

Comunicação

Dados provenientes de redes sociais: uma reflexão sobre os desafios e problemáticas relacionadas ao acesso e reuso

Data from social media: a reflection on the challenges and issues related to access and reuse

Datos provenientes de redes sociales: una reflexión sobre los desafíos y problemáticas relacionadas con el acceso y la reutilización

Frederico Ramos Oliveira*

Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict), Brasil

ORCID: <https://orcid.org/0000-0001-5653-5715>

Lattes: <http://lattes.cnpq.br/2791932279957633>

fredericooliveira@ibict.br

Graziela Barros Gomes

Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict), Brasil

ORCID: <https://orcid.org/0000-0001-8843-584X>

Lattes: <http://lattes.cnpq.br/2308190091007019>

graziela@ibict.br

Tainá Batista de Assis

Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict), Brasil

ORCID: <https://orcid.org/0000-0002-9176-4764>

Lattes: <http://lattes.cnpq.br/8018556894529318>

taina@ibict.br

Raíssa da Veiga de Meneses

Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict), Brasil

ORCID: <https://orcid.org/0000-0003-2503-9764>

Lattes: <http://lattes.cnpq.br/9129910362246718>

raissameneses@ibict.br

Larissa de Araújo Alves

Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict), Brasil

ORCID: <https://orcid.org/0000-0002-4757-1674>

Lattes: <http://lattes.cnpq.br/3532166143974051>

larissaalves@ibict.br

Cristiane Henrique Santos

Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict), Brasil

ORCID: <https://orcid.org/0009-0007-4437-1582>

Lattes: <http://lattes.cnpq.br/6194941001104710>

cristianehenrique@ibict.br

Tiago Emmanuel Nunes Braga

Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict), Brasil

ORCID: <https://orcid.org/0000-0001-6332-7965>

Lattes: <http://lattes.cnpq.br/8376134230259399>

tiagobraga@ibict.br

* Autor correspondente.

Resumo

Aborda os desafios relacionados à coleta, disponibilização e análise de dados provenientes de redes sociais, considerando o contexto da produção e comunicação científica marcado pela plataformização e datificação, bem como os princípios da Ciência Aberta. Trata-se de uma pesquisa de natureza exploratória e descritiva, que realiza o levantamento e a análise de conjuntos de dados oriundos de estudos que utilizam dados de redes sociais, disponibilizados nos repositórios Figshare, Zenodo, Deposita Dados e Aleia. Examina aspectos que comprometem o uso e o reúso desses dados, com ênfase em limitações relacionadas à documentação, padronização e acessibilidade. Discute os desafios para a adesão aos princípios FAIR (*Findable, Accessible, Interoperable, Reusable*) e CARE (*Collective Benefit, Authority to Control, Responsibility, Ethics*). Por fim, exemplifica ações e iniciativas do Instituto Brasileiro de Informação em Ciência e Tecnologia voltadas à estruturação de uma infraestrutura pública para coleta, processamento e análise de dados oriundos de redes sociais, com vistas ao fortalecimento da integridade, transparência e reprodutibilidade da pesquisa científica.

Palavras-chave:

Dados de pesquisa; Dados de redes sociais; Repositórios de dados; Reúso de dados.

Abstract

This study examines the challenges related to the collection, availability, and analysis of data derived from social media within the context of scholarly communication shaped by platformization, datafication, and Open Science principles. It is an exploratory and descriptive study that surveys and analyzes datasets from studies using social media data, available in the repositories Figshare, Zenodo, Deposita Dados, and Aleia. The analysis identifies key barriers to data use and reuse, particularly those related to documentation, standardization, and accessibility. It also discusses challenges associated with adherence to the FAIR (Findable, Accessible, Interoperable, Reusable) and CARE (Collective Benefit, Authority to Control, Responsibility, Ethics) principles. Finally, the study presents initiatives by the Brazilian Institute of Information in Science and Technology aimed at developing a public infrastructure for the collection, processing, and analysis of social media data, contributing to the strengthening of research integrity, transparency, and reproducibility.

Keywords:

Research data; Social media data; Data repositories; Data reuse.

