

Pecha kucha

Identificadores Persistentes e Auditabilidade na Ciência Aberta: Reconfigurações em Contextos de Mediação por Inteligência Artificial

Persistent Identifiers and Auditability in Open Science: Reconfigurations in Contexts of AI-Mediated Communication

Identificadores persistentes y auditabilidad en la ciencia abierta: reconfiguraciones en contextos de mediación por inteligencia artificial

Raphael Figueiredo Xavier*

FGV – Fundação Getulio Vargas / UNESP – Universidade Estadual Paulista, Brasil

ORCID: <https://orcid.org/0000-0003-1916-8308>

Lattes: <http://lattes.cnpq.br/5243574435048251>

raphael.xavier@fgv.br

raphael.xavier@unesp.br

Fabiano Ferreira de Castro

UFSCar – Universidade Federal de São Carlos / UNESP – Universidade Estadual Paulista, Brasil

ORCID: <https://orcid.org/0000-0002-8712-2654>

Lattes: <http://lattes.cnpq.br/7124931056289027>

fabianocastro@ufscar.br

*Autor Correspondente

Resumo

A crescente incorporação de sistemas de inteligência artificial na comunicação científica tem reconfigurado os modos de produção, organização e circulação do conhecimento, intensificando desafios relacionados à opacidade e à auditabilidade dos fluxos informacionais. No contexto da Ciência Aberta, essas transformações incidem sobre dimensões como visibilidade, autoridade e curadoria, tensionando as condições de transparência, reprodutibilidade e verificabilidade. Diante desse cenário, este estudo investiga de que modo os identificadores persistentes contribuem para a reconfiguração das condições de auditabilidade em ambientes mediados por inteligência artificial. Metodologicamente, adota-se uma

abordagem qualitativa de caráter teórico-analítico, baseada na articulação entre resultados de uma revisão exploratória da literatura e referenciais sobre identificadores persistentes e infraestruturas de informação científica. Argumenta-se que, embora não tornem explícitos os critérios que orientam os processos algorítmicos, os identificadores persistentes viabilizam a reconstrução dos fluxos informacionais ao estruturarem vínculos entre entidades científicas. Como contribuição, o estudo evidencia o papel dos identificadores persistentes como componentes de infraestrutura que sustentam a rastreabilidade de relações e percursos informacionais em contextos marcados por opacidade algorítmica. Conclui-se que, em ambientes informacionais cada vez mais automatizados, os identificadores persistentes ampliam as condições de auditabilidade ao possibilitar a identificação, a conexão e o acompanhamento de entidades ao longo do tempo.

Palavras-chave

Identificadores Persistentes; Auditabilidade; Ciência Aberta; Inteligência Artificial; Mediação Algorítmica; Infraestruturas de Informação

Abstract

The increasing incorporation of artificial intelligence systems into scholarly communication has reshaped the ways in which knowledge is produced, organized, and circulated, intensifying challenges related to opacity and auditability in information flows. In the context of Open Science, these transformations affect key dimensions such as visibility, authority, and curation, challenging conditions of transparency, reproducibility, and verifiability. Against this backdrop, this study investigates how persistent identifiers contribute to the reconfiguration of auditability conditions in AI-mediated environments. Methodologically, it adopts a qualitative theoretical-analytical approach, based on the articulation between the results of a previous exploratory literature review and theoretical frameworks on persistent identifiers and scholarly information infrastructures. The study argues that, although persistent identifiers do not make explicit the criteria underlying algorithmic processes, they enable the reconstruction of information flows by structuring links between scholarly entities. As a contribution, the study highlights the role of persistent identifiers as infrastructure components that support the traceability of relationships and informational pathways in contexts marked by algorithmic opacity. It concludes that, in increasingly automated information environments, persistent identifiers enhance auditability conditions by enabling the identification, connection, and tracking of entities over time.

