

Citação a dados abertos de pesquisa em normas técnicas: examinando a presença e os padrões nas normatizações científicas

Citing open research data in technical standards: examining the presence and patterns in scientific standardization

Sônia Oliveira Matos Moutinho¹, Maria Cláudia Cabrini Grácio², Carla Maria Hilário Carassa³, Daniel Martínez-Ávila⁴, Dietmar Wolfram⁵

¹ Universidade Estadual Paulista (UNESP), Marília – SP, Brasil. Doutoranda em Ciência da Informação. Instituto Federal do Piauí (IFPI), Teresina – PI, Brasil. ORCID: <https://orcid.org/0000-0002-8723-3637>

² Universidade Estadual Paulista (UNESP), Marília – SP, Brasil. Docente do Programa de Pós-Graduação em Ciência da Informação. ORCID: <https://orcid.org/0000-0002-8003-0386>

³ Universidade Estadual de Londrina (UEL), Londrina – PR, Brasil. Docente do Programa de Pós-Graduação em Ciência da Informação. ORCID: <https://orcid.org/0000-0002-2464-1502>

⁴ Universidad de León, León, Espanha. Professor Titular. Universidade Estadual Paulista (UNESP), Marília – SP, Brasil. Professor Permanente do Programa de Pós-Graduação em Ciência da Informação. ORCID: <https://orcid.org/0000-0003-2236-553X>

⁵ University of Wisconsin–Milwaukee (UWM), Milwaukee – WI, Estados Unidos. Diretor, Reitor Associado e Professor da School of Information Studies (iSchool). ORCID: <https://orcid.org/0000-0003-2236-553X>

Autor para correspondência/Mail to: Sônia Oliveira Matos Moutinho, soniamatos@ifpi.edu.br

Recebido/Submitted: 07 de novembro 2024; **Aceito/Approved:** 23 de junho de 2025



Copyright © 2025 Moutinho et al.. Todo o conteúdo da Revista (incluindo-se instruções, política editorial e modelos) está sob uma licença Creative Commons Atribuição 4.0 Internacional. Ao serem publicados por esta Revista, os artigos são de livre uso para compartilhar e adaptar e é preciso dar o crédito apropriado, prover um link para a licença e indicar se mudanças foram feitas. Mais informações em <http://revistas.ufpr.br/atoz/about/submissions#copyrightNotice>.

Resumo

Introdução: Considerando que os dados de pesquisas são cada vez mais reutilizados no cenário científico contemporâneo da Ciência Aberta, este artigo pretende analisar o fenômeno da citação de dados de pesquisas. Examinamos as recomendações sobre citação a dados de pesquisa e o reconhecimento e crédito aos desenvolvedores de dados nas normas internacionais que orientam a padronização das citações. Este estudo se caracteriza como uma investigação teórica. **Método:** Para alcançar o objetivo proposto, adotou-se a metodologia da revisão crítica de literatura (Grant e Booth (2009)) centrada nos principais eixos teóricos do reuso de dados e da citação formal de dados abertos de pesquisa. Em seguida, foi realizada uma análise das normas e estilo de padronização de citação das NBR's 10520/2023 e 6023/2018, Estilo APA, Chicago, Vancouver, ISO 690, Harvard e AMA. **Resultados:** Na análise das sete normas/estilos de padronização e citação analisadas, cinco apresentaram padronizações de citação de dados de pesquisa e duas não apresentaram qualquer recomendação. **Conclusão:** A disponibilidade de padronização de citação de dados e pesquisas em mais da metade dos padrões analisados é um indício de que os órgãos responsáveis estão gradativamente dando a devida importância à garantia de diretrizes claras para contribuir para pesquisas confiáveis, replicáveis e legítimas.

Palavras-chave: Comunicação Científica; Marketing Científico Digital; Redes Sociais; Revistas Científicas; Em Questão.

Abstract

Introduction: Considering that research data is increasingly reused in the contemporary scientific scenario of Open Science, this article aims to analyze the phenomenon of research data citation. We examine recommendations on citation of research data and the recognition and credit of data developers in international standards that guide the standardization of citations. **Method:** this study is characterized as a theoretical investigation. To achieve the proposed objective, the methodology of critical literature review (Grant e Booth (2009)) was adopted, centered on the main theoretical axes of data reuse and formal citation of open research data. Next, an analysis of the norms and citation standardization style of NBR's 10520/2023 and 6023/2018, APA Style, Chicago, Vancouver, ISO 690, Harvard and AMA was carried out. **Results:** In the analysis of the seven standards/styles of standardization and citation analyzed, five presented standardizations for citation of research data and two did not present any recommendations. **Conclusions:** The availability of standardization of data and research citations in at least more than half of the standards analyzed is an indication that the responsible bodies are gradually giving due importance to ensuring clear guidelines to contribute to reliable, replicable and legitimate research.

Keywords: Citation of research data; Data Sharing; Citation and reference standards.

INTRODUÇÃO

No contexto da pesquisa científica, o compartilhamento e reuso de dados de pesquisa surgem como práticas que visam contribuir com as iniciativas de ciência aberta. Ao permitir interação com dados de pesquisas publicizados, o compartilhamento e o reuso de dados visam otimizar recursos, acelerar descobertas e promover a transparência na construção do conhecimento, na mesma medida em que impulsionam o avanço científico por meio da construção colaborativa.

A Ciência da Informação e a Computação têm investigado o problema complexo que é a citação de dados de pesquisa, com suas implicações teóricas e técnicas. Muitos projetos de pesquisa nos últimos anos têm discutido a

importância da citação de dados e a definição de princípios e recomendações para sistemas de citação de dados Silvello (2018).

Discussões sobre o reconhecimento da autoria, bem como seus indicadores de citação, têm ganhado evidência ao tratar sobre os dados de pesquisa, dada a importância deste indicador para a validação e revalidação de estudos em andamento ou concluídos. Considerando que o ato de citar é, entre outros aspectos, uma forma de reconhecimento dos esforços dos pesquisadores, a citação de dados de pesquisa também merece atenção por parte dos comitês editoriais e pesquisadores, tal como a citação à literatura científica, que já possui normativas consolidadas no âmbito da comunicação científica.

A citação formal e correta de dados abertos de pesquisa é essencial para manter a integridade, seriedade e garantia ética das publicações. A citação informal aos dados de pesquisa pode acarretar problemas de credibilidade e confiabilidade na comunicação científica, uma vez que, quando mencionados de maneira informal, sem seguir as normas de citação adequadas, há risco de imprecisão ou até mesmo de deturpação das informações originais Park, You, e Wolfram (2018), implicando em problemas éticos e legais da atribuição do crédito de autoria dos dados que serviram de base para a construção da pesquisa, colocando em dúvida toda a honestidade e seriedade na condução da pesquisa, mesmo que a inadequação tenha ocorrido de forma não intencional.

Nesse sentido, os resultados de Park et al. (2018) chamam a atenção para a importância do estabelecimento de normas para a citação de dados de pesquisas pelas agências responsáveis pela padronização das citações em pesquisas, ao evidenciar o grande número de citação informal aos dados de pesquisa. Assim, observa-se uma carência de estudos que verifiquem as iniciativas das agências de fomento e Associações de Normas Técnicas nacionais e internacionais para estabelecimento de normas para citação de dados de pesquisas. Algumas discussões sobre o assunto são apresentadas por Cousijn et al. (2018), que enfatizam exemplos de algumas normas de referências bibliográficas internacionais.

Com base no exposto, pressupõe-se que as normas e estilos de padronização de citação e referências devem contribuir para a efetivação de políticas em benefício da difusão da prática de citação formal dos dados de pesquisa. As normas brasileiras NBR 10520/2023 e NBR 6023/2018, ambas atualizadas recentemente, e as normas Vancouver, Chicago, *American Psychological Association* (APA), Harvard, ISO 690 e a *American Medical Association* (AMA) são normatizações fundamentais no contexto brasileiro, em todas as áreas do conhecimento e, por isso, devem garantir de forma clara e eficiente as orientações quanto às práticas de citação, inclusive de dados de pesquisa.

Dessa forma, salienta-se a lacuna sobre estudos que identifiquem recomendações de como citar formalmente dados de pesquisa, dificultando a clareza dos pesquisadores no processo e reuso de dados, prejudicando todo o sistema de creditação e reconhecimento aos autores que optam por publicar seus dados de pesquisa. Assim, apresentam-se duas questões que norteiam este trabalho, a saber: como diferentes padrões e estilos têm reconhecido e promovido a citação de dados? Há iniciativas quanto à padronização de citação de dados pelas normas e estilos reconhecidos e utilizados, como Normas ABNT, APA, Chicago, Vancouver, ISO 690, Harvard e AMA?

Para responder a essas questões, este estudo teve como objetivo identificar como tem sido abordada a padronização da citação e referência a dados de pesquisa nas normas e estilos nacionais e internacionais, visto que são as indicadas na maioria das políticas editoriais e de escrita acadêmica. No âmbito da Ciência Aberta, a recomendação de modelos e exemplos de padronização de citação para devido uso e reuso de dados de pesquisa revela-se crucial para incentivar e garantir a transparência no desenvolvimento científico, garantindo a atribuição e o reconhecimento pela autoria e originalidade de dados de pesquisa, valendo-se dos princípios éticos da produção científica na nova dinâmica colaborativa em prol do desenvolvimento científico.

