

XX ENANCIB

21 a 25 Outubro/2019 – Florianópolis

A Ciência da Informação e a era da Ciência de Dados

ISSN 2177-3688

GT-7 – Produção e Comunicação da Informação em Ciência, Tecnologia & Inovação

PRESENÇA DO *DIGITAL OBJECT IDENTIFIER* (DOI) EM CURRÍCULOS DA PLATAFORMA LATTES

PRESENCE OF THE DIGITAL OBJECT IDENTIFIER (DOI) IN THE LATTES PLATFORM CVs

Iara Vidal - Fundação Getúlio Vargas
Fabio Castro Gouveia - Fundação Oswaldo Cruz

Modalidade: Resumo Expandido

Resumo: O *Digital Object Identifier* (DOI) é um identificador persistente utilizado por publicações científicas para garantir um “endereço” fixo para artigos e outros documentos. Este estudo descreve a presença de DOIs de artigos de periódico nos Currículos Lattes de pesquisadores com doutorado. A proporção de currículos com indicação de pelo menos um DOI varia significativamente entre as áreas de conhecimento, sendo mais expressiva nas grandes áreas Ciências Biológicas, Ciências da Saúde, e Ciências Exatas e da Terra. Na grande área Linguística, Letras e Artes, há indicação de DOI em aproximadamente $\frac{1}{4}$ dos currículos.

Palavras-Chave: *Digital Object Identifier*; DOI; Plataforma Lattes; Identificadores persistentes.

Abstract: The *Digital Object Identifier* (DOI) is a persistent identifier used by scientific publications to grant a fixed “address” for articles and other documents. This study describes the presence of journal article DOIs in the Lattes curricula belonging to researchers with a PhD degree. The proportion of curricula with at least one DOI varies significantly between different disciplines. It is more expressive in Biology, Health, Exact and Earth Sciences. For Linguistic, Literature and Arts, approximately $\frac{1}{4}$ of the curricula have at least one DOI.

Keywords: *Digital Object Identifier*; DOI; Lattes Platform; Persistent identifiers.

1 INTRODUÇÃO

A publicação eletrônica oferece diversas vantagens para a disseminação da informação científica, facilitando a realização de buscas em textos completos e a consulta simultânea a diferentes coleções e materiais. Traz também algumas dificuldades, algumas análogas às já existentes nos meios impressos (por exemplo, vieses de cobertura em bases de dados), e outras particulares aos meios eletrônicos. Um problema importante diz respeito à persistência dos endereços digitais.

Um artigo de periódico é identificado por seu título, autoria, data de publicação, e pelo fascículo em que foi publicado: volume, número, páginas. No meio eletrônico, esta identificação inclui um endereço digital, ou URL (sigla de *Uniform Resource Locator*, localizador uniforme de recursos). No entanto, ao contrário do número do fascículo e das páginas de um artigo impresso, URLs não são estáveis. Quando um *site* se modifica, reorganizando ou eliminando conteúdos, é comum que novas URLs sejam geradas, inutilizando as anteriores. Quem tentar utilizar estas URLs antigas irá se deparar com erros do tipo 404, que significam “página não encontrada”.

Uma solução para esse problema é a adoção de identificadores persistentes. Enquanto as URLs estão relacionadas à arquitetura interna dos *sites*, os identificadores persistentes servem para designar de forma unívoca um determinado objeto – artigo, figura, vídeo, etc. Por exemplo, imagine-se um artigo localizado na URL www.exemplo.org/123, ao qual foi atribuído o identificador persistente XYZ. Caso o artigo seja transferido para um outro endereço, por exemplo, www.exemplo.org/artigos/123, o identificador XYZ será atualizado com a nova URL. Assim, um identificador persistente deverá sempre apontar para o endereço correto de um dado objeto.