Resumen

Este estudio examina los desafíos relacionados con la recolección, disponibilidad y análisis de datos provenientes de redes sociales en el contexto de la comunicación científica, marcado por la plataformización, la datificación y los principios de la Ciencia Abierta. Se trata de una investigación de carácter exploratorio y descriptivo que realiza el levantamiento y análisis de conjuntos de datos derivados de estudios que utilizan datos de redes sociales, disponibles en los repositorios Figshare, Zenodo, Deposita Dados y Aleia. El análisis identifica barreras clave para el uso y la reutilización de estos datos, especialmente aquellas relacionadas con la documentación, la estandarización y la accesibilidad. Asimismo, se discuten los desafíos asociados a la adopción de los principios FAIR (Findable, Accessible, Interoperable, Reusable) y CARE (Collective Benefit, Authority to Control, Responsibility, Ethics). Finalmente, ilustra iniciativas del Instituto Brasileño de Información en Ciencia y Tecnología orientadas al desarrollo de una infraestructura pública para la recolección, procesamiento y análisis de datos de redes sociales, contribuyendo al fortalecimiento de la integridad, transparencia y reproducibilidad de la investigación científica.

Palabras clave:

Datos de investigación; Social media data; Data repositories; Data reuse.

Introdução

Na cibercultura, as plataformas digitais promovem a intensa dataficação da vida (Lemos, 2021), produzindo dados sobre o mundo. Esses dados são acompanhados de promessas: ao acessá-los, seria possível um retrato objetivo do mundo (D'Andréa, 2021). No entanto, os dados das redes sociais estão cada vez mais restritos, sendo imperativo pensar estratégias para o seu compartilhamento e para o reúso de dados de pesquisa oriundos dessas plataformas. De igual modo, importa discutir infraestruturas públicas de coleta, análise e processamento, a fim de fomentar investigações científicas que se fundamentam em tais dados. Nesse sentido, o objetivo desta pesquisa é compreender os desafios relacionados à coleta, disponibilização, acesso e análise de dados de redes sociais de *internet*, considerando os princípios da Ciência Aberta.

Plataformas organizam a interação do usuário, sendo pensadas para maximizar a coleta, o tratamento e o processamento de dados do usuário (Djick, Poell & De Wall, 2018, Zuboff, 2020). No atual contexto de plataformação¹, dataficação e performatividade algorítmica (Lemos, 2020), a informação circula mediada pelas plataformas, que produzem, coletam e consomem ininterruptamente dados pessoais e dados de uso (Lemos, 2021). Dentre as consequências disso, o redesenho da economia, com novas lógicas de trabalho e de extração do valor (Grohmann, 2020), assim como reflexos ambientais (Parrika, 2015; Lemos et al., 2020). Promove-se a ideia de que tal política de extração e de produção de dados oferece objetividade, embora sua coleta envolva interpretação e intervenção dos algoritmos (Crawford, 2021).

A produção científica não passa incólume pela plataformação. A promessa de acesso a um volume significativo de dados – cuja produção não é objetiva, mas atravessada por sistemas de classificação (Koopman, 2019; Santaella & Kaufman, 2021) –, deu às plataformas centralidade significativa na definição de metodologias (D'Andréa, 2021). O sucessivo fechamento das interfaces de programação de aplicações (*Application Programming Interface* ou APIs), desde o escândalo da *Cambridge Analytica*, quando o impacto dessas APIs foi decisivo, reorganiza as lógicas de produção da pesquisa científica e muda o valor atribuído aos dados de plataformas. No caso de estudos baseados em métricas alternativas, por exemplo, a dificuldade de acesso aos dados das plataformas exige a definição de outras estratégias de pesquisa (Barcelos et al., 2025). Afinal, o acesso às APIs é cada vez mais restrito e caro.

