

XXIV ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – XXIV ENANCIB

ISSN 2177-3688

GT 11 – Informação & Saúde

PROVENIÊNCIA DE DADOS EM SAÚDE: CONTRIBUIÇÕES E DESAFIOS

DATA PROVENANCE IN HEALTH: CONTRIBUTIONS AND CHALLENGES

Márcio José Sembay – Universidade Federal de Santa Catarina (UFSC)¹

Douglas Dyllon Jeronimo de Macedo – Universidade Federal de Santa Catarina (UFSC)²

Alexandre Augusto Gimenes Marquez Filho – Universidade Federal de Santa Catarina (UFSC)³

Modalidade: Trabalho Completo

Resumo: A proveniência de dados desempenha um papel crucial na área da saúde ao garantir a integridade e confiabilidade das informações empregadas em decisões clínicas e administrativas. Fortalece a confiança nos sistemas de informação em saúde e aprimora a qualidade dos serviços prestados. É pertinente discutir a proveniência de dados nesse contexto, destacando suas principais contribuições e desafios, a fim de orientar pesquisadores e profissionais de saúde. Este estudo teve como objetivo identificar e mapear as evidências científicas sobre a proveniência de dados nos diferentes contextos de saúde, apresentando as principais contribuições e desafios. Realizamos uma revisão integrativa utilizando como estratégia de busca os termos *data provenance and health*. Após analisar os resultados obtidos e revisar os textos completos, identificamos e selecionamos 11 artigos internacionais relevantes. Nossa pesquisa destaca a escassez de estudos sobre a proveniência de dados em saúde, embora tenham sido observados avanços na implementação de sistemas de proveniência em diferentes contextos de saúde na literatura analisada. Nesse entendimento, esta pesquisa representa uma contribuição significativa para o campo da Ciência da Informação. Sugerimos pesquisas futuras para explorar métodos, técnicas, modelos, metodologias e tecnologias inovadoras como blockchain e inteligência artificial que possam melhorar o uso da proveniência de dados em saúde e ampliar capacidades.

Palavras-chave: proveniência de dados; dados em saúde; saúde.

Abstract: Data provenance plays a crucial role in health by guaranteeing the integrity and reliability of the information used in clinical and administrative decisions. It strengthens trust in health information systems and improves the quality of the services provided. It is pertinent to discuss data provenance in this context, highlighting its main contributions and challenges, in order to guide researchers and

¹ marcio.sembay@posgrad.ufsc.br.

² douglas.macedo@ufsc.br.

³ alexandre.agmf@gmail.com.

health professionals. This study aimed to identify and map the scientific evidence on data provenance in different health contexts, presenting the main contributions and challenges. We carried out an integrative review using the terms data provenance and health as a search strategy. After analyzing the results obtained and reviewing the full texts, we identified and selected 11 relevant international articles. Our research highlights the scarcity of studies on data provenance in health, although advances in the implementation of provenance systems in different health contexts have been observed in the literature analyzed. With this in mind, this research represents a significant contribution to the field of Information Science. We suggest future research to explore innovative methods, techniques, models, methodologies and technologies such as blockchain and artificial intelligence that can improve the use of data provenance in health and expand capabilities.

Keywords: data provenance; health data; health.

1 INTRODUÇÃO

A literatura especializada em Ciência da Informação explica que as fontes primárias são caracterizadas pela ausência de organização prévia, sendo as primeiras fontes de dados que não foram submetidas a tratamento, o que permite ao pesquisador adaptá-las conforme suas necessidades específicas (Araújo; Mota, 2020). Nesse contexto, a Arquivologia, ao tratar da questão da proveniência de forma consolidada ao longo do tempo, define o termo “Proveniência” como um princípio arquivístico fundamental. De acordo com o “Dicionário de Biblioteconomia e Arquivologia”, esse princípio estabelece que os documentos ou arquivos provenientes de uma instituição, corporação, família ou indivíduo devem ser mantidos separados dos de outras origens, preservando-se sua integridade. Ademais, tal definição frequentemente abarca o princípio do respeito à ordem original dos documentos (Cunha; Cavalcanti, 2008).

Segundo Buneman, Khanna e Chiew (2001), a proveniência de dados é a documentação complementar de um determinado dado, que contém a descrição de “como”, “quando”, “onde” e “por que” ele foi obtido e “quem” o obteve. Ainda, conforme o estudo de Freund, Sembay e Macedo (2019), a proveniência de dados pode ser considerada um requisito importante para estabelecer confiabilidade e prover segurança em Sistemas de Informação (SI).

Já em um contexto de saúde, a proveniência de dados identifica as transformações pelas quais os dados passaram ao longo do tempo, desempenhando um papel crucial nos processos de gestão e rastreamento das informações de saúde. Isso garante a segurança dos dados e proporciona economia para a instituição de saúde, além de apoiar a tomada de decisões gerenciais e promover a melhoria contínua dos processos de gerenciamento de

dados (Sembay; Macedo; Dutra, 2020a; Sembay; Macedo; Marquez Filho, 2022; Sembay; Macedo; Marquez Filho, 2023; Sembay, 2023; Sembay; Macedo; Marquez Filho, 2024). Assim, é importante ressaltar que, segundo Sembay *et al.* (2023), poucas são as iniciativas descritas na literatura sobre o uso de métodos, técnicas, modelos e metodologias para o gerenciamento de dados de proveniência voltados ao contexto de saúde, principalmente no que diz respeito a tecnologias que utilizem proveniência de dados em Sistemas de Informação em Saúde (SIS), os quais são fundamentais para coletar, armazenar, gerenciar e transmitir informações de saúde, apoiando decisões clínicas, administrativas e políticas, além de melhorar a eficiência e qualidade dos serviços de saúde.

