

DADOS DE PROVENIÊNCIA GERENCIÁVEIS EM SISTEMAS DE INFORMAÇÃO EM SAÚDE¹

MANAGEABLE PROVENANCE DATA IN HEALTH INFORMATION SYSTEMS

Márcio José Sembay²

Douglas Dyllon Jeronimo de Macedo³

Alexandre Augusto Gimenes Marquez Filho⁴

Resumo: Para que os dados de proveniência possam ser gerenciados em Sistemas de Informação em Saúde, este estudo tem como objetivo propor um método genérico que possa contribuir para o gerenciamento de dados de proveniência nesses sistemas. Para isso, foi necessário estender o modelo de proveniência de dados W3C PROV, o qual compactua com um dos principais padrões de interoperabilidade de dados em saúde, o *Health Level 7* e *Fast Healthcare Interoperability Resources*. Assim, este estudo incentiva o uso de estratégias computacionais que possam ser adaptadas para diferentes Sistemas de Informação em Saúde. A metodologia de pesquisa deste estudo é aplicada, com abordagem quantitativa e qualitativa. Com a proposta do método deste estudo, destaca-se como resultados, possíveis contribuições significativas com relação aos processos de gerenciamento e rastreamento dos dados de proveniência em diferentes Sistemas de Informação em Saúde, considerado além da manutenção da segurança desses dados, a relação custo-benefício obtida para a instituição de saúde.

Palavras-Chave: Proveniência de Dados. Sistemas de Informação em Saúde. Informação em Saúde. HL7 FHIR. W3C PROV.

Abstract: *In order for provenance data to be managed in Health Information Systems, this study aims to propose a generic method that can contribute to the management of provenance data in these systems. To do so, it was necessary to extend the W3C PROV data provenance model, which compacts with one of the main standards for healthcare data interoperability, Health Level 7 and Fast Healthcare Interoperability Resources. Thus, this study encourages the use of computational strategies that can be adapted for different Health Information Systems. The research methodology of this study is applied, with quantitative and qualitative approach. With the proposal of the method of this study, the results highlight possible significant contributions regarding the processes of*

¹ Texto ampliado a partir do artigo submetido, avaliado, aprovado, apresentado e premiado no XXII ENANCIB XXII.

² Doutor em Ciência da Informação (PGCIN/UFSC). Docente do Centro Universitário UNIFACVEST. E-mail: marcio.sembay@posgrad.ufsc.br. Orcid: <https://orcid.org/0000-0002-7648-8861>.

³ Doutor em Engenharia e Gestão do Conhecimento (UFSC). Docente da Universidade Federal de Santa Catarina (UFSC). E-mail: douglas.macedo@ufsc.br. Orcid: <https://orcid.org/0000-0002-3237-4168>.

⁴ Mestre em Ciência da Informação (PGCIN/UFSC). Analista de Sistemas do Sistema Integrado de Telemedicina e Telessaúde da Universidade Federal de Santa Catarina (STT/UFSC). E-mail: alexandre.agmf@gmail.com. Orcid: <https://orcid.org/0000-0002-2656-9479>.

management and tracking of provenance data in different Health Information Systems, considering in addition to maintaining the security of these data, the cost-benefit ratio obtained for the health institution.

Keywords: *Data Provenance. Health Information Systems. Health Information.HL7 FHIR. W3C PROV.*

1 INTRODUÇÃO

Os Sistemas de Informação (SI) e as Tecnologias de Informação e Comunicação (TIC) fornecem um novo e inovador mecanismo de comunicação que estão mudando as formas como os negócios acontecem em todos os setores. Nesse sentido, os avanços das TIC têm permitido novas formas de prestação de serviços na área da saúde, possibilitando fornecer alta qualidade nos serviços prestados, promovendo melhores resultados por meio do acesso às informações de saúde (Ryu, 2012). Assim, a importância das diversas ferramentas tecnológicas de TIC voltadas para o ser humano levou à construção dos Sistemas de Informação em Saúde (SIS), definido por Haux (2006, p. 270, tradução nossa) como: “[...] sistemas de processamento de dados, informação e conhecimento em ambientes de cuidados de saúde”. Um exemplo de estrutura que utiliza os SIS de forma integrada, é a telemedicina. A telemedicina utiliza uma gama de equipamentos médicos e sistemas integrados possibilitando gerar diagnósticos e tratamentos independente da distância geográfica (Håkansson; Gavelin, 2000).

Nessa perspectiva, é importante ressaltar que os SIS geram e armazenam grandes volumes de dados que posteriormente necessitarão passar por análises em suas fontes de dados originais para possíveis tomadas de decisão. Compactuando isso com a proveniência de dados em SIS, torna-se de fato, um instrumento tecnológico essencial para contribuir na gestão dos dados de saúde nesses sistemas. Para Davidson e Freire (2008), a proveniência de dados diz respeito à origem dos dados, e pode também estar relacionada à auditoria, triagem e linhagem dos dados.

Nesse sentido, para que o processo de proveniência de dados possa ocorrer nos SIS, protocolos internacionais de interoperabilidade entre SIS, como é o caso do *Health Level 7 (HL7)*⁵ e *Fast Healthcare Interoperability Resources (FHIR)*⁶ baseado no modelo de proveniência da *World Wide Web Consortium (W3C)* mais conhecido como *W3C PROV*⁷, são essenciais e podem contribuir de forma significativa para o gerenciamento dos dados de proveniência em SIS.

Dessa forma, nasce a problemática deste estudo, propor um método com base na extensão do modelo de proveniência de dados *W3C PROV*, em favor de oportunizar melhores ações para gerenciar dados de proveniência em SIS. Para tanto, estabeleceu-se como objetivo: propor um método genérico que seja capaz de contribuir para o gerenciamento de dados de proveniência em SIS, baseado na extensão do modelo *W3C PROV*. Com base nisso, este estudo busca responder a seguinte questão de pesquisa: como um método de gerenciamento de dados de proveniência em SIS pode contribuir com a melhoria dos fluxos informacionais relacionados ao processo de identificação, rastreamento e reaproveitamento dos dados em repositórios descentralizados no domínio de saúde?

No que se refere aos aportes metodológicos, trata-se de uma pesquisa de natureza aplicada (Gerhardt; Silveira, 2009), cuja aplicação compete à proposta de um método genérico de gerenciamento de dados de proveniência em SIS baseada na extensão do modelo de proveniência de dados *W3C PROV*, prevendo uma estrutura tecnológica genérica que possa ser utilizada em diferentes SIS. Com relação à abordagem, classifica-se este estudo como quantitativo e qualitativo (Knechtel, 2014; Reneker, 1993). Quantitativo, devido aos dados de proveniência que podem ser armazenados e gerenciados em diferentes SIS descentralizados, e, qualitativo, por

⁵ Disponível em: <https://www.hl7.org/>. Acesso em: 8 jun. 2022.

