26º ENANCIB
Encontro Nacional de Pesquisa e Pós-Graduação em Ciência da Informação
A Amazônia na sociedade global em rede: assimetrias e potencialidades informacionais, tecnológicas e ambientais
Belém, Pará, 09 a 13 de novembro de 2026
Universidade Federal do Pará

ENCONTRO NACIONAL DE PESQUISA E PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO
ISSN 2177-3688

GT 8 – Dados, Informação e Tecnologia

IDENTIFICADORES PERSISTENTES COMO INFRAESTRUTURA DA CIÊNCIA ABERTA: UM OLHAR EMERGENTE PARA OS ECOSSISTEMAS DIGITAIS

PERSISTENT IDENTIFIERS AS INFRASTRUCTURE FOR OPEN SCIENCE: *AN EMERGING PERSPECTIVE ON DIGITAL ECOSYSTEMS*

Raphael Figueiredo Xavier – Fundação Getulio Vargas (FGV)/Universidade Estadual Paulista "Júlio de Mesquita Filho" (Unesp) Campus de Marília
Fabiano Ferreira de Castro – Universidade Federal de São Carlos (UFSCAR)/Universidade Estadual Paulista "Júlio de Mesquita Filho" (Unesp) Campus de Marília

Modalidade: Trabalho Completo

Resumo: A consolidação da Ciência Aberta tem ampliado o papel das infraestruturas informacionais na produção, circulação e preservação do conhecimento científico, atribuindo aos identificadores persistentes uma posição crescentemente central nesse ecossistema. Esta pesquisa investiga como a literatura científica tem discutido o papel dos identificadores persistentes no contexto da Ciência Aberta, por meio de uma revisão exploratória da literatura. Foram recuperados 342 registros nas bases *Web of Science* e *Scopus*, dos quais 125 compuseram o *corpus* final após triagem. A análise adotou estratégia indutiva, resultando na organização dos estudos em sete macroeixos temáticos, discutidos em três blocos analíticos: fundamentos conceituais e de infraestrutura; articulação com princípios FAIR, *FAIR Digital Objects* e *workflows* científicos; e dimensões de governança, institucionalização e comunicação científica. Os resultados evidenciam que os identificadores persistentes têm sido progressivamente reposicionados na literatura, de soluções técnicas para o problema da instabilidade de referências digitais para componentes estruturantes de infraestruturas sociotécnicas complexas, cuja efetividade depende de arranjos institucionais, modelos de governança e práticas coordenadas entre atores do ecossistema científico. O estudo contribui para uma compreensão integrada dos identificadores persistentes como infraestrutura transversal da Ciência Aberta, evidenciando convergências, ênfases e lacunas analíticas na literatura.

Palavras-chave: identificadores persistentes; ciência aberta; infraestrutura informacional; ecossistemas digitais.

Abstract: The consolidation of Open Science has expanded the role of informational infrastructures in the production, circulation, and preservation of scientific knowledge, assigning persistent identifiers an increasingly central position in this ecosystem. This research investigates how the scientific literature has discussed the role of persistent identifiers in the context of Open Science through an exploratory literature review. A total of 342 records were retrieved from the Web of Science and Scopus databases, of which 125 formed the final corpus after screening. The analysis adopted an inductive strategy, resulting in the organization of studies into seven thematic macro-axes, discussed in three analytical blocks: conceptual and infrastructural foundations; articulation with FAIR principles, FAIR Digital Objects, and scientific *workflows*; and dimensions of governance, institutionalization, and scholarly communication. The results show that persistent identifiers have been progressively repositioned in the literature, from technical solutions to the problem of digital reference instability to structural components of complex sociotechnical infrastructures, whose effectiveness depends on institutional arrangements, governance models, and coordinated practices among actors in the scientific ecosystem. The study contributes to an integrated understanding of persistent identifiers as transversal infrastructure of Open Science, highlighting convergences, emphases, and analytical gaps in the literature.
Keywords: persistent identifiers; open science; informational infrastructure; digital ecosystems.

1 INTRODUÇÃO

A consolidação da Ciência Aberta tem sido compreendida não apenas como um conjunto de práticas normativas voltadas à abertura de publicações e dados de pesquisa, mas como um fenômeno que articula, de forma indissociável, infraestruturas digitais, padrões técnicos, políticas institucionais e práticas científicas (Fecher; Friesike, 2014; Silva; Silveira, 2019). A produção, a circulação e o reuso do conhecimento científico passam a depender de arranjos técnicos e organizacionais capazes de garantir visibilidade, interoperabilidade, rastreabilidade e confiança ao longo de todo o ciclo de vida da pesquisa (Wittenburg, 2019).

A efetivação desses princípios depende de infraestruturas informacionais capazes de sustentar conexões persistentes e interoperáveis entre publicações, dados, pessoas, organizações e outros objetos de pesquisa (Macgregor; Lancho-Barrantes; Pennington, 2023; Wittenburg, 2019). Os identificadores persistentes (*Persistent Identifiers* - PIDs) emergem como elementos desse ecossistema, ao permitir a identificação inequívoca de entidades científicas, a integração entre sistemas heterogêneos e a construção de relações duráveis entre objetos informacionais ao longo do tempo (Rosen; Sánchez-Núñez; Belmar, 2025).