Desta forma, investigar a proveniência de dados na saúde é fundamental, pois, embora já explorado pela Arquivologia, a Ciência da Informação traz uma abordagem inovadora que aprimora o rastreamento, a origem e a evolução dos dados, enriquecendo o gerenciamento informacional em ambientes complexos como o setor de saúde. Segundo Ribeiro (2009), a Ciência da Informação, sendo uma ciência social, lida diretamente com indivíduos e contextos sociais, trabalhando com diferentes tipos de informações – clínicas, científicas e administrativas – consolidando-se como um campo de estudo relevante nesse cenário. Diante disso, surgiu a seguinte questão norteadora: Quais são as evidências científicas disponíveis na literatura sobre a proveniência de dados em um contexto geral de saúde? Definiu-se como objetivo desta pesquisa apresentar as principais contribuições e desafios encontrados nas evidências científicas disponíveis na literatura sobre a proveniência de dados nos diferentes contextos de saúde.

2 PROVENIÊNCIA DE DADOS

A proveniência de dados possui diversas definições, porém, neste estudo foram selecionadas aquelas que mais se aproximam e permitem alinhar-se aos aspectos interdisciplinares da Ciência da Computação e da Ciência da Informação.

Conforme destacado por Freire *et al.* (2008), a proveniência de dados é entendida como a informação que permite determinar a origem histórica de um dado, desde suas fontes originais, sendo essencial para garantir a reprodutibilidade dos resultados, compartilhamento e reuso do conhecimento pela comunidade científica. Pode abranger ainda, dados gerados manualmente ou automaticamente por sistemas informatizados.

Em diferentes contextos, a proveniência de dados apresenta características particulares que justificam sua aplicação. Assim, parte da taxonomia criada por Simmhan, Plale e Gannon (2005), apresentam as principais características da proveniência de dados que são fundamentais para categorizar os esforços de pesquisa, possibilitando a construção de sistemas de proveniência adaptados a diferentes formas de operação. Essas características são apresentadas no Quadro 1.

Quadro 1 – Principais características da Proveniência de Dados

Característica	Descrição
Qualidade dos dados	A linhagem pode ser utilizada para estimar a qualidade e confiabilidade dos dados com base nos dados originais e suas transformações, permitindo declarações de prova na derivação dos dados.
Trilhas de auditoria	A proveniência pode ser empregada para rastrear a trilha de auditoria e determinar o uso, além de detectar erros na geração de dados.
Receitas de replicação	Informações detalhadas de proveniência facilitam a repetição da derivação dos dados, contribuindo para a manutenção da atualização dos dados;
Atribuição	O pedigree dos dados pode estabelecer direitos autorais e propriedade dos dados, possibilitando citações adequadas e determinando responsabilidades em casos de dados incorretos.
Informacional	O uso genérico da linhagem dos dados é consultar com base em metadados para a descoberta de dados. Possibilita fornecer um contexto para interpretar os dados.

Fonte: Simmhan, Plale e Gannon (2005).

Portanto, as características descritas no Quadro 1 são essenciais para manter a integridade e a confiabilidade das informações, permitindo a validação de resultados e a proteção dos dados, o que é vital para a credibilidade e eficácia dos SI.

Por fim, é importante destacar ainda, que a proveniência de dados possui modelos desenvolvidos pelo *World Wide Web Consortium (W3C)* que estão incluídos na família PROV. Conforme Groth e Moreau (2013), os principais modelos são: PROV-PRIMER, PROV-O, PROV-XML, PROV-DM, PROV-N, PROV-AQ, PROV-DC, PROV-DICTIONARY, PROV-SEM e PROV-LINKS. Esses modelos padronizam a representação da proveniência de dados, descrevendo a origem, a história e as transformações que um dado sofre ao longo de seu ciclo de vida.

3 PROVENIÊNCIA DE DADOS EM SAÚDE

As aplicações da proveniência de dados estão passando por um crescente interesse de pesquisa em diversos cenários de saúde. As tecnologias aplicadas nessa área têm obtido resultados expressivos (Sembay; Macedo; Dutra, 2020b). Assim, nos mais variados cenários

de saúde que utilizam métodos, técnicas, modelos, metodologias e tecnologias que abordem a proveniência de dados em SIS, são essenciais, aprimorando a qualidade dos cuidados prestados (Sembay *et al.*, 2023). Nesse contexto, Ahmed *et al.* (2023) destacam que os registros de saúde inseridos nos SIS são cruciais para o tratamento de pacientes. Em países desenvolvidos e em desenvolvimento, os pacientes possuem registros eletrônicos de saúde que enfrentam desafios devido à complexidade e compartilhamento de dados sensíveis. Para Galvão e Ricarte (2019) vários países têm buscado soluções para melhorar a interoperabilidade entre sistemas existentes ou têm considerado a implementação de sistemas unificados para coleta de dados clínicos. Nesse sentido, Ahmed *et al.* (2023) ainda complementa que a necessidade urgente de acesso aos dados de saúde por meio de registros eletrônicos de saúde aumenta a relevância da proveniência dos dados no contexto de saúde para pesquisadores, analistas, pacientes e médicos.