⁶ Disponível em: <http://hl7.org/fhir/>. Acesso em: 8 jun. 2022.

⁷ Disponível em: <https://www.w3.org/TR/prov-primer/>. Acesso em: 8 jun. 2022.

motivo da proveniência de dados possuir características que qualificam os dados, sendo um dos principais fatores significativos que contribuem para os diferentes sistemas e cenários de saúde.

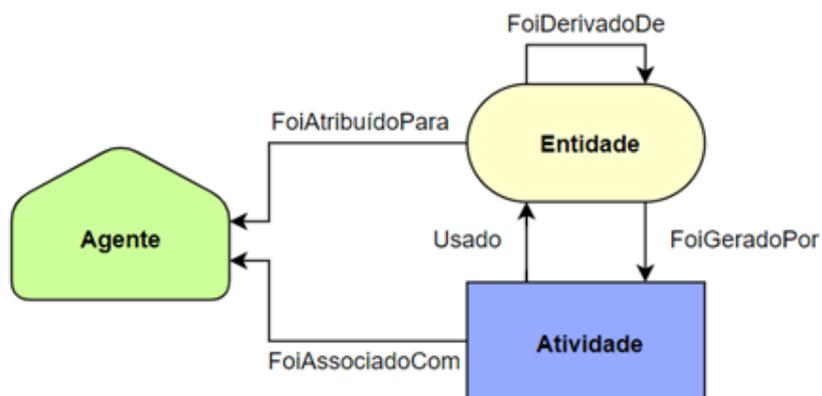
2 PROVENIÊNCIA DE DADOS

Para o contexto deste estudo, a proveniência de dados, pode ser definida como linhagem, genealogia ou pedigree, consiste em metadados que descrevem as origens de um item de dado e pode ser vista como uma documentação complementar de um determinado dado que contém a descrição de “como”, “quando”, “onde” e “por que” ele foi obtido e “quem” o obteve. refere-se a qualquer quantidade de informação, compreendendo todos os elementos e suas relações, que contribuem para a existência de um determinado dado (Buneman; Khanna; Tan, 2001; Pérez; Rubio; Sáenz-Adán, 2018). Assim, o gerenciamento de proveniência ganhou atenção significativa da comunidade de pesquisa e da indústria nas últimas décadas (Pérez; Rubio; Sáenz-Adán, 2018). Nesse sentido, a proveniência de dados pode ser considerada um requisito importante para estabelecer confiabilidade e prover segurança em SI, pois concentra-se na identificação das fontes de dados e nas transformações pelas quais passou ao longo do tempo (Freund; Sembay; Macedo, 2019).

Assim, a proveniência de dados está relacionada a diferentes cenários de aplicação. Conforme Simmhan, Plale e Gannon (2005), os vários cenários demandam muitos esforços de pesquisa que contribuem para que sistemas de proveniência possam ser construídos para operar de diversas formas, com diferentes características e operações. Para Sembay, Macedo e Dutra (2020) um dos cenários de aplicação da proveniência de dados que tem se destacado pelo grande volume de dados gerados e armazenados, é o cenário de saúde, sendo o foco deste estudo.

Dessa forma, a proveniência de dados no contexto da saúde vive um cenário crescente de pesquisas baseadas nos mais variados tipos de experimentos científicos, sendo que as tecnologias aplicadas nesta área vêm obtendo resultados expressivos (Sembay; Macedo; Dutra, 2021). Nesse contexto, a proveniência de dados apresenta vários modelos na literatura, mas para este estudo em específico, a Figura 1 apresenta o modelo de proveniência de dados W3C PROV traduzido de Gil e Miles (2013).

Figura 1: Modelo de proveniência de dados W3C PROV



Fonte: adaptada de Gil e Miles (2013, tradução nossa).

O modelo de proveniência de dados W3C PROV, apresentado na Figura 1, surgiu em 2013, destacando a proveniência como uma preocupação importante em várias aplicações (por exemplo, estabelecer responsabilidade, reprodutibilidade e confiabilidade da informação) (Groth; Moreau, 2013).

O W3C PROV apresenta três elementos: a) Entity (entidade): que pode ser algo físico, digital ou conceitual, real ou imaginário; b) Activity (atividade): algo que ocorre durante um período de tempo e age sobre ou com entidades; podendo incluir, consumir, processar, transformar, modificar, realocar, usar ou gerar entidades; e, c) Agent (agente): sendo alguma forma de responsabilidade por uma atividade que está

ocorrendo, pela existência de uma entidade ou pela atividade de outro agente (Gil; Miles, 2013).

Para Gil e Miles (2013), a estrutura principal do modelo W3C PROV depende da definição das entidades, atividades e agentes que estão envolvidos na produção de um dado ou coisa e em como eles estão relacionados, podendo haver tipos diferentes de relações como demonstrado na Figura 1, os quais são: *wasGeneratedBy* (foiGeradoPor), *wasAssociatedBy* (foiAssociadoCom), *wasAttributedTo* (foiAtribuídoPara), *wasDerivedFrom* (foiDerivadoDe) e *used* (usado).

A seguir, nos Quadros 1 e 2 são explicitados os elementos e as relações do modelo W3C PROV para que se tenha um entendimento de suas ligações.

Quadro 1: Elementos do modelo PROV

Elementos	Descrição
Entidade	É utilizada para representar qualquer tipo de entidade, como por exemplo, física, conceitual ou digital.
Agente	Refere-se aos agentes que podem influenciar na proveniência representada, como por exemplo, um software que realiza uma ação, ou um usuário.
Atividade	Relaciona-se às atividades, como por exemplo, processos ou ações.

Fonte: adaptado e traduzido de Gil e Miles (2013).

Quadro 2: Relações do modelo PROV

Elementos	Descrição
<i>foiAtribuídoPara</i>	Relaciona um agente a uma entidade a qual ele foi atribuído
<i>foiAssociadoCom</i>	Relaciona um agente a uma atividade da qual ele tem responsabilidade sobre
<i>usado</i>	Relaciona uma entidade a uma atividade que a usou
<i>foiGeradoPor</i>	Relaciona uma atividade a uma entidade pela qual foi gerada
<i>foiDerivadoDe</i>	Relaciona uma entidade a outra pela qual a mesma foi gerada

Fonte: adaptado e traduzido de Gil e Miles (2013).

Portanto, o W3C PROV é um modelo de dados de proveniência que apoia o intercâmbio interoperável de proveniência em ambientes heterogêneos, como por exemplo a *web* (Gil; Miles, 2013). Assim, para o contexto desta pesquisa, o modelo W3C PROV pode apoiar diferentes ambientes de saúde por meio de sua extensão.