Como método adotado, este estudo se caracteriza como uma investigação teórica. Para alcançar o objetivo proposto, adotou-se a metodologia da revisão crítica de literatura (Grant e Booth (2009)) centrada nos principais eixos teóricos do reuso de dados e da citação formal de dados abertos de pesquisa. Esse método envolveu uma análise minuciosa de textos clássicos e contemporâneos criteriosamente selecionados, baseando-se em sua relevância teórica e empírica para fornecer uma visão abrangente sobre a aplicação dessas teorias ao fenômeno da citação formal de dados abertos de pesquisa. Os critérios de seleção dos textos incluíram a profundidade da discussão teórica, a pertinência ao tema e o impacto acadêmico dos trabalhos no campo da ciência aberta e da ética científica. A revisão teórica foi dividida em duas partes principais, cada uma focando aspectos distintos que são descritos a seguir.

O MOVIMENTO DA CIÊNCIA ABERTA

O avanço das Tecnologias de Informação e Comunicação (TICs) e o novo rumo da Web, em geral, centrada na interatividade e colaboração na produção e compartilhamento de informação, tiveram um papel crucial na transformação dos padrões de comunicação nas comunidades científicas. Conforme definida pela *Open Knowledge Foundation* (OKF), a ciência aberta é entendida como um princípio fundamental que preconiza que “o conhecimento científico deve ser livre para que as pessoas o utilizem, reutilizem e distribuam sem restrições

legais” (Open Knowledge Foundation (2009)), não paginado, tradução nossa). Este conceito abrange uma gama diversificada de princípios, configurando-se como um “termo guarda-chuva” (Albagli (2015)) ou um ecossistema que engloba outras filosofias, como o movimento de acesso aberto (*Open Access*). Um desses princípios é a gestão de dados científicos de pesquisa, especificamente seu compartilhamento e reuso.

Nessa abordagem, acredita-se que os dados científicos são a força motriz da atividade científica, pois oferecem condições de afirmar, discordar, comprovar e refutar hipóteses por meio de experimentos em qualquer área do conhecimento e a partir de diversas naturezas de pesquisa. Para entender melhor a taxonomia da ciência aberta (*Open Science*) e como ela se estrutura, apresenta-se na Figura 1 uma estrutura de sua taxonomia:

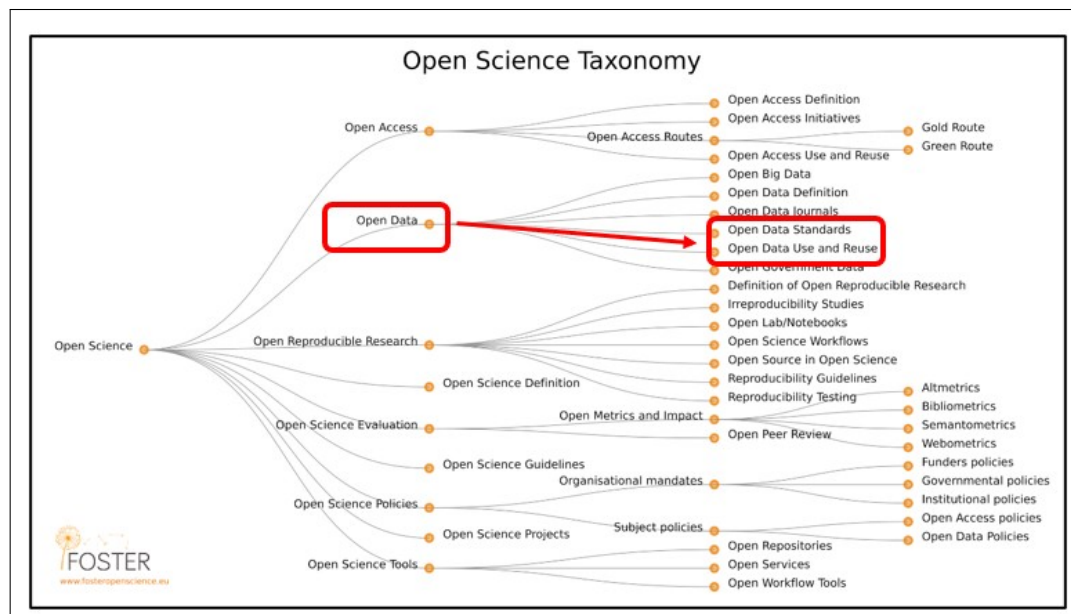


Figura 1. Taxonomia da Ciência Aberta

Fonte: Vélez Estévez (2023).

Vélez Estévez (2023) apresenta uma cadeia de segmentos abordados na proposta de ciência aberta que vão desde os aspectos mais populares e difundidos na ciência, como o acesso aberto e dados abertos, este discutido no artigo, até as vertentes de uso e avaliação de impacto desses dados. A gestão de dados científicos compreende diversos aspectos relacionados aos dados, como: gestão, compartilhamento, reuso, interoperabilidade, curadoria de dados, ciberinfraestrutura, serviços de apoio, licenciamento, aspectos comportamentais, políticas públicas de informação, entre outros (Piwowar, Day, e Fridsma (2007); ?; Borgman (2012); Albagli (2015); Oliveira e Silva (2016); Sayão e Sales (2016); Curty (2016); Curty, Crowston, Specht, Grant, e Dalton (2017)).

À medida que o movimento da Ciência Aberta se fortalece, surgem iniciativas que visam maior transparência das pesquisas financiadas pelas agências de fomento, como aquelas mencionadas por Park et al. (2018), ao evidenciar que as principais agências de financiamento, como a *National Science Foundation* (NSF), os *National Institutes of Health* (NIH) e o *National Cancer Institute* (NCI), exigem planos de gestão de dados de investigação que também facilitam a partilha de dados com outros investigadores. No Brasil, algumas agências de fomento à pesquisa, como a Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), têm recentemente solicitado essa informação para a submissão dos projetos de pesquisa, motivando os pesquisadores a refletirem sobre a responsabilidade social e ética na definição dos critérios de divulgação de seus resultados de pesquisa fomentados pelo governo. Para Albagli (2015), a exigência da disponibilidade de dados está diretamente ligada à sustentabilidade da pesquisa financiada com recursos públicos.

Da mesma forma que a discussão sobre o acesso aberto ganha força, gerando novos modelos de disseminação de resultados de pesquisas, agora mais ágeis, dinâmicos e organicamente próximos das comunidades científicas, esse movimento se estende aos dados científicos. Nesse contexto, os dados científicos constituem o elemento central que subsidia os resultados da pesquisa, pois são os dados coletados e utilizados para análise que serão, posteriormente, tratados e condensados em forma de resultados durante a construção de uma publicação oriunda de pesquisa científica.

COMPARTILHAMENTO, REUSO E CITAÇÃO DE DADOS DE PESQUISA

O compartilhamento de dados pode promover a transparência nas pesquisas, enquanto o reuso de dados gerados por outros permite a validação de descobertas feitas e o aumento da eficiência científica. (Gregory et al. (2023); Baker (2016); National Institutes of Health (2023); Pasquetto, Randles, e Borgman (2017)). Apesar de objetos

de dados abertos, como bancos de dados, coleções de dados e conjuntos de dados, serem reutilizados (Federer (2019)), a fonte desse reuso frequentemente é invisível ou difícil de ser localizada (Lane, Mulvany, e Nathan (2020); van de Sandt et al. (2019)). Os autores sublinham a importância fundamental do compartilhamento e do reuso de dados de pesquisa para promover a transparência, validar descobertas e aumentar a eficiência científica. Contudo, aponta um desafio significativo: mesmo sendo uma prática comum, o reuso desses dados frequentemente permanece “invisível” ou não rastreável, o que impede que a contribuição original seja devidamente reconhecida.

A relevância do reuso de dados de pesquisa reside precisamente na sua capacidade de acelerar o progresso científico. Ao reutilizar dados existentes, pesquisadores evitam a duplicação de esforços, economizam recursos e tempo e podem focar na análise e interpretação, validando ou expandindo o conhecimento já produzido. Além disso, o reuso fomenta a reprodutibilidade, permitindo que outros pesquisadores verifiquem os resultados de estudos anteriores e construam sobre eles, reforçando a integridade e a credibilidade da ciência. A “invisibilidade” do reuso, no entanto, mina esses benefícios, pois dificulta o crédito apropriado e a mensuração do verdadeiro impacto dos dados, desestimulando o compartilhamento.

Nesse ponto, os princípios FAIR são um conjunto de princípios orientadores para tornar os dados localizáveis, acessíveis, interoperáveis e reutilizáveis. Dentro do ecossistema da *eScience*, isso é um desafio, uma vez que as diversas instituições e centros de pesquisas, como o Centro de *eScience* da Holanda e o Centro Tecnológico Holandês para as Ciências da Vida (DTL) no Centro de Lorentz em Leiden, Holanda, surgiram para pensar e debater sobre como melhorar ainda mais esse ecossistema.