É importante destacar que os dados de saúde são usados por médicos, pesquisadores e analistas, e o crescente desempenho computacional ressalta a importância de registrar “quando”, “por que” e “quem” acessa esses dados. A proveniência de dados é essencial para garantir sua segurança. No entanto, sistemas de registros eletrônicos de saúde não capturam ou armazenam o processo de compartilhamento dessas informações. Nesse contexto, os sistemas de proveniência auxiliam no registro e no armazenamento das circunstâncias em que os dados foram criados ou compartilhados com outras entidades, respondendo perguntas sobre “quem” coletou os dados, “por que” e “quando” foram coletados, e “qual formato” ou parte dos dados foram envolvidos (Ahmed *et al.*, 2023).

Por fim, segundo Sembay, Macedo e Marquez Filho (2022), a proveniência de dados pode melhorar a segurança e eficiência dos processos de gerenciamento e rastreamento de dados de saúde, além de possibilitar a realização de auditorias para responsabilização de erros médicos ou violações de segurança, rastreando todo o histórico completo dos dados armazenados, seja em sistemas locais ou na nuvem.

4 ASPECTOS METODOLÓGICOS

Este estudo trata de uma revisão integrativa da literatura que de acordo com Torraco (2016), é um método abrangente de síntese de pesquisa que combina resultados de estudos teóricos e empíricos sobre um tema, visando resumir o conhecimento existente, propor novas

perspectivas, identificar lacunas na literatura e sugerir direções para pesquisas futuras. Desta forma, os passos aqui seguidos foram: a formulação da questão norteadora, a identificação da amostra por meio da busca na literatura, a categorização dos estudos, a avaliação e análise dos dados e a síntese dos resultados (Whittemore; Knafl, 2005).

A busca de publicações (artigos) nas bases de dados eletrônicas *MEDLINE/PubMed*⁴ e *LILACS*⁵ foi realizada no intervalo de (2013 a 2023) a fim de assegurar a contemporaneidade das informações dos últimos 10 anos. Realizamos o levantamento de dados nessas bases devido à sua ampla cobertura e alta credibilidade na área da saúde, garantindo acesso a uma vasta gama de estudos científicos de qualidade. Foram utilizados os seguintes termos para a estratégia de busca avançada adequados para cada base de dados: (*Data Provenance*) AND (*Health*) para a base de dados *MEDLINE/PubMed* e “*Data provenance*” AND “*Health*” para a base de dados *LILACS*.

Adotamos como critérios de inclusão, artigos publicados que abordassem a temática proposta em seus títulos, resumos e/ou descritores, corpo do texto, disponibilizados como textos completos no idioma inglês. Como critérios de exclusão, apesar da escassa produção acerca do tema, consideramos teses, dissertações, monografias e jornais.

Após a leitura dos artigos, a amostra final foi selecionada e as informações organizadas em um quadro sintético, com dados sobre autor, ano, título, objetivo e conclusão, apresentado na seção 5. A análise de conteúdo, baseada em Bardin (2004), que considera a frequência e o contexto em que os termos aparecem nos estudos, foi aplicada na subseção 5.1. Essa análise permitiu identificar os principais desafios e contribuições dos estudos revisados por meio da contabilização de termos mais recorrentes, consolidando os achados mais relevantes.

5 ANÁLISE DOS ESTUDOS ENCONTRADOS

Inicialmente, foram identificados por meio da busca avançada, 430 estudos no total, sendo encontrados na base de dados *MEDLINE/PubMed* (381 estudos) e *LILACS* (49 estudos), mas, como todos os estudos da *LILACS* já estavam incluídos na *MEDLINE/PubMed*, a análise focou-se nesta última. Após aplicar critérios de inclusão e exclusão, 11 artigos publicados

⁴ <https://pubmed.ncbi.nlm.nih.gov/>.

⁵ <https://lilacs.bvsalud.org/>.

XXIV Encontro Nacional de Pesquisa em Ciência da Informação – XXIV ENANCIB
Vitória-ES – 04 a 08 de novembro de 2024

entre 2018 e 2023 foram selecionados. Esses estudos foram analisados detalhadamente para sintetizar e destacar as conclusões dos autores, resultando na criação do Quadro 2.