No âmbito dos estudos de comunicação científica, os identificadores persistentes não só têm permitido recuperar documentos que ao longo do tempo mudaram de URL, como também facilitam a interoperabilidade entre diferentes bases de dados (SAYÃO, 2007). Plataformas como a Pubmed e o arXiv possuem identificadores internos (PMID, arXiv ID), mas há também identificadores independentes que podem ser adotados por diferentes plataformas, como por exemplo o sistema Handle. Um dos identificadores mais populares atualmente é o *Digital Object Identifier* (identificador digital de objeto), mais conhecido pela sigla DOI (BRITO et al., 2016), uma aplicação do sistema Handle mantida pela Fundação Internacional do DOI (*International DOI Foundation* ou IDF). De acordo com a IDF, até maio de 2019 aproximadamente 190 milhões de nomes DOIs já tinham sido atribuídos, e a cada ano são registradas mais de 5 bilhões de resoluções de DOI (INTERNATIONAL DOI FOUNDATION, 2019). Além disso, vem se verificando “o incremento do DOI nas publicações seriadas como forma de credibilidade de informações científicas e garantia dos direitos autorais” (PIRES; PEREIRA JUNIOR; GIRARD, 2017, p.547).

Mundialmente, o DOI se faz cada vez mais presente entre as revistas e plataformas de publicação científica. Um estudo sobre a presença de nomes DOI nas bases

internacionais Web of Science e Scopus mostrou que a proporção de itens com DOI vem crescendo em todas as áreas do conhecimento. No entanto, há diferenças significativas entre elas: em 2014, os índices chegavam a 90% nas áreas de Ciências e Ciências Sociais, mas atingiam apenas 50% entre periódicos de Artes e Ciências Humanas (GORRAIZ et al., 2016). No caso das revistas brasileiras, o principal contribuinte para a disseminação do DOI é a plataforma SciELO, membro da Crossref, que disponibiliza DOIs para os artigos desde 2008 (COLEÇÃO..., 2008). No entanto, um estudo centrado em revistas usuárias do sistema SEER (versão brasileira do *Open Journal System* – OJS) identificou que apenas 15% delas utilizavam nomes DOI (GUEDES; BRITO; SHINTAKU, 2013). Em 2014, a Associação Brasileira de Editores Científicos (ABEC) e o Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT) firmaram um contrato com a Crossref que facilita a aquisição de nomes DOI por parte de editores brasileiros, associados ou não à ABEC (BRITO et al., 2016). A parceria serve de impulso para a adoção do DOI no contexto brasileiro: mais de 170 mil nomes DOI foram registrados pela ABEC somente no decorrer de 2016, e em novembro de 2017 a ABEC já tinha cadastrado 487 prefixos e 1.129 periódicos nacionais (ROCHA, 2017).

Desde 2007, a partir de um acordo entre o CNPq e a IDF, a Plataforma Lattes aceita a inserção de nomes DOI no cadastro de artigos de periódico em currículos (TRABALHOS CERTIFICADOS, 2007). Ao inserir o DOI de um artigo, os metadados correspondentes são preenchidos automaticamente, facilitando o trabalho dos pesquisadores na manutenção dos currículos. A presença de DOIs nos Currículos Lattes, uma das bases mais representativas da produção científica nacional, abre um leque de possibilidades para a realização de estudos biblio- e cientométricos da ciência brasileira. Além de facilitar a identificação e deduplicação de publicações, o DOI, por ser mantido por uma organização internacional e independente, permite a validação externa das informações, etapa fundamental para estudos baseados em currículos preenchidos livremente pelos próprios pesquisadores, como é o caso na Plataforma Lattes (SILVA; SMIT, 2009). Assim, em teoria, o DOI permitiria realizar estudos mais abrangentes da produção científica nacional, indo além do que é indexado em bases como Web of Science, Scopus e SciELO.

Os DOIs da Plataforma Lattes têm servido como ferramenta adicional para trabalhos de coleta de dados e povoamento de repositórios institucionais, auxiliando especialmente na etapa de deduplicação de materiais (FERREIRA, 2016; MATIAS, 2015; VIDOTTI, 2015, entre outros). Outros estudos (por exemplo, ARAUJO; ALVES, 2019) também exploram sua

utilidade para estudos alométricos, que investigam a visibilidade de trabalhos científicos nas plataformas da *web* social (GOUVEIA, 2013). No entanto, falta uma visão panorâmica da presença de DOIs na Plataforma Lattes, identificando as áreas que mais utilizam o identificador e aquelas onde seu uso é ainda incipiente.