Keywords

Persistent Identifiers; Auditability; Open Science; Artificial Intelligence; Algorithmic Mediation; Information Infrastructures

Resumen

La creciente incorporación de sistemas de inteligencia artificial en la comunicación científica ha reconfigurado las formas de producción, organización y circulación del conocimiento, intensificando los desafíos relacionados con la opacidad y la auditabilidad de los flujos informacionales. En el contexto de la Ciencia Abierta, estas transformaciones inciden sobre dimensiones como la visibilidad, la autoridad y la curaduría, tensionando las condiciones de transparencia, reproducibilidad y verificabilidad. Ante este escenario, este estudio investiga de qué modo los identificadores persistentes contribuyen a la reconfiguración de las condiciones de auditabilidad en entornos mediados por inteligencia artificial. Metodológicamente, se adopta un enfoque cualitativo de carácter teórico-analítico, basado en la articulación entre los resultados de una revisión exploratoria de la literatura y marcos teóricos sobre identificadores persistentes e infraestructuras de información científica. Se sostiene que, aunque no hacen explícitos los criterios que orientan los procesos algorítmicos, los identificadores persistentes permiten la reconstrucción de los flujos informacionales al estructurar vínculos entre entidades científicas. Como contribución, el estudio destaca el papel de los identificadores persistentes como componentes de infraestructura que sostienen la trazabilidad de relaciones y trayectorias informacionales en contextos marcados por opacidad algorítmica. Se concluye que, en entornos informacionales cada vez más automatizados, los identificadores persistentes amplían las condiciones de auditabilidad al permitir la identificación, la conexión y el seguimiento de entidades a lo largo del tiempo.

Palabras clave

Identificadores persistentes; Auditabilidad; Ciencia Abierta; Inteligencia artificial; Mediación algorítmica; Infraestructuras de información

1. Introdução

A incorporação de sistemas de inteligência artificial generativa (IAg) na comunicação científica reconfigura os modos de produção, organização e circulação do conhecimento. Essa mudança traz desafios de opacidade e de auditabilidade nos fluxos informacionais (Ahn, 2024; Hosseini et al., 2025; Veiga, 2025). Sistemas de IAg influenciam critérios de visibilidade, autoridade e curadoria — funções historicamente exercidas por curadoria humana. Raji et al. (2020) mostram que a ausência de mecanismos de documentação e responsabilização cria uma lacuna: decisões automatizadas tornam-se difíceis de auditar e de atribuir.

Os identificadores persistentes (Persistent Identifiers – PIDs) oferecem um caminho distinto. Eles identificam, ligam e integram entidades do ecossistema de pesquisa, viabilizando identificação unívoca, resolução persistente e vinculação estruturada entre registros (Bunakov, 2020; Wittenburg, 2019).

Este estudo parte da seguinte questão: de que modo os identificadores persistentes contribuem para a reconfiguração das condições de auditabilidade dos fluxos científicos em contextos de mediação por inteligência artificial? O objetivo é analisar como as propriedades estruturais dos PIDs — identificação

unívoca, resolução persistente, vinculação entre entidades — mitigam riscos de opacidade e favorecem a rastreabilidade dos fluxos informacionais na Ciência Aberta.

A hipótese é a seguinte: os PIDs não eliminam a opacidade dos sistemas de inteligência artificial. Eles reconfiguram as condições de auditabilidade porque introduzem camadas estruturadas de identificação, conexão e registro entre entidades científicas. Esse deslocamento de perspectiva — de uma leitura técnico-especializada sobre PIDs para uma abordagem informacional e de infraestrutura — é a principal contribuição deste estudo, associando os PIDs às condições de transparência e governança da informação científica.

Metodologicamente, o estudo articula dois movimentos. Primeiro, recupera resultados de uma pesquisa exploratória anterior (Xavier et al., 2025), da qual emergiram três eixos analíticos: visibilidade automatizada, autoridade algorítmica e deslocamento da curadoria. Segundo, mobiliza referenciais teóricos sobre identificadores persistentes e infraestruturas de informação científica.

2. Mediação algorítmica e identificadores persistentes

Em estudo anterior, Xavier et al. (2025) identificaram três eixos para compreender os impactos da IA generativa (IAg) na Ciência Aberta: visibilidade automatizada, autoridade algorítmica e deslocamento da curadoria informacional. Os três eixos organizam a discussão desta seção.

