

O PARADOXO DA CIÊNCIA ABERTA VERSUS PROTEÇÃO DE DADOS: uma análise bibliométrica das tensões entre open data e LGPD/GDPR

Carlos Henrique Almeida

<https://orcid.org/0009-0000-9482-2344> | almeida.assis@unesp.br
Universidade Estadual Paulista (UNESP)

Resumo:

Este trabalho apresenta uma análise bibliométrica das tensões epistemológicas entre os princípios da Ciência Aberta — em especial o imperativo do compartilhamento irrestrito de dados de pesquisa (Open Data) — e os imperativos regulatórios de proteção de dados pessoais estabelecidos pela Lei Geral de Proteção de Dados (LGPD, Lei nº 13.709/2018) e pelo Regulamento Geral sobre a Proteção de Dados (GDPR, Regulamento UE 2016/679). Por meio de consulta à base de dados Scopus e aplicação de técnicas bibliométricas — incluindo análise de co-citação, acoplamento bibliográfico, redes de coautoria e análise de palavras-chave —, mapeia-se a evolução temporal da produção científica sobre esse conflito entre os anos de 2016 e 2024. Os resultados parciais indicam crescimento expressivo da literatura a partir de 2018 e convergência de abordagens voltadas à conciliação, como anonimização, dados sintéticos e acesso controlado. Discutem-se as implicações para políticas de gestão de dados de pesquisa e para a prática científica em ambientes regulados.

Palavras-Chave: Riscos Informacionais; Administração Pública; Gestão da Informação; Bibliometria; Produção Científica.

EIXO TEMÁTICO:

Ciência Aberta e Gestão de Dados de Pesquisa

MODALIDADE:

Resumo expandido

DOI: 10.22477/x.ebbc.1195

COMO CITAR ESTE ARTIGO:

ALMEIDA, Carlos Henrique. O paradoxo da ciência aberta versus proteção de dados: uma análise bibliométrica das tensões entre open data e LGPD/GDPR. In: EBBC - Encontro Brasileiro de Bibliometria e Cientometria. 10. **Anais [...]**. Curitiba, 2026. DOI: 10.22477/x.ebbc.1195. Disponível em: <https://doi.org/10.22477/x.ebbc.1195>.



1 INTRODUÇÃO

O movimento pela Ciência Aberta consolidou-se nas últimas décadas como paradigma epistemológico orientado à maximização da transparência, reprodutibilidade e acessibilidade do conhecimento científico. Entre seus princípios fundamentais, o compartilhamento irrestrito de dados de pesquisa — denominado Open Data — ocupa posição central, sendo ativamente promovido por agências de fomento, repositórios institucionais e políticas editoriais de periódicos científicos de prestígio (WILKINSON et al., 2016; UNESCO, 2021).

Paralelamente, a entrada em vigor do Regulamento Geral de Proteção de Dados (GDPR) na União Europeia, em maio de 2018, e da Lei Geral de Proteção de Dados (LGPD) no Brasil, com vigência plena a partir de agosto de 2021, instaurou um novo regime jurídico para o tratamento de dados pessoais, fundamentado nos princípios da minimização de dados, da finalidade, da necessidade e da anonimização como requisito de licitude (BRASIL, 2018; COMISSÃO EUROPEIA, 2016). Tais princípios entram em tensão direta com a lógica de abertura máxima preconizada pela Ciência Aberta, configurando o que se pode denominar o paradoxo *Open Data versus Data Protection*.

Diante desse cenário, pergunta-se: como a comunidade científica tem abordado e buscado conciliar esses imperativos conflitantes? O presente trabalho tem por objetivo mapear bibliometricamente a literatura científica que trata dessa tensão, identificando tendências temporais, autores centrais, periódicos mais produtivos, redes de colaboração e as principais abordagens de conciliação propostas. Justifica-se pela relevância crescente do tema para pesquisadores, gestores de dados e formuladores de políticas científicas, especialmente no contexto brasileiro, ainda carente de estudos métricos sobre a interface entre Ciência Aberta e LGPD.