Em vista disso, o objetivo deste trabalho é descrever a presença do DOI nos currículos de doutores presentes na Plataforma Lattes, a fim de apurar a viabilidade de utilizar este identificador em estudos da produção científica brasileira. Trata-se de um extrato de uma pesquisa em andamento que visa levantar a produção bibliográfica de pesquisadores doutores vinculados a instituições brasileiras de ensino e pesquisa, com o objetivo de obter indicadores biblio-, cito- e alométricos e mapear a rede da produção destas instituições a partir de suas produções individuais. Pretende-se assim contribuir para os estudos de métricas da informação fornecendo contexto quanto ao uso do DOI por diferentes áreas permitindo subsidiar estudos futuros.

2 METODOLOGIA

Este estudo descritivo tem o objetivo de analisar a presença de nomes DOI em currículos de doutores registrados na Plataforma Lattes. A coleta de dados da Plataforma Lattes teve lugar durante os meses de setembro e outubro de 2018, a partir de uma listagem de 307.383 currículos de pessoas com doutorado. Os currículos completos, em formato HTML, foram obtidos com auxílio do programa scriptLattes (MENA-CHALCO; CÉSAR JÚNIOR, 2009). 22 itens não puderam ser baixados nem consultados diretamente na base Lattes, sendo descartados da análise. Assim, trabalhamos efetivamente com 307.361 currículos.

As áreas de conhecimento dos artigos foram determinadas a partir das indicações dos doutores na seção “Áreas de Atuação” de seus respectivos Currículos Lattes, preenchida de acordo com a tabela do CNPq contendo nove grandes áreas, a saber: Ciências Agrárias, Ciências Biológicas, Ciências da Saúde, Ciências Exatas e da Terra, Ciências Humanas, Ciências Sociais Aplicadas, Engenharias, Linguística, Letras e Artes, e Outras. Embora os pesquisadores possam apontar múltiplas áreas, optamos por selecionar apenas a primeira Grande Área indicada em cada currículo. Não foi possível identificar uma área do conhecimento para 22.422 currículos, 7,30% do total.

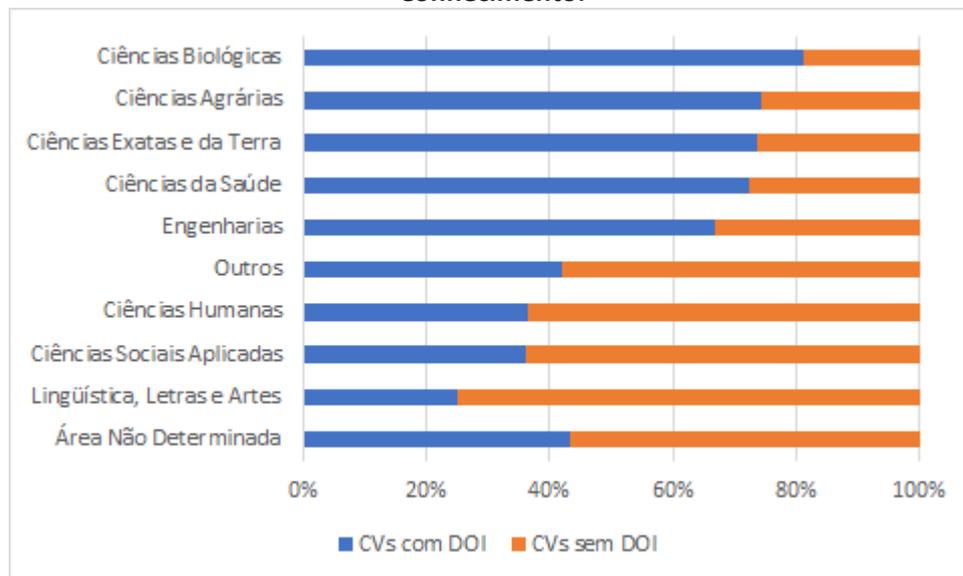
Para extrair os nomes DOI e as áreas de conhecimento de cada currículo, utilizamos rotinas de busca no editor de textos Notepad++, organizando posteriormente os dados em

planilhas com o software Excel. De 307.361 currículos analisados, 180.168 (58,62%) continham pelo menos um DOI. Encontramos 2.155.015 nomes DOIs cadastrados no campo específico para DOIs da seção Artigos de Periódicos destes currículos, número que corresponde a 39,88% do total bruto de Artigos de Periódicos identificados (n=5.404.116).

