizar os termos entre aspas e esse número diminuiu para 131 trabalhos. Desses, apenas 15 trabalhos pertinentes à temática proposta foram identificados após a análise.

A busca por *theory of data citing* retornou 110.000 resultados. Buscando pelo termo entre aspas, esse número diminuiu para um resultado, sendo o texto do Silvello (2018). O termo *data citation* recuperou 277 mil resultados, por isso optou-se pela expressão entre aspas, que recuperou 6.170 resultados. Foram analisados os primeiros 100 resultados e constou-se que, embora fossem trabalhos interessantes e relevantes para a temática, não eram da América Latina. Como o Google Acadêmico não possibilita usar o filtro por regiões e acaba se limitando ao filtro de tempo e de idioma, sendo que o primeiro já estava aplicado e o segundo se limitava ao filtro pelo idioma português brasileiro, uma vez que a expressão estava em inglês, a busca pelos termos em inglês no Google Acadêmico acabou sendo inviável e foi descartada. Quanto às expressões em espanhol, somente a expressão *cita de datos* se mostrou útil, recuperando 5.130 resultados, que diminuíram para 16 quando foi realizada a pesquisa com o termo entre aspas. Desses 16 resultados, apenas um deles se mostrou útil (Quadro 7).

**Quadro 7** - Resultados do Google Acadêmico.

1 de 2

Instituição de Ensino Superior	Periódico/Evento	Autor	Título do trabalho	Ano	Termo utilizado
Universidade Federal de Santa Catarina	Encontros Bibli	Silveira <i>et al.</i> (2020)	Citação de dados científicos: <i>scoping review</i>	2020	"Citação de dados"
N/A	Liinc em Revista	Costa e Cunha (2019)	A necessidade de uma política nacional para a gestão de dados de pesquisa no Brasil	2019	"Citação de dados"
N/A	Liinc em Revista	Dias, Anjos e Araújo (2019)	A gestão dos dados de pesquisa no âmbito da comunidade dos pesquisadores vinculados aos programas de pós-graduação brasileiros na área da Ciência da Informação: desvendando as práticas e percepções associadas ao uso e reuso de dados.	2019	"Citação de dados"
Universidade Federal do Rio de Janeiro	I Encontro Da Rede Sudeste De Repositórios Institucionais	Santos (2019)	Competência em Informação para a gestão de dados de pesquisa	2019	"Citação de dados"

Quadro 7 - Resultados do Google Acadêmico.