3 SISTEMAS DE INFORMAÇÃO EM SAÚDE

Os SIS são cada vez mais usados em todo o mundo para melhorar a eficiência hospitalar, a qualidade do atendimento e a satisfação do paciente. Esses sistemas, introduzidos na década de 1960, são utilizados em quase todas as nações, dependendo do desenvolvimento tecnológico do país (Ben-Assuli, 2015; Cesnik; Kidd, 2010). Os SIS também podem integrar requisitos técnicos e profissionais necessários para a coleta, registro, processamento, análise, apresentação e difusão de informações fundamentais para aperfeiçoar a eficácia e a eficiência dos serviços de saúde (Haux, 2006; Melo; Mota; Araújo, 2022).

Nesse sentido, ressalta-se ainda que, os SIS estão sendo implementados em todos os aspectos da saúde, desde a administração até os sistemas de apoio às decisões clínicas. A geração e armazenamento de grandes volumes de dados em seus repositórios descentralizados tornam esses processos desafiadores no que diz respeito ao gerenciamento desses dados (Sembay; Macedo, 2022). A telemedicina é um exemplo que se utiliza de diversos SIS integrados em sua estrutura, gerando grandes volumes de dados que requerem técnicas e ferramentas computacionais avançadas para o gerenciamento desses dados. A telemedicina é caracterizada como ciência aberta e em constante evolução, incorpora novos avanços em tecnologia se adaptando às mudanças e necessidades de saúde e contextos das sociedades (Ryu, 2012).

Já em um contexto brasileiro, agregada a esses avanços, a expansão do acesso às tecnologias de informação e informática, que facilitou e ampliou o uso de dados gerados pelos SIS, e o baixo custo e a disponibilização de um grande volume de dados com reduzido tempo entre ocorrência do evento e seu registro, expandiu o interesse pela utilização dos sistemas de informação em saúde de base nacional. Nessa lógica, os bancos de dados dos SIS, disponibilizados na internet pelo Departamento de

Informática do Sistema Único de Saúde (SUS) (Datasus) do Ministério da Saúde (MS), tornaram-se fontes importantes e muito utilizadas para a produção do conhecimento sobre a situação de saúde, além de aplicadas no planejamento, organização, operação e avaliação de ações, serviços, programas e políticas (Victoria *et al.*, 2011).

Já no que compete sobre os diferentes tipos de SIS, são aqui descritos os que mais se destacam em relação a este estudo no contexto das aplicações da proveniência de dados em âmbito nacional e mundial, quais sejam:

- Registro Eletrônico de Saúde/Electronic Health Record (EHR): descreve o conceito de uma coleta abrangente, interinstitucional e longitudinal de dados de saúde do paciente. Isso, portanto, inclui dados que não são apenas particularmente relevantes para a avaliação médica do tratamento de um sujeito, mas também para a saúde do sujeito em geral (Hoerbst; Ammenwerth, 2010).
- Sistema de Informação Hospitalar/Hospital Information System (HIS): O HIS pode ser identificado como um sistema de informação computadorizado, instalado em um ambiente hospitalar, com o objetivo de registrar informações sobre os pacientes, de tal forma que possam ser compartilhadas por todos os setores do hospital que delas necessitem (Johanston, 1993; Shortliffe; Perreault, 2000).
- Sistema de Informação em Radiologia/Radiology Information System (RIS): surgiu com a implantação dos computadores nos hospitais quando foi reconhecido que eles poderiam ser utilizados como auxílio na área da radiologia (Bakker, 1991).
- Sistema de Comunicação e Armazenamento de Imagens/Picture Archiving and Communication System (PACS): de maneira simples e resumida, Law e Zhou (2003) definem o PACS como um sistema de tecnologia de informação responsável pela transmissão e armazenamento das imagens médicas,

afirmando que um PACS consiste basicamente nos componentes de interface para HIS/RIS, modalidades de imagens, Digital Imaging and Communications in Medicine (DICOM⁸), controle de armazenamento e estações de visualização.

No Brasil, destaca-se o Prontuário Eletrônico do Paciente (PEP), um dos documentos eletrônicos mais importantes presentes nos SIS brasileiros:

- O Prontuário Eletrônico do Paciente (PEP): Segundo Albuquerque *et al.* (2017), a automação de processos em hospitais é uma prática mundial, seja para fins operacionais, administrativos ou assistenciais. O software responsável pelos registros de atividades do paciente no Brasil é realizado pelo (PEP), sendo um documento eletrônico constituído por um conjunto de informações referentes a tratamentos, cuidados aplicados, gestão de informação e comunicação referentes ao paciente.

Por fim, no que diz respeito aos processos de interoperabilidade de dados existentes entre os sistemas, cabe ainda ressaltar que, em relação aos SIS, esses são processos necessários e essenciais, estabelecidos de acordo com um padrão definido para o compartilhamento de informações dentro de um cenário de saúde.

4 INTEROPERABILIDADE PARA GERENCIAMENTO DOS DADOS DE PROVENIÊNCIA EM SISTEMAS DE INFORMAÇÃO EM SAÚDE: HL7 FHIR BASEADO EM W3C PROV

A proveniência de dados e a interoperabilidade são requisitos fundamentais para que os SIS possam garantir não somente o intercâmbio efetivo e eficiente dos dados de saúde, mas também, contribuir para definir as fontes de dados confiáveis (Sembay; Macedo, Marquez Filho, 2022a). Dessa forma, esforços de vários anos em padrões

⁸ Disponível em: <https://www.DICOMstandard.org/>. Acesso em: 8 jun. 2022.

internacionais como HL7 FHIR, o qual possui definições dos recursos de proveniência com base no modelo W3C PROV, confirmam esses requisitos (Sembay; Macedo, Marquez Filho, 2022b).

Sendo assim, destaca-se a preocupação em relação à utilização de padrões internacionais de interoperabilidade de dados entre SIS. Existem diversos padrões, mas para este estudo indica-se o HL7 FHIR. Dessa forma, vale destacar que, o HL7 FHIR usa a proveniência como um recurso, indicando significância clínica em termos de confiança na autenticidade, confiabilidade, integridade e estágio do ciclo de vida dos dados de saúde (HL7 International Inc, 2023b). Assim, o HL7 FHIR é baseado no modelo de proveniência de dados W3C PROV, o qual fornece design e meios de implementação para compartilhar atributos e recursos de proveniência semanticamente interoperáveis (HL7 International Inc, 2023b).

Ressalta-se ainda, que o HL7 FHIR baseado no W3C PROV, possui consonância com o padrão HL7 FHIR 5Ws⁹, o qual executa os processos referentes aos recursos de proveniência¹⁰. Assim, no padrão HL7 FHIR 5Ws, todos os recursos de proveniência incluem informações de atribuição. Classicamente, essas informações incluem: Who?, What?, When?, Where?, Why?, atribuindo o nome desse padrão (5Ws) (HL7 International Inc, 2023a).