Quadro 2 – Sintetização dos estudos selecionados

Base de dados: MEDLINE/PUBMED	
1	Autor (es)/Ano: Holub <i>et al.</i> (2018). Título: Enhancing Reuse of Data and Biological Material in Medical Research: From FAIR to FAIR-Health. Objetivo: Discutir a subutilização de dados e materiais biológicos de biorepositórios, propondo a extensão dos Princípios FAIR (Findability, Accessibility, Interoperability, and Reusability) para o enriquecimento de dados e privacidade. Conclusão: A adoção dos princípios FAIR-Health melhora a rastreabilidade e a transparência dos dados e materiais biológicos na pesquisa médica, facilitando a proveniência de dados de saúde.
2	Autor (es)/Ano: Xu <i>et al.</i> (2018). Título: Application of Data Provenance in Healthcare Analytics Software: Information Visualisation of User Activities. Objetivo: Explorar a aplicação da proveniência de dados em softwares de análise de saúde, com foco na visualização das atividades dos usuários, visando aumentar a auditabilidade e transparência. Conclusão: A implementação da proveniência de dados em softwares empresariais de análise de saúde apresenta desafios como a complexidade e o volume dos dados, porém, a integração de relatórios de proveniência no software Atmolytics facilita a compreensão e uso dos dados pelos especialistas.
3	Autor (es)/Ano: Alterovitz <i>et al.</i> (2018). Título: Enabling precision medicine via standard communication of HTS provenance, analysis, and results. Objetivo: visa estabelecer um padrão para a comunicação de dados de “High-Throughput Sequencing (HTS)”, visando facilitar a aplicação da medicina de precisão por meio da padronização da proveniência. Conclusão: A padronização na comunicação de dados de HTS é essencial para a validade e reprodutibilidade científica, exigindo a captura precisa da proveniência e dos detalhes da análise.
4	Autor (es)/Ano: Sahoo <i>et al.</i> (2019). Título: ProvCaRe: Characterizing Scientific Reproducibility of Biomedical Research Studies using Semantic Provenance Metadata. Objetivo: visa caracterizar a reprodutibilidade científica em estudos biomédicos utilizando metadados semânticos de proveniência por meio do ProvCaRe. Conclusão: A melhoria na disponibilidade de metadados de proveniência é crucial para apoiar a reprodutibilidade científica e o uso da plataforma ProvCaRe, que utiliza a especificação W3C PROV para modelar, extrair e analisar metadados de proveniência de estudos biomédicos.
5	Autor (es)/Ano: Rahman <i>et al.</i> (2020). Título: Secure and Provenance Enhanced Internet of Health Things Framework: A Blockchain Managed Federated Learning Approach. Objetivo: Desenvolver um framework seguro para Internet of Healthcare Things (IoHT), melhorando a proveniência de dados e utiliza aprendizado federado gerenciado por blockchain para garantir a privacidade, segurança e integridade dos dados. Conclusão: O framework proposto gerencia eficazmente a proveniência de dados e melhora a segurança e privacidade dos dados do IoHT, combinando aprendizado federado e tecnologia blockchain, demonstrando grande potencial para uma adoção mais ampla em aplicações de gestão de saúde.
6	Autor (es)/Ano: Gierend <i>et al.</i> (2021). Título: Approaches and Criteria for Provenance in Biomedical Data Sets and Workflows: Protocol for a Scoping Review. Objetivo: investigar as abordagens e critérios existentes para rastreamento de proveniência, além de identificar lacunas de conhecimento no domínio de dados científicos biomédicos sensíveis. Conclusão: neste estudo são revistas as abordagens atuais e identificados os desafios no rastreamento da proveniência em ciências biomédicas, promovendo transparência, rastreabilidade, reprodutibilidade e qualidade nos fluxos de trabalho e pesquisas clínicas.
7	Autor (es)/Ano: Johns <i>et al.</i> (2023). Título: Data Provenance in Biomedical Research: Scoping Review Objetivo: fornecer uma visão estruturada das tecnologias de proveniência de dados na pesquisa biomédica e promover futuras pesquisas para adoção mais ampla dessas tecnologias. Conclusão: pouco progresso foi feito na área biomédica em tecnologias de proveniência de dados, crucial para melhorar a reprodutibilidade. Há falta de soluções claras para capturar dados de proveniência e limitações na análise e visualização desses dados também foram fatores observados.

XXIV Encontro Nacional de Pesquisa em Ciência da Informação – XXIV ENANCIB
Vitória-ES – 04 a 08 de novembro de 2024

Base de dados: MEDLINE/PUBMED	
8	Autor (es)/Ano: Sembay <i>et al.</i> (2023). Título: Provenance Data Management in Health Information Systems: A Systematic Literature Review. Objetivo: realizar uma revisão sistemática da literatura sobre o gerenciamento de dados de proveniência em sistemas de informação de saúde identificando abordagens, tecnologias e metodologias utilizadas para capturar, armazenar e utilizar dados de proveniência nesse contexto específico. Conclusão: o gerenciamento de dados de proveniência em SIS enfrenta desafios importantes apesar do progresso em tecnologias como modelos específicos de dados e captura automatizada. Lacunas incluem padronização, integração com sistemas existentes e eficácia na utilização desses dados para melhorar a qualidade e transparência das informações de saúde.
9	Autor (es)/Ano: Ahmed <i>et al.</i> (2023). Título: Data Provenance in Healthcare: Approaches, Challenges, and Future Directions. Objetivo: explorar técnicas atuais, desafios e futuras direções da proveniência de dados na saúde. Conclusão: ressalta os desafios significativos da proveniência de dados na saúde, incluindo falta de padronização na captura e representação dos dados, e problemas de interoperabilidade entre SI. Apesar de avanços como modelos específicos de dados e sistemas automatizados de captura, é crucial progredir em segurança, privacidade e integração de dados de fontes diversas.
10	Autor (es)/Ano: Black <i>et al.</i> (2023). Título: Understanding data provenance when using electronic medical records for research: Lessons learned from the Deliver Primary Healthcare Information (DELPHI) database. Objetivo: explorar a proveniência de dados em registros médicos eletrônicos para pesquisa, analisando a coleta, armazenamento e uso no banco de dados canadense Deliver Primary Healthcare Information (DELPHI), identificando ainda como a proveniência garante qualidade, transparência e confiabilidade na pesquisa médica. Conclusão: a compreensão da proveniência de dados em pesquisa com registros médicos eletrônicos é crucial, destacando desafios na captura precisa e documentação dos dados. Diretrizes claras e boas práticas, além de educar os pesquisadores sobre a importância da proveniência fortalece a confiança nos resultados e facilita o uso ético dos registros médicos eletrônicos na saúde pública e medicina.
11	Autor (es)/Ano: Gierend <i>et al.</i> (2023). Título: Traceable Research Data Sharing in a German Medical Data Integration Center With FAIR (Findability, Accessibility, Interoperability, and Reusability)-Geared Provenance Implementation: Proof-of-Concept Study. Objetivo: demonstrar um estudo sobre a implementação de proveniência orientada pelos princípios FAIR em um centro alemão de integração de dados médicos, estabelecendo métodos e tecnologias para compartilhamento de dados de pesquisa melhorando assim a transparência e a confiabilidade na gestão de dados médicos. Conclusão: tecnologias e métodos apropriados podem melhorar a transparência, rastreabilidade e reutilização dos dados médicos de pesquisa focados em padrões claros e diretrizes para a proveniência de dados, facilitando a colaboração entre instituições e aumentando a confiança nos resultados da pesquisa.