2 de 2

Instituição de Ensino Superior	Periódico/Evento	Autor	Título do trabalho	Ano	Termo utilizado
Instituto de Comunicação e Informação Científica e Tecnológica em Saúde	Revista Ciência da Informação	Veiga <i>et al.</i> (2019)	Compartilhamento de dados de pesquisa na Fiocruz: diagnóstico e percepção do pesquisador	2019	"Citação de dados"
Instituto de Comunicação e Informação Científica e Tecnológica em Saúde	N/A - Projeto de especialização	Ramos (2018)	Data papers e o sistema de recompensas na avaliação da ciência	2018	"Citação de dados"
Universidade Federal da Bahia	PontodeAcesso	Sayão e Sales (2018)	Subsídios para a Construção de um modelo de avaliação de sistemas de gestão de dados de pesquisa	2018	"Citação de dados"
Instituto de Comunicação e Informação Científica e Tecnológica em Saúde	N/A - Projeto de especialização	Araújo (2019)	Metadados e a ampliação do acesso aos dados de pesquisa nos Repositórios digitais do campo da saúde	2019	"Citação de dados"
Universidade Federal do Rio Grande do Sul	N/A - Relatório técnico e de pesquisa	Rocha <i>et al.</i> (2018)	Acesso aberto a dados de pesquisa no Brasil: soluções tecnológicas para compartilhamento de dados no Brasil: relatório	2018	"Citação de dados"
Universidade Federal do Paraná	N/A - Capítulo de livro	Torino; Roa-Martínez e Vidotti (2020)	Dados de pesquisa: disponibilização ou publicação?	2020	"Citação de dados"
Universidade de Brasília	N/A - TCC	Silva (2019)	Gestão de dados científicos sob a ótica da ciência aberta: uma análise de planos de gestão de dados europeus	2019	"Citação de dados"
Universidade de Brasília	N/A - TCC	Bonetti (2019)	Serviços de gestão de dados de pesquisa em bibliotecas universitárias brasileiras	2019	"Citação de dados"
Universidade Federal de Minas Gerais	N/A - Dissertação de mestrado	Resende (2019)	A curadoria de dados científicos na ciência da informação: levantamento do cenário nacional	2019	"Citação de dados"
Universidade Federal de Santa Catarina	N/A - Capítulo de livro	Santos e Nassi-Calò (2020)	Gestão editorial: tendências e desafios na transição para a ciência aberta	2020	"Citação de dados"
Universidade Federal de Minas Gerais	N/A - Tese de doutorado	Almeida (2019)	Suporte à gestão de dados de pesquisa: uma ampliação dos serviços oferecidos pelas bibliotecas	2019	"Citação de dados"
Instituto de Comunicação e Informação Científica e Tecnológica em Saúde	N/A - Tese de doutorado	Guanaes (2020)	Abertura e compartilhamento de dados de pesquisa subjacentes a artigos científicos: questões do direito autoral	2020	"Citação de dados"
Udelar	N/A - Tese de mestrado	Sienra (2018)	<i>Comportamiento informativo colaborativo y prácticas al compartir datos sobre biodiversidad marina</i>	2018	"Cita de datos"

Fonte: Elaborado pelos autores (2021).

Na seção a seguir é descrita a análise dos trabalhos identificados como relevantes, com o objetivo de identificar de que forma os trabalhos recuperados discutem a citação de dados e contribuem para a discussão acerca de teorias da citação de dados.

## Discussão

Em linhas gerais, observou-se uma prevalência de trabalhos que discutiam as práticas de citação e gestão de dados, em alguns casos sendo parte do mesmo objeto de estudo. Pouco foi encontrado acerca da necessidade de existência de uma teoria sobre a citação de dados ou que propusesse a dita teoria. Ainda assim, alguns trabalhos que tratavam especificamente da citação de dados e de seus aspectos enquanto boas práticas científicas abordam temas importantes e pertinentes à problemática de uma teoria da citação dos dados. Destaca-se ainda a predominância das produções brasileiras nos resultados encontrados, sendo que 17 dos 19 trabalhos estavam em português. A seguir serão analisados os trabalhos considerados relevantes.

Explorando como as bibliotecas e funções informacionais podem auxiliar na gestão de dados de pesquisa oferecendo serviços referentes a essa gestão para pesquisadores e estudantes, Almeida (2019) devota uma seção inteira para tratar da citação de dados. Trazendo diretrizes internacionais sobre o tema, ela retoma a importância da citação de dados como boa prática científica, além de dar ênfase ao papel da citação de dados na reprodutibilidade das pesquisas. A autora ainda aborda aspectos técnicos necessários à citação, como o uso de identificadores persistentes e a problemática que surge pela ausência de uma sintaxe padronizada para que sejam feitas essas citações. Conclui a seção sobre citação de dados trazendo os princípios propostos pelo *Data Citation Synthesis Group* para o desenvolvimento de práticas e ferramentas de citação de dados.