3 RESULTADOS E DISCUSSÃO

No cômputo geral, a maioria (58,62%) dos currículos analisados possui indicação de pelo menos um nome DOI. No entanto, esta proporção varia significativamente conforme a área de conhecimento, como demonstra o Gráfico 1.

Gráfico 1: Proporção de Currículos Lattes com pelo menos um DOI em cada Grande Área do Conhecimento.



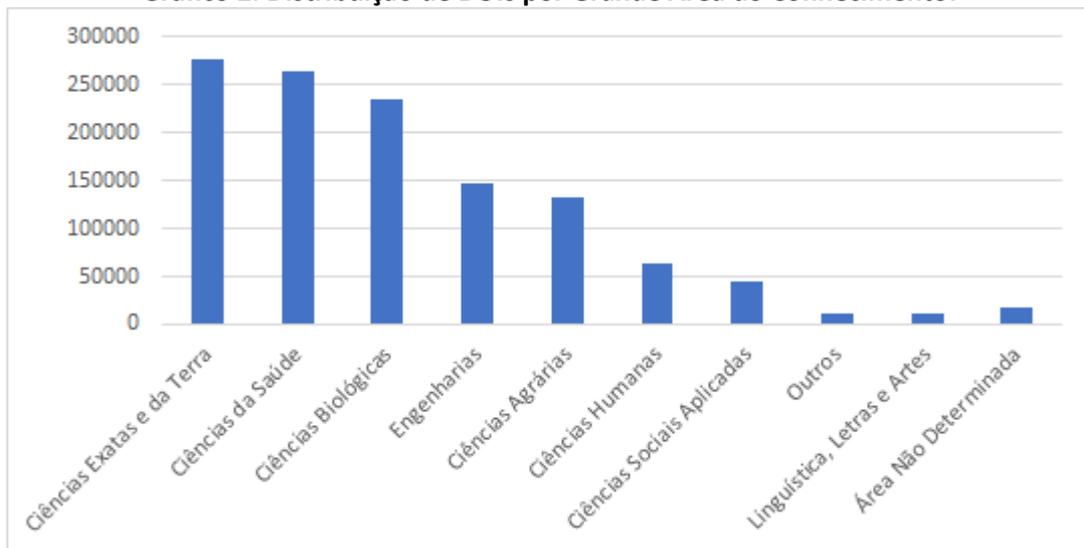
Fonte: Dados da pesquisa.

A maioria dos currículos nas Ciências Biológicas, Ciências Agrárias, Ciências Exatas e da Terra, Ciências da Saúde e Engenharias apresentam pelo menos um nome DOI. A proporção de currículos com DOI nestas áreas é superior à média geral, e chega a ultrapassar os 80% para as Ciências Biológicas. No outro extremo, Ciências Humanas e Ciências Sociais Aplicadas apresentam índices de currículos com DOI ligeiramente superiores a 36%, enquanto na grande área Linguística, Letras e Artes menos de 25% dos currículos possuem alguma indicação de DOI.

As diferenças entre as áreas são visíveis também considerando o número absoluto de DOIs. Ao todo, foram encontrados 2.155.015 nomes DOI nos currículos analisados. Após o descarte de duplicatas, restaram 1.002.117 nomes DOI únicos. O Gráfico 2 mostra a

quantidade de nomes DOI registrados por Grande Área do Conhecimento. Cabe ressaltar que as áreas não são mutuamente exclusivas. DOIs presentes em 2 ou mais currículos da mesma área foram contados apenas uma vez, e aqueles presentes em currículos de diferentes áreas foram contados uma vez para cada uma delas. A categoria “Área Não Determinada” inclui apenas os casos em que nenhum dos currículos associados a um determinado DOI apresentava indicação de área.

Gráfico 2: Distribuição de DOIs por Grande Área do Conhecimento.



Fonte: Dados da pesquisa.