Dessas discussões, emergiu a noção de que, por meio da definição e do apoio generalizado de um conjunto mínimo de princípios e práticas orientadores acordados pela comunidade, provedores de dados e consumidores de dados – tanto de máquinas quanto de humanos – poderiam mais facilmente descobrir, acessar, interoperar e reutilizar de forma sensata, com citação adequada, as vastas quantidades de informações geradas pela ciência contemporânea intensiva em dados. Para melhor compreensão, apresentamos no Quadro 1 os quatro princípios FAIR aplicados aos dados de investigação:

Princípios FAIR	Orientações
Findable (Localizáveis)	A atribuição de um identificador único persistente aos (meta)dados. A descrição dos dados com metadados pormenorizados. O registo ou a indexação dos (meta)dados em um recurso pesquisável; A inclusão do identificador nos metadados.
Accessible (Acessíveis)	Os (meta)dados são recuperáveis através do seu identificador, mediante um protocolo de comunicações normalizado: - O protocolo de comunicações é aberto, gratuito e universalmente implementável. - O protocolo de comunicações permite um procedimento de autenticação e autorização, quando necessário. Os metadados permanecem acessíveis, mesmo se os dados já não estiverem disponíveis.
Interoperable (Interoperáveis)	- Os (meta)dados usam uma linguagem formal, acessível, partilhada e de ampla aplicabilidade para a representação do conhecimento. - Os (meta)dados usam vocabulários que seguem os princípios FAIR. - Os (meta)dados incluem referências qualificadas a outros (meta)dados.
Reusable (Reutilizáveis)	- Os (meta)dados têm uma pluralidade de atributos precisos e relevantes. - Os (meta)dados são disponibilizados com uma licença clara e acessível de uso dos dados. - Os (meta)dados são associados à sua proveniência.

Tabela 1. Princípios orientadores FAIR.

Fonte: Banerjee (2025).

Esses princípios e práticas simples devem permitir uma ampla gama de comportamentos integrativos e exploratórios, e apoiar uma ampla gama de escolhas e implementações de tecnologia, assim como o Protocolo de Internet (IP) forneceu uma camada mínima – a “cintura” de uma ampulheta – que permitiu a criação de uma vasta gama de ferramentas de provisão, consumo e visualização de dados na Internet.

Entre os benefícios do reuso de dados, destacam-se a otimização do tempo e o trabalho da coleta, economia de armazenamento e redução do número de acessos à memória lenta do computador. Além disso, Piwowar e Vision (2013) destacam como um importante benefício do reuso de dados o aumento na taxa de citação. Porém, isso só será possível se houver a devida citação formal, destacada por Park et al. (2018) como elemento essencial para a devida creditação e recompensa para os criadores primários de dados de pesquisa.

Apesar dos benefícios do reuso dos dados de pesquisa, diversos pesquisadores como Piwowar e Vision (2013), Albagli (2015), Oliveira e Silva (2016), Sayão e Sales (2016), Curty (2016), Curty et al. (2017), Park et al. (2018), Martins e Perlin (2020) e Lima (2020) vêm destacando os desafios a serem enfrentados. Entre eles, discute-se no presente artigo a questão do reconhecimento adequado dos criadores originais dos dados a serem reutilizados e a atribuição de créditos, principalmente no contexto da citação de dados de pesquisa.

As citações de dados abertos de pesquisa, tanto no corpo do texto do trabalho de pesquisa quanto nas listas de referência, são consideradas fundamentais para a localização e reuso de dados, além de incentivarem o compartilhamento (Lowenberg, Chodacki, Fenner, Kemp, e Jones (2019)) e permitirem o devido crédito aos pesquisadores que disponibilizaram esses dados para reuso.

Dito isso, a citação de dados de pesquisa é moldada por fatores que vão desde padrões de citação formalizados e recomendações até normas disciplinares. Embora as citações de dados possam funcionar como um sinal de que os dados foram compartilhados, as citações também significam que os dados foram (re)utilizados (Silvello (2018)). Portanto, a reutilização de dados pode ser vista como um importante precursor da citação de dados que pode ser definida como a ação de se referir a objetos de dados de pesquisa em lista de referências, o que acreditamos também possa ser chamado de referenciar dados de pesquisa. Esse tipo de citação é o que Park et al. (2018) denominam de citação formal de dados.

A menção a dados de pesquisa refere-se a objetos de dados ao longo de uma publicação. São citações que podem ser descritas ao longo das seções do corpo de uma publicação, em notas de rodapé, material suplementar, figuras e agradecimentos. Park et al. (2018) chamam também de citação informal de dados, e Mayo, Vision, e Hull (2016) as denominam de citação intratextual. A necessidade de mecanismos de citação formal para garantia do devido crédito aos criadores de dados e para localização e exame dos dados utilizados nas investigações é um requisito de longa data (Silvello (2018)). Além disso, a citação de dados é um método amplamente recomendado para estabelecer a acessibilidade aos dados arquivados, um dos requisitos para uma boa prática científica (ser FAIR) (Cousijn et al. (2018)).

Ademais, há uma forte demanda para dar às bases de dados o mesmo status acadêmico das referências tradicionais e para definir uma metodologia para a citação de dados, uma vez que a conexão entre os resultados científicos e os dados que fornecem suas evidências está intimamente ligada à citação de dados, aspecto fundamental da ética na pesquisa científica. Nesse cenário, alguns editores científicos têm considerado as políticas de dados e diretrizes aos autores para incluir citações de dados nas listas de referências, uma vez que os trabalhos e as publicações científicas dependem cada vez mais de bases de dados curadas (Silvello (2018)).

Cousijn et al. (2018) apresentam um roteiro para implementar a citação de dados conforme a Declaração Conjunta de Princípios de Citação de Dados (JDDCP). O roteiro é desdobrado em duas perspectivas: para editores e para autores. Para o editor, eles propõem um ponto de partida ao implementarem a citação de dados compatível com JDDCP. Entre as recomendações aos editores sobre como citar os dados, eles listam: “Capturar a citação de dados na lista de referências no ponto de submissão de forma estruturada; [...] entregar metadados de citação de dados para CrossRef”. (Cousijn et al. (2018), p. 3).

Em outro estudo, usando uma amostra de citações de dados do *OpenAlex*, Krause et al. (2023) investigaram a relação entre criadores e citantes de conjuntos de dados. Os autores observaram que a maioria dos conjuntos de dados não possui citações registradas e que, entre os citados, a maioria possui apenas uma única citação. Os autores destacam que compreender onde e como ocorre a partilha de dados é essencial para o desenvolvimento de práticas de investigação aberta.

Gregory et al. (2023) investigaram não só as práticas e motivações atuais dos pesquisadores para reutilizar e citar dados, mas também suas preferências sobre como gostariam que seus próprios dados fossem citados em diferentes disciplinas apresentadas na *Web Of Science*. Em suas conclusões, eles apontam que todas as disciplinas investigadas reutilizam dados para diversos fins em pesquisa e ensino. No entanto, o estudo revelou que esta reutilização se reflete por meio de uma variedade de mecanismos nas publicações, incluindo menções de dados e citações indiretas à literatura relacionada. Ao mesmo tempo, a maioria dos reutilizadores de dados que respondeu ao questionamento da pesquisa afirmou que cita dados em listas de referência, sugerindo que os investigadores podem ter interpretações diferentes do que significa “citar” dados e que podem ter diferentes interpretações.

As citações indiretas de dados de pesquisa referem-se aos dados que incluem uma referência a outras publicações relacionadas (ou seja, um artigo que analisa os dados ou um documento de dados). Para uma melhor compreensão, descrevemos no Quadro 2 as definições.

Termos	Definição	Localização na Publicação	Outras terminologias
Citação de Dados de Pesquisa	Refere-se a objetos de dados em listas de referências. Depende de diretrizes, normas, estilos, formatos, integridade e uso de identificadores persistentes (PDI's).	Bibliografias, trabalhos citados, listas de referências. As listas de referências podem constar publicações tradicionais (artigos de periódicos, capítulos de livros, livros, trabalhos submetidos a eventos). Ou em outras entidades, como postagens de blog, projetos de programas de ensino.	Citação formal de dados (Park et al. (2018)). Citações com variados graus de completude, ou seja, aquelas que não possuem PDI's (Moss e Lyle (2018)).
Menção a Dados de Pesquisa	Refere-se a objetos de dados ao longo da publicação.	Ao longo das seções do corpo de uma publicação, notas de rodapé, material suplementar, figuras, agradecimentos, entre outros.	Citação informal de dados (Park et al. (2018)). Citação intratextual (Mayo et al. (2016)).
Citação indireta de dados de Pesquisa	Refere-se aos dados de pesquisa, incluindo uma referência a outras publicações relacionadas (por exemplo, um artigo que analisa os dados ou documentos de dados).	Em, por exemplo, bibliografias, trabalhos citados, listas de referências.	Citações indiretas (van de Sandt et al. (2019)).

Tabela 2. Definição e explicação dos termos de citação de dados abertos de pesquisa.

Fonte: Adaptado de Gregory et al. (2023).

RESULTADOS

Esta seção está desdobrada em oito subseções, em que cada uma apresenta o resultado da análise das normas, assim como das iniciativas encontradas com recomendações relativas à citação formal de dados de pesquisa.

Normas NBR 10520/2023 e NBR 6023/2018

No Brasil, há uma diferença no entendimento do que seja citação ao longo do texto e citação na lista de referência. A Associação Brasileira de Normas Técnicas (ABNT), instituição responsável por emitir as Normas Brasileiras (NBR), recomenda duas normas separadas, mas seu uso é relacionado, não deve ser dissociado, são as normas NBR 10520/2023, referente à norma de citação, e a NBR 6023/2018, referente à norma de referenciação. Ao citar um documento no texto, deve-se fazer uso da norma de citação e, para constar na lista de referências, deve-se ter como parâmetro a norma de referências. Ao analisar as Normas brasileiras NBR 10520/2023 e NBR 6023/2018, constatamos que elas não possuem qualquer item que se refira à recomendação de padronização de citação a dados de pesquisa ou de arquivo de dados de pesquisa.