Na visibilidade automatizada, sistemas de recomendação e síntese definem, de modo pouco transparente, quais conteúdos circulam. Hosseini et al. (2025) apontam o risco de esses sistemas reproduzirem vieses e favorecerem padrões já consolidados de publicação. Königstorfer et al. (2024) chegam a um resultado próximo: a automação da seleção informacional tende a privilegiar autores, periódicos e áreas já dominantes na literatura.

Na autoridade algorítmica, assistentes baseados em modelos de linguagem sintetizam conteúdos e sugerem referências. Parte da argumentação científica passa a ser produzida por sistemas cujos critérios não são plenamente verificáveis, o que Ballester (2023) chama de "autoridade difusa". Veiga (2025) descreve o mesmo efeito: decisões informacionais influenciadas por mecanismos automatizados pouco documentados comprometem a transparência e a reprodutibilidade.

No deslocamento da curadoria informacional, funções antes exercidas por bibliotecas, editoras e repositórios passam para plataformas digitais e sistemas algorítmicos. Rockembach (2024) e Silveira e Silva (2024) descrevem essas plataformas como novos gatekeepers da informação científica: elas decidem, ainda que de forma implícita, o que circula e ganha legitimidade.

Esses três eixos têm um efeito em comum, que Xavier et al. (2025) já indicavam: tornam mais difícil reconstruir os critérios que orientam a seleção, a validação e a organização da informação científica. Raji et al. (2020) descrevem essa dificuldade como uma lacuna de responsabilidade: decisões automatizadas sem documentação adequada tornam-se difíceis de auditar. É essa lacuna, e não os três eixos

isoladamente, que abre espaço para os identificadores persistentes: eles não corrigem a opacidade da visibilidade automatizada, da autoridade algorítmica ou da curadoria deslocada. Oferecem, em vez disso, um tipo distinto de rastro, um registro estável de identidade e de relação entre entidades, que essas três dinâmicas, por si, não produzem.

Os identificadores persistentes (Persistent Identifiers – PIDs) nasceram como solução técnica para identificar objetos digitais de forma inequívoca (Paskin, 1999, 2005). Com o tempo, passaram a articular também pessoas, organizações e outras entidades do ciclo de pesquisa: o DOI identifica publicações e dados, o ORCID identifica autores, o ROR identifica instituições. Essa expansão fez o valor dos PIDs migrar da identificação isolada para a capacidade de estruturar redes de relações entre entidades científicas.

Três propriedades sustentam essa capacidade: identificação unívoca de entidades, resolução persistente ao longo do tempo e vinculação estruturada entre registros (Klump et al., 2023; Rosen et al., 2025). Wittenburg (2019) e Bunakov (2020) mostram como essas propriedades permitem construir grafos de conhecimento que ligam publicações, dados, autores e financiamentos. Nas propostas de FAIR Digital Objects, os PIDs passam a integrar a própria arquitetura dos objetos digitais (Schwardmann, 2020; Wilkinson et al., 2016).

Os PIDs não tornam explícitos os critérios por trás dos processos algorítmicos descritos nos três eixos acima. Seu funcionamento depende de metadados, serviços de resolução e arranjos de governança (Schwardmann, 2020; Wilkinson et al., 2016; Wittenburg, 2019). Ainda assim, ao ligar objetos, agentes e contextos de uso ao longo do tempo, eles preservam algo que a mediação algorítmica tende a apagar: a possibilidade de reconstruir, depois do fato, quem produziu o quê, quando e em conexão com o quê.

3. Reconfiguração da auditabilidade

Xavier et. al (2025) definiram auditabilidade como a possibilidade de explicitar critérios, justificar decisões e reconstruir racionalmente os processos de produção do conhecimento. Em contextos mediados por sistemas de aprendizagem de máquina, essa definição encontra um limite prático: os processos decisórios operam em níveis de complexidade que dificultam explicação e compreensão plena (Ahn, 2024; Raji et al., 2020). Chamamos isso de crise da auditabilidade explicativa — a mediação algorítmica desloca a seleção e a priorização da informação para sistemas cujos mecanismos não são plenamente acessíveis (Hosseini et al., 2025; Veiga, 2025).