Fonte: Elaborado pelos autores (2024).

O Quadro 2 evidencia o contexto atual da proveniência de dados em saúde, conforme encontrado nos estudos selecionados na base de dados *MEDLINE/PubMed*. Essas evidências foram amplamente discutidas na subseção 5.1, onde as principais contribuições e desafios foram sintetizados e organizados com base na análise de conteúdo de Bardin (2004). Posteriormente, utilizando-se ainda dessa metodologia, foram consolidados os aspectos mais recorrentes descritos nos estudos, proporcionando uma visão clara e concisa dos principais pontos destacados na literatura.

5.1 Principais contribuições e desafios da proveniência de dados em saúde

Os estudos recentes apresentados no Quadro 2 sobre proveniência de dados na área da saúde, revelam um conjunto significativo de contribuições comuns e desafios compartilhados entre os diferentes autores investigados.

Assim, utilizando a metodologia de análise de conteúdo de Bardin (2004), as Figuras 1 e 2 sintetizam as principais contribuições e desafios relacionados ao uso da proveniência de dados em saúde, juntamente com os respectivos autores que os mencionam em suas pesquisas. Esses dados oferecem insights valiosos para fortalecer a eficiência e a confiabilidade dos SIS, promovendo avanços contínuos na gestão de dados de proveniência na área da saúde.

Figura 1 – Principais contribuições



Fonte: Elaborado pelos autores (2024).

Em relação às principais contribuições da proveniência de dados em saúde destacadas na Figura 1:

- a transparência e rastreabilidade dos dados aumentam a confiança e eficiência na pesquisa médica;

- a documentação detalhada da proveniência melhora a qualidade dos dados e garante a confiabilidade das descobertas;
- a gestão eficaz da proveniência facilita a integração de dados e fornece insights valiosos no suporte à pesquisa médica;
- a replicabilidade assegura resultados verificáveis e robustos;
- a proveniência também garante a auditoria e o suporte à tomada de decisões, o que facilita decisões clínicas fundamentadas, além de assegurar conformidade com padrões de pesquisa, essenciais para avanços científicos contínuos.

Os estudos revisados ainda destacam desafios significativos em comum, conforme apresentados na Figura 2.

Figura 2 – Principais desafios



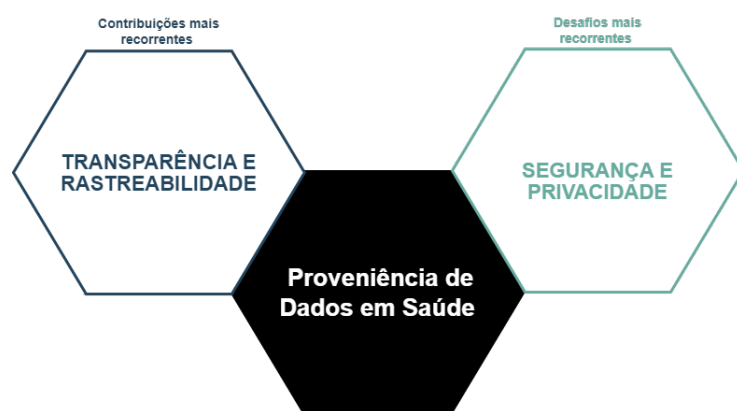
Fonte: Elaborado pelos autores (2024).

Em relação aos principais desafios da proveniência de dados em saúde destacados na Figura 2:

- a complexidade de implementação dos sistemas de proveniência inclui também a complexidade de soluções que mantem a eficiência operacional e a segurança dos dados sensíveis;
- a segurança e privacidade dos dados são cruciais para cumprir regulamentações e manter a confiança;
- no que compete a integração e interoperabilidade de sistemas de proveniência, esses apresentam dificuldades na troca de dados entre sistemas heterogêneos, necessitando harmonização com outras infraestruturas de saúde;
- a escassez de tecnologias adequadas para implementação de sistemas de proveniência é um obstáculo significativo, assim como os custos operacionais (overheads) associados;
- a adoção de sistemas de proveniência é vital para assegurar transparência e confiabilidade nos dados de saúde usados em pesquisa e prática clínica.