Nesse contexto, a proveniência de dados no HL7 FHIR tem base no modelo de proveniência de dados W3C PROV, em que os mapeamentos são fornecidos. O W3C PROV fornece meios de design e implementação para compartilhar atributos de proveniência semanticamente interoperáveis. Além disso, os principais órgãos de saúde, como Integrating the Healthcare Enterprise (IHE) e HL7 FHIR, apoiam o W3C PROV (Margheri *et al.*, 2020). O W3C PROV, é a representação padrão de fato para

⁹ Disponível em: <http://hl7.org/fhir/fivews.html>. Acesso em: 8 jun. 2022.

¹⁰ Disponível em: <https://build.fhir.org/provenance-mappings.html>. Acesso em: 8 jun. 2022.

informações de proveniência interoperáveis adotada pelo HL7 FHIR (Prasser *et al.*, 2018).

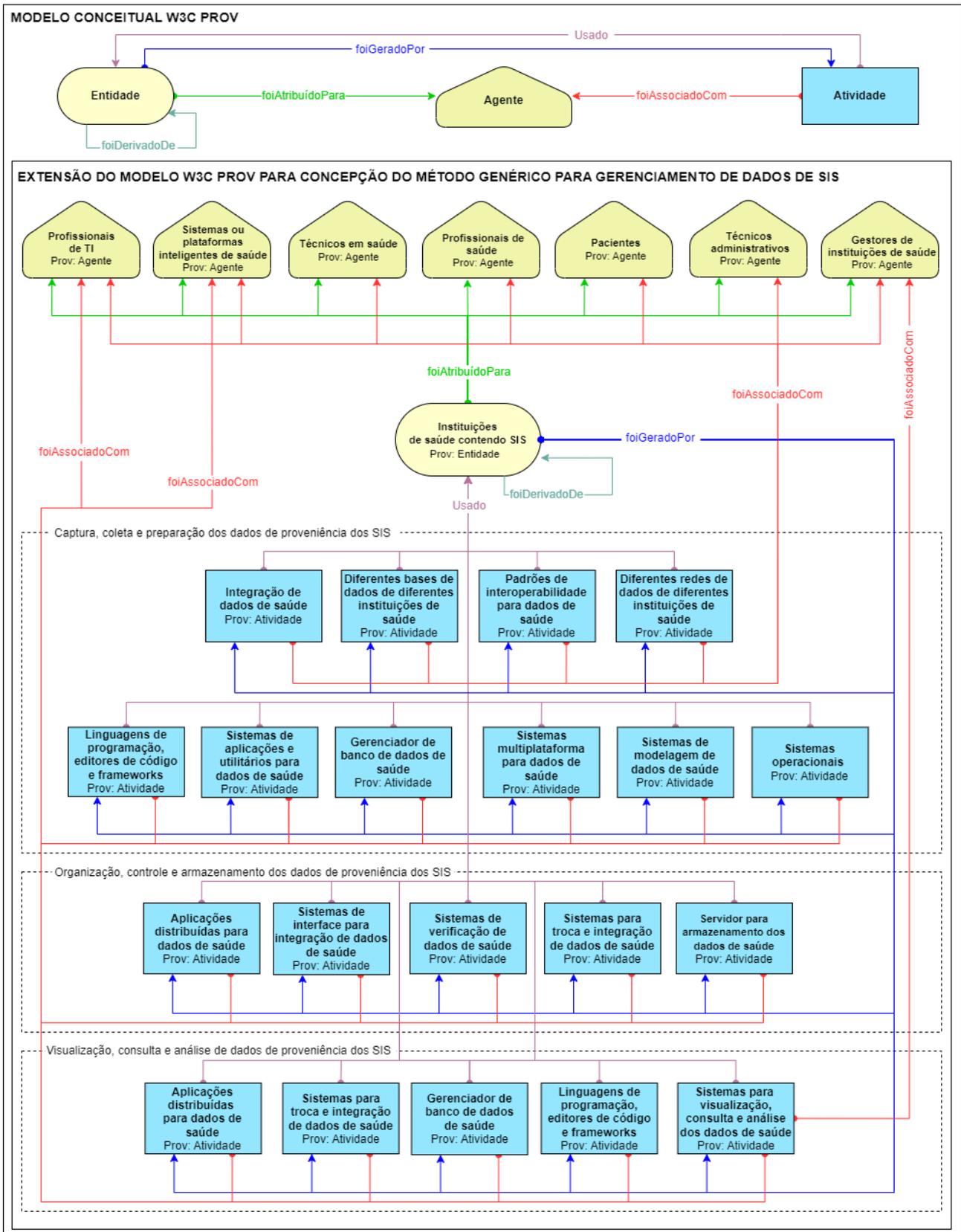
Portanto, o HL7 FHIR trata a proveniência de dados com base no W3C PROV com intuito de contribuir para a interoperabilidade dos dados em SIS, o que de fato, mostra um importante processo para o gerenciamento dos dados de proveniência em SIS.

5 PROPOSTA DO MÉTODO DE GERENCIAMENTO DE DADOS DE PROVENIÊNCIA EM SISTEMAS DE INFORMAÇÃO EM SAÚDE

Levando em consideração o contexto sobre o tema deste estudo, com base no levantamento bibliográfico e uma Revisão Sistemática de Literatura (RSL) desenvolvida por Sembay *et al.* (2023), os quais nortearam as discussões para a concepção do método de gerenciamento de dados de proveniência em SIS, algumas etapas foram necessárias para a construção dos processos necessários.

Primeiramente foi preciso compreender o objetivo do método com base nos componentes e relações ideais do modelo de proveniência de dados W3C PROV. Para isso, foi realizada a extensão do W3C PROV para o contexto genérico de SIS para este estudo, observando os níveis de abstração necessários que configuram a captura, coleta, preparação, organização, controle, armazenamento, visualização, consulta e análise dos dados de proveniência gerados pelos diferentes SIS descentralizados. O intuito foi descrever os possíveis processos que podem ser adaptados para qualquer cenário de saúde mediante integração e interoperabilidade de dados entre diferentes SIS. Dessa forma, apresenta-se, na Figura 2, a extensão do modelo de proveniência de dados W3C PROV com seus respectivos componentes e relações para o contexto de SIS de forma genérica. É importante ressaltar que essa extensão se baseou no estudo de Kock-Schoppenhauer *et al.* (2018), que descrevem uma extensão do modelo W3C PROV voltado para o contexto de saúde.

Figura 2: Extensão do modelo W3C PROV para concepção do método



Fonte: elaborada pelos autores.

