

Sistemas de informação em saúde: proposta de um método de gerenciamento de dados de proveniência no instanciamento do modelo W3C PROV-DM

Health information systems: proposal of a provenance data management method in the instantiation of the W3C PROV-DM model

Márcio José Sembay

Universidade Federal de Santa Catarina (UFSC), Brasil.

E-mail: marcio.sembay@posgrad.ufsc.br

Douglas Dyllon Jeronimo de Macedo

Universidade Federal de Santa Catarina (UFSC), Brasil.

E-mail: douglas.macedo@ufsc.br

RESUMO

Sistemas de Informação em Saúde (SIS) estão sendo implementados em todos os aspectos da saúde, desde a administração aos sistemas de apoio às decisões clínicas. A geração e armazenamento de grandes volumes de dados em seus repositórios descentralizados tornam esses processos desafios no que diz respeito ao gerenciamento desses dados. As abordagens da proveniência de dados nesse quesito ajudam a responder perguntas como: “por quê?”, “como?”, “onde?”, “quando?”, “por quem?” e “para quê?” os dados foram produzidos nos SIS, contribuindo para diversos benefícios, principalmente no quesito de rastreabilidade dos dados de proveniência para que possam ser gerenciados dentro desses sistemas. Nesse contexto, o presente trabalho é uma pesquisa bibliográfica de natureza qualitativa baseada no *Provenance Data Model* (PROV-DM) recomendado pela *World Wide Web Consortium* (W3C), onde propomos um método para gerenciamento dos dados de proveniência em SIS. Considerando a temática deste estudo, a questão de pesquisa aqui

Como citar: Sembay, M. J.; & Macedo, D. D. J. (2022). 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*, volume 2 (pp. 192-201). Tallinn, Estonia: ColNes Publishing. DOI: 10.47909/anis.978-9916-9760-3-6.101.

Copyright: © 2022, The author(s). This is an open access work distributed under the terms of the CC BY-NC 4.0 license which permits copying and redistributing the material in any medium or format, adapting, transforming and building upon the material as long as the license terms are followed.

apresentada é: como desenvolver um método para o gerenciamento dos dados de proveniência em SIS no instanciamento do modelo W3C PROV-DM? Um resultado esperado é o desenvolvimento de um protótipo para validar o método aqui proposto em um cenário real de saúde, apresentando, de fato, o benefício do uso do método.

Palavras-chave: Sistemas de Informação em Saúde; Proveniência de Dados; Gerenciamento de Dados de Proveniência; W3C PROV-DM; Método.

ABSTRACT

Health Information Systems (HIS) are being implemented in all aspects of health, from administration to clinical decision support systems. The generation and storage of large volumes of data in their decentralized repositories make these processes challenging when it comes to managing this data. The data provenance approaches in this regard help to answer questions such as: why? How? Where? When? by whom? And for what? The data were produced in the HIS, contributing to several benefits, especially in terms of traceability of the data of origin so that they can be managed within these systems. In this context, the present work is bibliographical research of qualitative nature based on the Provenance Data Model (PROV-DM) recommended by the World Wide Web Consortium (W3C), where we propose a method for management of provenance data in HIS. Considering the theme of this study, the research question presented here is: how to develop a method for the management of provenance data in HIS in the instantiation of the W3C PROV-DM model? An expected result is the development of a prototype to validate the method proposed here in a real health scenario, showing, in fact, the benefit of using the method.

Keywords: Health Information Systems; Data Provenance; Provenance Data Management; W3C PROV-DM; Method.

INTRODUÇÃO

O USO DE SISTEMAS de Informação (SI) em centros de saúde possui grande potencial “para melhorar a relação custo-eficácia, qualidade e acessibilidade aos cuidados em saúde” (Chiasson & Davidson, 2004, p. 156). Nesse sentido, podemos destacar os Sistemas de Informação em Saúde (SIS), que são caracterizados como um conjunto de componentes que atuam de forma integrada, por meio de mecanismos de coleta, processamento, análise e transmissão da informação necessária e oportuna com o objetivo de auxiliar no processo decisório em saúde. Esses sistemas contribuem também na produção de conhecimentos sobre as condições de vida e saúde da população, permitindo a otimização dos processos gerenciais e assistenciais, inclusive no cuidado direto ao paciente (Garcia & Reis, 2016; Araújo, 2017; Ferreira, Oliveira, Marques, Lima, Barbosa, Castro & Guimarães, 2020).

