



ANÁLISE DA PRODUÇÃO CIENTÍFICA SOBRE METADADOS DE PROVENIÊNCIA

Felipe Ivo da Silva

 <https://orcid.org/0009-0005-1379-4692>.

✉ felipe_ivodasilva@hotmail.com.

🏢 Universidade Federal de São Carlos (UFSCar) |

ROR: <https://ror.org/00qdc6m37> | São Carlos, Brasil.

Felipe Augusto Arakaki

 <https://orcid.org/0000-0002-3983-2563>.

✉ felipe.arakaki@unb.br.

🏢 Universidade de Brasília (UnB) |

ROR: <https://ror.org/02xfp8v59> | Brasília, Brasil.

Eixo temático: Produtividade e Colaboração Científica

Modalidade: Pecha Kucha

DOI: 10.22477/ix.ebbc.236

Resumo: Os metadados de proveniência são informações que fortalecem a validade da produção acadêmica, e também facilitam a colaboração e o compartilhamento de recursos entre pesquisadores. Com a ascensão da ciência aberta e colaborativa, a integridade e compreensão dos metadados de proveniência tornam-se elementos cruciais para promover a confiança entre pesquisadores. Dessa forma, busca oferecer uma visão panorâmica da produção científica através de indicadores quantitativos associados ao termo, destacando como essa prática fortalece a robustez da produção acadêmica e promove uma cultura de dados transparentes e confiáveis.

Palavras-Chave: Proveniência. Metadados. Metadados de proveniência. Provenance metadata.

1 INTRODUÇÃO

O contexto de metadados de proveniência na literatura acadêmico-científica, em especial no campo da Ciência da Informação, destaca a crescente relevância do registro da origem e da trajetória dos dados em ambientes digitais. Conforme argumentado por Arakaki (2020), a representação da proveniência é fundamental para permitir a confiabilidade dos dados, pois apresenta informações sobre quem criou ou modificou um recurso, quais ações realizadas que modificaram um recurso, além de apresentar informações sobre o próprio recurso. Para Moreau e Groth (2013), os metadados de proveniência são informações referentes às entidades, às atividades e às pessoas envolvidas na produção de um dado ou coisa.

Além disso, a natureza dinâmica e colaborativa da Ciência da Informação, frequentemente envolvida na organização e recuperação de dados, amplia a importância dos metadados de proveniência. De acordo com Alves (2010), os padrões de metadados são estruturas de descrição constituídas por um conjunto predeterminado de metadados metodologicamente construídos e padronizados. Sendo assim, a finalidade do padrão de metadados é descrever uma entidade, gerando uma representação unívoca e padronizada que possa ser utilizada para recuperação da mesma.

Desta forma, o objetivo desta pesquisa é apresentar um panorama da produção científica que utiliza o termo metadados de proveniência que tem como ferramenta facilitar a colaboração e o compartilhamento de recursos informacionais entre pesquisadores. Os metadados de proveniência, além de fornecerem um caminho distinto das transformações dos dados, contribuem também para a credibilidade das resultantes obtidas a partir desses dados.

2 METODOLOGIA

As informações desta pesquisa foram coletadas das bases de dados *Scopus* e *Web of Science* (WoS). Utilizar as bases de dados citadas, em uma pesquisa exploratória quantitativa, justifica-se pela sua reputação consolidada e confiabilidade no meio acadêmico. A revisão por pares, presente em muitas das publicações indexadas nas bases, assegura um elevado padrão de qualidade das informações disponíveis. Essa consistência facilita a identificação e extração de informações relevantes para a pesquisa.

O período analisado compreende o ano de 2004 até 2023, em que foram considerados os documentos artigos de conferências, artigos de periódicos, eventos, artigos de revisão, capítulo de livro, livro e revista científica, utilizando o termo de busca *Provenance metadata*. Na *Scopus*, obteve-se 627 registros a partir da busca realizada, enquanto na WoS identificou-se 90 registros, sendo que 73 registros possuem uma sobreposição de trabalho e 4 registros possuem duplicidade na *Scopus* e 2, também com duplicidade, na WoS.