Diante desse limite, a auditoria muda de função: já não se apoia na transparência dos critérios decisórios, mas na rastreabilidade dos vínculos entre objetos, agentes e contextos de uso. Auditar deixa de significar explicar decisões e passa a significar reconstruir os encadeamentos que as sustentam. É exatamente esse tipo de vínculo que os identificadores persistentes, discutidos na seção anterior, preservam.

Essa reconstrução se operacionaliza em seis operações, derivadas diretamente das três propriedades dos PIDs, não de uma taxonomia prévia de práticas de auditoria: verificar a existência de um objeto, desambiguar entidades, reconstruir relações entre elementos do ecossistema de pesquisa, contextualizar

registros em redes mais amplas, acompanhar versões e rastrear procedência. A Tabela 1 sintetiza essas operações com exemplos de aplicação.

Tabela 1

Operações de Auditoria Viabilizadas por Identificadores Persistentes

Operações de Auditoria	Capacidade viabilizada por PIDs	Exemplos de Aplicação
Verificar a existência de um objeto	Identificação persistente	DOI
Desambiguar entidades	Identificação unívoca	ORCID
Reconstruir relações entre entidades	Vinculação estruturada entre Registros	DOI + relações <i>DataCite</i>
Contextualizar um objeto no ecossistema científico	Integração em grafos de conhecimento	<i>OpenAlex</i> ; <i>PID Graph</i>
Acompanhar versões e atualizações	Persistência e versionamento de registros	<i>Crossmark</i>
Rastrear procedência e encadeamento informacional	Encadeamento relacional entre entidades e metadados	DOI + metadados associados

Nota. Elaborado pelos autores (2026)

Articuladas aos três eixos da seção anterior, essas operações atuam de formas diferentes sobre cada tipo de opacidade: na visibilidade automatizada, permitem identificar e contextualizar os objetos que circulam nos sistemas; na autoridade algorítmica, permitem verificar autoria e proveniência; no deslocamento da curadoria, permitem reconstruir os encadeamentos que estruturam a organização da informação.

Essas capacidades têm um limite claro. Elas não equivalem à transparência dos processos que estruturam a mediação algorítmica, dependem da qualidade dos metadados, dos padrões adotados e dos arranjos de governança das infraestruturas de PIDs. Rastreabilidade dos fluxos e inteligibilidade dos processos são coisas distintas: a auditabilidade discutida aqui opera sob condições parciais e situadas.

4. Considerações finais

Este estudo argumentou que os identificadores persistentes reconfiguram as condições de auditabilidade em contextos mediados por inteligência artificial. Eles não tornam explícitos os critérios que orientam os processos algorítmicos, mas introduzem camadas estruturadas de identificação, conexão e registro entre entidades científicas, o que basta para viabilizar a reconstrução dos fluxos informacionais.

A contribuição central deste trabalho está no deslocamento analítico da auditabilidade: de uma capacidade de explicitar decisões para uma possibilidade de reconstruir os encadeamentos que estruturam a circulação do conhecimento. A auditabilidade deixa de depender exclusivamente da transparência dos processos e passa a se apoiar na rastreabilidade das relações entre objetos, agentes e contextos de uso.

Essa leitura não elimina as limitações da mediação algorítmica. Os PIDs não tornam os processos decisórios inteligíveis, nem substituem a necessidade de governança e supervisão nos sistemas de inteligência artificial. Como agenda futura, a ser desenvolvida em versão estendida deste trabalho, cabe investigar a operacionalização dessas capacidades em casos concretos, bem como suas implicações para políticas públicas de Ciência Aberta e para a governança de infraestruturas de PIDs em larga escala.

Agradecimentos

Os autores agradecem ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

Conflito de Interesses

Os autores declaram não haver conflitos de interesses.