Inicialmente concebidos para enfrentar o problema da instabilidade de referências e da

perda de acesso a objetos digitais, os PIDs passaram a ser incorporados a arquiteturas informacionais mais amplas, viabilizando a vinculação entre diferentes tipos de objetos de pesquisa e a automação de processos associados à comunicação e à gestão da informação científica. Identificadores como o *Digital Object Identifier* (DOI), o *Open Researcher and Contributor ID* (ORCID), o *Research Organization Registry* (ROR), o *International Generic Sample Number* (IGSN) e o *Research Activity Identifier* (RAiD) exemplificam essa ampliação funcional, ao possibilitar a identificação persistente não apenas de publicações, mas também de pessoas, de organizações, de amostras, de instrumentos, de projetos e de eventos científicos (Ferguson *et al.*, 2019). No contexto brasileiro, essa centralidade institucional se reflete no lançamento, em outubro de 2025, do Guia de Interoperabilidade da Pós-Graduação pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), documento que orienta a padronização de identificadores como DOI, ORCID e ROR na integração de dados entre sistemas institucionais e a Plataforma Sucupira (CAPES, 2025).

Apesar dessa crescente centralidade, a literatura científica sobre PIDs distribui-se por enfoques conceituais, arquiteturais, normativos e aplicados que raramente dialogam entre si, sem que se identifique, até o momento, um esforço de sistematização capaz de integrar essas abordagens em uma leitura conjunta. Essa lacuna dificulta a compreensão integrada dos identificadores persistentes como infraestrutura do ecossistema científico, bem como a identificação de convergências, lacunas e padrões de abordagem. Desse modo, a questão norteadora consiste em analisar como a literatura científica tem discutido o papel dos identificadores persistentes no contexto da Ciência Aberta.

Diante desse problema, este estudo investiga como a literatura científica tem discutido o papel dos identificadores persistentes no contexto da Ciência Aberta, por meio de um mapeamento exploratório da literatura organizado em macroeixos temáticos. Ao mapear essas abordagens, busca-se contribuir para uma compreensão mais integrada dos PIDs como infraestrutura transversal da Ciência Aberta, evidenciando tanto suas potencialidades quanto os desafios associados à sua adoção, governança e uso efetivo nas práticas científicas.

2 ENQUADRAMENTO CONCEITUAL: CIÊNCIA ABERTA, INFRAESTRUTURAS E IDENTIFICAÇÃO PERSISTENTE

A Ciência Aberta é compreendida, nesta pesquisa, numa perspectiva sociotécnica que enfatiza a interdependência entre dimensões técnicas, institucionais e sociais na produção e na

circulação do conhecimento científico (Fecher; Friesike, 2014; Silva; Silveira, 2019). Essa base conceitual estabelece o enquadramento a partir do qual os PIDs são analisados como componentes estruturantes das infraestruturas da Ciência Aberta, cuja função extrapola a mera persistência de referências, envolvendo a integração entre sistemas, a articulação de diferentes tipos de objetos científicos e o suporte à automação de fluxos informacionais (Rosen; Sánchez-Núñez; Belmar, 2025; Wittenburg, 2019).

2.1 CIÊNCIA ABERTA COMO ECOSSISTEMA SOCIOTÉCNICO

A Ciência Aberta tem sido abordada, nas últimas décadas, a partir de perspectivas que variam desde enfoques normativos, centrados em princípios como abertura, transparência e compartilhamento, até abordagens mais críticas e analíticas, que enfatizam as condições materiais, institucionais e técnicas necessárias para sua implementação efetiva (Fecher; Friesike, 2014).

A abertura científica não se reduz à disponibilização de conteúdo, mas depende da existência de infraestruturas informacionais capazes de sustentar práticas como o compartilhamento de dados, a reprodutibilidade da pesquisa, a atribuição adequada de autoria e financiamento, bem como a preservação e o reuso de objetos científicos ao longo do tempo. Tais infraestruturas envolvem padrões técnicos, serviços digitais, arranjos institucionais e modelos de governança que operam de maneira articulada, condicionando tanto as possibilidades quanto os limites da Ciência Aberta na prática (Silva; Silveira, 2019).

Wittenburg (2019) argumenta que a consolidação da Ciência Aberta está diretamente associada à capacidade de conectar, de forma persistente e interoperável, diferentes tipos de entidades científicas ao longo do ciclo de vida da pesquisa, numa conectividade que não é apenas técnica, mas também social e institucional, dependendo da adoção coordenada de padrões, de identificadores e de serviços por comunidades científicas, instituições de pesquisa, agências de fomento e provedores de infraestrutura. Suas infraestruturas incorporam, portanto, escolhas arquiteturais, políticas e normativas que influenciam práticas científicas, modos de avaliação e fluxos de informação, contexto no qual os PIDs passam a ser analisados não como soluções técnicas pontuais, mas como componentes estruturantes das infraestruturas que sustentam a Ciência Aberta.

2.2 IDENTIFICADORES PERSISTENTES COMO INFRAESTRUTURA DA CIÊNCIA ABERTA

Entre os diversos componentes das infraestruturas informacionais científicas, os PIDs destacam-se como mecanismos transversais responsáveis por garantir a identificação inequívoca, durável e interoperável de entidades científicas no ambiente digital. Inicialmente concebidos para mitigar problemas de instabilidade de *links* e perda de acesso a objetos digitais, os PIDs passaram a ocupar uma posição estrutural nas arquiteturas informacionais contemporâneas, atuando como pontos de ancoragem para a integração entre sistemas heterogêneos e para a construção de relações estáveis entre diferentes tipos de objetos de pesquisa (Rosen; Sánchez-Núñez; Belmar, 2025).