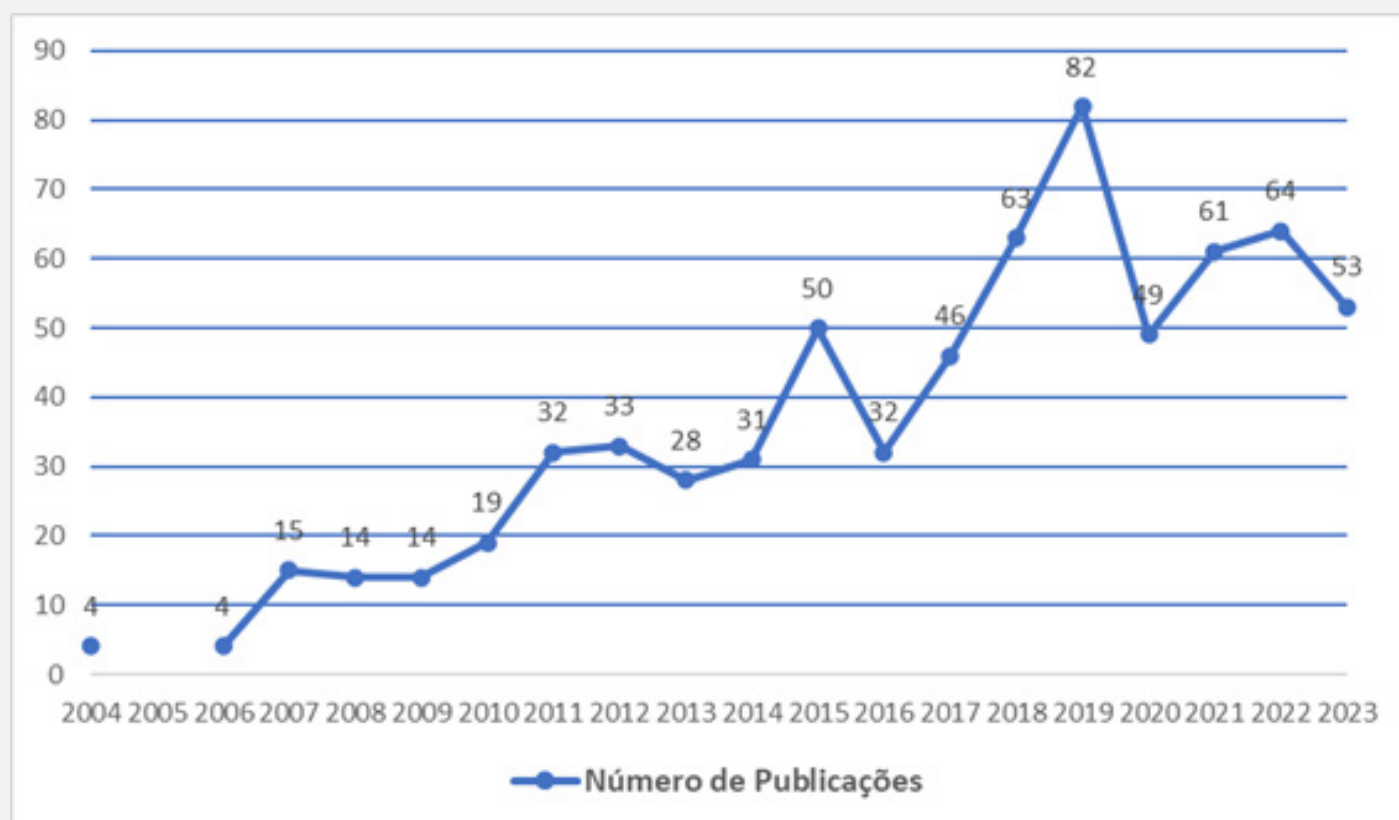
Para concluir os objetivos desta pesquisa, foi feito um tratamento que verificou se havia duplicidade de informação no documento, com dados extraídos em formato *Microsoft Excel®*, onde destacou-se os registros em DOI, autor, título e identificador das bases. Em seguida, os dados foram organizados e confrontados com as ferramentas de verificação de duplicidade e de repetição, cada qual com suas

especificidades, de modo a proporcionar uma melhor visualização dos resultados.