Desenvolvendo pesquisa similar a essa, mas maior em escopo, Silveira *et al.* (2020) executaram uma revisão bibliográfica acerca da temática da citação de dados, trazendo importantes contribuições, tanto em caráter quantitativo e de revisão, quanto no âmbito qualitativo, desenvolvendo suas conclusões a partir do que analisaram. Expõem diferentes realidades acerca do uso e da citação de dados nas diferentes áreas do conhecimento, trazendo os resultados encontrados em sua revisão. Acerca das motivações que levam os pesquisadores a citarem conjuntos de dados científicos, afirmam: “A motivação principal ao citar dados científicos é a possibilidade de obter mais colaboração, rastreabilidade e, conseqüentemente, impacto e visibilidade” (Silveira *et al.*, 2020, p. 8). Em contraponto, também expõem as motivações que levam os pesquisadores a não divulgarem seus dados para citação, variando desde questões técnicas (falta de recursos ou ferramentas, grandes volumes de dados) até pessoais (preocupações acerca de danos à reputação acadêmica no caso de refutações, *etc*), conforme exposto na literatura revista. Concluem reiterando a importância de haver condições técnicas que permitam o compartilhamento padronizado dos dados e políticas que satisfaçam as necessidades de uma área do conhecimento e que permitam a utilização interdisciplinar dos dados.

O trabalho de Silva e Silveira (2019a) traz a importância dos dados abertos como dimensão imprescindível no contexto da Ciência Aberta. Os autores afirmam que os dados científicos brutos, quando disponibilizados de maneira estruturada em bibliotecas de dados, tornam-se “fontes primárias” (Silva; Silveira, 2019a, p. 5) para os pesquisadores. Trazem ainda recomendações técnicas para o compartilhamento e futuro reuso desses dados, denotando a importância do comprometimento do pesquisador para com a abertura e transparência do processo científico.

O trabalho de Machuca-Martínez (2020) traz a perspectiva da publicação dos artigos na forma de *data papers*; ou seja, artigos específicos de dados. Consoante a outros autores abordados anteriormente, afirma que os dados científicos deixam de ser uma informação secundária, de suporte, para se tornarem o foco de muitas publicações, com grandes implicações para “[...] a visibilidade, reuso de informação, citação e reconhecimento, uso de recursos públicos, entre outros aspectos” (Machuca-Martínez, 2020, p. 1, tradução nossa).

Costa e Cunha (2019), em seu trabalho sobre propostas para uma política nacional brasileira de gestão de dados, abordam aspectos técnicos acerca da disponibilização dos dados para reuso e citação. Reiteram a necessidade de que existam condições tecnológicas que permitam essa dinâmica de compartilhamento e reuso, mas apontam também que essa demanda de *hardware* se torna uma problemática cada vez menor à medida que a tecnologia avança. Destacam também a importância de se terem claras as condições éticas e legais em torno dos dados disponibilizados para uso. Trazem dados acerca da citação dos dados e da presença crescente dos conjuntos de dados citados na seção de referências dos trabalhos.

O trabalho de Dias, Anjos e Araújo (2019), que questiona pesquisadores atuantes da CI acerca de práticas de uso e reuso de dados, traz uma contribuição valiosa para a discussão da citação de dados. A maioria dos respondentes disse concordar fortemente com a afirmação “[...] é importante que meus dados sejam citados quando utilizados por outros pesquisadores”, o que, na visão dos autores, envolve o recebimento de crédito e denota que “[...] políticas nesse sentido devem ser desenvolvidas para estimular os pesquisadores a disponibilizarem os seus dados” (Dias; Anjos; Araújo, 2019, p. 22).

Santos (2019), em seu trabalho apresentado no I Encontro da Rede Sudeste de Repositórios Institucionais, aborda a questão das “competências de dados” por parte dos pesquisadores, trazendo à tona o papel dos profissionais mediadores da informação (bibliotecários, arquivistas, gerentes de dados) na conscientização dos pesquisadores quanto à importância das boas práticas em dados. Destaca o papel do bibliotecário de dados como profissional atuante no ambiente acadêmico, promovendo as práticas de compartilhamento, reuso e citação dos dados junto ao corpo pesquisador da instituição na qual está inserido.