Pode-se elencar desafios relacionados à coleta, ao tratamento e à análise dos dados; à comparabilidade entre tipos de dados produzidos por cada plataforma; ao manejo justo e ético da coleta de dados; à representação e ao depósito dos dados em repositórios; à reprodutibilidade das pesquisas que usam dados de redes sociais *online* e; por fim, à proteção de pesquisadores. No que tange aos desafios da coleta e análise de dados, o desenvolvimento de métodos digitais de pesquisa e recursos de raspagem buscam responder ao fechamento das APIs. Há de se considerar, contudo, que os dados das plataformas são produzidos (Koopman, 2019; Amore, 2020) do mesmo modo que as estratégias específicas de coleta

¹ É a “penetração de infraestruturas, processos econômicos e estruturas governamentais de plataformas em diferentes setores econômicos e esferas da vida” (Poell et al., 2020, p. 5).

podem interferir na integridade e na análise dos dados. A estratégia de raspagem utilizada, a seleção de perfis ou termos de busca, a forma de exportação dos dados: tudo isso implica em desafios para o tratamento, análise e reuso dos dados coletados das plataformas sociais.

Tais desafios podem ser observados quando se acessa *datasets* de pesquisas que utilizam dados de redes sociais que estão disponíveis no Figshare (<https://figshare.com/>) e no Zenodo (<https://zenodo.org/>). O levantamento desenvolvido no presente artigo demonstrou que a maior parte dos dados compartilhados não permite reuso e não é composta pelos dados originais. O habitual observado é o compartilhamento de dados de pesquisa como mera comprovação da execução da metodologia, não pensando no seu reuso. Desse modo, os *datasets* disponíveis são compostos, em sua maioria, por dados já processados.

Em atenção a esses desafios, e em consonância com a sua missão e trajetória institucional, o Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict) tem também atuado na concepção e implantação de infraestruturas informacionais e computacionais para a coleta, armazenamento, tratamento, análise e acesso a dados provenientes de plataformas digitais. Nesse sentido, destaca-se o *software* Capitu (<https://capitu.minerva.ibict.br/>), desenvolvido e coordenado pelo Instituto, que possibilita a coleta de dados de redes sociais e cria um robusto *data lake* com esses dados coletados. A infraestrutura encontra-se em fase de reorganização interna e a perspectiva é que o *software* esteja amplamente disponível para uso da comunidade acadêmica, como um serviço de informação voltado ao acompanhamento, em tempo real, do resultado da coleta de dados a partir da definição prévia dos critérios de busca.

A exemplo do uso prático dessa infraestrutura do Ibict, foi possível o desenvolvimento de produtos de análises sobre as narrativas desinformativas no campo das políticas públicas. Esses produtos foram concebidos em formatos de relatórios e os conjuntos de dados públicos provenientes das plataformas digitais que subsidiaram as análises foram depositados no repositório de dados do Ibict, o Aleia (<https://aleia.ibict.br/>).

A concepção de uma infraestrutura pública aberta para que pesquisadores possam realizar suas pesquisas, contribui para a celeridade das análises, a economicidade dos estudos a partir do incentivo ao reuso desses dados. No entanto, reforçam-se os desafios no tratamento e acesso a dados públicos provenientes das plataformas digitais, especialmente no que tange a respeito dos princípios FAIR (*Findable, Accessible, Interoperable, Reusable*) e CARE (*Collective Benefit, Authority to Control, Responsibility, Ethics*).

Desafios para compartilhamento e reuso de dados de redes sociais

Questões técnicas, jurídicas e éticas se relacionam à coleta, ao processamento, ao uso, ao compartilhamento e ao reuso de dados de redes sociais. No que tange à coleta, o fechamento das APIs força, muitas vezes, o usuário a adotar a raspagem de dados, ainda que tal estratégia possa implicar em vieses (resultantes da intervenção de interfaces e algoritmos) e problemas judiciais. A coleta por meio de API, por sua vez, não implica necessariamente no acesso a todos os dados: não se pode garantir que os

sistemas coletam, sempre, os mesmos *posts*. Questões relacionadas foram discutidas amplamente na literatura (D’Andréa, 2021).

Diante do acesso dificultado a esses dados, pensar o compartilhamento de tais dados é essencial. Pode-se apontar três desafios principais: (a) formato; (b) descrição adequada; e (c) questões legais. Em relação ao formato, o uso de formatos legíveis por máquina (*csv*, *json*, dentre outros) é essencial. Há de se considerar, contudo, que a coleta estabelece uma formatação de tal dado. Deste modo, se um pesquisador coletar tuítes, por exemplo, a partir de requisição à API e de raspagem de dados, simultaneamente, terá dois modelos distintos de planilha. Assim, nem sempre é possível garantir o reuso em estudos longitudinais, já que estratégias de coleta podem variar – resultando em estruturas distintas de dados.