Nesse íterim, a Figura 3 apresenta as principais contribuições e desafios mais recorrentes identificados nos estudos analisados, utilizando a mesma análise de conteúdo descrita por Bardin (2004), previamente aplicada nas Figuras 1 e 2. Assim, a Figura 3 oferece um panorama abrangente sobre a proveniência de dados em saúde, conforme relatado nos estudos selecionados.

Figura 3 – Contribuições e desafios mais recorrentes



Fonte: Elaborado pelos autores (2024).

Conforme apresentado na Figura 3, a análise de 11 estudos sobre a proveniência de dados em saúde revelou que as principais contribuições são a transparência e a

rastreabilidade, enfatizadas em oito estudos. Os desafios mais comuns incluem segurança e privacidade, abordados em todos os estudos. Isso destaca a importância desses aspectos na implementação de sistemas de proveniência, visando garantir precisão e qualidade na pesquisa e prática clínica. Os autores concordam que a proveniência de dados é essencial para a pesquisa em saúde, embora inovações sejam necessárias para superar os desafios. O estudo, embora limitado às bases de dados *MEDLINE/PubMed* e *LILACS*, oferece insights valiosos para a Ciência da Informação e práticas de proveniência em saúde.

6 CONSIDERAÇÕES FINAIS

A proveniência de dados é crucial na saúde, rastreando a origem e transformação dos dados para fortalecer a gestão de recursos e promover transparência em seus processos. Além disso, a pesquisa nesse campo desempenha um papel interdisciplinar na Ciência da Informação, fornecendo insights sobre fluxos informacionais essenciais para a Ciência da Computação. Isso facilita a implementação e melhoria de SIS que asseguram a proveniência de dados de forma ética e eficaz, impulsionando inovações e avanços contínuos na pesquisa em saúde.

Esta pesquisa atinge seu objetivo ao destacar as principais contribuições e desafios da proveniência de dados em saúde nos últimos dez anos. As contribuições principais são a transparência e rastreabilidade dos dados, que aumentam a confiança e eficiência nos processos de saúde. Os desafios incluem a segurança e privacidade dos dados sensíveis, necessitando de medidas robustas para proteção e conformidade com regulamentações. Essas descobertas beneficiam pesquisadores e profissionais de saúde, melhorando práticas clínicas e reforçando a confiança e transparência nos processos de saúde.

Para avançar no uso da proveniência de dados na área da saúde, é crucial explorar métodos inovadores, técnicas robustas, modelos adaptáveis, metodologias eficientes e tecnologias emergentes, além de investigações na área de inteligência artificial e blockchain que podem contribuir para aumentar a confiança nos SIS.

REFERÊNCIAS

AHMED, M.; DAR, A. R.; HELFERT, M.; KHAN, A.; KIM, J. Data provenance in healthcare: approaches, challenges, and future directions. **Sensors (Basel)**, Basel, v. 23, n. 14, p. 6495,

jul. 2023. DOI: <http://doi.org/10.3390/s23146495>. Disponível em: <https://www.mdpi.com/1424-8220/23/14/6495>. Acesso em: 10 jul. 2024.

ALTEROVITZ, G.; DEAN, D.; GOBLE, C.; CRUSOE, M. R.; SOILAND-REYES, S.; BELL, A.; HAYES, A.; SURESH, A.; PURKAYASTHA, A.; KING, C. H.; TAYLOR, D.; JOHANSON, E.; THOMPSON, E. E.; DONALDSON, E.; MORIZONO, H.; TSANG, H.; VORA, J. K.; GOECKS, J.; YAO, J.; ALMEIDA, J. S.; KEENEY, J.; ADDEPALLI, K.; KRAMPIS, K.; SMITH, K. M.; GUO, L.; WALDERHAUG, M.; SCHITO, M.; EZEWUDO, M.; GUIMERA, N.; WALSH, P.; KAHSAÝ, R.; GOTTIPATI, S.; RODWELL, T. C.; BLOOM, T.; LAI, Y.; SIMONYAN, V.; MAZUMDER, R. Enabling precision medicine via standard communication of HTS provenance, analysis, and results. **PLoS Biology**, San Francisco, v. 16, n. 12, e3000099, p. 1-14, dez. 2018. DOI: <http://doi.org/10.1371/journal.pbio.3000099>. Disponível em: <https://journals.plos.org/plosbiology/article/file?id=10.1371/journal.pbio.3000099&type=printable>. Acesso em: 10 jul. 2024.

ARAÚJO, N. C. de; MOTA, F. R. L. Prontuário de paciente: questões éticas. **Informação e Pauta**, Fortaleza, v. 5, n. esp., p. 52-67, mar. 2020. DOI: <http://doi.org/10.36517/2525-3468.ip.v5iespecial1.2020.43512.52-67>. Disponível em: <http://www.periodicos.ufc.br/informacaoempauta/article/view/43512/100125>. Acesso em: 10 jul. 2024.

BARDIN, L. **Análise de conteúdo**. Lisboa: Edições 70, 2004.

BLACK, J. E.; TERRY, A. L.; CEJIC, S.; FREEMAN, T.; LIZOTTE, D.; MCKAY, S.; SPEECHLEY, M.; RYAN, B. Understanding data provenance when using electronic medical records for research: lessons learned from the Deliver Primary Healthcare Information (DELPHI) database. **International Journal Population Data Science**, Wales, v. 8, n. 5, p. 1-8, set. 2023. DOI: 10.23889/ijpds.v8i5.2177. Disponível em: <https://ijpds.org/article/view/2177/4930>. Acesso em: 10 jul. 2024.