Essa ampliação funcional acompanha a própria transformação do escopo da comunicação científica. A identificação persistente, antes fortemente associada a publicações acadêmicas, foi progressivamente estendida para abranger pessoas, organizações, amostras físicas, projetos e atividades de pesquisa, entre outras entidades, refletindo a crescente necessidade de representar de forma estruturada e interoperável a complexidade dos processos científicos contemporâneos.

Ao permitir a vinculação persistente entre essas entidades, os PIDs viabilizam a construção de redes de relações que sustentam funcionalidades centrais da Ciência Aberta: a rastreabilidade da produção científica, a atribuição de autoria e financiamento, a integração de fluxos de trabalho distribuídos e a automação de processos de gestão e avaliação da pesquisa. Os PIDs operam não apenas como identificadores isolados, mas como elementos estruturantes de infraestruturas relacionais, frequentemente materializadas em grafos de conhecimento e serviços de interoperabilidade que conectam diferentes sistemas e plataformas (Wittenburg, 2019). Essa posição transversal os vincula diretamente aos princípios *Findability, Accessibility, Interoperability, and Reuse* (FAIR) e, mais recentemente, aos FAIR *Digital Objects*, nos quais os PIDs, associados a metadados estruturados e serviços de resolução, são concebidos como elementos constitutivos de objetos científicos acionáveis por máquinas (Rosen; Sánchez-Núñez; Belmar, 2025).

Estudos dedicados à análise das infraestruturas de identificação científica apontam que a eficácia dos PIDs não pode ser compreendida exclusivamente a partir de sua dimensão técnica. Sua implementação e funcionamento dependem de decisões arquiteturais, modelos de governança, políticas institucionais e práticas de adoção pelas comunidades científicas (Dappert; Farquhar; Kotarski; Hewlet, 2017; Macgregor; Lancho-Barrantes; Pennington, 2023; Wittenburg, 2019). Aspectos como a qualidade e a completude dos metadados associados, a

sustentabilidade dos serviços de resolução, a interoperabilidade entre esquemas de identificação e o alinhamento entre atores institucionais distintos influenciam diretamente o potencial dos PIDs como infraestrutura da Ciência Aberta (Rosen; Sánchez-Núñez; Belmar, 2025).

Compreender os identificadores persistentes como infraestrutura transversal implica, portanto, reconhecê-los como elementos situados na interseção entre tecnologia, organização e prática científica, premissa que orienta a investigação desenvolvida nas seções seguintes.

3 METODOLOGIA

Este estudo adota uma revisão exploratória da literatura com o objetivo de mapear, organizar e analisar como a produção científica tem discutido o papel dos identificadores persistentes no contexto da Ciência Aberta. Conforme Galvão e Ricarte (2019), revisões de caráter exploratório são especialmente adequadas para campos amplos e heterogêneos, nos quais coexistem abordagens qualitativas, quantitativas e mistas e cujo objetivo não é comparar resultados empíricos de forma direta, mas identificar tendências, lacunas e padrões de abordagem, condições que caracterizam precisamente o campo aqui investigado e que justificam o emprego do termo "mapeamento exploratório" para designar essa função estruturante da revisão.

3.1 DELINEAMENTO E ESTRATÉGIA DE BUSCA

A questão orientadora dessa revisão pauta-se em analisar como a literatura científica tem discutido o papel dos identificadores persistentes no contexto da Ciência Aberta. Optou-se por uma formulação ampla, capaz de abarcar diferentes níveis de análise: abordagens conceituais, arquiteturais, normativas e aplicadas.

As buscas foram realizadas nas bases *Web of Science* (WoS) e Scopus, escolhidas por sua abrangência multidisciplinar e representatividade na literatura internacional sobre comunicação científica e infraestruturas de pesquisa. Na *Web of Science*, foi empregada a seguinte *string*: *("persistent identifier" OR "persistent digital identifier") AND ("open science" OR metadata OR "scholarly communication")*. Na *Scopus*, manteve-se a mesma lógica conceitual com adaptação sintática: *TITLE-ABS-KEY ("persistent identifier" OR "persistent digital identifier") AND TITLE-ABS-KEY ("open science" OR metadata OR "scholarly communication")*. Os resultados foram exportados e consolidados em planilha única para tratamento.

3.2 SELEÇÃO, TRIAGEM E ORGANIZAÇÃO DO *CORPUS*

A busca nas duas bases recuperou 342 registros: 108 na *Scopus* e 234 na *Web of Science*. Os registros foram consolidados e submetidos à deduplicação com base no DOI e, quando necessário, na verificação manual de títulos e autores. A triagem subsequente baseou-se na análise de título, resumo e palavras-chave, orientada pela seguinte pergunta analítica: o estudo discute os identificadores persistentes como conceito, infraestrutura ou elemento relevante da Ciência Aberta, e não apenas como detalhe operacional ou acessório? Os registros foram classificados em três categorias - “incluir”, “excluir” e “limítrofe” -, sendo incorporados ao *corpus* final apenas os 125 classificados como "incluir", com discussão substantiva sobre os PIDs no contexto da Ciência Aberta.

A análise dos 125 estudos adotou estratégia indutiva, orientada pela leitura de títulos, resumos e palavras-chave. A cada estudo foi atribuído um eixo temático descritivo que sintetizava seu foco analítico principal, formulado em linguagem próxima à dos próprios autores, sem padronização terminológica prévia, nem imposição de esquema classificatório fechado. Em uma primeira etapa de consolidação, eixos semanticamente muito próximos foram fundidos, reduzindo o conjunto a aproximadamente 100 eixos distintos, número que evidencia a heterogeneidade do campo e a ausência de vocabulário analítico consolidado.