Pesquisa anterior (Oliveira, 2023) realizou a coleta de *tweet* sobre o *kit gay* publicados em um período de dez anos (2010-2020). Embora tenha sido utilizada a mesma ferramenta, outra questão interferiu nos dados: em 2010, o X (anteriormente chamado de Twitter) não tinha *retweet* nativo e os usuários mencionavam conteúdos de outros usuários por meio de estruturas como “RT @usuário” ou “via @usuário”. É uma realidade distinta de 2020, quando foi realizada a coleta. Diante disso, optou-se por desconsiderar os *retweet*, ainda que se entenda sua relevância como endosso, críticas costumavam ser expressadas por meio da ferramenta *quote*.

Orienta-se, ainda, a disponibilização de uma planilha por rede ou por termo. Versões consolidadas que reúnem diversas coletas podem ser apresentadas, desde que se deixe explícito os dados de cada coleta / rede. Afinal, há relação direta entre a interface de cada plataforma e os dados extraídos. Ainda no que tange ao formato, cabe apontar que retirar os conteúdos da interface na qual circulam pode apagar intervenções relevantes das plataformas, que não podem ser desconsideradas (Oliveira & Lemos, 2025).

No que se refere à descrição dos dados, é importante garantir que a metodologia seja apresentada de forma transparente, assim como permitir, sempre que possível, o acesso ao dado como extraído. Dados processados e analisados podem ser inseridos conjuntamente, desde que como complemento para o entendimento do *dataset*. Sugere-se que uma nota metodológica seja apresentada, indicando período de coleta, redes coletadas, termos de pesquisa, possíveis usos dos dados e resultados já identificados. Para ampliar as possibilidades de uso das coletas, é importante garantir que quem as utilize entenda sua produção.

Por fim, há diversas questões legais a serem discutidas. Dados disponibilizados, assim como coletados, nem sempre são anonimizados, o que pode implicar em dilemas éticos e graves questões jurídicas. Apagar a autoria, contudo, impede observar a formação de redes de autores, observar *clusters* semânticos relacionados a grupos de usuários, dentre diversas outras análises. Por sua vez, instituições mantenedoras de repositórios com dados de redes sociais podem ser alvo de processo das plataformas, que querem impedir o compartilhamento dos dados. A insegurança jurídica e os dilemas éticos, portanto, são desafios a serem enfrentados.

Procedimentos metodológicos

Considerando o objetivo geral, foi desenvolvida uma pesquisa exploratória e descritiva, com abordagem qualitativa. A primeira etapa da investigação consistiu em um levantamento de *datasets* de pesquisas que utilizam dados de redes sociais que estão disponíveis no Figshare e no Zenodo. A proposta desta fase é a compreensão de como tais dados são compartilhados: o tipo de dado compartilhado, o tipo de acesso disponibilizado e se há anonimização. Para tanto, consultou-se o Figshare e o Zenodo com o termo “social media data”. A busca pelas bases de dados foi realizada em 25 de março de 2026, a partir do termo supracitado. Reconhece-se que tal termo pode implicar em resultados pouco relevantes, o que se buscou contornar acessando todos os *datasets* e fazendo a análise manualmente.

Foram identificados, em um primeiro momento, 209 bases de dados no Figshare e 51 no Zenodo. Foram considerados *datasets* com dados oriundos de plataformas de redes sociais, ainda que processados. Sendo assim, foram excluídas bases com dados que não eram oriundos de redes sociais, duplicatas, bases de modelos estatísticos ou compostas por artigos científicos usados em revisão de literatura. Foram incluídos, portanto, apenas bases de dados que, na descrição, explicitam a coleta de dados de plataformas de redes sociais. Desta forma, o *corpus* foi composto por 181 bases no Figshare e 51 no Zenodo.