BUNEMAN, P.; KHANNA, S. E; CHIEW, W. Why and where: a characterization of data provenance. In: VAN DEN BUSSCHE, J.; VIANU, V. (ed.). **Database theory: ICDT 2001**. Berlin; Heidelberg: Springer, 2001. p. 316-330. (Lecture Notes in Computer Science, v. 1973). Disponível em: <https://pdfs.semanticscholar.org/b647/2ee11749ef70713bfb0e322a9ec27523ed88.pdf>. Acesso em: 2 maio 2019.

CUNHA, M. B. da; CAVALCANTI, C. R. de O. **Dicionário de biblioteconomia e arquivologia**. 3. ed. rev. e ampl. Brasília: Briquet de Lemos, 2008. Disponível em: <http://www.rlbea.unb.br/jspui/handle/10482/34113>. Acesso em: 24 jun. 2024.

FREIRE, J.; KOOP, D.; SANTOS, E.; SILVA, C. T. Provenance for computational tasks: a survey. **Journal Computing in Science and Engineering**, [s. l.], v. 10, n. 3, p. 11-21, 2008. DOI: <http://doi.org/10.1109/MCSE.2008.79>. Disponível em: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.147.3801&rep=rep1&type=pdf>. Acesso em: 20 jun. 2024.

GALVÃO, M. C. B.; RICARTE, I. L. M. O prontuário eletrônico do paciente em escala nacional. **InCID: revista de Ciência da Informação e Documentação**, Ribeirão Preto, v. 10, n. 1, p. 244-264, 2019. DOI: <http://doi.org/10.11606/issn.2178-2075.v10i1p244-264>. Disponível em: <https://www.revistas.usp.br/incid/article/view/153003/153415>. Acesso em: 6 jun. 2024.

GIEREND, K.; KRÜGER, F.; WALTEMATH, D.; FÜNFELD, M.; GANSLANDT, T.; ZELEKE, A. A. Approaches and criteria for provenance in biomedical data sets and workflows: Protocol for a scoping review. **JMIR Research Protocols**, Toronto, v. 10, n. 11, e31750, nov. 2021. DOI: <http://doi.org/10.2196/31750>. Disponível em: <https://www.researchprotocols.org/2021/11/e31750>. Acesso em: 10 jul. 2024.

GIEREND, K.; WALTEMATH, D.; GANSLANDT, T.; SIEGEL, F. Traceable research data sharing in a German medical data integration center with FAIR (Findability, Accessibility, Interoperability, and Reusability)-geared provenance implementation: proof-of-concept study. **JMIR Formative Research**, Toronto, v. 7, e50027, 7 dez. 2023. DOI: <http://doi.org/10.2196/50027>. Disponível em: <https://formative.jmir.org/2023/1/e50027>. Acesso em: 10 jul. 2024.

GROTH, P; MOREAU, L. **PROV-Overview**: an overview of the PROV family of documents. [Wakefield]: W3C, 2013. Disponível em: <https://www.w3.org/TR/prov-overview/>. Acesso em: 20 jun. 2024.

HOLUB, P.; KOHLMAYER, F.; PRASSER, F.; MAYRHOFER, M. T.; SCHLÜNDER, I.; MARTIN, G. M.; CASATI, S.; KOUMAKIS, L.; WUTTE, A.; KOZERA, Ł.; STRAPAGIEL, D.; ANTON, G.; ZANETTI, G.; SEZERMAN, O. U.; MENDY, M.; VALÍK, D.; LAVITRANO, M.; DAGHER, G.; ZATLOUKAL, K.; VAN OMMEN, G. B.; LITTON, J. E. Enhancing reuse of data and biological material in medical research: from FAIR to FAIR-Health. **Biopreservation and Biobanking**, New Rochelle, v. 16, n. 2, p. 97-105, abr. 2018. DOI: <http://doi.org/10.1089/bio.2017.0110>. Disponível em: <https://www.liebertpub.com/doi/10.1089/bio.2017.0110>. Acesso em: 10 jul. 2024.

JOHNS, M.; MEURERS, T.; WIRTH, F. N.; HABER, A. C.; MÜLLER, A.; HALILOVIC, M.; BALZER, F.; PRASSER, F. Data provenance in biomedical research: scoping review. **Journal of medical Internet research**, Toronto, v. 25, e42289, mar. 2023. DOI: <http://doi.org/10.2196/42289>. Disponível em: <https://www.jmir.org/2023/1/e42289>. Acesso em: 10 jul. 2024.

RAHMAN, M. A.; HOSSAIN, M. S.; ISLAM, M. S.; ALRAJEH, N. A.; MUHAMMAD, G. Secure and provenance enhanced internet of health things framework: a blockchain managed federated learning approach. **IEEE Access**, [s. l.], v. 8, p. 205071-205087, nov. 2020. DOI: <http://doi.org/10.1109/ACCESS.2020.3037474>. Disponível em: <https://ieeexplore.ieee.org/document/9256355>. Acesso em: 10 jul. 2024.

RIBEIRO, F. Medicina e Ciência da Informação: uma abordagem integradora e interdisciplinar. In: DUARTE, Z.; FARIAS, L. **A medicina na era da informação**. Salvador: EDUFBA, 2009. p. 111–125.