O trabalho de Veiga *et al.* (2019) também trata sobre a opinião dos pesquisadores quanto ao uso de dados. A partir da perspectiva dos pesquisadores, reitera que a ausência de recursos técnicos que permitam o compartilhamento e a citação deles serve como obstáculo para que os pesquisadores adotem práticas coletivas de uso de dados.

Sayão e Sales (2018) oferecem sugestões de como elaborar e avaliar sistemas de gestão de dados, ressaltando a importância do aporte técnico necessário para o funcionamento desses sistemas. Nesse aporte, incluem-se mecanismos padronizados que permitam a citação de conjuntos de dados e atribuição de créditos aos autores desses conjuntos, facilitando ainda seu acesso e compartilhamento.

O trabalho de Rocha *et al.* (2018) traz uma colocação interessante ao comparar as citações de dados com as citações tradicionais: “A citação de dados é mais ampla que a citação de publicações científicas, pois questiona referências mais granulares aos dados, incluindo subconjuntos de observações, variáveis ou outros componentes, assim como subconjuntos de um conjunto maior de dados” (Rocha *et al.*, 2018, p. 10). Além disso, os autores também abordam aspectos técnicos da citação apresentando, por exemplo, o esquema de metadados de citação do *Dataverse*, desenvolvido para descrever e documentar conjuntos de dados (Rocha *et al.*, 2018) e profundas análises dos diferentes *softwares* existentes para gestão de dados abertos.

Torino, Roa-Martínez e Vidotti (2020) trazem um importante debate acerca da diferença entre disponibilizar os dados e realmente publicá-los. Afirmam que a simples disponibilização dos dados, embora importante, não supre as necessidades do fluxo de comunicação científica contemporâneo. Isso se dá porque a simples disponibilização muitas vezes carece de informações sobre a procedência dos dados, contextualização de sua coleta ou produção e detalhes sobre sua finalidade. Nesse sentido, reiteram o valor do *data paper*<sup>3</sup> como veículo de publicação dos dados de pesquisa de forma mais completa.

Resende (2019) discutiu a perspectiva dos pesquisadores da área de Ciência da Informação em relação à prática de uso e citação de dados de pesquisa por meio de um questionário respondido por 97 pesquisadores.

<sup>3</sup> *Data papers*, segundo Torino, Roa-Martínez e Vidotti (2020), são um tipo específico de periódico que se distingue por se dedicar à publicação de artigos de dados de pesquisa.

No estudo da autora, 77% dos respondentes afirmaram considerar que estudos que publicam dados de pesquisa são mais citados, ao mesmo tempo que 70% afirmaram ainda manter seus dados de pesquisa em computadores pessoais. Quarenta e nove por cento dos respondentes afirmaram citar dados científicos. Resende (2019) argumenta que a importância de citar dados de pesquisa se justifica por serem produtos de pesquisa de primeira classe e que esses dados devem estar disponíveis e acessíveis em domínio público.

O trabalho de Santos e Nassi-calò (2020) apontou que a citação de dados ainda é muito menos frequente quando comparada à citação tradicional de trabalhos acadêmicos e apresentou diferentes níveis para a citação de dados baseados nas diretrizes de *Transparency and Openness Promotion* (TOP) para instruir periódicos, editoras e agências de fomento: Nível 0 (periódico encoraja a citação de dados, códigos e materiais ou não diz nada); Nível 1 - Expor (periódico descreve como citar dados, códigos e materiais nas instruções aos autores, com regras claras e exemplos); Nível 2 - Exigir (artigo fornece citação apropriada para dados, códigos e materiais de acordo com as instruções aos autores) e Nível 3 - Comprovar (artigo não é publicado até fornecer citação apropriada para dados, códigos e materiais de acordo com as instruções aos autores) (Santos; Nassi-Calò, 2020).

Nessa perspectiva, Santos e Nassi-calò (2020) afirmam que normas e diretrizes devem ser estabelecidas por periódicos, diretores e agências de fomento com o objetivo de estimular a citação de dados de pesquisa entre a comunidade científica, de maneira semelhante à que ocorre com a citação tradicional de artigos, recompensando os autores pela disponibilização dos dados de seus estudos.