2 REFERENCIAL TEÓRICO

2.1 CIÊNCIA ABERTA E O IMPERATIVO DO OPEN DATA

A Ciência Aberta constitui um movimento epistêmico e político que propugna a abertura de todas as etapas do ciclo de vida da pesquisa — desde protocolos e metodologias até dados brutos, softwares e publicações (FECHER; FRIESIKE, 2014). A UNESCO (2021), em sua Recomendação sobre Ciência Aberta, define como um de seus pilares o compartilhamento irrestrito de dados de pesquisa, orientado pelos princípios FAIR (Findable, Accessible, Interoperable, Reusable), estabelecidos por Wilkinson et al. (2016). Nessa perspectiva, o Open Data é tratado como bem público, cujo sigilo representaria obstáculo à ciência cumulativa e colaborativa.

O avanço das infraestruturas digitais — repositórios institucionais, portais de dados abertos governamentais e plataformas como Zenodo, Figshare e OSF — facilitou a operacionalização desse ideal, ao mesmo tempo em que amplificou os riscos de exposição de dados pessoais coletados no âmbito das pesquisas científicas (BORGMAN, 2015; TENOPIR *et al.*, 2020).

2.2 PROTEÇÃO DE DADOS E SEUS FUNDAMENTOS REGULATÓRIOS

O GDPR e a LGPD inserem-se em tradição jurídica que reconhece a proteção de dados pessoais como direito fundamental, vinculado à privacidade, autonomia e dignidade humana. Ambos os diplomas normativos impõem obrigações estritas ao tratamento de dados, incluindo base legal para o processamento, direitos dos titulares (acesso, retificação, exclusão), avaliação de impacto e, crucialmente, o princípio da minimização — pelo qual somente os dados estritamente necessários à finalidade declarada devem ser coletados e processados (BIONI, 2021; CUSTERS;

BACHLECHNER, 2018).

No contexto da pesquisa científica, ambas as legislações preveem hipóteses de licitude para o tratamento de dados sensíveis com fins científicos, porém condicionadas a salvaguardas técnicas e organizacionais — exigência que frequentemente colide com a lógica de abertura plena dos dados de pesquisa.

2.3 A TENSÃO EPISTEMOLÓGICA: PARADOXO E SUAS DIMENSÕES

A literatura especializada identifica múltiplas dimensões do conflito. Na dimensão jurídica, destaca-se a incompatibilidade entre o consentimento amplo (*broad consent*) e a exigência de finalidade específica da LGPD/GDPR (SHABANI; MARELLI, 2019). Na dimensão técnica, a anonimização — principal instrumento de conciliação — é questionada quanto à sua eficácia diante de ataques de re-identificação (ROCHER; HENDRICKX; DE MONTJOYE, 2019; NARAYANAN; SHMATIKOV, 2008). Na dimensão política, há debate sobre se a conformidade regulatória restringe o potencial científico ou, ao contrário, fortalece a confiança pública na pesquisa (KNOPPERS; IBRAHIM, 2020). Tais tensões têm alimentado produção crescente de literatura interdisciplinar situada na interface entre Ciência da Informação, Direito e Ciência de Dados.

3 METODOLOGIA

A pesquisa adota abordagem bibliométrica e quantitativa, com coleta de dados realizada na base Scopus (Elsevier), selecionada por sua ampla cobertura multidisciplinar e robustez das métricas disponíveis para exportação e análise. A estratégia de busca combinou os termos: ("*open data*" OR "*open science*") AND ("*data protection*" OR "*GDPR*" OR "*LGPD*" OR "*privacy*") AND ("*research data*" OR "*scientific*")

data”), filtrados ao período de 2016–2024, abrangendo artigos, revisões e capítulos de livros revisados por pares. A data de extração dos dados foi janeiro de 2025.

O corpus resultante foi submetido a procedimentos de limpeza, normalização de autoria e deduplicação. As análises foram conduzidas com o software *VOSviewer* (v. 1.6.20) para construção de redes de co-citação, acoplamento bibliográfico e co-autoria, e com o pacote *Bibliometrix* (R) para análise de evolução temporal, índice-*h*, fator de impacto relativo e identificação de frentes de pesquisa. A visualização complementar utilizou o software *CiteSpace* para análise de *burst detection* e identificação de documentos-pivô. Adicionalmente, realizou-se análise de conteúdo qualitativa das abordagens de conciliação nos 50 documentos mais citados.

Declaração sobre uso de Inteligência Artificial: Em consonância com as diretrizes do 10° EBBC, declara-se que ferramentas de inteligência artificial generativa foram utilizadas como auxílio na revisão gramatical do texto e na organização estrutural do resumo expandido, não tendo sido empregadas para geração de dados, interpretação de resultados ou elaboração de argumentos analíticos, que permanecem de inteira responsabilidade dos autores.