Na Figura 2 apresenta-se o modelo de proveniência de dados W3C PROV em sua estrutura conceitual e logo após é realizada a extensão deste modelo para o contexto genérico de SIS para servir de base na concepção do método deste estudo. Desse modo, observa-se que na Figura 2 foram previstos os agentes identificados na extensão do modelo W3C PROV para a concepção do método deste estudo, sendo eles:

- Profissionais de Tecnologia da Informação (TI) e sistemas ou plataformas inteligentes de saúde: profissionais voltados ao gerenciamento de dados, engenheiros de software, analistas de sistemas, entre outros, que podem acessar todos os processos técnicos e operacionais necessários nos fluxos de dados existentes;
- Técnicos em saúde: técnicos de enfermagem, técnicos em farmácia, técnicos em radiologia, entre outros técnicos da área de saúde que podem apenas inserir dados de saúde referentes aos pacientes no fluxo dos serviços de saúde;
- Profissionais de saúde: médicos de diferentes especialidades, fisioterapeutas, farmacêuticos, entre outros profissionais de saúde que podem tanto acessar os dados como alterá-los dependendo do tratamento de saúde do paciente;
- Técnicos administrativos: podem apenas inserir e acessar os dados de agendamento dos procedimentos médicos solicitados pelos pacientes; e,
- Gestores de instituições de saúde: profissionais os quais tomam decisões gerenciais conforme informações enviadas pelos técnicos administrativos, e decisões técnicas e operacionais conforme solicitações dos profissionais de TI.

Nota-se que para cada agente inserido no contexto da extensão proposta, conta com autorizações diferenciadas em relação à inserção, acesso e uso dos dados de saúde.

Em seguida, na Figura 2 apresenta-se:

- Instituições de saúde contendo diferentes SIS: compartilham os dados, buscando benefícios em relação à tomada de decisão médica no auxílio direto ao paciente em suas consultas, tratamentos e exames por meio do gerenciamento dos dados de proveniência. Dos diferentes SIS destas instituições, são selecionados os dados que serão armazenados no servidor de dados de proveniência local do método, a fim de aplicar os conceitos de proveniência de dados na prática, com intuito de oferecer segurança a saúde das populações. Nesse quesito, importante ressaltar que ainda se leva em consideração o cumprimento dos requisitos da Lei Geral de Proteção de Dados, nº 13.709, de 14 de agosto de 2018 (Brasil, 2018). Os SIS específicos para esse estudo foram: EHR, HIS, RIS, PACS, PEP, sendo possível incorporar outros SIS que sejam necessários ao método. Esses SIS consistem em um importante ecossistema de saúde de uma instituição de saúde, com intuito de coletar, processar e organizar os principais dados de saúde de uma população, servindo de base para a tomada de decisões no planejamento dos serviços de diferentes contextos de saúde.

Logo após, na Figura 2, são apresentados três módulos que sustentam as relações entre Agente–Entidade–Atividade, os quais são: captura, coleta e preparação dos dados de proveniência dos SIS; Organização, controle e armazenamento dos dados de proveniência dos SIS; e, Visualização, consulta e análise de dados de proveniência dos SIS, os quais, também serviram de base para a estrutura conceitual do método. São observadas também, que algumas atividades se repetem, pois são constituintes e necessárias em diferentes módulos, a fim de executar funções computacionais compatíveis com os diferentes SIS.

Assim, para o módulo de captura, coleta e preparação dos dados de proveniência dos SIS, propõem-se o uso de tecnologias computacionais que contribuam para o gerenciamento de dados de proveniência em SIS, as quais, conforme as autorizações

dos agentes anteriormente descritas, são expressamente acessadas pelos Profissionais de TI ou pelos Sistemas ou Plataformas Inteligentes de Saúde, sendo elas:

- Integração de dados de saúde: estratégias computacionais no uso de linguagens de programação que executam scripts para reunir dados de saúde de diferentes fontes com intuito de produzir informações relevantes para uma instituição de saúde, possibilitando centralizá-los;
- Uso de diferentes bases de dados de diferentes instituições de saúde: ferramentas tecnológicas que organizam bancos de dados de saúde de diversos pacientes de diversas modalidades médicas em formato digital, sendo possível ser acessada por médicos, pesquisadores de saúde, assim como outros profissionais de saúde autorizados;
- Padrões de interoperabilidade para dados de saúde: conjunto de políticas e especificações técnicas e tecnológicas que realizam o intercâmbio de dados de saúde entre os diferentes SIS, seja na esfera Federal, Municipal, Distrital ou Estadual;
- Diferentes redes de dados de diferentes instituições de saúde: são redes de dados de saúde ligadas a diferentes SIS, possuindo autorizações específicas para diferentes agentes envolvidos nas rotinas médicas;
- Linguagens de programação, editores de códigos e frameworks: são softwares (ambientes programáveis) para criação e edição de códigos em aplicações de algoritmos para estruturação dos dados de saúde;
- Sistemas de aplicações e utilitários para dados de saúde: Levando em consideração os padrões de interoperabilidades de dados de saúde utilizados, a possibilidade de conjuntos de dados de saúde e imagens médicas de forma gerenciável e escalável;
- Gerenciador de banco de dados de saúde: plataformas que ajudam a organizar

e armazenar os dados de saúde de forma centralizada no método, indispensável para o gerenciamento eficiente desses dados gerados pelos diferentes SIS de diferentes instituições de saúde, mantendo-os com base nos requisitos de confidencialidade, integridade e disponibilidade;

- Sistemas multiplataformas para dados de saúde: mecanismos de interface multiplataforma que permitam a comunicação entre SIS distintos e o envio bidirecional de vários tipos de mensagens relacionadas aos padrões de interoperabilidade de dados utilizados pelas instituições de saúde;
- Sistemas de modelagem dos dados de saúde: tecnologias e padrões para troca de dados de saúde entre sistemas que tenham escopo internacional, nacional, regional ou local para uso genérico nos cenários de saúde, com intuito de contribuir para que recursos voltados a troca desses dados entre diferentes SIS de diferentes instituições de saúde, possam ter êxito;
- Sistemas operacionais: plataformas para a execução das tarefas, tanto em interface gráfica como por programas shell¹¹ executados em uma interface de linha de comando para controlar esses sistemas.

Já para o módulo de Organização, controle e armazenamento dos dados de proveniência dos SIS, é proposto o uso de tecnologias computacionais que contribuam para o gerenciamento de dados de proveniência em SIS, sendo essas tecnologias expressamente acessadas pelos mesmos agentes do módulo anterior, sendo elas:

- Aplicações distribuídas para dados de saúde: levando em consideração que os dados de saúde podem estar descentralizados no contexto dos SIS, e diversas estratégias computacionais precisam ser executadas em diferentes redes de dados de saúde em diferentes bancos de dados distribuídos, o método prevê a

¹¹ Programa especial disponibilizado ao usuário para que este possa se comunicar diretamente com o sistema operacional.