Ciências Exatas e da Terra, Ciências da Saúde, e Ciências Biológicas são as áreas com maior presença de nomes DOI, enquanto Linguística, Letras e Artes novamente aparecem no extremo oposto. Estes resultados são compatíveis com os encontrados no estudo de Gorraiz e outros sobre a presença de DOIs nas bases internacionais Web of Science e Scopus (2016), sinalizando que há diferenças na taxa de adesão ao sistema DOI entre periódicos de diferentes áreas. Deve-se considerar também as diferenças na produtividade e nas práticas de comunicação científica entre as áreas: Ciências Exatas e da Terra, Ciências da Saúde, e Ciências Biológicas estão entre as áreas que mais produzem artigos de periódicos, veículo relativamente menos expressivo entre as Humanidades.

5 CONSIDERAÇÕES FINAIS

Este trabalho apresenta resultados preliminares de um estudo em andamento que busca descrever a presença do DOI nos currículos Lattes de pesquisadores com doutorado. Verifica-se que, embora a maioria dos currículos possua pelo menos uma indicação de DOI,

há variações significativas entre as diferentes áreas do conhecimento. Isto pode se dever, dentre outras questões, à diferença por área na adoção do DOI pelos periódicos. Outra questão pode estar no fato da base Lattes ser preenchida pelos pesquisadores, portanto artigos que tenham DOI podem não ter tido o identificador fornecido, especialmente em se tratando de produções antigas. Ainda assim, mesmo na área de Linguística, Letras e Artes, em que o artigo de periódico é relativamente menos importante para a comunicação, é possível encontrar DOIs em aproximadamente ¼ dos currículos. Estes resultados demonstram a popularidade obtida pelo identificador DOI entre a comunidade científica brasileira, e sinalizam que os DOIs presentes na Plataforma Lattes podem ser considerados um corpus representativo da produção científica nacional, especialmente nas áreas de Ciências Biológicas, Ciências da Saúde, e Ciências Exatas e da Terra.

Nas etapas seguintes, será realizada a validação dos metadados referentes aos DOIs do Lattes, permitindo identificar os periódicos mais frequentes e mapear a distribuição temporal destes DOIs, entre outras possibilidades. Pretendemos assim contribuir para estudos futuros que possam se servir do DOI para análises da produção científica nacional.

AGRADECIMENTOS

Os autores gostariam de agradecer o apoio do Conselho Nacional de Desenvolvimento Científico e Tecnológico via processo 430982/2018-6 da Chamada Universal MCTIC/CNPq nº28/2018.

REFERÊNCIAS

ARAUJO, Ronaldo Ferreira; ALVES, Marcelo. Indicadores altmétricos da produção de pesquisadores brasileiros: análise dos bolsistas de produtividade do CNPq. **Revista Ibero-Americana de Ciência da Informação**, v. 12, n. 3, p. 850-862, 13 set. 2019.

BRITO, Ronnie Fagundes de *et al.* **Guia do usuário do Digital Object Identifier**. Brasília, DF: IBICT, 2016. Disponível em: <https://doi.org/10.21452/978-85-7013-112-6>. Acesso em: 2 ago. 2019.

COLEÇÃO SciELO Brasil passa a incluir o tratamento completo do número DOI. **Comunidade Virtual de Comunicação Científica em Saúde**, 28 jan. 2008. Disponível em: http://cvirtual-ccs.bvsalud.org/tiki-read_article.php?articleId=236. Acesso em: 11 ago. 2019.

DAMASIO, Edilson. CrossRef, DOI (Digital Object Identifier) e serviços: estudo comparativo luso-brasileiro. **InCID: Revista de Ciência da Informação e Documentação**, v. 4, n. 2, p. 126-142, 2013. Disponível em: <https://doi.org/10.11606/issn.2178-2075.v4i2p126-142>.

Acesso em: 10 ago. 2019.

FERREIRA, Elisabete. **Um método de coleta e classificação de metadados de produção científica em repositórios digitais institucionais**. 2016. 64 f. Dissertação (Mestrado) - Universidade Federal do Paraná, Curitiba, 2016.

GORRAIZ, Juan *et al.* Availability of digital object identifiers (DOIs) in Web of Science and Scopus. **Journal of Informetrics**, v. 10, n. 1, p. 98-109, 2016. Disponível em: <https://doi.org/10.1016/j.joi.2015.11.008>. Acesso em: 10 ago. 2019.