A ausência dessa recomendação para padronização de citação formal de dados de pesquisa interfere em todos os benefícios que a citação de dados de pesquisa permite, uma vez que a maioria das universidades brasileiras elabora os seus manuais de trabalhos acadêmicos baseados nas normas da ABNT. As consequências da ausência dessa recomendação podem desestimular a citação de dados formais de pesquisa nas pesquisas elaboradas pelos alunos e consequentemente, o não depósito dos dados de pesquisas. Esse fato incentiva a citação informal de dados, conforme Park et al. (2018) explicam:

Se os dados são citados de passagem no texto dos artigos, sem uma referência formal na seção de referência dos artigos, os compartilhadores de dados provavelmente não saberão que seus dados compartilhados foram reutilizados porque as citações informais não serão registradas em índices de citações, prejudicando inclusive na não contagem dessa citação nas métricas de citação geradas automaticamente pelos robôs de busca na indexação dos artigos (Park et al. (2018), p. 3, trad. nossa).

Ao trazer o problema da citação informal de dados de pesquisa, Park et al. (2018) referem-se a situações em que esses dados são mencionados no texto de artigos acadêmicos de forma casual ou incidental, sem que haja uma referência padronizada e formalizada listada na seção de referências do artigo. Quando citados dessa maneira, os pesquisadores que compartilharem esses dados podem não estar cientes de que são dados reutilizados. Isso acontece porque as citações informais não são registradas em bases de dados de citações. Dessa forma, as métricas de citação geradas automaticamente por softwares que rastreiam as menções de artigos e dados em outros trabalhos não são capturadas e registradas, o que pode subestimar o impacto e a relevância dos dados compartilhados.

A subestimação das citações nas métricas de impacto pode ter sérias implicações para a avaliação da qualidade e relevância da pesquisa, bem como para a avaliação da produtividade de pesquisadores e instituições, pois se acredita que esses dados podem ser essenciais para a validação de resultados e para a replicação de experimentos. Sendo assim, podemos perceber várias ações que buscam padronizar e garantir que todas as menções de dados compartilhados sejam registradas e contadas corretamente, desde o desenvolvimento de diretrizes mais claras para citações, como a promoção do uso de ferramentas de gestão de referências e a conscientização entre os pesquisadores sobre a importância de citar corretamente os dados compartilhados.

A citação destaca a importância da formalidade e da precisão nas práticas de citação visando garantir que o compartilhamento e a reutilização de dados sejam reconhecidos e mensurados adequadamente, ao passo que se recomenda a atualização de normas e estilos com a devida recomendação de como citar e referenciar formalmente os dados de pesquisa compartilhados.

Norma/Estilo APA da Associação Americana de Psicologia - American Psychological Association

O estilo APA (também conhecido como formato ou norma APA) é um estilo e formato de escrita para documentos acadêmicos, como artigos de periódicos acadêmicos e livros. É comumente utilizado para citar fontes no campo das ciências comportamentais e sociais. É descrito no guia de estilo da *American Psychological Association* (APA), intitulado Manual de Publicação da *American Psychological Association*. Suas diretrizes foram desenvolvidas para auxiliar a compreensão da leitura nas ciências sociais e comportamentais, para clareza de comunicação e para “escolha de palavras que possam melhor reduzir o viés na linguagem” ([American Psychological Association \(2020\)](#)).

Em sua última atualização, o Manual de Estilo APA incluiu exemplos de citação e referência a conjunto de dados no capítulo 10, subitem 10.9 – Conjunto de dados recomenda fornecer citações para conjuntos de dados quando o citante tiver realizado análises secundárias de dados arquivados publicamente ou arquivado seus próprios dados sendo apresentados pela primeira vez no trabalho atual e exemplifica:

Como citar conjunto de dados de pesquisa recomendado pelo estilo APA:

- **Citação entre parênteses:** (Memon & Carley, 2020; O'Donohue, 2017)
- **Citação narrativa:** Memon e Carley (2020) e O'Donohue (2017)

Como referenciar em estilo APA ([American Psychological Association \(2020\)](#)):

Memon, S. A., & Carley, K. M. (2020). *CMU-MisCov19: A novel twitter dataset for characterizing COVID-19 misinformation* [Data set]. 5th International Workshop on Mining Actionable Insights from Social Networks (MAISoN) at CIKM 2020, Online. Zenodo. <https://doi.org/10.5281/zenodo.4024154>

O'Donohue, W. (2017). *Content analysis of undergraduate psychology textbooks* (ICPSR 21600; Version V1) [Data set]. ICPSR. <https://doi.org/10.3886/ICPSR36966.v1>

O guia possui explicações para o reuso de dados para que se tenha o cuidado ao citar dados ou estatísticas existentes, pois o recomendado é citar a publicação na qual os dados foram publicados (por exemplo, um artigo de revista, relatório ou página da Web), em vez do conjunto de dados em si, evitando, assim, possíveis equívocos.



Figura 2. Recomendação para citação formal de dados de pesquisa segundo a norma APA.
Fonte: Guia APA Style (American Psychological Association (2020)).

Explica que a data na referência é o ano de publicação da versão dos dados utilizada. Indica ao compartilhador de dados que se forneça o título do conjunto de dados em itálico e descreve, como regra, que qualquer identificador numérico e número de versão para os dados devem estar entre parênteses sem itálico, separados por ponto-e-vírgula, que a descrição entre colchetes é flexível (por exemplo, “[Conjunto de dados]”, “[Conjunto de dados e livro de códigos]”). Ainda, indica o registro do editor do conjunto de dados no elemento de origem. A Inter University Consortium for Political and Social Research (ICPSR) é um exemplo comum. No Brasil, seria Data.gov, LattesData entre outros.

Norma/Estilo Vancouver

O estilo/norma Vancouver é um conjunto de regras para publicação de manuscritos no âmbito das Ciências da Saúde. O sistema de Vancouver toma esse nome de uma reunião celebrada em Vancouver (Canadá) em 1978, que levou à criação do Comitê Internacional de Editores de Revistas Médicas (ICMJE). Assim como outros estilos, à medida que a escrita científica ganha novos elementos e fontes, ficam passíveis de atualização também a norma Vancouver. O guia fornece informações sobre como formatar referências no estilo Vancouver. As referências são numeradas sequencialmente, seguindo a ordem em que aparecem pela primeira vez no texto. A lista de referências deve ser colocada no final do artigo. Trata-se de uma norma aberta e acessível ao público.

Na norma Vancouver, a citação se dá pela inclusão de números em parênteses. As citações são numeradas e a lista de referências segue organizada na mesma ordem. O padrão recomendado para citação formal de dados de pesquisa se apresenta:

Surname AA/Organisation/Agency. Title [Internet]. Version. Place of publication: Publisher; Year. [updated date; cited date]. Available from: URL

Como citar: [1]

Exemplo de referência:

1. Andersen LM. Data for group analyses in the Frontiers Research Topic: From raw MEG/EEG to publication: How to perform MEG/EEG group analysis with free academic software [Internet]. Version 2. Geneva: Zenodo; 2017. [cited 2019 Apr 3]. Available from: <http://doi.org/10.5281/zenodo.11347762>

Conjunto de dados ^

Template

Surname AA/Organisation/Agency. Title [Internet]. Version. Place of publication: Publisher; Year. [updated date; cited date]. Available from: URL

Example

Palin K, Pitkänen E, Turunen M, Sahu B, Pihlajamaa P, Kivioja T et al. Contribution of allelic imbalance to colorectal cancer [Internet]. Geneva: Zenodo; 2018. [cited 2019 Apr 3]. Available from: <http://doi.org/10.5281/zenodo.1222172>

Andersen LM. Data for group analyses in the Frontiers Research Topic: From raw MEG/EEG to publication: How to perform MEG/EEG group analysis with free academic software [Internet]. Version 2. Geneva: Zenodo; 2017. [cited 2019 Apr 3]. Available from: <http://doi.org/10.5281/zenodo.1134776>

World Health Organization. Adolescent birth rate. Data by country [Internet]. Geneva: Global Health Observatory data repository; [date unknown]. [updated 2018 Apr 9; cited 2019 Apr 4]. Available from: <http://apps.who.int/gho/data/node.main.REPADO39?lang=en>

Figura 3. Recomendação para citação formal de dados de pesquisa segundo a norma Vancouver.

Fonte: Guia de referências Vancouver Style (Karolinska Institutet (2025).

Complementando as orientações, a norma recomenda que, para referências que possuem até seis criadores, todos são indicados na lista de referências. Se houver mais de seis criadores, é indicado por “et al.”, conforme ilustra-se na seguinte referência:

Como citar: [2]

Exemplo:

2. Doornbos MM, Groenhout ER, Hotz GK, Brandsen C, Cusveller B, Flikkema M, et al. Transforming care: a Christian vision of nursing practice. Grand Rapids, Michigan: William B. Eerdmans Publishing Company; 2005.