No que tange à análise, identificou-se o tipo de coleta (raspagem, acesso por meio de APIs ou outros), se os dados eram anonimizados, se estão disponíveis em acesso aberto, bem como comentários gerais sobre as bases de dados analisadas. Os artigos relacionados foram acessados, a fim de se identificar o tipo de coleta que permitiu o acesso aos dados de redes sociais. Tal levantamento se justifica quando se considera os desafios impostos à pesquisa em redes sociais com o fechamento das APIs (D’Andréa, 2021). A partir da compreensão de como tais dados são compartilhados, é possível entender como a comunidade científica prevê (ou não) seu reúso. Fez-se o *download* das bases, a fim de se observar se os dados estavam anonimizados, eram processados ou incluíam dados pessoais.

Em etapa complementar, no mesmo período, foram analisados repositórios de dados mantidos pelo Ibict, especialmente o Deposita Dados e o Aleia. A busca teve como objetivo identificar *datasets* relacionados a investigações que se fundamentam em dados de redes sociais, com compartilhamento dos dados. Foram encontradas 5 bases de dados, estando uma no Deposita Dados e quatro no Aleia. Parte-se, assim, para a apresentação dos dados disponíveis em tais repositórios. Por fim, apresenta-se o *software* Capitu e as ações do Ibict para construção de uma infraestrutura pública de coleta, processamento e análise de dados oriundos de redes sociais.

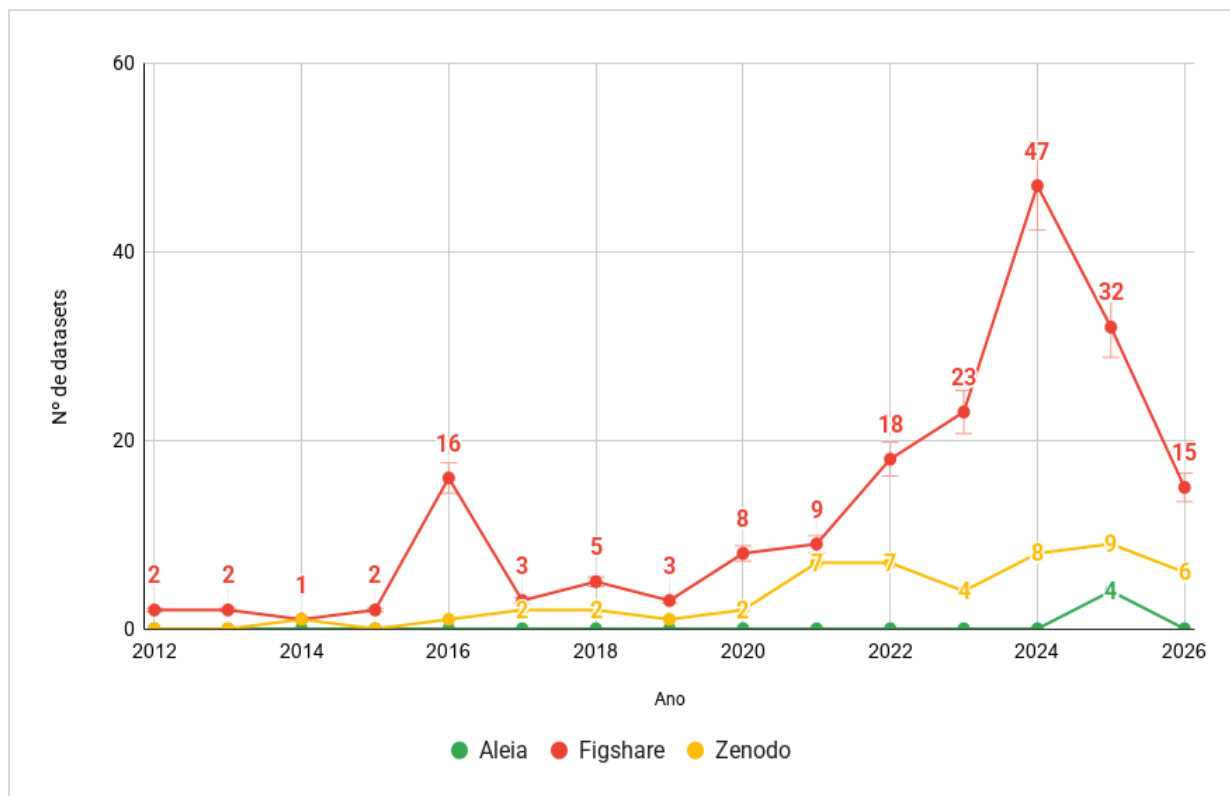
Ressalta-se que se trata de uma pesquisa em desenvolvimento, sendo o levantamento aqui apresentado preliminar. Em próximas etapas do estudo, busca-se identificar a tipologia de conteúdos oriundos de redes sociais disponíveis nos repositórios *online* (o inventário já desenvolvido demonstra que a maioria dos conteúdos são tuítes), as condições para reúso (se há descrição de *queries* e estratégias de coleta, por exemplo), assim como a proposição de um modelo para reúso de dados de redes sociais.