GOUVEIA, Fabio Castro. Almetria: métricas de produção científica para além das citações. **Liinc em Revista**, v. 9, n. 1, p. 214-227, 2013. Disponível em: <https://doi.org/10.18617/liinc.v9i1.569>. Acesso em: 10 ago. 2019.

GUEDES, Mariana Giubertti; BRITO, Ronnie Fagundes De; SHINTAKU, Milton. Análise da utilização do DOI no Brasil em periódicos de acesso aberto. In: CONGRESSO BRASILEIRO DE BIBLIOTECONOMIA, DOCUMENTAÇÃO E CIÊNCIA DA INFORMAÇÃO, 25., 2013, Florianópolis. **Anais...** Florianópolis: FEBAB, 2013. p. 1669-1684. Disponível em: <http://portal.febab.org.br/anais/article/view/1377>. Acesso em: 03 ago. 2019.

INTERNATIONAL DOI FOUNDATION. **Factsheet: key facts on Digital Object Identifier system**. [S.l.], 2019. Disponível em: <https://www.doi.org/factsheets/DOIKeyFacts.html>. Acesso em: 9 ago. 2019.

MATIAS, Mesailde Souza de Oliveira. **Base referencial para o povoamento de repositórios institucionais: coleta automatizada de metadados da Plataforma Lattes**. 2015. 86 f. Dissertação (Mestrado) - Universidade Federal de São Carlos, São Carlos, 2015.

MENA-CHALCO, Jesús Pascual; CÉSAR JÚNIOR, Roberto Marcondes. ScriptLattes: an open-source knowledge extraction system from the Lattes platform. **Journal of the Brazilian Computer Society**, v. 15, n. 4, p. 31-39, 2009. Disponível em: <https://doi.org/10.1007/BF03194511>. Acesso em: 06 ago. 2019.

PIRES, Erik André de Nazaré; PEREIRA JÚNIOR, Antônio Afonso; GIRARD, Carla Daniella Teixeira; et al. O Digital Object Identifier (DOI) em periódicos científicos eletrônicos de comunicação e informação. **RDBCi: Revista Digital de Biblioteconomia e Ciência da Informação**, v. 15, n. 3, p. 533, 2017.

ROCHA, Leandro. Abertura do ABEC Meeting 2017 contou com a presença do secretário estadual de Ciência, Tecnologia e Ensino Superior do Paraná. **Associação Brasileira de Editores Científicos - ABEC**, 2017. Disponível em: <https://www.abecbrasil.org.br/novo/2017/11/abertura-do-abec-meeting-2017-contou-com-a-presenca-do-secretario-estadual-de-ciencia-tecnologia-e-ensino-superior-do-parana/>. Acesso em: 02 ago. 2019.

SAYÃO, Luís Fernando. Interoperabilidade das bibliotecas digitais: o papel dos sistemas de identificadores persistentes - URN, PURL, DOI, Handle System, CrossRef e OpenURL.

Transinformação, v. 19, n. 1, p. 65-82, 2007. Disponível em:
<https://doi.org/10.1590/S0103-37862007000100006>. Acesso em: 2 ago. 2019.

SILVA, Fábio Mascarenhas; SMIT, Johanna Wilhelmina. Organização da informação em sistemas eletrônicos abertos de informação científica & tecnológica: análise da Plataforma Lattes. **Perspectivas em Ciência da Informação**, v. 14, n. 1, p. 77-98, 2009. Disponível em:
<http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/720>. Acesso em: 10 ago. 2019.

TRABALHOS certificados. **Agência FAPESP**, 20 jun. 2007. Disponível em:
http://agencia.fapesp.br/trabalhos_certificados/7317/. Acesso em: 8 ago. 2019.

VIDOTTI, Silvana Aparecida Borsetti Gregorio. Coleta de dados a partir dos currículos da Plataforma Lattes: procedimentos utilizados no Repositório Institucional UNESP. **PontodeAcesso**, v. 9, n. 3, 2015. Disponível em: <https://doi.org/10.9771/1981-6766rpa.v9i3.15164>. Acesso em: 01 out. 2019.

.