Em uma segunda etapa, os 100 eixos foram agrupados por afinidade conceitual em macroeixos analíticos, considerando dois critérios combinados: o objeto de análise predominante no estudo e o modo de abordagem empregado - conceitual, empírico, normativo ou aplicado. Esse agrupamento teve caráter interpretativo, não automático ou estatístico, e resultou em sete macroeixos. Cada registro foi associado ao macroeixo que melhor representava seu foco predominante, sem que isso implique categorias mutuamente exclusivas. O Quadro 1 apresenta os sete macroeixos e a distribuição dos estudos:

Quadro 1 - Definição dos macroeixos e número de ocorrências

Macroeixo	Descrição sintética	Número de estudos
M1: Fundamentos conceituais, tipologias e arquitetura de PIDs	Abordagens conceituais e arquiteturais dos PIDs	11
M2: Metadados, interoperabilidade e serviços de resolução	Integração, normalização e resolução de PIDs	22
M3: FAIR/FAIR *Digital Objects* e dados como objetos acionáveis	PIDs no contexto FAIR, FDO e *machine-actionability*	13

M4: *Workflows*, curadoria, proveniência e reprodutibilidade	Uso operacional em fluxos e ciclos de vida científicos	14
M5: Governança, políticas e institucionalização da Ciência Aberta	Institucionalização, políticas e coordenação	27
M6: Tipos de PIDs e aplicações por entidade (pessoas, organizações, amostras, etc.)	ORCID, DOI, ROR, IGSN, RAiD etc.	21
M7: Comunicação científica, visibilidade, citação e preservação	Visibilidade, citação e preservação	17

Fonte: elaboração própria (2026).

A distribuição apresentada no Quadro 1 não deve ser interpretada como medida de relevância entre os macroeixos, mas como indicativo das ênfases analíticas presentes na literatura. A opção por sete macroeixos foi deliberada: a granularidade intermediária preserva distinções conceituais relevantes entre abordagens que, embora relacionadas, mobilizam objetos de análise e perspectivas distintas. Para fins de discussão, os sete macroeixos são organizados nas seções seguintes em três blocos temáticos, reunidos segundo a dimensão analítica predominante: fundacional e infraestrutural (M1 e M2, seção 4.1), operacional e de dados (M3 e M4, seção 4.2) e institucional, política e comunicacional (M5, M6 e M7, seção 4.3), de modo a favorecer a legibilidade da discussão sem comprometer a fidelidade analítica ao *corpus*.

4 RESULTADOS E DISCUSSÃO

Retomando o problema de pesquisa formulado a partir da ausência de um esforço de sistematização, a análise evidencia um campo marcado por elevada diversidade conceitual, técnica e institucional no tratamento dos identificadores persistentes no contexto da Ciência Aberta. Os 125 estudos examinados mobilizam os PIDs a partir de múltiplas perspectivas, articulando desde discussões fundacionais sobre princípios conceituais e arquiteturais até análises voltadas à sua aplicação em infraestruturas informacionais, *workflows* científicos, políticas institucionais e práticas de comunicação e preservação da produção científica. Essa pluralidade reflete tanto a ampliação progressiva do escopo funcional dos identificadores persistentes quanto a inexistência de um modelo analítico único e consensual para enquadrar seu papel no ecossistema científico contemporâneo.

A distribuição dos estudos pelos sete macroeixos revela ênfases significativas: governança, políticas e institucionalização (M5), metadados e interoperabilidade (M2) e tipos

de PIDs por entidade (M6) concentram a maior parte dos trabalhos, indicando que a literatura tem dedicado atenção especial às dimensões organizacionais e infraestruturais dos identificadores persistentes em larga escala. Macroeixos como fundamentos conceituais (M1) e FAIR *Digital Objects* (M3) apresentam menor volume relativo, ainda que exerçam papel estruturante no campo: seus trabalhos tendem a formular modelos e princípios que fundamentam desenvolvimentos técnicos e institucionais posteriores. Em termos temporais, observa-se uma progressão das discussões fundacionais sobre persistência e identificação digital, predominantes nos trabalhos mais antigos, para uma produção crescente voltada à integração dos PIDs em infraestruturas de pesquisa, *workflows* e ecossistemas institucionais, acompanhando a consolidação da agenda da Ciência Aberta e dos princípios FAIR.

4.1 FUNDAMENTOS CONCEITUAIS, DE ARQUITETURA E DE INFRAESTRUTURA DOS IDENTIFICADORES PERSISTENTES (M1 E M2)

Uma parcela significativa da literatura analisada dedica-se a discutir os identificadores persistentes a partir de uma perspectiva fundacional, concentrando-se em seus princípios conceituais, em suas arquiteturas técnicas e em sua articulação com infraestruturas de metadados. Esses estudos tratam os PIDs não apenas como soluções pontuais para o problema da persistência de *links*, mas como componentes estruturais de modelos mais amplos de organização, circulação e preservação da informação científica.

No plano conceitual, parte da literatura aborda os sistemas de identificação persistente a partir de sua função como infraestrutura da comunicação científica, com destaque para o DOI como mecanismo de referência estável e independente da localização dos objetos digitais. Esses trabalhos enfatizam que o valor dos PIDs não reside apenas no identificador em si, mas nas conexões e nos serviços que eles viabilizam no ecossistema informacional (Gould, 2022). Essa perspectiva é complementada por propostas que destacam a necessidade de camadas mínimas de metadados associadas aos identificadores, como o chamado *metadata kernel*, concebido para sustentar a permanência, a inteligibilidade e a gestão de objetos digitais ao longo do tempo (Kunze, 2002).