Norma/Estilo Chicago

Ao analisar a norma/estilo Chicago, em *The Chicago Manual of Style Online* (2017) constata-se que a citação formal de dados e pesquisa não foi ainda contemplada em suas recomendações. Nenhum item que se refira a uma recomendação de padronização de citação de conjunto de dados de pesquisa ou de arquivo de dados de pesquisa foi localizado. A ausência dessa recomendação para padronização de citação formal de dados de pesquisa interfere nas iniciativas em prol de uma ciência aberta, verificável e transparente, interferindo em todos os benefícios que a citação de dados de pesquisa possibilita. Além das normas analisadas, no decorrer dessa pesquisa encontramos três importantes documentos que são diretrizes e padrões de citação de dados de pesquisa em outras iniciativas, além das normas analisadas.

Norma/Estilo Harvard

O estilo Harvard (também chamado de autor-ano, autor-data ou parentético) é um estilo de citação muito semelhante ao da APA. Apesar do nome, o estilo Harvard não tem conexão oficial com a Universidade de Harvard, mas com o uso popularizado do zoólogo Edward Laurens Mark, diretor do laboratório zoológico de Harvard até 1921. O estilo Harvard não tem uma norma ou guia oficial publicado, no entanto, diversas universidades e organizações têm publicado recomendações e exemplos para diferentes tipos de materiais. A University College London (2025) apresenta a seguinte informação para citar dados:

Como citar: (Author Surname/Organisation Year)

Author names. Date (in round brackets). Title of data (in single quotation marks). Available at: URL (or doi if available). (Accessed: date) (not required when doi used).

Exemplo:

Como citar: (Institute Social and Economic Research 2020).

Institute Social and Economic Research. (2020). 'Understanding Society: Waves 1–10, 2009–2019 and Harmonised BHPS: Waves 1–18, 1991–2009'. Available at: 10.5255/UKDA-SN-6614-14.

A University of Southern Queensland (2025) também apresenta exemplos e orientações para dados de pesquisa (*datasets*):

Como citar: (Author Surname/Organisation Year)

Referência:

Organisation Year of publication, Title, Data Sheet No. XXXX, viewed Day Month Year, <<https://xxxxx>>.

Exemplo:

Citação no texto: (3M Company 2016)

3M Company 2016, Jupiter nickel metal hydride batteries, Safety Data Sheet No. 32-7009-7, viewed 16 June 2021, <<https://multimedia.3m.com/mws/mediawebserver?mwsId>>.

Norma/Estilo AMA

AMA é o estilo da American Medical Association, desenvolvido pelos editores da JAMA (Journal of the American Medical Association) e publicado na atualidade pela Oxford University Press. A última edição do guia inclui uma atualização de maio de 2020 sobre citação de repositórios de dados, no capítulo 3.11.4.2.2 Data Repository. O estilo AMA (2020) da Oxford University Press indica que, ao citar dados de um repositório, a fonte original dos dados deve ser citada em:

Exemplo de referência:

Cutter AD, Gray JC. Data from: Ephemeral ecological speciation and the latitudinal biodiversity gradient. *Dryad Digital Repository*. Deposited August 17, 2016. doi:10.5061/dryad.734v9

Também indica, em relação a citações de pacotes de dados (data packages), em que o pacote de dados ou o conjunto de dados deve ser citado na publicação original para vincular a publicação e os dados. Exemplos:

Francuzik W. Data from: Skin microbiome in atopic dermatitis; 16S gene sequence data. *figshare*. 2016. doi:10.6084/m9.figshare.4028943.

Levy I, Maor Y, Mahroum N, et al. Data from: Missed opportunities for earlier diagnosis of HIV in patients that presented with advanced HIV disease: a retrospective cohort study. *Dryad Digital Repository*. 2016. doi:10.5061/dryad.73c003.11.5

Ao anunciar a atualização dos repositórios de dados, é indicado que a referência deva incluir as informações da fonte original dos dados primeiro, seguidas do nome do depositário, a data do depósito dos dados (o repositório), e o DOI (Digital Object Identifier) do conjunto de dados. Veja o exemplo:

DeLeon TT, Almquist D, Kipp BR, et al. Data from: Assessment of clinical outcomes with immune checkpoint inhibitor therapy in melanoma patients with CDKN2A and TP53 pathogenic mutations. *Dryad Digital Repository*. Deposited March 12, 2020. doi:10.5061/dryad.m0cfxpp0g

Norma/Estilo ISO 690/2021

A ISO 690:2021 ([International Organization for Standardization \(2021\)](#)) é a norma internacional que estabelece recomendações para referências bibliográficas e citações de recursos de informação, desenvolvida pela *International Organization for Standardization*. Enquanto a versão anterior de 2010 da norma não mencionava dados de pesquisa, a versão de 2021 inclui uma seção inteira dedicada a “*research datasets*” (na seção 8.13).

Especificamente, a norma aborda apenas conjuntos de dados eletrônicos. Devido à grande diversidade dos conjuntos de dados, não são fornecidas diretrizes detalhadas sobre como citar categorias específicas, com exceção das entrevistas de pesquisa, que são tratadas em outra seção da norma. No entanto, a ISO 690:2021 lista os elementos da referência que são:

- **Sempre obrigatórios:** título, formato e tipo de recurso, data de publicação e identificador persistente.
- **Obrigatórios em determinadas circunstâncias:** nome do criador, requisitos do sistema, nome do repositório/host, distribuidor ou host online, relacionamentos e tamanho do arquivo.
- **Opcionais:** identificadores padronizados de identidades públicas dos criadores, edição e versão, direitos autorais, licença, fonte dos dados, instruções de uso, proveniência, URL e outras informações que possam ser adicionadas em qualquer posição apropriada.

A norma fornece exemplos e explicações sobre alguns dos elementos mencionados, como a licença, a autoria, os identificadores persistentes, os recursos vinculados, o tamanho do arquivo, os conjuntos de dados e as publicações, a proveniência, a edição e versão, e substitutos (como representações gráficas em vez dos dados reais). Quanto às pessoas responsáveis pelo conjunto de dados citado e aos recursos relacionados, a ISO 690 menciona exemplos como investigadores, técnicos, gestores de dados, compiladores, editores e programadores. A norma indica que devem ser fornecidos os nomes e funções dos criadores, conforme mencionados na documentação do conjunto de dados de pesquisa, bem como o nome completo do projeto ou organização responsável (incluindo o identificador da organização – ISNI, ORCID ou ROR – quando possível). Em relação ao identificador preferencial para o conjunto de dados, a norma cita o DOI.

Por último, a ISO 690-2021 também menciona outras recomendações de citação de conjuntos de dados, como a guia do Digital Curation Centre, os princípios de citação de dados recomendados na Declaração dos Princípios da Citação de Dados pelo *Data Citation Synthesis Group* (2014), a plataforma de gestão de dados *Dryad* (2025) e Projeto *Dataverse* (2025). Das normas consultadas, acredita-se ser a mais completa em termos de exemplos e possibilidades de citação de dados de pesquisa.

Exemplo:

PORTES, Alejandro and Rubén G. RUMBAUT. *Children of Immigrants. Longitudinal Study (CILS) 1991–2006 ICPSR20520*, version 2 (Jan 23, 2012). Dataset. Ann Arbor [US-MI]: Inter-university Consortium for Political and Social Research (distributor), 2012. Available from: <https://doi.org/10.3886/ICPSR20520.v2>. [82.7 MB]. [accessed 2018-05-06].

A Norma ISO 690 representa um marco fundamental na padronização de referências bibliográficas, oferecendo diretrizes internacionalmente reconhecidas que promovem a consistência, a precisão e a credibilidade na citação de fontes em trabalhos acadêmicos e científicos. Sua relevância transcende fronteiras disciplinares e geográficas, facilitando a comunicação científica global ao estabelecer critérios uniformes para referenciar desde publicações tradicionais até recursos digitais emergentes, incluindo conjuntos de dados de pesquisa, websites e materiais multimídia. Os benefícios da norma incluem a redução de ambiguidades na identificação de fontes, o fortalecimento da integridade acadêmica por meio do reconhecimento adequado da propriedade intelectual, a facilitação da verificação e reprodutibilidade de pesquisas e a promoção da transparência científica. Além disso, a ISO 690 (2021) contribui significativamente para a prevenção do plágio, oferece suporte à gestão eficiente de referências bibliográficas e adapta-se continuamente às transformações do panorama informacional contemporâneo, garantindo que pesquisadores, estudantes e profissionais disponham de ferramentas padronizadas e confiáveis para o desenvolvimento de trabalhos científicos de alta qualidade.

O outros documentos com recomendações relativas à citação formal de dados de pesquisa

Esta seção está desdobrada em três iniciativas, uma brasileira e duas internacionais, sendo uma com parceria subsequente no Brasil, a saber: iniciativa *SciELO*, *DataCite* e Projeto *Dataverse* colaborando com o *DataLattes*.

Iniciativa SciELO

Dada a importância crescente dos dados de pesquisa como contribuição intelectual original de pesquisadores, encontramos outros documentos além das normas que possuem recomendações para padronização de citação

formal de dados de pesquisa. O Guia de citação de dados de pesquisa da base de dados *SciELO* (Scientific Electronic Library Online (2020)) traz vários exemplos de como referenciar uma citação de dados de pesquisa (Figura 4).