Guañes (2020) aborda a citação de dados de pesquisa com foco nas questões do direito autoral por meio da devida citação das fontes originais (autores) e destaca o papel das editoras, periódicos e repositórios em incentivar a prática de citação legal de dados de pesquisa, entre várias outras práticas relacionadas à gestão de dados de pesquisa para reuso.

Já Sienra (2018) discorre sobre a citação de dados como uma das boas práticas colaborativas, com potencial de se tornar uma norma de atribuição acadêmica, auxiliando na descoberta e reutilização de dados de pesquisas que possuem interesses em comum. O trabalho da autora foi o único a citar a teoria da citação de dados por meio de Silvello (2018), apontando a urgência para o desenvolvimento de discussões consistentes dos dados de pesquisa como objetos primários de investigação.

## Conclusão

Embora existam trabalhos que contribuem significativamente para a discussão teórica da citação de dados, ainda é notável a ausência de trabalhos que sequer mencionam uma teoria da citação de dados. Constatou-se que existem temáticas recorrentes nos trabalhos que discutem a citação de dados, dentre as quais foram identificadas: a gestão de dados, aspectos éticos e legais da citação de dados e aspectos psicológicos concernentes aos pesquisadores que citam e disponibilizam dados de pesquisa. No contexto da gestão de dados, a citação de dados geralmente é discutida como uma finalidade que só se torna possível por meio de um esforço de periódicos, editoras, repositórios e pesquisadores em planejar, organizar, descrever e disponibilizar dados de pesquisa em acesso aberto.

Na temática dos aspectos éticos e legais, a citação de dados é discutida como uma boa prática cuja finalidade é prestigiar e dar os devidos créditos ao autor dos dados de pesquisa, valorizando seu esforço. Já com relação aos trabalhos que discutem aspectos psicológicos dos pesquisadores que publicam dados científicos, identificou-se que esses geralmente abordam aspectos como as necessidades e motivações pessoais, como a possibilidade de ser mais citado e ter maior reconhecimento entre pares ao compartilhar seus dados, bem como fatores que fazem com que o pesquisador se sinta inibido em publicar seus dados em domínio público, como, por exemplo, o reuso dos



seus dados sem a devida citação ou críticas, por parte da comunidade científica, à interpretação feita dos dados. Observou-se também um maior número de trabalhos em português e produzidos por pesquisadores brasileiros, demonstrando que a discussão se encontra em um nível mais avançado no Brasil.

Ao término deste trabalho, espera-se ter sido oferecido um estudo que contribua para o aprofundamento do tema da citação de dados e da teoria que a justifique. Embora este estudo não tenha a pretensão de ser exaustivo, buscou-se descrever as etapas metodológicas e os resultados obtidos o mais detalhadamente possível, com a finalidade de possibilitar que pesquisadores interessados possam reutilizar a metodologia no futuro, em seus próprios trabalhos. Dessa forma, espera-se que este artigo sirva como um bom trabalho introdutório sobre o tema, que exponha a pertinência da problemática estudada e contribua para as discussões na área da comunicação científica inclusiva e da Ciência Aberta.

## Agradecimentos

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Processos n°88887.620415/2021-00 e n°88887.620280/2021-00).

## Colaboradores

C. C. OLIVEIRA e M. C. SILVA colaboraram com a concepção e desenho, análise e interpretação dos dados e revisão. C. M. G. PAVÃO e F. C. C. SILVA colaboraram com a concepção e desenho, análise e interpretação dos dados e revisão e aprovação da versão final do artigo. A. M. M. MOURA e T. H. B. BARROS colaboraram com a análise e interpretação dos dados e revisão e aprovação da versão final do artigo.