Essas discussões são aprofundadas por estudos que problematizam a própria noção de persistência, enfatizando que a durabilidade dos identificadores não constitui um atributo meramente técnico, mas resulta de compromissos institucionais, políticas explícitas e expectativas de serviço associadas aos sistemas de identificação. Nesse contexto, propostas

como os *persistence statements* buscam tornar-se explícitas as promessas de estabilidade e responsabilidade assumidas pelos provedores de PIDs, evidenciando que a persistência é construída social e organizacionalmente, e não garantida apenas pela tecnologia (Kunze *et al.*, 2017).

Outra vertente da literatura chama atenção para os limites conceituais e técnicos dos PIDs quando aplicados a objetos dinâmicos, versões sucessivas de dados ou recursos computacionais mutáveis. Nesses casos, são discutidas alternativas e extensões conceituais voltadas a lidar com temporalidade, versionamento e granularidade de objetos digitais dinâmicos, tensionando modelos tradicionais de identificação persistente e aproximando os PIDs de arquiteturas baseadas em grafos de conhecimento e objetos digitais interligados (Silveira; Pruski, 2022).

No que se refere à infraestrutura de metadados e à interoperabilidade, os estudos associados ao macroeixo M2 evidenciam a centralidade da relação entre PIDs e esquemas de descrição. A eficácia dos identificadores depende diretamente da qualidade, da normalização e da interoperabilidade dos metadados associados, reforçando a ideia de que identificação e descrição constituem dimensões indissociáveis das infraestruturas informacionais científicas (Klump *et al.*, 2023). Esses estudos também discutem serviços e arquiteturas voltados à gestão, à resolução e à escalabilidade dos identificadores persistentes, destacando que tais serviços devem ser concebidos de forma distribuída, interoperável e alinhada a padrões consolidados da *Web*. Um conjunto relevante de trabalhos explora ainda o papel dos PIDs na construção de infraestruturas baseadas em *linked data* e arquiteturas semânticas relacionais, nas quais os identificadores funcionam como âncoras estáveis para a conexão entre diferentes entidades científicas, viabilizando consultas, integração semântica e serviços avançados de descoberta em portais de Ciência Aberta (Bunakov, 2020; Klump *et al.*, 2023).

4.2 IDENTIFICADORES PERSISTENTES, *FAIR DIGITAL OBJECTS* E *WORKFLOWS* CIENTÍFICOS (M3 E M4)

Uma parte expressiva da literatura analisada concentra-se na articulação entre identificadores persistentes, os princípios FAIR e a operacionalização de dados e objetos de pesquisa como componentes acionáveis das infraestruturas científicas. Os estudos reunidos nos macroeixos M3 e M4 deslocam o foco dos PIDs enquanto mecanismos isolados de identificação para compreendê-los como elementos estruturantes de arquiteturas orientadas à automação,

à integração de sistemas e à reprodutibilidade dos processos científicos, especialmente em contextos de ciência intensiva em dados e ambientes computacionais distribuídos (Weigel; Kindermann; Lautenschlager, 2014; Wittenburg, 2019).

A formulação dos princípios FAIR por Wilkinson *et al.* (2016) estabelece que dados científicos devem ser encontráveis, acessíveis, interoperáveis e reutilizáveis. Nesse enquadramento, os identificadores persistentes assumem papel central, particularmente no princípio da *findability*, ao requerer que dados e metadados estejam associados a identificadores globais, únicos e persistentes. A literatura analisada amplia essa compreensão ao argumentar que os PIDs não devem ser compreendidos apenas como requisitos formais de aderência aos princípios FAIR, mas como componentes técnicos indispensáveis à construção de objetos digitais capazes de operar de forma integrada, automatizada e confiável em ecossistemas computacionais complexos (Praetzellis *et al.*, 2023; Wittenburg, 2019).

Wittenburg (2019) propõe compreender os identificadores persistentes como elementos fundacionais da noção de *Digital Objects*, destacando que a identificação persistente constitui o ponto de ancoragem a partir do qual metadados, serviços e relações podem ser organizados de maneira sistemática. Essa abordagem desloca o debate dos PIDs de uma função meramente referencial para uma compreensão infraestrutural, na qual a persistência da identificação sustenta a automação de processos, a interoperabilidade entre sistemas e a circulação confiável de objetos digitais ao longo do tempo. Essa perspectiva é aprofundada no debate sobre os *FAIR Digital Objects* (FDOs), nos quais identificadores persistentes, metadados estruturados e protocolos de acesso são concebidos como elementos constitutivos de uma mesma unidade informacional, entidades identificáveis, acionáveis e interoperáveis por máquinas, que viabilizam processos automatizados de descoberta, acesso, validação e reutilização de dados em escala (Schwardmann, 2020; Weigel; Kindermann; Lautenschlager, 2014).

No contexto operacional, estudos do macroeixo M4 analisam como a atribuição sistemática de identificadores persistentes a dados, versões, *softwares*, instrumentos e resultados intermediários permite estruturar fluxos de pesquisa reprodutíveis, nos quais etapas analíticas, dependências e transformações permanecem documentadas e rastreáveis ao longo do tempo (Haak, 2014; Praetzellis *et al.*, 2023). Nesses trabalhos, os PIDs são mobilizados como mecanismos centrais de proveniência e atribuição, conectando diferentes entidades científicas e sustentando práticas de transparência, reúso e validação dos resultados de pesquisa. A

incorporação de PIDs em ambientes FAIR e em *workflows* científicos, contudo, não é neutra nem automática: seu potencial depende de adoção consistente, metadados de qualidade, infraestrutura técnica sustentável e alinhamento entre políticas institucionais, ferramentas computacionais e práticas das comunidades científicas (Haak, 2014; Praetzellis *et al.*, 2023).