- utilização de aplicações distribuídas para compartilhamento de dados de saúde entre diferentes SIS, preocupando-se na centralização desses dados para poder armazená-los e organizá-los;
- Sistemas de interface para integrações de dados de saúde: com a grande volume de dados gerados pelos SIS, a necessidade de centralizá-los e ordená-los de forma lógica e automatizada, torna-se um processo fundamental. Para isso, recomenda-se o uso de sistemas compatíveis com os padrões de interoperabilidade de dados de saúde, as quais facilitam e otimizam os serviços e melhoram a integração dos diferentes SIS;
 - Sistemas de verificação de dados de saúde: sistemas que realizem a verificação dos dados de saúde, para que esses dados sejam reconhecidos entre as diferentes instituições de saúde que irão realizar os processos de integração e interoperabilidades de seus dados com o método;
 - Sistemas para troca e integração de dados de saúde: o método depende da utilização de padrões de interoperabilidade entre diferentes SIS, possibilitando a troca de dados de saúde nesses sistemas. Desta forma, oportuniza também, a integração entre diferentes SIS, contribuindo para que a troca de dados de saúde aconteça de forma eficiente e contínua;
 - Servidor para armazenamento de dados de saúde: levando em consideração o grande volume de dados produzidos nos setores de saúde, o método preocupa-se principalmente com o gerenciamento dos dados de proveniência dos SIS que se encontram descentralizados. Nesse sentido, como o intuito do método é centralizar os dados de proveniência dos SIS, recomenda-se utilizar-se de um servidor interno (servidor físico local) ou servidor externo (armazenamento em nuvem), ambos centralizando os dados e seguindo as leis de proteção de dados vigentes.

Para o módulo de Visualização, consulta e análise de dados de proveniência dos SIS, é proposto o uso de tecnologias computacionais que contribuam para o gerenciamento de dados de proveniência em SIS, sendo que, a maioria dessas tecnologias se repetem com o mesmo acesso pelos agentes do módulo anterior. Assim, a única tecnologia nova nesse módulo é:

- Sistemas para visualização, consulta e análise dos dados de saúde: recomenda-se o uso de sistemas que integrem os dados de proveniência contidos no fluxo do método em banco de dados para fins de análises em um contexto de sistemas de aplicação *web* multiplataforma para visualização interativa de dados de saúde. Dessa maneira, esse processo contribui para a geração de tabelas, gráficos e relatórios, conforme a necessidade dos Gestores de Instituições de Saúde previstos no método.

Nessa perspectiva, a proposta dessas tecnologias para uso no método, tem o intuito de permitir que toda a interoperabilidade e integração dos dados de saúde entre diferentes SIS, além da cooperação de todos os profissionais envolvidos, possam contribuir para que os dados de proveniência dos fluxos informacionais existentes possam ser gerenciados. Ainda, no que se refere às propostas de tecnologias em relação à interoperabilidade de dados, incentiva-se o uso do HL7 FHIR, pelo fato de este estar associado internacionalmente ao modelo de proveniência de dados W3C PROV. Porém, nada impede que seja utilizado outro tipo de padrão de interoperabilidade conforme a necessidade.

Assim sendo, para melhor compreensão dos fluxos compreendidos na Figura 2, o Quadro 3, exposto a seguir, descreve as relações existentes da extensão do modelo de proveniência de dados W3C PROV para a concepção do método aqui proposto.

Quadro 3: Descrição das relações da extensão W3C PROV para concepção do método proposto

Elementos	Descrição
FoiAtribuídoPara	Esta relação está voltada às instituições de saúde contendo SIS (Entidade)
foiAssociadoCom	Esta relação está voltada às atividades e agentes. Cada agente pode realizar uma determinada atividade conforme sua responsabilidade e autorização. As atividades contidas nos três módulos, sendo: captura, coleta e preparação dos dados de proveniência em SIS; organização, controle e armazenamento dos dados de proveniência em SIS; e, visualização, consulta e análises de dados de proveniência em SIS, são associadas basicamente aos agentes, profissionais de TI e sistemas ou plataformas inteligentes de saúde, por se tratar de processos técnicos tecnológicos, sendo analisadas as atribuições e autorizações necessárias. Exceto os agentes gestores de instituições de saúde, que possuem acesso a qualquer módulo do método, os agentes técnicos em saúde, profissionais de saúde e técnicos administrativos, podem inserir e visualizar os dados conforme restrições de acesso e os agentes pacientes, podem apenas visualizar informações pertinentes a receitas e laudos médicos.
foiGeradoPor	Esta relação ocorre entre entidades no quesito de observar as atividades geradas pelas instituições de saúde que contêm SIS. Nesse sentido, instituições de saúde contendo SIS (Entidades), geram todas as atividades expostas na Figura 2 e já mencionadas anteriormente na relação foiAssociadoCom. Em resumo, essas instituições geram atividades direta ou indiretamente que contribuem para os três módulos de atividades do método. Assim, essas atividades são arquitetadas em estruturas tecnológicas e executadas pelos agentes responsáveis, contribuindo ainda para os processos de rastreabilidade das fontes de dados de saúde de origem para fins de tomada de decisão nos diferentes SIS.
foiDerivadoDe	Esta relação está voltada apenas de entidade para entidade, ou seja, na existência de diferentes entidades.
usado	Esta relação está voltada ao uso das atividades por parte das entidades. Assim, como descrito anteriormente na relação foiGeradoPor, em que entidades geram atividades, a relação usada afirma que cada atividade tem seu uso na entidade de forma direta ou indireta, ou seja, são executadas em fluxos de estratégias computacionais necessárias para contribuir para a execução dos três módulos de atividades do método.

Fonte: elaborado pelos autores.

Com base na Figura 2 e nas descrições do Quadro 3, pôde-se fundamentar a proposta da concepção do método genérico com o intuito de preservar os fundamentos do modelo de proveniência de dados W3C PROV.

5 CONSIDERAÇÕES FINAIS

Neste estudo discutiu-se sobre a proveniência de dados, interoperabilidade de dados e SIS, com uma perspectiva voltada ao gerenciamento de dados de proveniência em SIS. Foi proposto, a partir das discussões apresentadas, um método genérico baseado na extensão do modelo de proveniência de dados, o W3C PROV, compreendendo os principais agentes, atividades e entidades juntamente com as relações do modelo para atender diferentes cenários de saúde, especificamente os SIS. Ressalta-se ainda, que ao propor o método com base na extensão do W3C PROV, o mesmo proporciona toda a compreensão e utilização do padrão de interoperabilidade HL7 FHIR no que se refere às tecnologias que envolvem o método proposto nesse quesito, o que de fato é internacionalmente utilizado na indústria de sistemas médicos. Com isso, o objetivo deste estudo foi alcançado.

Nesse contexto, o método genérico proposto neste estudo pode contribuir de forma significativa para o gerenciamento de dados de proveniência em SIS para a melhoria dos fluxos informacionais relacionados aos processos que envolvem a identificação, rastreamento e reaproveitamento dos dados de proveniência em diferentes SIS que se apresentam descentralizados.