## Referências

- Abadal, E. Ciencia abierta: un modelo con piezas por encajar. *Arbor*, v. 197, n. 799, 2021. Disponível em: <https://arbor.revistas.csic.es/index.php/arbor/article/view/2403>. Acesso em: 24 jul. 2021.
- Almeida, F. G. Suporte à gestão de dados de pesquisa: uma ampliação dos serviços oferecidos pelas bibliotecas. 2019. 250 f. Tese (Doutorado em Ciência da Informação) - Universidade Federal de Minas Gerais, Belo Horizonte, 2019. Disponível em: <http://hdl.handle.net/1843/31596>. Acesso em: 28 jul. 2021.
- Alvarez, G. R.; Caregnato, S. E. Preprints na comunicação científica da Física de Altas Energias: análise das submissões no repositório arXiv (2010-2015). *Perspectivas em Ciência da Informação*, v. 22, n. 2, 2017. Doi: <https://doi.org/10.1590/1981-5344/2830>
- Barbosa, M. C. B.; Pavan, C. Financiamento público no Brasil para a publicação de artigos em acesso aberto: alguns apontamentos. *Em Questão*, v. 23, n. 2, p.120-145, 2017. Disponível em: <http://hdl.handle.net/10183/159734>. Acesso em 24 jul. 2021.
- Borges, M. M.; Sanz Casado, E. (coord.). *Sob a lente da Ciência Aberta: olhares de Portugal, Espanha e Brasil*. Coimbra: Imprensa da Universidade de Coimbra, 2021. Doi: <https://doi.org/10.14195/978-989-26-2022-0>
- Boulton, G. *et al.* Science as a public enterprise: the case for open data. *The Lancet*, v. 377, n. 9778, p. 1633-1635, 2011. Doi: [https://doi.org/10.1016/S0140-6736\(11\)60647-8](https://doi.org/10.1016/S0140-6736(11)60647-8)
- Budapest Open Access Initiative. *Iniciativa de Budapeste pelo Acesso Aberto*. Budapeste, 2002. Disponível em: <https://www.budapestopenaccessinitiative.org/translations/portuguese-translation>. Acesso em: 24 jul. 2021.
- Costa, M. M.; Cunha, M. B. A necessidade de uma política nacional para a gestão de dados de pesquisa no Brasil. *Liinc em Revista*, v. 15, n. 2, p. 287-309, 2019. Disponível em: <http://revista.ibict.br/liinc/article/view/4763>. Acesso em: 25 jul. 2021.
- Cousijn, H. *et al.* A data citation roadmap for scientific publishers. *Scientific Data*, v. 5, 180259, 2018. Disponível em: <http://www.nature.com/articles/sdata2018259>. Acesso em: 22 jul. 2021.
- Cunha, M. B.; Costa, M. M. Fontes de informação sobre gestão de dados de pesquisa. *Informação & Sociedade: Estudos*, v. 30, n. 4, 2020. Doi: <https://doi.org/10.22478/ufpb.1809-4783.2020v30n4.57183>
- Dias, G. A.; Anjos, R. L.; Araújo, D. G. A gestão dos dados de pesquisa no âmbito da comunidade dos pesquisadores vinculados aos programas de pós-graduação brasileiros na área da Ciência da Informação: desvendando as práticas e percepções associadas ao uso e reuso de dados. *Liinc em Revista*, v. 15, n. 2, p. 5-31, 2019. Disponível em: <http://eprints.rclis.org/40179/>. Acesso em: 25 jul. 2021.
- Directory of Open Access Journals. Online, 2021. Disponível em: <https://doaj.org/search/journals>. Acesso em: 24 jul. 2021.

Fecher, B.; Friesike, S. Open science: one term, five schools of thought. *Opening science*, 2013. Doi: <http://dx.doi.org/10.2139/ssrn.2272036>

Fernández Pinto, M. Open Science for private interests? How the logic of open science contributes to the commercialization of research. *Frontiers in Research Metrics and Analytics*, v. 5, n. 588331, 2020. Disponível em: <https://doi.org/10.3389/frma.2020.588331>. Acesso em: 24 jul. 2021.