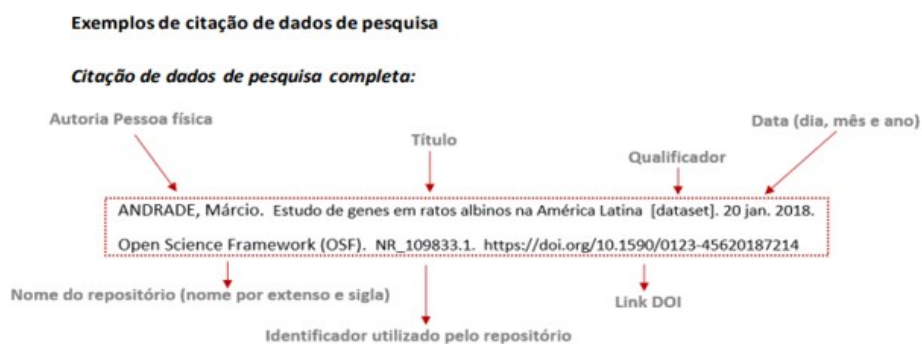


Figura 4. Recomendação para citação formal de dados de pesquisa segundo a norma Vancouver.

Fonte: Guia de citação de dados de pesquisa Style (Scientific Electronic Library Online (2020)).

A iniciativa SciELO foi umas das pioneiras na recomendação e padrão para citação de conjunto de dados no Brasil, sendo a opção como recomendação de citação e dados de pesquisa no âmbito brasileiro.

Projeto DataCite e DOI Citation Formatter

Trata-se de uma comunidade global que compartilha um interesse comum: garantir que os resultados e recursos de pesquisa estejam disponíveis e conectados abertamente para que sua reutilização possa avançar o conhecimento entre disciplinas, tanto na atualidade quanto no futuro. O objetivo dessa iniciativa é tornar a pesquisa mais eficaz com metadados que conectam resultados e recursos de pesquisa. Esses dados podem ser desde amostras e dados de imagens a *preprints*. Permitem a criação e o gerenciamento de identificadores persistentes (IP's), possibilitando a integração de serviços para melhoria dos fluxos de trabalhos de pesquisa para facilitar a descoberta e reutilização de resultados e recursos de pesquisas.

Com relação à citação de dados e de referências em geral, essa iniciativa possuía um recurso chamado *DataCite Citation Formatter*. Porém desde janeiro de 2025 esse recurso passou a ser gerenciado pela Fundação DOI, trata-se de um recurso que por meio da descrição de metadados de um DOI, usa as informações para construir uma citação seguindo mais de 5.000 estilos de citação disponibilizados pelo projeto citationstyles.org. Esse serviço funciona em todas as Agências de Registro, incluindo *Crossref*, *mEdra* etc. É importante destacar que esse recurso está aberto a toda comunidade de pesquisadores e foi projetado para apoiar a construção de listas de referências sobre uma interface simples e eficaz. Para gerar a referência, basta inserir o DOI da publicação e escolher o estilo e o idioma.

DOI Foundation VISIT DOI.ORG

DOI CITATION FORMATTER

https://doi.org/10.48432/CAESHU
type or paste a known DOI. For example 10.1145/2783446.2783605

apa
Select style: Begin typing (e.g. Chicago or IEEE) or use the drop-down menu.

pt-BR
Select language: Begin typing (e.g. en GB for English, Great Britain) or use the drop-down menu.

SUBMIT

Santil, A. (2025). Dados estruturados sobre o estoque de carbono no solo da região Centro-Sul do Paraná. [Conjunto de dados]. Redape, https://doi.org/10.48432/CAESHU

Copy to clipboard

Figura 5. Visualização da geração de citação pelo DOI Citation Formatter

Fonte: DOI(DOI Citation Formatter (2025)).

Ao inserir o DOI referente a um arquivo de dados de pesquisa, o recurso nos apresenta a citação nos diversos estilos e formatos. Deve-se ter o cuidado de verificar se o estilo escolhido possui previsão e padronização de

citação para dados de pesquisa. Ao realizar um teste com os estilos que não possuem essa previsão, a referência é gerada sem a informação [dataset], ou [Conjunto de dados] como podemos verificar no exemplo da referência oriunda do DOI <https://doi.org/10.48432/CAE5NJ>:

Santi, A. (2025). *Dados estruturados sobre o estoque de carbono no solo da região Centro-Sul do Paraná*. [Conjunto de dados]. Redape. <https://doi.org/10.48432/CAE5NJ>

Projeto Dataverse e LattesData

Trata-se de uma iniciativa da Universidade de Harvard que uniu um grupo de pessoas para a construção de um aplicativo web de código aberto para compartilhar, preservar, citar, explorar e analisar dados de pesquisa, de modo a facilitar a disponibilização de dados para outros pesquisadores que possam depositar e compartilhar o trabalho de pesquisadores com mais facilidade. Assim, permite-se que pesquisadores, periódicos, autores de dados, editores, distribuidores de dados e instituições afiliadas recebam crédito acadêmico e visibilidade na web (King (2007)).



Figura 6. Imagem ao acessar um arquivo de dados de pesquisa do LattesData.

Fonte: DOI(LattesData (2023)).

O projeto *Dataverse* disponibiliza de forma gratuita a instalação do software que hospeda vários arquivos virtuais chamados coleções *Dataverse*. A visão central por trás do Projeto *Dataverse* é automatizar grande parte do trabalho do arquivista profissional e fornecer serviços e distribuir crédito para o criador de dados. Antes do Projeto *Dataverse*, os pesquisadores eram forçados a escolher entre receber crédito por seus dados, controlando a distribuição por conta própria, mas sem garantia de preservação de longo prazo, ou ter garantia de preservação de longo prazo, enviando-os para um arquivo profissional, mas sem garantia do recebimento de créditos.

Com o projeto *Dataverse*, é possível colocar uma coleção *Dataverse* (um arquivo virtual) em qualquer site e manter o formato, a marca e a URL do seu site, juntamente com uma citação acadêmica para os dados que lhe dão crédito total e visibilidade na web (Conselho Nacional de Pesquisa, Desenvolvimento e Tecnologia (2022)). Dessa forma, qualquer instituição pode implementar o seu repositório de dados de pesquisa sem ter que mobilizar uma grande equipe do setor de Tecnologia da Informação (TI), bastando aderir ao projeto *Dataverse* e obter de maneira rápida e segura, com respaldo institucional e garantias de preservação de longo prazo.

Em junho de 2022, o Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) lançou, em parceria com o projeto *Dataverse* e o Instituto Brasileiro de Ciência e Tecnologia (IBICT), o seu repositório oficial de dados abertos: o LattesData (2023), com o objetivo de disponibilizar uma ferramenta oficial aos pesquisadores beneficiários que depositarem os conjuntos de dados que surgiram como resultado dos projetos desenvolvidos

com recursos do CNPq. O papel do IBICT foi fundamental na concepção e operacionalização do *LattesData*. Como parceiro do CNPq na criação deste repositório, o IBICT foi corresponsável pela definição da arquitetura da informação do sistema, pela elaboração de sua política de uso e pela estruturação dos metadados que garantem a organização e a encontrabilidade dos conjuntos de dados. Além disso, o Instituto assegura a governança dos dados, crucial para a integridade e a confiabilidade do repositório. Complementarmente, o IBICT atua ativamente na promoção e divulgação do *LattesData*, desempenhando um papel estratégico na disseminação da cultura da Ciência Aberta em todo o Brasil.

O repositório segue a padronização para citação de dados de pesquisa dentro dos requisitos internacionais, conforme demonstrado na Figura 6. Há, dentro do projeto *DataVerse*, uma preocupação em seguir as boas práticas científicas e em garantir a padronização da citação de conjuntos de dados de pesquisa com base na Declaração Conjunta de Princípios de Citação de Dados da *Citation Synthesis Group*, uma síntese de todos os princípios e iniciativas anteriormente existentes sobre citação de dados, visando facilitar a publicação de dados pelos pesquisadores e a obtenção do crédito, bem como seu reconhecimento. Dessa forma, dentro do repositório *LattesData*, a citação é gerada automaticamente, fruto do empenho de mais de 20 anos da equipe de profissionais que trabalharam a citação de dados dentro do projeto *Dataverse* (King (2007); Altman e Crosas (2013)).

Em consequência, considera-se que a conscientização e o uso maior de citações formais de dados de pesquisa terão um aumento significativo à medida que normas/bases/repositórios de dados de pesquisa, pesquisadores, editores de periódicos e instituições de pesquisa compreendam a importância e a relevância do uso e reuso da citação formal de dados de pesquisa. Nesse contexto, e considerando a persistente lacuna em algumas normas e estilos de citação, a norma ISO 690:2021 é recomendada para a citação de dados de pesquisa. Conforme as conclusões deste estudo, sua abrangência a estabelece como uma referência para a padronização da citação formal de dados tanto no Brasil quanto em âmbito internacional.