CRedit – Contribuições dos Autores

Raphael Figueiredo Xavier | Concetualização, Escrita – redação original

Fabiano Ferreira de Castro | Escrita – revisão e edição, Supervisão

Referências

- Ahn, S. (2024). The transformative impact of large language models on medical writing and publishing: current applications, challenges and future directions. *The Korean Journal of Physiology & Pharmacology*, 28(5), 393–401. <https://doi.org/10.4196/kjpp.2024.28.5.393>
- Ballester, P. L. (2023). Open science and software assistance: Commentary on «artificial intelligence can generate fraudulent but authentic-looking scientific medical articles: Pandora’s box has been opened». *Journal of Medical Internet Research*, 25, e49323. <https://doi.org/10.2196/49323>
- Bunakov, V. (2020). Open Science portal based on knowledge graph. *Supplementary Proceedings of the XXII International Conference on Data Analytics and Management in Data Intensive Domains (DAMDID/RCDL 2020)*, 134–141. <https://ceur-ws.org/Vol-2790/paper13.pdf>
- Hosseini, M., Horbach, S. P. J. M., Holmes, K., & Ross-Hellauer, T. (2025). Open Science at the generative AI turn: An exploratory analysis of challenges and opportunities. *Quantitative Science Studies*, 6, 22–45. https://doi.org/10.1162/qss_a_00337
- Klump, J., Fils, D., Devaraju, A., Ramdeen, S., Robertson, J., Wyborn, L., & Lehnert, K. (2023). Scaling identifiers and their metadata to gigascale: An architecture to tackle the challenges of volume and variety. *Data Science Journal*, 22(1). <https://doi.org/10.5334/dsj-2023-005>
- Paskin, N. (1999). Toward unique identifiers. *Proceedings of the IEEE. Institute of Electrical and Electronics Engineers*, 87(7), 1208–1227. <https://doi.org/10.1109/5.771073>
- Paskin, Norman. (2005). Digital Object Identifiers for scientific data. *Data Science Journal*, 4, 12–20. <https://doi.org/10.2481/dsj.4.12>
- Raji, I. D., Smart, A., White, R. N., Mitchell, M., Gebru, T., Hutchinson, B., Smith-Loud, J., Theron, D., & Barnes, P. (2020). Closing the AI accountability gap: Defining an end-to-end framework for internal

- algorithmic auditing. *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 33–44. <https://doi.org/10.1145/3351095.3372873>
- Rockembach, M. (2024). Ciência Aberta e Inteligência Artificial: Desafios éticos e transparência em modelos generativos. *Ciência da Informação*, 53(3). <https://doi.org/10.18225/ci.inf.v53i3.7227>
- Rosen, I. A., Sánchez-Núñez, P., & Belmar, R. H. (2025). *Open Science governance: the role of persistent identifiers and metadata standards*. https://doi.org/10.31219/osf.io/9h564_v3
- Schwardmann, U. (2020). Digital objects – FAIR digital objects: Which services are required? *Data Science Journal*, 19. <https://doi.org/10.5334/dsj-2020-015>
- Silveira, L., & Silva, F. C. C. (2024). El futuro compartido de la ciencia abierta y la inteligencia artificial: Traducción al español. *Revista Electrónica Educare*, 28(Suppl. 1), 1–6. <https://doi.org/10.15359/ree.28-s.20907>
- Veiga, A. (2025). Ethical guidelines for the use of generative artificial intelligence and artificial intelligence-assisted tools in scholarly publishing: a thematic analysis. *Science Editing*, 12(1), 28–34. <https://doi.org/10.6087/kcse.352>
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., ... Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3(1), 160018. <https://doi.org/10.1038/sdata.2016.18>
- Wittenburg, P. (2019). From persistent identifiers to digital objects to make data science more efficient. *Data Intelligence*, 1(1), 6–21. https://doi.org/10.1162/dint_a_00004
- Xavier, R. F., Castro, F. F., & Alves, R. C. V. (2025). Reconfigurações informacionais na era da inteligência artificial generativa: riscos e disputas em torno da ciência aberta. *XXV Encontro Nacional de Pesquisa e Pós-graduação em Ciência da Informação*, 5(1) https://enancib.ancib.org/enancib/pt_BR/article/view/38/38