Observou-se, que é necessário aprofundar o método proposto neste estudo a partir de uma aplicação prática em um cenário de saúde real, como, por exemplo, em um sistema de telemedicina. Isso pela necessidade, que de fato, possa ser discutido com mais afinco ações que podem ser realizadas por diferentes profissionais além dos identificados para o método aqui proposto, em busca de encontrar outras especificidades que ainda sejam necessárias para aperfeiçoá-lo, levando em consideração os grandes volumes de dados gerados pelos diferentes SIS.

Com este estudo, considera-se que o gerenciamento de dados de proveniência em SIS pode contribuir para além desse processo, por exemplo, mantendo a segurança

desses dados e influenciando no custo-benefício da instituição de saúde, sendo útil para tomadas de decisão.

REFERÊNCIAS

ALBUQUERQUE, Erika Akemi Yanaguibash; ALBUQUERQUE, Gabriela de Araújo; SOUZA, Luisa Christina de; SANTOS, Suéllen Silva dos; RÊGO, Yasmynni Lyssa Silva. Prontuário eletrônico do paciente e certificação de software em saúde: avanços que visam maior segurança dos dados médicos. **Revista Brasileira de Inovação Tecnológica em Saúde**, [s. l.], v. 7, n. 2, 2017. DOI: <https://doi.org/10.18816/r-bits.v7i2.11074>. Disponível em: <https://periodicos.ufrn.br/reb/article/view/11074/8973>. Acesso em: 6 abr. 2023.

BAKKER, Albert. HIS and RIS and PACS. *In*: HUANG, H. K.; RATIB, Osman; BAKKER, Albert; WITTE, Gerd (ed.). **Picture Archiving and Communications System (PACS) in Medicine**. Berlin: Springer, 1991. v. 74, p. 157-162. (Nato Asi Series).

BEN-ASSULI, Ofir. Electronic health records, adoption, quality of care, legal and privacy issues and their implementation in emergency departments. **Health Policy**, Amsterdam, v. 119, n. 3, p. 287-297, 2015. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/25483873/>. Acesso em: 6 abr. 2023.

BRASIL. **Lei nº 13.709, de 14 de agosto de 2018**. Dispõe sobre a proteção de dados pessoais e altera a Lei nº 12.965, de 23 de abril de 2014 (Marco Civil da Internet). Brasília, DF: Presidência da República, 2018. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/L13709.htm. Acesso em: 29 abr. 2022.

BUNEMAN, Peter; KHANNA, Sanjeev; TAN, Wang-Chiew. Why and where: a characterization of data provenance. *In*: INTERNATIONAL CONFERENCE LONDON, 8., 2001, London. **Proceedings** [...]. Berlin: Springer, 2001. p. 316-330. Disponível em: https://link.springer.com/chapter/10.1007/3-540-44503-X_20. Acesso em: 2 maio 2023.

CESNIK, Branko; KIDD, Michael R. History of health informatics: a global perspective. **Studies in Health Technology and Informatics**, Amsterdam, v. 151, p. 3-8, 2010. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/20407147/>. Acesso em: 2 maio 2023.

DAVIDSON, Susan; FREIRE, Juliana. Provenance and scientific workflows: challenges and opportunities. *In: ACM SIGMOD INTERNATIONAL CONFERENCE ON MANAGEMENT OF DATA*, Vancouver, 2008. **Proceedings** [...]. New York: Association for Computing Machinery, 2008. p. 1345-1350. Disponível em: <https://dl.acm.org/doi/10.1145/1376616.1376772>. Acesso em: 2 maio 2023.

FREUND, Gislaine Parra; SEMBAY, Márcio José; MACEDO, Douglas Dyllon Jeronimo de. Proveniência de dados e segurança da informação: relações interdisciplinares no domínio da Ciência da Informação. **Revista Ibero-Americana de Ciência da Informação**, Brasília, DF, v. 12, n. 3, p. 807-825, 2019. Disponível em: <https://periodicos.unb.br/index.php/RICI/article/view/21203/23548>. Acesso em: 20 maio 2023.

GERHARDT, Tatiana Engel; SILVEIRA, Denise Tolfo (org.). **Métodos de pesquisa**. Porto Alegre: UFRGS, 2009.

GIL, Yolanda; MILES, Simon (ed.). PROV Model Primer. **W3C Working Group Note**, [s. l.], n. 30, 2013. Disponível em: <https://www.w3.org/TR/prov-primer/>. Acesso em: 20 maio 2022.

GROTH, Paul; MOREAU, Luc (ed.). PROV-Overview: an overview of the PROV family of documents. **W3C Working Group Note**, [s. l.], n. 30, 2013. Disponível em: <https://eprints.soton.ac.uk/356854/>. Acesso em: 15 abr. 2023.

HÅKANSSON, Stefan; GAVELIN, Carin. What do we really know about the cost-effectiveness of telemedicine. **Journal of Telemedicine and Telecare**, London, v. 6, suppl. 1, p. 133-136, 2000. Disponível em: <https://www.ncbi.nlm.nih.gov/pubmed/10793998>. Acesso em: 21 nov. 2021.

HAUX, Reinhold. Health information systems-past, present, future. **International Journal of Medical Informatics**, Ireland, v. 75, n. 3-4, p. 268-281, 2006. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/16169771/>. Acesso em: 21 nov. 2021.

HL7 INTERNATIONAL INC. 5Ws Pattern. **HL7**, [s. l.], Dec. 18, 2023a. Disponível em: <https://build.fhir.org/fivews.html>. Acesso em: 10 jan. 2023.

HL7 INTERNATIONAL INC. Provenance. **HL7**, [s. l.], Mar. 26, 2023b. Disponível em: <https://www.hl7.org/fhir/provenance.html>. Acesso em: 11 jan. 2023.

HOERBST, Alexander; AMMENWERTH, Elske. Electronic health records: a systematic review on quality requirements. **Methods of Information in Medicine**, Stuttgart, v. 49, n. 4, p. 320-336, 2010. Disponível em:

<https://pubmed.ncbi.nlm.nih.gov/20603687/>. Acesso em: 10 jan. 2023.

JOHANSTON, Halley. Sistemas de informação hospitalar: presente e futuro. **Revista Informédica**, São Paulo, v. 1, n. 2, p. 5-9, 1993. Disponível em:

<http://www.informaticamedica.org.br/informed/halley.htm>. Acesso em: 10 jan. 2023.

KNECHTEL, Maria do Rosário. **Metodologia da pesquisa em educação**: uma abordagem teórico-prática dialogada. Curitiba: Intersaberes, 2014.