SAHOO, S. S.; VALDEZ, J.; KIM, M.; RUESCHMAN, M.; REDLINE, S. ProvCaRe: characterizing scientific reproducibility of biomedical research studies using semantic provenance metadata. **International Journal of Medical Informatics**, Amsterdam, v. 121, p. 10-18, Jan. 2019. DOI: <http://doi.org/10.1016/j.ijmedinf.2018.10.009>. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S1386505618302697?via%3Dihub>. Acesso em: 10 jul. 2024.

SEMBAY, M. J.; MACEDO, D. D. J. de; JÚNIOR, L. P.; BRAGA, R. M. M.; SARASA-CABEZUELO, A. Provenance data management in health information systems: a systematic literature review.

Journal Personalized Medicine, Basel, v. 13, n. 6, p. 991, jun. 2023. DOI: <http://doi.org/10.3390/jpm13060991>. Disponível em: <https://www.mdpi.com/2075-4426/13/6/991>. Acesso em: 10 jul. 2024.

SEMBAY, M. J.; MACEDO, D. J.; DUTRA, M. L. A Method for collecting provenance data: a case study in a Brazilian hemotherapy center. *In: INTERNATIONAL CONFERENCE ON DATA AND INFORMATION IN ONLINE ENVIRONMENTS, DIONE 2020, 1., 2020, Florianópolis. Proceedings [...]*. [S. l.]: Springer International Publishing, 2020a. p. 89-102. DOI: http://doi.org/10.1007/978-3-030-50072-6_8. Disponível em: https://link.springer.com/chapter/10.1007/978-3-030-50072-6_8#citeas. Acesso em: 10 jul. 2024.

SEMBAY, M. J.; MACEDO, D. D. J. de; DUTRA, M. L. A proposed approach for provenance data gathering. **Mobile Networks & Applications**, New York, v. 26, p. 1–13, 2 Sep. 2020b. DOI: <http://doi.org/10.1007/s11036-020-01648-7>. Disponível em: <https://link.springer.com/article/10.1007/s11036-020-01648-7>. Acesso em: 10 jul. 2024.

SEMBAY, M. J.; MACEDO, D. D. J. de; MARQUEZ FILHO, A. A. G. Proveniência de dados em Sistemas de Informação em Saúde. *In: ENCONTRO NACIONAL DE PESQUISA E PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO, 22., Porto Alegre. Anais [...]*. Porto Alegre: ANCIB, 2022. p. 1-11. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/200577>. Acesso em: 20 jun. 2024

SEMBAY, M. J.; MACEDO, D. D. J. de; MARQUEZ FILHO, A. A. G. Dados de Proveniência Gerenciáveis em Sistemas de Informação em Saúde. **Tendências da Pesquisa Brasileira em Ciência da Informação**, Rio de Janeiro, v. 16, p. 1-26, 2023. Disponível em: <https://ancib.org/revistas/index.php/tpbci/article/view/637>. Acesso em: 10 jul. 2024.

SEMBAY, M. J. **PROV-Health**: método para gerenciamento de dados de proveniência em sistemas de informação em saúde. 2023. 311 f. Tese (Doutorado em Ciência da Informação) – Programa de Pós-Graduação em Ciência da Informação, Centro de Ciências da Educação, Universidade Federal de Santa Catarina, Florianópolis, 2023. Disponível em: <https://repositorio.ufsc.br/handle/123456789/251582>. Acesso em: 10 jun. 2024.

SEMBAY, M. J.; MACEDO, D. D. J. de; MARQUEZ FILHO, A. A. G. Gerenciamento de dados de proveniência em telemedicina e telessaúde: possíveis abordagens e perspectivas à luz da Ciência da Informação em Saúde no Brasil. **Asklepion: informação em saúde**, Rio de Janeiro, v. 3, n. 1, p. e–89, 2024. DOI: 10.21728/asklepion.2024v3n1e-89. Disponível em: <https://asklepionrevista.info/asklepion/article/view/89>. Acesso em: 25 jun. 2024.

SIMMHAN, Y. L.; PLALE, B.; GANNON, D. A survey of data provenance techniques. *In: TECHNICAL Report TR-618*. Indiana: Computer Science Department, Indiana University, 2005. p. 1-25. Disponível em: <http://ftp.extreme.indiana.edu/pub/techreports/TR618.pdf>. Acesso em: 24 maio 2018.

TORRACO, R. J. Writing integrative literature reviews: using the past and present to explore the future. **Human Resource Development Review**, Thousand Oaks, v. 15, n. 4, p. 404-428, 2016. DOI: <http://doi.org/10.1177/1534484316671606>. Disponível em: <https://journals.sagepub.com/doi/abs/10.1177/1534484316671606>. Acesso em: 11 jun. 2024.

WHITTEMORE, R.; KNAFL, K. The integrative review: updated methodology. **Journal of advanced nursing**, Portland, v. 52, n. 5, p. 546-553, 2005. DOI: <http://doi.org/10.1111/j.1365-2648.2005.03621.x>. Disponível em: <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-2648.2005.03621.x>. Acesso em: 14 jun. 2024.

XU, S.; ROGERS, T.; FAIRWEATHER, E.; GLENN, A.; CURRAN, J.; CURCIN, V. Application of data provenance in healthcare analytics software: information visualisation of user activities. **AMIA joint summits on translational science**, Bethesda, v. 2017, p. 263-272, maio 2018. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5977650/>. Acesso em: 24 jun. 2024.