Resultados e discussão

O compartilhamento de dados de redes sociais em repositórios como Figshare e Zenodo geralmente é um recurso usado pelos autores para comprovar que aplicaram a metodologia descrita em seus textos. Afinal, poucos se preocupam em dar condições de reuso dos dados: parte significativa dos *datasets* não lista os termos de busca ou é composta por dados já processados. O acesso ao conteúdo dos *posts* é muitas vezes restrito ao texto da postagem (retirando dados importantes como data de publicação e, até mesmo, autoria). Em outros momentos, somente o *id* das postagens é indicado, enquanto muitos *datasets* se restringem a dados já processados. O levantamento pode ser acessado em Oliveira et al., 2026.

Figura 1

Distribuição de datasets por repositório de dados (2012-2026)



Nota: Dados da pesquisa (2026).

Ao todo, foram analisados 281 conjuntos de dados, com base na coleta das seguintes variáveis qualitativas: repositório de origem, ano, *link* de acesso, descrição, tipo de coleta, disponibilização em acesso aberto e indicação de anonimização dos dados. A Figura 1 apresenta o quantitativo de *datasets* identificados no período de 2012 a 2026.

Observa-se que a totalidade de conjunto de dados coletados pelo estudo, nos repositórios Figshare, Zenodo, Deposita Dados e Aleia, apresenta identificador persistente relacionado, fato que corrobora positivamente com o proposto pelas diretrizes do princípio *Findable*, que sinaliza a garantia de encontrabilidade dos dados e metadados diante dos identificadores únicos (Sales et al., 2021).

A análise dos *datasets* encontrados no levantamento demonstra um crescimento significativo no número de *datasets* publicados a partir de 2021, tendo seu pico no ano de 2024. Nota-se que 221 dos conjuntos de dados, dentre os 241 componentes da amostra da pesquisa, foram publicados em acesso aberto. Ainda quanto ao tipo de acesso aos conjuntos de dados analisados, foi identificada a possibilidade parcial de acesso em 10 *datasets*, mediante cadastro ou solicitação de acesso aos dados. Nesse sentido, faz-se válida a pontuação de que a exigência de cadastro e a existência de *datasets* fechados, a depender do caso considerado, podem gerar desdobramentos quanto à reprodutibilidade e reuso dos dados coletados.

Em relação aos aspectos éticos dos dados analisados, observa-se que muitos *datasets* não informam se os dados coletados passaram por processo de anonimização, além dos casos em que foi possível identificar explicitamente que os dados não foram anonimizados. Esse panorama levanta preocupações no que tange à proteção de dados e privacidade dos indivíduos.

Com o levantamento, foi possível identificar um único caso de reuso de dados: Bahamonde et al. (2018) utilizam dois *datasets* de *tweets* produzidos por outros pesquisadores. Sua pesquisa investiga se a similaridade de temas publicados no X (Twitter) por jornais chilenos estaria relacionada à propriedade desses veículos. No entanto, os dados que compartilham em seu texto incluem somente a lista de veículos da imprensa e seu *username* no X.

Há de se apontar, no entanto, que o reuso de *datasets* só é possível por se tratar de uma publicação de 2018: desde o fechamento da API, o X não permite compartilhamento dos *tweets* coletados, apenas do *id* das postagens. Por isso, o *dataset* “COVID-19 Twitter Dataset”, do Figshare (<https://doi.org/10.6084/m9.figshare.16713448>), deixa explícito que o conteúdo não pode ser compartilhado em função dos termos de uso da API.