4.3 GOVERNANÇA, INSTITUCIONALIZAÇÃO E EFEITOS DOS IDENTIFICADORES PERSISTENTES NA COMUNICAÇÃO CIENTÍFICA (M5, M6 E M7)

Os estudos reunidos nos macroeixos M5, M6 e M7 deslocam a discussão sobre os identificadores persistentes para além de sua dimensão técnico-operacional, enfatizando seus efeitos institucionais, organizacionais e normativos na comunicação científica. Nesse conjunto, os PIDs são analisados como componentes de infraestruturas sociotécnicas cuja efetividade depende de arranjos de governança, processos de institucionalização e alinhamento entre atores, políticas e *workflows* de pesquisa.

No macroeixo M5, a literatura concentra-se na construção e na coordenação de ecossistemas de identificadores persistentes em escala sistêmica. Dappert, Farquhar, Kotarski e Hewlet (2017) analisam os PIDs como componentes centrais de infraestruturas digitais de pesquisa distribuídas, argumentando que sua efetividade não decorre apenas de propriedades técnicas, mas da articulação entre camadas tecnológicas, serviços interoperáveis, estruturas de governança e compromissos institucionais explícitos - a persistência como resultado coletivo, sustentado por políticas claras e responsabilidades assumidas por organizações de registro, repositórios, editoras e bibliotecas. Bardi *et al.* (2023) retomam essa perspectiva ao evidenciar que a integração de PIDs em infraestruturas de repositórios orientadas à Ciência Aberta ocorre em contextos marcados por assimetrias institucionais, diferenças de capacidade técnica e dependência de infraestruturas globais de identificação, impondo desafios relevantes à sustentabilidade e à governança no longo prazo.

O macroeixo M6 desloca o foco para os processos de institucionalização da Ciência Aberta e para o papel dos PIDs na reorganização das práticas científicas. Plomp (2020) destaca que a incorporação sistemática de identificadores é fundamental para ampliar a encontrabilidade, a acessibilidade, a interoperabilidade e a reutilização de diferentes objetos científicos. Brown, Demeranville e Meadows (2016) demonstram como identificadores como ORCID e DOI possibilitam a automação de processos editoriais e redistribuem responsabilidades entre pesquisadores, editoras, bibliotecas e organizações de infraestrutura. Nesse contexto,

Piwowar *et al.* (2018) oferecem evidência empírica relevante: sua análise em larga escala demonstra que a visibilidade e o alcance da produção científica estão diretamente condicionados pela qualidade das infraestruturas de identificação e acesso, reforçando que decisões sobre adoção de PIDs têm consequências mensuráveis sobre a circulação do conhecimento científico.

Os estudos do macroeixo M7 evidenciam que os efeitos dos PIDs extrapolam a identificação técnica de objetos digitais, alcançando dimensões como visibilidade, citação, preservação e avaliação da produção científica - atuando como dispositivos estratégicos que condicionam processos que vão da descoberta à preservação de longo prazo e cuja efetividade depende de modelos de governança capazes de sustentar práticas abertas, coordenadas e duráveis no ecossistema da Ciência Aberta.

Em conjunto, os três blocos analíticos evidenciam dois movimentos. No âmbito da Ciência Aberta, os PIDs deixam de operar como mecanismos técnicos pontuais e passam a constituir uma infraestrutura transversal de governança, automação e comunicação. Para a Ciência da Informação, esse deslocamento aproxima os PIDs de questões constitutivas do campo, como infraestrutura informacional, comunicação científica e organização da informação, embora esses identificadores ainda ocupem espaço pouco consolidado em sua agenda de pesquisa.

5 CONSIDERAÇÕES FINAIS

Este estudo investigou como a literatura científica tem discutido o papel dos identificadores persistentes no contexto da Ciência Aberta, por meio de um mapeamento exploratório organizado em macroeixos temáticos. A revisão de 125 estudos permitiu mapear um campo marcado por elevada heterogeneidade conceitual, técnica e institucional, tornando inteligível uma produção dispersa por meio de um instrumento heurístico capaz de revelar recorrências, ênfases e zonas de concentração analítica.

Os resultados evidenciam que os identificadores persistentes têm sido progressivamente reposicionados na literatura: de soluções técnicas para o problema da instabilidade de *links* para componentes centrais das infraestruturas da Ciência Aberta. A análise dos macroeixos revelou que os PIDs são mobilizados como elementos estruturantes de arquiteturas informacionais complexas, articulando metadados, princípios FAIR, *workflows* científicos, políticas institucionais, práticas de governança e processos de comunicação e

preservação da produção científica. Esse deslocamento evidencia que a persistência não pode ser compreendida como uma propriedade puramente técnica, mas como um resultado sociotécnico, sustentado por arranjos institucionais, compromissos organizacionais e serviços de longo prazo.