ANÁLISE E DISCUSSÃO DOS RESULTADOS

Das sete normas/estilos de padronização de citação analisadas, cinco apresentaram padronização de citação e dados de pesquisas, foram elas: normas APA, Vancouver, Harvard, AMA e ISO-690. É encorajador ver que os estilos APA, Vancouver, Harvard, AMA e ISO-690 abordaram citações e referências de dados. Por outro lado, não é surpreendente que o estilo de Chicago ainda não tenha abordado a questão da citação de dados, uma vez que tende a ser mais usado por periódicos da área de humanidades, em que o uso, compartilhamento e o reuso de dados são menos prováveis. A promoção de formatos de citação de dados pelas iniciativas *SciELO*, *DataCite/DOI Citation Formatter* e *Dataverse* fornecem formas adicionais importantes de promover a citação de dados. Considera-se que a disponibilidade de padronização nessas normas internacionais é um indicativo de que essas instituições estão cientes da importância de se ter diretrizes claras para obtenção de pesquisas confiáveis, replicáveis e legítimas.

Na análise das normas/estilos, observou-se a ausência da referência quanto ao embasamento da formulação das diretrizes. Por outro lado, em iniciativas como o projeto *Dataverse*, há a referência à Declaração Conjunta de Princípios de Citação de Dados. O cumprimento aos 8 princípios da declaração promove um incentivo a uma cultura mais ampla de compartilhamento na comunidade acadêmica. Ao proporcionar diretrizes simples e acessíveis, os pesquisadores são encorajados a disponibilizar seus próprios conjuntos de dados, contribuindo para a construção de um banco de dados mais extenso e diversificado. Esse fenômeno contribui para a ampliação do potencial de descobertas significativas e de inovações em diferentes campos do conhecimento.

Neste ensejo, diretrizes bem definidas como as desenvolvidas e implementadas no projeto *Dataverse* e *DataCite* podem contribuir sobremaneira para que a citação formal de dados de pesquisa esteja cada vez mais presente em todas as instâncias de padronização de citações, pois a citação e o reuso de dados são essenciais para proteger a integridade e a autenticidade da pesquisa. Elas ajudam a evitar problemas relacionados à má interpretação ou uso inadequado dos dados, mitigando a disseminação de informações incorretas ou enganosas na literatura científica. Outro aspecto importante para promover a citação formal de dados é a necessidade de fontes de indexação como o DCI para documentar efetivamente as citações de dados. Os autores podem estar citando dados formalmente, mas se as fontes de indexação não indexarem inúmeros repositórios de dados, as citações ainda não serão reconhecidas.

CONCLUSÕES

A análise realizada revela avanços significativos na incorporação da padronização de citação de dados em estilos amplamente utilizados, como APA, Vancouver, Harvard, AMA e ISO-690, refletindo uma conscientização crescente sobre a importância dos dados na ciência aberta, replicável e confiável. A existência de diretrizes para citação de dados nessas normas internacionais demonstra um movimento consistente rumo à transparência e à reprodutibilidade nas pesquisas, pilares fundamentais para o avanço científico em diferentes áreas do conhecimento.

O cumprimento desses princípios fortalece uma cultura de abertura, inclusão e colaboração, facilitando não somente o compartilhamento, mas também o aproveitamento de dados por pesquisadores em escala global.

É crucial reconhecer que o presente estudo, ao investigar os padrões de citação de dados, pode apresentar um viés de seleção inerente à sua metodologia. A análise concentrou-se primariamente em normas e/ou estilos reconhecidos internacionalmente e em iniciativas de instituições reconhecidas mundialmente como referenciais que já possuem políticas explícitas de dados. Essa abordagem, embora necessária para a delimitação do escopo e viabilidade da pesquisa, inerentemente exclui formas de citação que não são formalmente registradas ou indexadas, como as citações informais em notas de rodapé ou no corpo do texto dos artigos. Conforme destacado por [Park et al. \(2018\)](#) (tradução nossa), a citação informal de dados ainda é predominante na literatura acadêmica. Portanto, os resultados aqui apresentados devem ser interpretados com essa limitação em mente, focando nos padrões observados dentro da amostra analisada e não necessariamente extrapolando para todas as formas de reuso e citação de dados.

Conclui-se que, dentre as normas/estilos analisados, a norma ISO 690 é a mais completa, com previsão de inclusão das pessoas responsáveis pelo conjunto de dados citado e recursos relacionados. Prevê, por exemplo, investigadores, técnicos, gerentes de dados, compiladores, editores e programadores. Nomes e funções dos criadores mencionados na documentação do conjunto de dados de pesquisa e o nome completo do projeto ou organização responsável.

No entanto, o aprimoramento desse ecossistema requer esforços coordenados entre os provedores de normas, repositórios de dados e sistemas de indexação, como o DCI. Para que a citação formal de dados tenha impacto real, é fundamental que sistemas de indexação ampliem sua cobertura, garantindo visibilidade e reconhecimento às citações formalmente realizadas. Assim, ganha força o ciclo virtuoso de transparência, integridade e inovação, promovendo pesquisas mais sólidas, confiáveis e socialmente relevantes.

Dito isso, apontam-se algumas pesquisas que contribuirão para os avanços dos estudos de modo a aprofundar a compreensão sobre a adesão, o impacto, as barreiras e os aprimoramentos necessários nos processos e instrumentos ligados à padronização e à citação de dados, auxiliando a consolidação de práticas mais robustas e efetivas para a Ciência Aberta. Tais estudos podem verificar a incorporação dessas diretrizes nas políticas editoriais, os desafios enfrentados por editores, revisores e autores em diversas disciplinas e o papel fundamental que as editoras, editores chefes e equipes editoriais desempenham na promoção da citação formal de dados de pesquisa ao exigir sua padronização por meio das políticas editoriais.

Recomenda-se também estudos sobre o impacto da citação formal de dados no reconhecimento e nas métricas acadêmicas dos trabalhos publicados, bem como pesquisas comparativas sobre a cultura de compartilhamento e citação de dados entre diferentes disciplinas, buscando entender motivações e barreiras à adoção dessas práticas entre pesquisadores. Em suma, a padronização da citação de dados é uma etapa crucial para consolidar a abertura e o avanço da ciência, devendo ser continuamente incentivada, aprimorada e monitorada por todos os atores do ecossistema científico.

AGRADECIMENTOS

O presente trabalho foi realizado com apoio do Programa Institucional de Internacionalização (PrInt) da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001 e da Fundação de Amparo à Pesquisa do Estado do Piauí (FAPEPI) - referente ao Edital nº 001/2022. Processo nº 00110.000156/2022-67

ELEMENTOS	APA	VANCOUVER	HARVARD	AMA	ISO 690:2021
Nome do(s) criador(es)	Autor/organização no início	Autor/organização no início. Primeiros seis autores/organização, seguido de <i>et al.</i>	Autor/organização no início	Autor/organização no início	Nome(s) do(s) criador(es); funções podem ser especificadas
Identificadores das identidades públicas dos criadores	–	–	–	–	Opcional
Ano/Data de publicação	Entre parênteses após o autor	Após local/editor	Após autor/organização, entre parênteses	Após o nome do repositório ou data de depósito	Obrigatório
Título do dataset	Itálico	Destaque opcional logo após o autor	Entre aspas simples ou sem aspas	Após autores, precedido de "Data from:" ou "Data package"	Obrigatório, entre os elementos essenciais
Descrição adicional	Tipo de recurso (ex.: [Data set])	[Internet] após o título	Variável (ex.: "dataset")	Pode constar antes ou após o título	Tipo ou descrição do recurso obrigatória
Edição e versão	Após o título, se relevante	Após o título	Pode aparecer como V1, ano	Pode aparecer, se fornecido	Opcional, mas recomendado
Evento/Publicação	Evento ou publicação relacionada	Não obrigatório	Raramente incluído	Não usual	Opcional
Repositório/Editora	Nome do repositório	Local/editora/repositório	Indicado após "Available at:"	Nome do repositório explicitado	Distribuidor, host ou repositório eletrônico
Identificador/Link	DOI ou URL	DOI ou URL	DOI ou URL obrigatório	DOI obrigatório ao final	Identificador persistente obrigatório
Data de acesso	Não obrigatória, salvo exceções	Obrigatória para recursos online	Obrigatória quando não há DOI	Não usual quando DOI está presente	Obrigatória quando não há identificador persistente
Ordem dos elementos	Autor – (Ano) – Título – Tipo – Repositório – DOI/URL	Autor – Título – Tipo – Local – Ano – Data de acesso – DOI/URL	Autor – (Ano) – Título – Disponível em – DOI/URL	Autor – Título – Repositório – Data de depósito – DOI	Título; Tipo; Data; Identificador persistente; demais conforme necessidade
Formatação	Ênfase no título em itálico	Padronização concisa	Estrutura simples e flexível	Uso recorrente de "Data from:"	Estrutura flexível com elementos obrigatórios

Apêndice 3. Quadro comparativo das normas analisadas e os elementos de metadados para conjuntos de dados de pesquisa.

Fonte: Dados da Pesquisa, 2024.