No caso do Deposita Dados, somente um *dataset* foi encontrado. Trata-se de coleta que apenas identificava se um pesquisador possuía ou não perfil no Facebook. Os dados compartilhados não incluem nomes de usuário, mas também não podem ser reutilizados por sua natureza. No Aleia, as bases de dados disponibilizadas estão acessíveis mediante cadastro e incluem os dados brutos de coletas específicas. O acesso restrito, no entanto, se dá em função da natureza dos dados, que inclui conteúdos não anonimizados. Nesse contexto, as crescentes restrições de acesso a dados provenientes de plataformas digitais, considerando aspectos de bloqueio completo ou na limitação de APIs, evidenciam assimetrias de recursos para a realização de coletas desta tipologia de dados de pesquisa e reforçam a necessidade de consolidação de infraestruturas públicas de pesquisa orientadas à coleta, curadoria e compartilhamento de dados.

Do ponto de vista da soberania científica, entendida como implantação de estratégias e ferramentas eficazes para assegurar a autonomia, consistência, o acesso e a precisão da informação, infraestruturas públicas voltadas para dados de pesquisa são fundamentais para mitigar desigualdades no acesso à informação, promover a transparência e assegurar condições mais equitativas para o desenvolvimento científico, em consonância com os princípios da Ciência Aberta. Além de favorecerem a preservação e a documentação dos dados, essas iniciativas potencializam o reuso qualificado de dados primários, ampliando a reprodutibilidade dos estudos e viabilizando análises longitudinais e comparativas sobre fenômenos mediados por plataformas digitais.

Tendo em vista o cenário posto, o Ibict propõe a Capitu, *software* que possibilita aos pesquisadores o desenvolvimento de pesquisas a partir da coleta de dados provenientes de redes sociais e seu acompanhamento em tempo real, a partir da definição de critérios de busca, por meio de termos e palavras-chave, bem como a definição de recorte temporal de interesse da coleta. Ressalta-se que a presente pesquisa serve de insumo para a proposição de política de reuso de dados de redes sociais no âmbito do projeto Capitu.

Considerações

O acompanhamento do debate público que circula nas plataformas digitais se mostra como uma ação essencial também no campo da pesquisa científica, uma vez que essas plataformas se tornam novos espaços de interação, disseminação e publicização de resultados da ciência em diversas disciplinas. Para além disso, esse acompanhamento se torna também chave para a identificação de narrativas desinformativas que colocam em xeque as descobertas científicas.

Todavia, as barreiras identificadas para a obtenção, tratamento, disponibilização e reuso dos dados públicos provenientes das plataformas digitais dificultam as análises, como por exemplo, a profundidade da discussão do problema de pesquisa e também a correlação entre dados. Em última instância, essas barreiras são opostas aos próprios objetivos de universalização do acesso aos dados da World Wide Web, proposta por Tim-Berners Lee em 1989, e não condizem com as práticas defendidas por conceitos como o FAIR e o CARE.

Nesse sentido, este trabalho contribui diretamente para impulsionar a reflexão e debate quanto à importância da consideração dessa tipologia de dados de pesquisa na agenda da Ciência Aberta como uma forma de fortalecer o acesso e o reuso dos dados de pesquisa proveniente de plataformas digitais e fomentar reflexões associadas à soberania informacional a partir da proposição e fornecimento de infraestruturas abertas de pesquisa.

Conflito de interesses

Os autores declaram não haver conflitos de interesses.

Disponibilização dos Dados de Investigação

Os dados estão disponibilizados no link: <https://zenodo.org/records/19359118>

CRedit – Contribuições dos Autores

Frederico Ramos Oliveira | Conceitualização, Escrita – redação original, Supervisão, Investigação

Graziela Barros Gomes | Conceitualização, Escrita – redação original, Investigação

Tainá Batista de Assis | Escrita – revisão e edição

Raíssa da Veiga de Meneses | Escrita – revisão e edição

Larissa de Araújo Alves | Escrita – revisão e edição

Cristiane Henrique Santos | Escrita – revisão e edição, Investigação

Tiago Emmanuel Nunes Braga | Escrita – revisão e edição, Supervisão

Referências

Amoore, L. (2020). *Cloud ethics: algorithms and the attributes of ourselves and others*. Duke University Pres. <https://doi.org/10.1215/9781478009276>