Guaaes, P. C. V. *Abertura e compartilhamento de dados de pesquisa subjacentes a artigos científicos: questões do direito autorial*. 2020. 234 f. Tese (Doutorado em Ciência da Informação) - Instituto Brasileiro de Informação em Ciência e Tecnologia, Rio de Janeiro, 2020. Disponível em: <https://www.arca.fiocruz.br/handle/icict/46188> Acesso em: 26 jul. 2021.

Kuramoto, H. Acesso Livre: uma solução adotada em todo o globo; porém, no Brasil parece existir uma indefinição. *RECIIS*, v. 8, n. 2, p.166-179, 2014. Disponível em: <http://eprints.clis.org/23478/1/935-6274-1-PB.pdf>. Acesso em 24 jul. 2021.

Leydesdorff, L. Theories of citation? *Scientometrics*, v. 43, n. 1, p. 5-25, 1998. Disponível em: <https://www.leydesdorff.net/citation/>. Acesso em: 24 jul. 2021.

Machuca-Martínez, F. Importance of scientific data and its publication as data paper. *Ingeniería y Competitividad*, v. 22, n. 1, p.8843, 2020. Disponível em: <https://www.redalyc.org/jatsRepo/2913/291365764009/html/>. Acesso em: 24 jul. 2021.

Marconi, M. A.; Lakatos, E. M. *Fundamentos de Metodologia Científica*. 5.ed. São Paulo: Atlas, 2003.

Murray-Rust, P. Open Data in Science. *Nature Precedings*, 2008. Doi: <https://doi.org/10.1038/npre.2008.1526.1>

Ngomo, A.-C. N. et al. Introduction to linked data and its life cycle on the web. In: *Reasoning Web: semantic Technologies for the Web of Data*. [S.l.]: Springer, 2011. Disponível em: [https://link.springer.com/chapter/10.1007/978-3-319-10587-1\\_1](https://link.springer.com/chapter/10.1007/978-3-319-10587-1_1). Acesso em: 20 ago. 2021.

Resende, L. C. *A curadoria de dados científicos na Ciência da Informação: levantamento do cenário nacional*. 2019. 134 f. Dissertação (Mestrado em Gestão e Organização do Conhecimento) - Universidade Federal de Minas Gerais, Belo Horizonte, 2019. Disponível em: <https://repositorio.ufmg.br/handle/1843/32413>. Acesso em: 26 jul. 2021.

Rocha, R. P. et al. *Acesso aberto a dados de pesquisa no Brasil: soluções tecnológicas: relatório 2018*. Disponível em: <http://hdl.handle.net/10183/185126>. Acesso em: 24 jul. 2021.

Santos, M. C. C. Competência em Informação para a gestão de dados de pesquisa. In: Encontro da Rede Sudeste de Repositórios Institucionais, 1., 2019, Rio de Janeiro. *Anais* [...]. Rio de Janeiro: Fiocruz/Icict/UFRJ, 2019. Disponível em: <https://www.arca.fiocruz.br/handle/icict/33336>. Acesso em: 26 jul. 2021.

Santos, S. M.; Nassi-Calò, L. Gestão editorial: tendências e desafios na transição para a ciência aberta. In: Silveira, L.; Silva, F. C. C. (org.). *Gestão editorial de periódicos científicos: tendências e boas práticas*. Florianópolis: Edições do Bosque/UFSC, 2020. Doi: <https://doi.org/10.5007/978-65-87206-08-0>

Sayão, L. F.; Sales, L. F. Dados abertos de pesquisa: ampliando o conceito de acesso livre. *RECIIS*, v. 8, n. 2, p. 76-92, 2014. Disponível em: <https://www.reciis.icict.fiocruz.br/index.php/reciis/article/view/611>. Acesso em: 22 jul. 2021.