Se faz importante ressaltar também que os SIS têm a capacidade de gerar e armazenar grandes volumes de dados, o que de fato é um salto positivo para a saúde, mas, ao mesmo tempo, o gerenciamento desses dados tornou-se um desafio. Dessa forma, destaca-se o conceito de proveniência de dados, que diz respeito à origem ou procedência dos dados, e pode também estar relacionado à auditoria, triagem e linhagem dos dados utilizados pelos SI. Além disso, a proveniência dos dados identifica transformações pelas quais os dados passaram ao longo do tempo (Davidson & Freire, 2008; Sembay; Macedo & Dutra, 2020a).

A proveniência de dados pode ser aplicada em ambientes digitais, armazenando os dados em bancos de dados e registrando seus históricos de processos manuais. Auxilia também na determinação da autenticidade e consequente valorização de um dado, por meio de algumas perguntas que servem como filtro de rastreabilidade desses dados, tais como: “o quê?”, “quando?”, “onde?”, “como?”, “quem?”, “qual?”, “por quê?” e “para quê”. Nesse sentido, o principal fundamento da proveniência de dados é recolher os indícios e documentações quanto ao tempo, lugar e, se for o caso, à pessoa responsável pela criação, produção, descoberta ou inserção do dado. (Freire, Koop, Santos, & Silva, 2008; Tan, 2008; Simmhan, Plale & Gannon, 2006).

Conforme o estudo de Freund, Sembay e Macedo (2019), a proveniência de dados pode ser considerada também um requisito importante para estabelecer confiabilidade e prover segurança em SI. Nesse contexto, é importante ressaltar ainda que a proveniência de dados está relacionada a diferentes cenários de aplicação, como é o caso da saúde, foco deste estudo. O uso de dados de proveniência no contexto da saúde vive um cenário crescente de pesquisa com base nos mais variados tipos de experimentos científicos, e as tecnologias aplicadas nessa área vêm obtendo resultados expressivos (Sembay, Macedo & Dutra, 2020b).

Outro ponto importante da proveniência de dados, conforme os autores Groth e Moreau (2013), é a existência dos modelos de proveniência de dados recomendados pela W3C, que podem ser adaptados e aplicados nos mais variados cenários informacionais, sendo que destacamos para este estudo o modelo PROV-DM. Desta forma, este estudo tenta responder a seguinte questão: como desenvolver um método para o gerenciamento dos dados de proveniência em SIS no instanciamento do modelo W3C PROV-DM? Portanto, este estudo faz parte de uma investigação mais ampla em andamento na forma de uma tese de doutorado referente ao gerenciamento de dados de proveniência em SIS a ser validado em um cenário real de saúde.

1. OBJETIVOS

Este estudo tem como objetivo propor um método de gerenciamento de dados de proveniência em SIS no instanciamento do modelo W3C PROV-DM.

2. PROCEDIMENTOS METODOLÓGICOS

No presente estudo realizamos uma pesquisa bibliográfica de natureza qualitativa acerca da temática do gerenciamento dos dados de proveniência em SIS. Foram utilizados também os conceitos do modelo de proveniência de dados PROV-DM para realização do instanciamento e modelagem do método aqui proposto. Segundo a W3C (2013), o modelo PROV-DM, tem como principal função descrever as pessoas, entidades e atividades envolvidas na produção de um dado. Além disso, o modelo cria as condições para que a proveniência seja demonstrada e trocada entre diferentes sistemas.

O PROV-DM é um modelo genérico para descrever informações detalhadas sobre os dados e processos envolvidos no ciclo de vida de objetos de dados e que contém um conjunto de especificações da família PROV , projetado para promover a fácil exploração dos nós e relações para modelar domínios específicos (Belhajjame, B'Far, Cheney, Coppens, Creswell, Gil, Groth, Klyne, Lebo, McCusker, Miles, Myers, & Sahoo, 2013; Moreau, Groth, Cheney, Lebo, 2015).