KOCK-SCHOPPENHAUER, Ann-Kristin; HARTUNG, Lina; ULRICH, Hannes; DUHM-HARBECK, Petra; INGENERF, Josef. Practical Extension of Provenance to Healthcare Data Based on the W3C PROV Standard. **Studies in Health Technology and Informatics**, Amsterdam, v. 253, p. 28-32, 2018. Disponível em:

<https://pubmed.ncbi.nlm.nih.gov/30147034/>. Acesso em: 10 jan. 2023.

LAW, Maria Y. Y.; ZHOU, Zheng. New Direction in PACS Education and Training.

Computerized Medical Imaging and Graphics, New York, v. 27, n. 2-3, p. 147-156, 2003. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/12620305/>. Acesso em: 10 jan. 2023.

MARGHERI, Andrea; MASI, Massimiliano; MILADI, Abdallah; SASSONE, Vladimiro; ROSENZWEIG, Jason. Decentralised provenance for healthcare data. **International Journal Medical Informatics**, Ireland, v. 141, p. 104197, 2020. Disponível em:

<https://pubmed.ncbi.nlm.nih.gov/32540775/>. Acesso em: 10 jan. 2023.

MELO, Edleusa Torres de; MOTA, Francisca Rosaline Leite; ARAÚJO, Nelma Camêlo. Organização da informação no Instituto Médico Legal Estácio de Lima-Maceió/AL.

Revista de Saúde Digital e Tecnologias Educacionais, Fortaleza, v. 7, n. esp. 3, p. 185-203, 2022. Disponível em:

https://repositorio.ufc.br/bitstream/riufc/64044/1/2022_art_etmelo.pdf. Acesso em: 20 maio 2022.

PÉREZ, Beatriz; RUBIO, Julio; SÁENZ-ADÁN, Carlos. A systematic review of

provenance systems. **Knowledge and Information Systems**, [s. l.], v. 57, p. 495-543,

2018. Disponível em: <https://link.springer.com/article/10.1007/s10115-018-1164-3>. Acesso em: 20 maio 2022.

PRASSER, Fabian; KOHLBACHER, Oliver; MANSMANN, Ulrich; BAUER, Bernhard; KUHN, Klaus A. **Data Integration for Future Medicine (DIFUTURE):** an architectural and methodological overview. **Methods of Information in Medicine**, Stuttgart, v. 57, suppl. 1, p. 57-65, 2018. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6178202/#>. Acesso em: 18 maio 2022.

RENEKER, Maxine H. A qualitative study of information seeking among members of an academic community: methodological issues and problems. **Library Quarterly**, [s. l.], v. 63, n. 4, p. 487-507, 1993. Disponível em: <https://www.journals.uchicago.edu/doi/abs/10.1086/602623?journalCode=lq>. Acesso em: 18 maio 2022.

RYU, Seewon. Telemedicine: opportunities and developments in member states: report on the second global survey on eHealth 2009 (Global Observatory for eHealth Series, Volume 2). **Healthcare informatics research**, Seoul, v. 18, n. 2, p. 153-155, 2012. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3402558/>. Acesso em: 21 abr. 2022.

SEMBAY, Márcio José; MACEDO, Douglas Dyllon Jeronimo de. Sistemas de informação em saúde: proposta de um método de gerenciamento de dados de proveniência no instanciamento do modelo W3C PROV-DM. **Advanced Notes in Information Science**, [s. l.], v. 2, p. 192-201, 2022. Disponível em: <https://anis.prometrics.org/index.php/a/article/view/22/23>. Acesso em: 18 maio 2022.

SEMBAY, Márcio José; MACEDO, Douglas Dyllon Jeronimo de; DUTRA, Moisés Lima. A method for collecting provenance data: a case study in a Brazilian hemotherapy center. In: INTERNATIONAL CONFERENCE ON DATA AND INFORMATION IN ONLINE, 2020. **Proceedings** [...]. Cham: Springer, 2020. p. 89-102. Disponível em: https://link.springer.com/chapter/10.1007/978-3-030-50072-6_8#citeas. Acesso em: 18 maio 2022.

SEMBAY, Márcio José; MACEDO, Douglas Dyllon Jeronimo de; DUTRA, Moisés Lima. A proposed approach for provenance data gathering. **Mobile Networks and Applications**, [s. l.], v. 26, n. 1, p. 304-318, 2021. Disponível em: <https://link.springer.com/article/10.1007/s11036-020-01648-7>. Acesso em: 18 maio

2022.

SEMBAY, Márcio José; MACEDO, Douglas Dyllon Jeronimo de; MARQUEZ FILHO, Alexandre Augusto Gimenes. Proveniência de dados em Sistemas de Informação em Saúde. *In: ENCONTRO NACIONAL DE PESQUISA E PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO*, 22., Porto Alegre, 2022. **Anais [...]**. Porto Alegre: UFRGS, 2022a. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/200577>. Acesso em: 29 mar. 2023.

SEMBAY, Márcio José; MACEDO, Douglas Dyllon Jeronimo de; MARQUEZ FILHO, Alexandre Augusto Gimenes. HL7 FHIR baseado em W3C PROV para alcançar a proveniência de dados em sistemas de informação em saúde. *In: WORKSHOP DE INFORMAÇÃO, DADOS E TECNOLOGIA*, 5., 2022, Vitória. **Anais [...]**. Vitória: UFES, 2022b.

SEMBAY, Márcio José; MACEDO, Douglas Dyllon Jeronimo de; PIOLI JÚNIOR, Laércio; BRAGA, Regina Maria Maciel; SARASA-CABEZUELO, Antonio. Provenance data management in health information systems: a systematic literature review. **Journal of Personalized Medicine**, Basel, v. 13, n. 6, p. 991, 2023. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/37373980/>. Acesso em: 29 ago. 2023.

SHORTLIFFE, Edward H.; PERREAUULT, Leslie E. (ed.). **Medical Informatics: computer applications in health care and biomedicine**. 2 ed. New York: Spring Science, 2000.

SIMMHAN, Yogesh L.; PLALE, Beth; GANNON, Dennis. **A survey of data provenance techniques**: technical report TR-618. *ACM SIGMOD Record*, [s. l.], p. 1-25, 2005. Disponível em: <https://legacy.cs.indiana.edu/ftp/techreports/TR618.pdf>. Acesso em: 24 maio 2018.

VICTORA, Cesar Gomes; AQUINO, Estela Maria Motta Lima Leão de; LEAL, Maria do Carmo; MONTEIRO, Carlos Augusto; BARROS, Fernando Celso Lopes Fernandes de; SZWARCOWALD, Celia Landmann. Saúde de mães e crianças no Brasil: progressos e desafios. **The Lancet**, London, n. esp., p. 32-46, 2011.



✉ tpbci@ancib.org

📷 [@anciboficial](https://www.instagram.com/anciboficial)

🐦 [@ancib_brasil](https://twitter.com/ancib_brasil)