A pesquisa contribui, do ponto de vista teórico, ao oferecer uma leitura integrada da literatura sobre identificadores persistentes, evidenciando como diferentes tradições de pesquisa coexistem e se articulam no debate contemporâneo sobre Ciência Aberta. Do ponto de vista prático, os achados sugerem que decisões sobre adoção e governança de PIDs devem ser tratadas como decisões estratégicas, com envolvimento coordenado de bibliotecas, repositórios, editoras, agências de fomento e organizações de registro - uma vez que os efeitos dos identificadores persistentes extrapolam a identificação de objetos e impactam visibilidade, interoperabilidade, avaliação e preservação da produção científica.

Como toda revisão exploratória, este estudo apresenta limitações. Uma delas, inerente à natureza indutiva e interpretativa da análise temática adotada, é que outro pesquisador trabalhando com o mesmo corpus poderia chegar a agrupamentos distintos. Toda classificação é, nesse sentido, uma redução deliberada da complexidade: o valor de um mapa não está em coincidir ponto a ponto com o território, mas em torná-lo navegável.

Pesquisas futuras podem aprofundar outros aspectos identificados neste trabalho tais como estudos de caso, investigações empíricas sobre a implementação de PIDs em contextos institucionais específicos, ou análises comparativas entre infraestruturas de identificação, por exemplo, entre modelos centralizados e descentralizados de governança de PIDs, comparação especialmente relevante no contexto brasileiro diante de iniciativas como o dARK (*Decentralized Archival Resource Key*), desenvolvido como alternativa descentralizada aos sistemas tradicionais de identificação persistente (Segundo *et al*., 2022).

Por fim, a pesquisa sustenta que compreender os identificadores persistentes como infraestruturas sociotécnicas é fundamental para avançar no debate sobre Ciência Aberta. Mais do que ferramentas técnicas, os PIDs configuram-se como dispositivos centrais na reorganização contemporânea da produção, da circulação e da preservação do conhecimento científico, exigindo abordagens analíticas capazes de articular técnica, política e institucionalidade.

REFERÊNCIAS

BARDI, Alessia; MANGHI, Paolo; MANNOCCI, Andrea; OTTONELLO, Enrico *et al*. A primer on open science-driven repository platforms. *In*: GAROUFALLOU, Emmanouel; VLACHIDIS, Andreas (ed.). **Metadata and semantic research**. Cham: Springer Nature Switzerland, 2023. p. 222-234. (Communications in Computer and Information Science, v. 1789). Disponível em: https://link.springer.com/chapter/10.1007/978-3-031-39141-5_19. Acesso em: 10 fev. 2026.

BROWN, Josh; DEMERANVILLE, Tom; MEADOWS, Alice. Open access in context: connecting authors, publications and workflows using ORCID identifiers. **Publications**, Basel, v. 4, n. 4, p. 1-8, set. 2016. Disponível em: https://www.mdpi.com/2304-6775/4/4/30. Acesso em: 10 fev. 2026.

BUNAKOV, Vasily. Open science portal based on knowledge graph. *In*: DATA ANALYTICS AND MANAGEMENT IN DATA INTENSIVE DOMAINS, 2020, Voronezh. **Anais** [...]. Aachen: CEUR-WS, 2020. p. 134-141. (CEUR Workshop Proceedings, v. 2790). Disponível em: https://ceur-ws.org/Vol-2790/paper13.pdf. Acesso em: 4 fev. 2026.

COORDENAÇÃO DE APERFEIÇOAMENTO DE PESSOAL DE NÍVEL SUPERIOR (Brasil). **Guia de interoperabilidade da pós-graduação**. Brasília, DF: CAPES, 2025. Disponível em: https://sucupira.capes.gov.br/guia-de-interoperabilidade/introducao/guia-pratico. Acesso em: 11 set. 2026.

DAPPERT, Angela; FARQUHAR,Adam; KOTARSKI, Rachael; HEWLET, Kirstie. Connecting the persistent identifier ecosystem: building the technical and human infrastructure for open research. **Data Science Journal**, Londres, v. 16, p. 28-44, jun. 2017. Disponível em: https://datascience.codata.org/articles/dsj-2017-028. Acesso em: 10 fev. 2026.

FECHER, Benedikt; FRIESIKE, Sascha. Open science: one term, five schools of thought. *In*: BARTLING, Sönke; FRIESIKE, Sascha (ed.). **Opening science**: the evolving guide on how the internet is changing research, collaboration and scholarly publishing. Cham: Springer, 2014. p. 17-47. Disponível em: https://link.springer.com/chapter/10.1007/978-3-319-00026-8_2. Acesso em: 17 out. 2025.

FERGUSON, Christine; McENTRYE, Jo; BUNAKOV, Vasily; LAMBERT, Simon *et al*. **D3.1 survey of current PID services landscape**: revised. [*S. l.*]: Zenodo, 2019. Disponível em: https://doi.org/10.5281/zenodo.3554254. Acesso em: 11 set. 2026.

GALVÃO, Maria Cristiane Barbosa; RICARTE, Ivan Luiz Marques. Revisão sistemática da literatura: conceituação, produção e publicação. **LOGEION: Filosofia da Informação**, Rio de Janeiro, v. 6, n. 1, p. 57-73, set. 2019. Disponível em: https://revista.ibict.br/fiinf/article/view/4835. Acesso em: 27 jan. 2026.

GOULD, Maria. People, places, and things: persistent identifiers in the scholarly communication landscape. **College & Research Libraries News**, Chicago, v. 83, n. 9, p. 398-402, 2022. Disponível em: https://crln.acrl.org/index.php/crlnews/article/view/25638/33548. Acesso em: 10 fev. 2026.