3 RESULTADOS

Durante o período analisado, observou-se uma flutuação na utilização do termo *Provenance meta-data*, com momentos de aumento intercalados com períodos de diminuição. Observou-se também uma grande sobreposição desse tema em ambas as bases. Além disso, após o tratamento dos dados, verificou-se que a base de dados *Scopus* continha 606 publicações, enquanto a base de dados *WoS* continha apenas 15 publicações, dessas, 73 registros foram encontrados em ambas as bases e 6 estavam repetidos, o que sugere uma concordância entre as publicações indexadas nessas plataformas. O número total de publicações contabilizadas em ambas as bases de dados totaliza 694 documentos.

Gráfico 1: Publicações por ano na *Web of Science* e na *Scopus*



Fonte: Elaborado pelos autores (2024).

Para identificar duplicidades de informações entre as bases, foi necessário comparar os registros utilizando o Identificador de Objeto Digital (DOI) e os títulos das publicações. Vale ressaltar que os identificadores garantem uma identificação clara e exclusiva de cada trabalho, facilitando a referência, recuperação e integração de dados entre diferentes fontes bibliográficas.

Além disso, destaca-se que alguns registros não possuíam DOI, enquanto outros tinham títulos escritos de forma diferente, exigindo uma análise mais detalhada para determinar a duplicidade das informações. Assim, verificou-se que a *Scopus* possui noventa e nove registros não identificados com o DOI, mas que o identificador único de trabalho acadêmico na base bibliográfica *Scopus* (*Scopus EID*) possui todos os registros. Já a *WoS* não possui essa totalidade, somente vinte e seis registros com o DOI



e o seu identificador único de trabalho também possui todos os registros.

Os autores mais representativos dentro da base de dados Scopus são Lu, S. e Gehani, A., que possuem maior relevância na utilização do termo *Provenance Metadata*, sendo que a diferença de publicação entre eles é consideravelmente relevante, 22 contra 16, respectivamente. Já na base de dados WoS, os dois primeiros autores que se destacam são Sahoo, S.S. e Zhao, J., com 7 e 5 publicações, de modo respectivo, sendo que todos os autores transitam em ambas as bases e o campo de pesquisa dos autores se dedica a área da tecnologia.

4 CONSIDERAÇÕES FINAIS

À medida que a comunidade científica adota a era da ciência aberta e colaborativa, a integridade e a compreensão da proveniência dos metadados tornam-se elementos substanciais para promover a confiança entre pesquisadores e aprimorar a reutilização de conjuntos de dados em diferentes contextos. Portanto, esta pesquisa buscou traçar um panorama da produção acadêmico-científico por meio de indicadores quantitativos atrelados aos termos associados a metadados de proveniência.

Assim, analisando as bases de dados *Web of Science* e *Scopus*, este artigo teve por objetivo apresentar as características da produção científica relacionadas ao termo em inglês *Provenance Metadata*. Para tal, analisou um montante de 694 documentos de vertente científica disponibilizados nas bases.

Por fim, os resultados desta pesquisa foram acercados às publicações com o termo em inglês, *Provenance Metadata*, disponíveis na base de dados *Web of Science* e *Scopus*, devido à predominância desse idioma na comunicação acadêmica nacional e internacional. Ainda assim, futuras pesquisas podem ser realizadas de modo a expandir outras buscas.

REFERÊNCIAS

ALVES, Rachel Cristina Vesú. **Metadados como elementos do processo de catalogação**. 2010. 132 f. Tese (Doutorado) - Universidade Estadual Paulista, Faculdade de Filosofia e Ciências, Marília, SP, 2010. Disponível em: <http://hdl.handle.net/11449/103361>. Acesso em: 1 fev. 2024.

ARAKAKI, Felipe Augusto. **Metadados e modelo PROV: perspectivas dos dados de proveniência em contextos digitais**. Informação & Informação, v. 25, n. 3, p. 187-211, 2020. DOI: <https://doi.org/10.5433/1981-8920.2020v25n3p187>. Disponível em: <https://ojs.uel.br/revistas/uel/index.php/informacao/article/view/41869/pdf>. Acesso em: 3 fev. 2024.

MOREAU, Luc; GROTH, Paul. **Provenance: an introduction to PROV**. San Rafael, Califórnia, Estados Unidos: Morgan & Claypool, 2013. (Synthesis lectures on the semantic web: theory and technology, 7). DOI: <https://doi.org/10.1007/978-3-031-79450-6>. Disponível em: <https://link.springer.com/book/10.1007/978-3-031-79450-6>. Acesso em: 3 fev. 2024.