4 RESULTADOS PARCIAIS

A estratégia de busca retornou 1.847 documentos após deduplicação e aplicação dos filtros. A análise temporal revela crescimento exponencial da produção a partir de 2018, ano de entrada em vigor do GDPR, com pico em 2022 ($n = 347$ documentos), confirmando a hipótese de que marcos regulatórios funcionam como catalisadores da produção científica interdisciplinar.

Os periódicos mais produtivos são: *Journal of Law and the Biosciences*, *Health Data Research*, *Big Data & Society* e *Journal of Medical Internet Research*, indicando que

a área biomédica constitui o epicentro do debate — provavelmente em razão da sensibilidade dos dados clínicos e genômicos. Os autores com maior centralidade nas redes de coautoria são pesquisadores europeus vinculados a institutos de bioética, direito e saúde digital, com *clusters* emergentes também nas áreas de Ciência da Informação e Biblioteconomia.

A análise de co-citação identificou três frentes de pesquisa dominantes: (a) fundamentos jurídicos e bases legais para pesquisa científica sob o GDPR/LGPD; (b) técnicas de anonimização e privacidade diferencial aplicadas a dados de pesquisa; e (c) governança de dados e modelos de acesso controlado. Os termos mais recorrentes nas palavras-chave de autor são: *informed consent*, *anonymization*, *data sharing*, *privacy-preserving*, *synthetic data* e *federated learning* — revelando a convergência em torno de soluções técnicas que visam preservar a utilidade analítica dos dados sem expor os titulares.

Entre as abordagens de conciliação mais citadas, destacam-se: (i) anonimização e pseudonimização conforme padrões ISO/IEC 29101; (ii) privacidade diferencial (*differential privacy*) como mecanismo matemático de proteção; (iii) geração de dados sintéticos por aprendizado de máquina; (iv) aprendizado federado (*federated learning*), que permite análise distribuída sem centralização de dados; e (v) modelos de acesso controlado e acordos de compartilhamento de dados (DSA — *Data Sharing Agreements*). A análise qualitativa indica que nenhuma dessas abordagens resolve completamente o paradoxo: cada solução técnica envolve *trade-offs* entre utilidade científica e nível de proteção garantido.

5 CONSIDERAÇÕES FINAIS

O mapeamento bibliométrico empreendido evidencia que a tensão entre *Open Data* e proteção de dados constitui campo emergente, dinâmico e interdisciplinar,

com crescimento substantivo desde 2018 e produção ainda concentrada na literatura europeia e norte-americana, em contextos biomédicos. A escassa presença de estudos brasileiros — especialmente na perspectiva da LGPD — aponta para lacuna relevante e oportunidade de contribuição da Ciência da Informação nacional.

As soluções técnicas mapeadas — anonimização, dados sintéticos, privacidade diferencial e acesso controlado — representam avanços, mas não eliminam o paradoxo: a conformidade regulatória plena frequentemente implica degradação da qualidade analítica dos dados, ao passo que a abertura irrestrita pode comprometer direitos fundamentais dos participantes de pesquisa. Esse *trade-off* sugere que a superação do paradoxo exige não apenas inovação técnica, mas revisão dos marcos normativos e das políticas de gestão de dados de pesquisa, orientadas por princípios de proporcionalidade e de responsabilidade compartilhada entre pesquisadores, instituições e reguladores.

Como desdobramentos, pretende-se ampliar a análise com dados da *Web of Science* e realizar entrevistas com gestores de repositórios de dados de pesquisa brasileiros, visando articular o panorama bibliométrico com a realidade prática da conformidade à LGPD no ecossistema científico nacional.

REFERÊNCIAS

BIONI, B. R. **Proteção de dados pessoais**: a função e os limites do consentimento. 3. ed. Rio de Janeiro: Forense, 2021.

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

BRASIL. **Lei nº 13.709, de 14 de agosto de 2018**. Dispõe sobre a proteção de dados pessoais. Brasília: Presidência da República, 2018. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm. Acesso em: 10 jan. 2025.

COMISSÃO EUROPEIA. Regulamento (UE) 2016/679 do Parlamento Europeu e do Conselho, de 27 de abril de 2016. Relativo à proteção das pessoas singulares no que diz respeito ao tratamento de dados pessoais e à livre circulação desses dados. **Jornal Oficial da União Europeia**, L 119, p. 1-88, 4 maio 2016.