HAAK, Laurel L. Persistent identifiers can improve provenance and attribution and encourage sharing of research results. **Information Services & Use**, Amsterdam, v. 34, n. 1-2, p. 93-96, 2014. Disponível em: https://content.iospress.com/articles/information-services-and-use/isu140736. Acesso em: 20 jun. 2026.

KLUMP, Jens *et al*. Scaling identifiers and their metadata to gigascale: an architecture to tackle the challenges of volume and variety. **Data Science Journal**, Londres, v. 22, p. 1-17, mar. 2023. Disponível em: https://datascience.codata.org/articles/dsj-2023-005. Acesso em: 4 fev. 2026.

KUNZE, John A. A metadata kernel for electronic permanence. **Journal of Digital Information**, [*s. l.*], v. 2, n. 2, p. 1-10, jan. 2002. Disponível em: https://jodi-ojs-tdl.tdl.org/jodi/article/view/jodi-38. Acesso em: 4 fev. 2026.

KUNZE, John A. *et al*. Persistence statements: describing digital stickiness. **Data Science Journal**, Londres, v. 16, p. 1-11, ago. 2017. Disponível em: https://datascience.codata.org/articles/dsj-2017-039. Acesso em: 4 fev. 2026.

MACGREGOR, George; LANCHO-BARRANTES, Barbara S.; PENNINGTON, Diane Rasmussen. Measuring the concept of PID literacy: user perceptions and understanding of PIDs in support of open scholarly infrastructure. **Open Information Science**, Berlin, v. 7, n. 1, art. 20220142, mar. 2023. Disponível em: https://www.degruyterbrill.com/document/doi/10.1515/opis-2022-0142/html. Acesso em: 22 jan. 2026.

PIWOWAR, Heather *et al*. The state of OA: a large-scale analysis of the prevalence and impact of open access articles. **PeerJ**, [*s. l.*], v. 6, art. e4375, p. 1-23, fev. 2018. Disponível em: https://peerj.com/articles/4375/. Acesso em: 13 nov. 2025.

PLOMP, Esther. Going digital: persistent identifiers for research samples, resources and instruments. **Data Science Journal**, Londres, v. 19, p. 1-8, dez. 2020. Disponível em: https://datascience.codata.org/articles/dsj-2020-046. Acesso em: 10 fev. 2026.

PRAETZELLIS, Maria *et al*. A programmatic and scalable approach to making data management machine-actionable. **Data Science Journal**, Londres, v. 22, p. 1-9, ago. 2023. Disponível em: https://datascience.codata.org/articles/10.5334/dsj-2023-026. Acesso em: 4 fev. 2026.

ROSEN, Isabel Abedrapo; SÁNCHEZ-NÚÑEZ, Pablo; BELMAR, Ricardo Hartley. **Open science governance**: the role of persistent identifiers and metadata standards. [*S. l.*]: OSF Preprints, out. 2025. Preprint. Disponível em: https://osf.io/preprints/osf/9h564_v3. Acesso em: 22 jan. 2026.

SCHWARDMANN, Ulrich. Digital objects – FAIR digital objects: which services are required? **Data Science Journal**, Londres, v. 19, p. 1-6, 2020. Disponível em: https://datascience.codata.org/articles/dsj-2020-015. Acesso em: 4 fev. 2026.

SEGUNDO, Washington *et al*. **dARK**: a decentralized blockchain implementation of ARK persistent identifiers (1.0). [*S. l.*]: Zenodo, 2022. Disponível em: https://doi.org/10.5281/zenodo.7442743. Acesso em: 11 set. 2026.

SILVA, Fabiano Couto Corrêa da; SILVEIRA, Lúcia da. O ecossistema da ciência aberta. **Transinformação**, Campinas, v. 31, art. e190001, 2019. Disponível em: https://www.scielo.br/j/tinf/a/dJ89vRg94Qxtf6Y7M49Hztr/?lang=pt. Acesso em: 13 nov. 2025.

SILVEIRA, Marcos da; PRUSKI, Cédric. Persistent identifiers and their limitations in a dynamic web. *In*: INTERNATIONAL CONFERENCE ON SEMANTIC SYSTEMS (SEMANTICS), 18., 2022, Viena. **Anais** [...]. Aachen: CEUR-WS, 2022. p. 1-6. (CEUR Workshop Proceedings, v. 3235). Disponível em: https://ceur-ws.org/Vol-3235/paper24.pdf. Acesso em: 4 fev. 2026.

WEIGEL, Tobias; KINDERMANN, Stephan; LAUTENSCHLAGER, Michael. Actionable persistent identifier collections. **Data Science Journal**, [*s. l.*], v. 12, p. 191-206, jan. 2014. Disponível em: https://datascience.codata.org/articles/dsj.12-058. Acesso em: 4 fev. 2026.

WILKINSON, Mark D.; DUMONTIER, Michel; AALBERSBERG, IJsbrand Jan; APPLETON, Gabrielle *et al*. The FAIR guiding principles for scientific data management and stewardship. **Scientific Data**, [*s. l.*], v. 3, art. 160018, mar. 2016. Disponível em: https://www.nature.com/articles/sdata201618. Acesso em: 4 fev. 2026.

WITTENBURG, Peter. From persistent identifiers to digital objects to make data science more efficient. **Data Intelligence**, Cambridge, MA, v. 1, n. 1, p. 6-20, jun. 2019. Disponível em: https://direct.mit.edu/dint/article/1/1/6/9975/From-Persistent-Identifiers-to-Digital-Objects-to. Acesso em: 22 jan. 2026.