REFERÊNCIAS

- Albagli, S. (2015). Ciência aberta em questão. In S. Albagli, M. L. Maciel, & A. H. Abdo (Eds.), *Ciência aberta, questões abertas* (p. 9–26). IBICT and UNIRIO. Recuperado de http://livroaberto.ibict.br/bitstream/1/1060/1/Ciencia%20aberta_questoes%20abertas_PORTUGUES_DIGITAL%20%285%29.pdf
- Altman, M., & Crosas, M. (2013). The evolution of data citation: from principles to implementation. *IASSIST Quarterly*, 37, 62–70. Recuperado de https://iassistquarterly.com/public/pdfs/iqvol371_4_altman.pdf
- American Psychological Association. (2020). *Publication manual of the american psychological association* (7a. ed.). APA. Recuperado de <https://apastyle.apa.org/products/publication-manual-7th-edition>
- Baker, M. (2016). 1,500 scientists lift the lid on reproducibility. *Nature*, 533, 452–454. doi: 10.1038/533452a
- Banerjee, S. (2025). *A guide to the fair principles: best practices for data handling*. Recuperado de <https://openscience.eu/article/infrastructure/guide-fair-principles>
- Borgman, C. L. (2012). The conundrum of sharing research data. *Journal of the American Society for Information Science and Technology*, 63(6), 1059–1078. Recuperado de <https://onlinelibrary.wiley.com/doi/abs/10.1002/asi.22634>
- Conselho Nacional de Pesquisa, Desenvolvimento e Tecnologia. (2022). *Lançamento do repositório de dados de pesquisa lattesdata*. Recuperado de https://www.youtube.com/watch?v=qO_Q9TaB0CE (Vídeo)
- Cousijn, H., Kenall, A., Ganley, E., Harrison, M., Kernohan, D., Lemberger, T., ... Clark, T. (2018). A data citation roadmap for scientific publishers. *Scientific Data*, 5. doi: 10.1038/sdata.2018.259
- Curty, R. G. (2016). Factors influencing research data reuse in the social sciences. *International Journal of Digital Curation*, 11(1), 96–117. doi: 10.2218/ijdc.v11i1.401
- Curty, R. G., Crowston, K., Specht, A., Grant, B. W., & Dalton, E. D. (2017). Attitudes and norms affecting scientists data reuse. *PLOS ONE*, 12(12). doi: 10.1371/journal.pone.0189288
- DOI Citation Formatter. (2025). *Doi citation formatter*. Recuperado de <https://citation.doi.org/>
- Dryad. (2025). *Dryad digital repository*. Recuperado de <https://datadryad.org/>
- Federer, L. M. (2019). *Who, what, when, where, and why? quantifying and understanding biomedical data reuse* (Tese de doutorado, University of Maryland). doi: 10.13016/60jd-9hux
- Grant, M. J., & Booth, A. (2009). A typology of reviews. *Health Information and Libraries Journal*, 26(2), 91–108. doi: 10.1111/j.1471-1842.2009.00848.x
- Gregory, K., Ninkov, A., Ripp, C., Roblin, E., Peters, I., & Haustein, S. (2023). Tracing data: a survey investigating disciplinary differences in data citation. *Quantitative Science Studies*, 4(3). doi: 10.1162/qss_a_00264
- International Organization for Standardization. (2021). *Iso 690: Information and documentation*. Recuperado de <https://www.iso.org/standard/72642.html>
- Karolinska Institutet. (2025). *Referencing guide for vancouver*. Recuperado de <https://kib.ki.se/en/write-cite/writing-references-apa-vancouver/reference-guides/reference-guide-vancouver>
- King, G. (2007). An introduction to the dataverse network as an infrastructure for data sharing. *Sociological Methods and Research*, 36, 173–199. Recuperado de <https://gking.harvard.edu/files/abs/dvn-abs.shtml>
- Krause, G., et al. (2023). *Who re-uses data? a bibliometric analysis of dataset citations*. doi: 10.48550/arXiv.2308.04379
- Lane, J. I., Mulvany, I., & Nathan, P. (Eds.). (2020). *Rich search and discovery for research datasets*. SAGE Publications. Recuperado de <https://study.sagepub.com/richcontext/student-resources/e-book-files/web-pdf>
- LattesData. (2023). *Repositório lattesdata*. Recuperado de <https://lattesdata.cnpq.br/dataverse/lattesdata>
- Lima, K. C. R. d. (2020). *Impacto percebido pelos pesquisadores quanto ao reuso de dados científicos* (Dissertação de mestrado, Universidade Federal do Paraná). Recuperado de <https://acervodigital.ufpr.br/handle/1884/70823>
- Lowenberg, D., Chodacki, J., Fenner, M., Kemp, J., & Jones, M. B. (2019). *Open data metrics: lighting the fire*. doi: 10.5281/zenodo.3525349
- Martins, H. C., & Perlin, M. S. (2020). Call for papers data reuse. *Revista de Administração Contemporânea*. doi: 10.5281/zenodo.3858031
- Mayo, C., Vision, T. J., & Hull, E. A. (2016). The location of the citation. *International Journal of Digital Curation*, 11(1). doi: 10.2218/ijdc.v11i1.400
- Moss, E., & Lyle, J. (2018). *Opaque data citation*. Recuperado de <https://deepblue.lib.umich.edu/handle/2027.42/142393> (Poster ICPSR)
- National Institutes of Health. (2023). *Final nih policy for data management and sharing*. Recuperado de <https://grants.nih.gov/grants/guide/notice-files/NOT-OD-21-013.html>
- Oliveira, A. C. S. d., & Silva, E. M. d. (2016). Ciência aberta: dimensões para um novo fazer científico. *Informação & Informação*, 21(2), 5–39. Recuperado de <https://ojs.uel.br/revistas/uel/index.php/informacao/article/view/27666/20113>
- Open Knowledge Foundation. (2009). *Is it open data?* Recuperado de <https://isitopendata.org/>
- Park, H., You, S., & Wolfram, D. (2018). Informal data citation for data sharing and reuse. *Journal of the Association for Information Science and Technology*, 69(11), 1346–1354. doi: 10.1002/asi.24049
- Pasquetto, I. V., Randles, B. M., & Borgman, C. L. (2017). On the reuse of scientific data. *Data Science Journal*, 16(8). doi: 10.5334/dsj-2017-008
- Piwowar, H. A., Day, R. S., & Fridsma, D. B. (2007). Sharing detailed research data is associated with increased citation rate. *PLOS ONE*, 2(3). doi: 10.1371/journal.pone.0000308
- Piwowar, H. A., & Vision, T. J. (2013). Data reuse and the open data citation advantage. *PeerJ*, 1, e175. doi: 10.7717/peerj.175
- Sayão, L. F., & Sales, L. F. (2016). Algumas considerações sobre os repositórios digitais de dados de pesquisa. *Informação & Informação*, 21(2), 90–115. doi: 10.5433/1981-8920.2016v21n2p90
- Scientific Electronic Library Online. (2020). *Guia de citação*

de dados de pesquisa scielo. Recuperado de https://wp.scielo.org/wp-content/uploads/guia-de-citacao-de-dados_pt.pdf

Silvello, G. (2018). Theory and practice of data citation. *Journal of the Association for Information Science and Technology*, 69(1), 1–24. doi: 10.1002/asi.23917

van de Sandt, S., Nielsen, L. H., Ioannidis, A., Muench, A., Henneken, E., Accomazzi, A., ... Dallmeier-Tiessen, S. (2019). *Practice meets principle*. doi: 10.48550/arXiv.1911.00295

Vélez Estévez, A. (2023). *Definición y taxonomía de open science*. Recuperado de [https://researcheu.sea-eu.org/wp-content/uploads/2023/01/](https://researcheu.sea-eu.org/wp-content/uploads/2023/01/Presentaci_n-reSEArch-EU.pdf)Presentaci_n-reSEArch-EU.pdf

Como citar este artigo (APA):

Moutinho, S. O. M., Grácio, M. C. C., Carassa, C. M. H., Martínez-Ávila, D., & Wolfram, D. (2025). Citação a dados abertos de pesquisa em normas técnicas: examinando a presença e os padrões nas normatizações científicas. *AtoZ: novas práticas em informação e conhecimento*, 14, 1 – 19. Recuperado de: <http://dx.doi.org/10.5380/atoz.v14.97450>

NOTAS DA OBRA E CONFORMIDADE COM A CIÊNCIA ABERTA

CONTRIBUIÇÃO DE AUTORIA

	Sonia Oliveira Matos Moutinho; Maria Cláudia Cabriní Grácio; Daniel Martínez-Ávila; Dietmar Wolfram	Sonia Oliveira Matos Moutinho; Maria Cláudia Cabriní Grácio; Daniel Martínez-Ávila; Carla Maria Hilário Carassa	Sonia Oliveira Matos Moutinho; Maria Cláudia Cabriní Grácio; Carla Maria Hilário Carassa	Sonia Oliveira Matos Moutinho; Maria Cláudia Cabriní Grácio; Daniel Martínez-Ávila; Carla Maria Hilário Carassa	Nome Completo
Papéis e contribuições					
Concepção do manuscrito	X				
Escrita do manuscrito		X			
Metodologia			X		
Curadoria dos dados				NA*	NA
Discussão dos resultados				X	
Análise dos dados					

*Não se aplica.

EQUIPE EDITORIAL

Editora/Editor Chefe

Paula Carina de Araújo (<https://orcid.org/0000-0003-4608-752X>)

Editora/Editor Associada/Associado Júnior

Karolayne Costa Rodrigues de Lima (<https://orcid.org/0000-0002-6311-8482>)

Editora/Editor de Texto Responsável

Suzana Zulpo (<https://orcid.org/0000-0002-6311-8482>)

Seção de Apoio às Publicações Científicas Periódicas - Sistema de Bibliotecas (SiBi) da Universidade Federal do Paraná - UFPR

Editora/Editor de Layout

Tiago Batista Pedra (<https://orcid.org/0009-0000-7385-7273>)