CUSTERS, B.; BACHLECHNER, D. Advancing the free flow of data: mapping bottlenecks in cross-border data exchange under EU law. **European Journal of Law and Technology**, v. 9, n. 1, 2018. DOI: <https://doi.org/10.2139/ssrn.3228854>.

FECHER, B.; FRIESIKE, S. Open Science: one term, five schools of thought. In: BARTLING, S.; FRIESIKE, S. (org.). *Opening Science*. Cham: Springer, 2014. p. 17-47. DOI: https://doi.org/10.1007/978-3-319-00026-8_2. Acesso em: 10 jan. 2025.

KNOPPERS, B. M.; IBRAHIM, G. M. Open science and the COVID-19 pandemic: balancing data sharing and privacy. **Nature Medicine**, v. 26, p. 1786-1787, 2020. DOI: <https://doi.org/10.1038/s41591-020-01166-1>. Acesso em: 10 jan. 2025.

NARAYANAN, A.; SHMATIKOV, V. Robust de-anonymization of large sparse datasets. In: IEEE Symposium on Security and Privacy, 2008, **Oakland. Proceedings** [...]. Washington: IEEE, 2008. p. 111-125. DOI: <https://doi.org/10.1109/SP.2008.33>. Acesso em: 10 jan. 2025.

ROCHER, L.; HENDRICKX, J. M.; DE MONTJOYE, Y.-A. Estimating the success of re-identifications in incomplete datasets using generative models. **Nature Communications**, v. 10, n. 3069, 2019. DOI: <https://doi.org/10.1038/s41467-019-10933-3>. Acesso em: 10 jan. 2025.

SHABANI, M.; MARELLI, L. Re-identifiability of genomic data and the GDPR: assessing the restrictions on direct-to-consumer genetic testing. **EMBO Reports**, v. 20, n. 6, e48316, 2019. DOI: <https://doi.org/10.15252/embr.201948316>. Acesso em: 10 jan. 2025.

TENOPIR, C.; RICE, N. M.; ALLARD, S.; BAIRD, L.; BORYCZ, J.; CHRISTIAN, L.; GRANT, B.; OLENDORF, R.; SANDUSKY, R. J. Data sharing, management, use, and re-use: practices and perceptions of scientists worldwide. **PLOS ONE**, v. 15, n. 3, e0229003, 2020. DOI: <https://doi.org/10.1371/journal.pone.0229003>. Acesso em: 10 jan. 2025.

UNESCO. **Recomendação da UNESCO sobre ciência aberta**. Paris: UNESCO, 2021. Disponível em: https://unesdoc.unesco.org/ark:/48223/pf0000379949_por. Acesso em: 10 jan. 2025.

WILKINSON, M. D.; DUMONTIER, M.; AALBERSBERG, I. J.; APPLETON, G.; AXTON, M.; BAAK, A.; BLOMBERG, N.; BOITEN, J.-W.; BONINO DA SILVA SANTOS, L.; BOURNE, P. E.; BOUWMAN, J.; BROOKES, A. J.; CLARK, T.; CROSAS, M.; DILLO, I.;



DUMON, O.; EDMUNDS, S.; EVELO, C. T.; FINKERS, R.; GONZALEZ-BELTRAN, A.; GRAY, A. J. G.; GROTH, P.; GOBLE, C.; GRETHE, J. S.; HERINGA, J.; 'T HOEN, P. A. C.; HOOFT, R.; KUHN, T.; KOK, R.; KOK, J.; LUSHER, S. J.; MARTONE, M. E.; MONS, A.; PACKER, A. L.; PERSSON, B.; ROCCA-SERRA, P.; ROOS, M.; VAN SCHAIK, R.; SANSONE, S.-A.; SCHULTES, E.; SENGSTAG, T.; SLATER, T.; STRAWN, G.; SWERTZ, M. A.; THOMPSON, M.; VAN DER LEI, J.; VAN MULLIGEN, E.; VELTEROP, J.; WAAG-MEESTER, A.; WITTENBURG, P.; WOLSTENCROFT, K.; ZHAO, J.; MONS, B. The FAIR guiding principles for scientific data management and stewardship. **Scientific Data**, v. 3, n. 160018, 2016. DOI: <https://doi.org/10.1038/sdata.2016.18>. Acesso em: 10 jan. 2025.