Bahamonde, J., Bollen, J., Elejalde, E., Ferres, L., & Poblete, B. (2018). Power structure in Chilean news media. *PLOS ONE*, 13(6), e0197150. <https://doi.org/10.1371/journal.pone.0197150>

Barcelos, J., Oliveira, F., & Souza, M. (2025). Almetria: Desafios éticos, técnicos e políticos diante da plataformação. *Ciência da Informação*, 54(2), 1-14. <https://doi.org/10.18225/ci.inf.v54i2.7232>

Crawford, K. (2021). *The Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. Yale University Press. <https://doi.org/10.2307/j.ctv1ghv45t>

D'Andréa, C. (2021). Para além dos dados coletados: Políticas das APIs nas plataformas de mídias digitais. *MATRIZES*, 15(1), 103–122. <https://doi.org/10.11606/issn.1982-8160.v15i1p103-122>

Dijck, V., Poell, T., & Waal, M. (2018). *The platform society: Public values in a connective world*. Oxford university press.

Grohmann, R. (2020). Plataformização do trabalho: entre dataficação, financeirização e racionalidade neoliberal. *Revista Eletrônica Internacional De Economia Política Da Informação Da Comunicação E Da Cultura*, 22(1), 106–122. <https://periodicos.ufs.br/eptic/article/view/12188>

Koopman, C. (2019). *How we became our data: a genealogy of the informational person*. University of Chicago Press. <https://press.uchicago.edu/ucp/books/book/chicago/H/bo38181810.html>

Lemos, A. (2020). Plataformas, Dataficação e Performatividade Algorítmica (PDPA): desafios atuais da cibercultura. In N. Prata, & S. Caldas (Org.), *Fluxos comunicacionais e crise da democracia* (pp 107-116). Intercom.

Lemos, A. (2021). Dataficação da vida. *Civitas: Revista de Ciências Sociais*, 21(2), 193-202. <https://revistaseletronicas.pucrs.br/civitas/article/view/39638>

Lemos, A., Bittencourt, E., & Santos, J. G. B. (2020). Fake news as fake politics: the digital materialities of YouTube misinformation videos about Brazilian oil spill catastrophe. *Media, Culture & Society*, 43(3), 886-905. <https://journals.sagepub.com/doi/full/10.1177/0163443720977301>

Oliveira, F. R. (2023). *As fake news e a produção jornalística de referências* [Tese de pós-graduação na Universidade Federal da Bahia]. https://bdtd.ibict.br/vufind/Record/UFBA-2_02474ecd06b6031bfe8a5778b16e18

Oliveira, F., & Lemos, A. (2025). Como as plataformas produzem desinformação: interfaces, algoritmos e conteúdos falsos. *Contracampo*, 44(2), 1-23. <https://doi.org/10.22409/v95zqw38>

Oliveira, F., Meneses, R. da V. de, Gomes, G., Henrique, C., Braga, T., Assis, T., & Alves, L. de A. (2026). *Data from social media: a reflection on the challenges and issues related to access and reuse*. Zenodo. <https://zenodo.org/records/19354501>

Parrika, J. (2015). *A geology of media*. University of Minnesota Press.

Poell, T., Nieborg, D., & Dijck, J. V. (2020). Plataformização. *Fronteiras - estudos midiáticos*, 22(1), 2–10. <https://doi.org/10.4013/fem.2020.221.01>

Sales, L., Veiga, V., Henning, P., & Sayão, L. (2021). *Princípios FAIR aplicados à gestão de dados de pesquisa*. Ibict. <http://ridi.ibict.br/handle/123456789/1182>

Santaella, L., & Kaufman, D. (2021). Os dados estão nos engolindo? *Civitas - Revista de Ciências Sociais*, 21(2), 214–223. <https://doi.org/10.15448/1984-7289.2021.2.39640>

Zuboff, S. (2020). *A era do capitalismo de vigilância: a luta por um futuro humano na nova fronteira de poder*. Intrínseca.