Sayão, L. F.; Sales, L. F. Subsídios para a construção de um modelo de avaliação de sistemas de gestão de dados de pesquisa. *PontodeAcesso*, v. 12, n. 3, p. 801-108, 2018. Disponível

em: <https://periodicos.ufba.br/index.php/revistaici/article/view/28965> Acesso em: 24 jul. 2021.

Sienra, C. A. *Comportamiento informativo colaborativo y prácticas al compartir datos sobre biodiversidad marina*. 2018. 130 f. Dissertação (Mestrado em Informação e Comunicação) - Universidad de la República, Uruguay, 2018. Disponível em: <https://www.colibri.udelar.edu.uy/jspui/handle/20.500.12008/19468>. Acesso em: 26 jul. 2021.

Silva, F. C. C.; Silveira, L. O ecossistema da Ciência Aberta. *Transinformação*, v. 31, e190001, 2019a. Doi: <http://dx.doi.org/10.1590/2318-0889201931e190001>

Silva, F. C. C.; Silveira, L. O que é e qual a importância da Ciência Aberta? *SciELO em Perspectiva: Humanas*, 2019b. Disponível em: <https://humanas.blog.scielo.org/blog/2019/12/02/o-que-e-e-qual-a-importancia-da-ciencia-aberta/> Acesso em: 24 jul. 2021.

Silveira, L. et al. Citação de dados científicos: scoping review. *Encontros Bibli: Revista eletrônica de Biblioteconomia e Ciência da informação*, v. 25, 2020. Disponível em: <https://doi.org/10.5007/1518-2924.2020.e72153>. Acesso em: 24 jul. 2021.

Silvello, G. Theory and Practice of Data Citation. *Journal of the Association for Information Science and Technology*, v. 69, n. 1, p. 1-24, 2018. Disponível em: <https://arxiv.org/ftp/arxiv/papers/1706/1706.07976.pdf>. Acesso em: 24 jul. 2021.

Sousa, A. S.; Oliveira, G. S.; Alves, L. H. A pesquisa bibliográfica: princípios e fundamentos. *Cadernos da Fucamp*, v. 20, n. 43, p. 64-83, 2021. Disponível em: <https://www.fucamp.edu.br/editora/index.php/cadernos/article/view/2336>. Acesso em: 24 jul. 2021.

Torino, E.; Roa-Martinez, S. M.; Vidotti, S. A. B. G. Dados de pesquisa: disponibilização ou publicação? In: Shintaku, M.; Sales, L. F.; Costa, M. (org.). *Tópicos sobre dados abertos para editores científicos*. Botucatu, SP: ABEC, 2020, p. 183-201. Disponível em: <https://repositorio.utfrp.edu.br/jspui/handle/1/4725>. Acesso em: 26 jul. 2021.

United Nations Educational, Scientific and Cultural Organization. *UNESCO's Open Access (OA) Curriculum is now online*. Online, 2015. Disponível em: <https://en.unesco.org/news/unescos-open-access-oa-curriculum-now-online>. Acesso em: 24 jul. 2021.

Veiga, V. S. O. et al. Compartilhamento de dados de pesquisa na Fiocruz: diagnóstico e percepção do pesquisador. *Ciência da Informação*, v. 48 n. 3 (Supl.), p. 319-328, 2019. Disponível em: <https://www.arca.fiocruz.br/handle/icict/47937> Acesso em: 26 jul. 2021.

Vision, T. Open data and the social contract of scientific publishing. *BioScience*, v. 60, n. 5, p. 329-331, 2010. Disponível em: <https://doi.org/10.1525/bio.2010.60.5.2>. Acesso em: 24 jul. 2021.

Xavier, R. C. M.; Costa, R. O. A ciência como mercadoria. *Revista Eletrônica Internacional de Economia Política da Informação, da Comunicação e da Cultura*, v. 12, n. 1, 2010. Disponível em: <https://seer.ufs.br/index.php/eptic/article/view/8217>. Acesso em: 2 jul. 2021.