O modelo é dividido em três componentes que contém tanto os elementos como as relações possíveis entre eles (W3C, 2013): (1) Entidade (*Entity*): podem representar qualquer objeto (real ou imaginário); (2) Atividade (*Activity*): representam os processos que usam e geram entidades; e (3) Agente (*Agent*): influenciam, direta ou indiretamente, a execução das atividades, recebem atribuições de outros agentes e podem ter algum tipo de ligação (posse, direitos etc.). A Figura 1 apresenta o modelo PROV-DM recomendado pela W3C (2013).

No modelo de proveniência de dados PROV-DM apresentado na Figura 1 é possível modelar a proveniência como um gráfico acíclico dirigido, no qual os três tipos de nós (*Entity*, *Activity*, *Agent*) possuem sete arestas descritas em cada tipo de relação conforme especificado pela W3C (2013):

- (i) *ActedOnBehalfOf*: “Usado para delegar autoridade e responsabilidade entre os Agentes”;
- (ii) *WasAssociatedWith*: “Atividades associadas com o Agente”;
- (iii) *WasAttributeTo*: “Entidade foi atribuída ao Agente”;

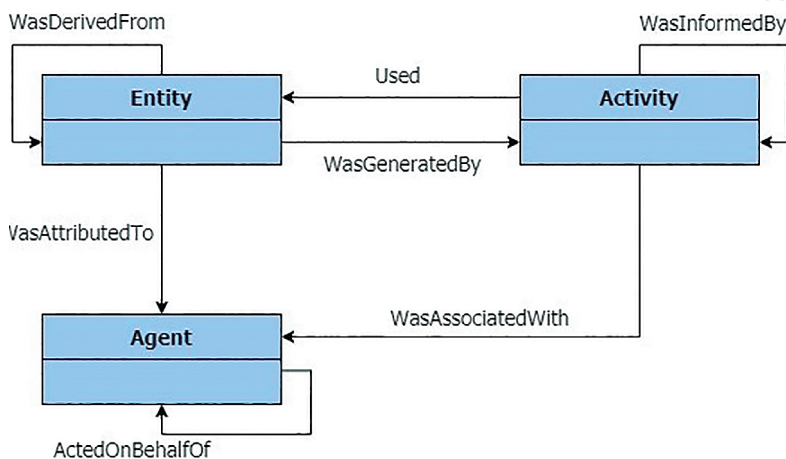


Figura 1. Modelo PROV-DM. **Fonte:** W3C (2013)

- (iv) *WasDerivedFrom*: “(uma) Entidade foi derivada de (outra) Entidade”;
- (v) *WasGeneratedBy*: “Entidade foi gerada por atividade”;
- (vi) *WasInformedBy*: “Atividade foi informada por atividade”;
- (vii) *Used*: “Entidade usada por atividade”.

O modelo PROV-DM, foi escolhido com base na literatura internacional para servir de instanciamento para o método aqui proposto, por se tratar de um dos modelos que mais detalham os processos que podem ser aplicados no gerenciamento de dados de proveniência em SIS. Nesse sentido, o objetivo é tornar mais clara a ideia sobre o assunto em estudo, além de criar estruturas que possam ser colocadas em prática.

Desta forma, os procedimentos deste estudo serão avaliados em aplicação prática, utilizando uma estratégia computacional para o gerenciamento dos dados de proveniência em SIS integrados em uma rede de telemedicina no Brasil. Essa será a próxima etapa no que compete ao andamento da tese de doutorado no Programa de Pós-Graduação em Ciência da Informação da Universidade Federal de Santa Catarina (PGCIN/UFSC).

3. MÉTODO PROPOSTO

Realizando um instanciamento das estruturas do modelo de proveniência de dados PROV-DM, e no intuito de descrever os possíveis processos para o método aqui proposto para este estudo, apresenta-se na Figura 2 a modelagem com seus respectivos nós (módulos) e arestas em suas relações.

Conforme a modelagem do método proposto (Figura 2), pode-se observar que o mesmo contém módulos que especificam a execução do método, fazendo referência aos componentes instanciados do modelo PROV-DM (Agent, Entity e Activity) e suas respectivas relações. Desta forma, para melhor compreender o método aqui proposto, apresenta-se o passo a passo do seu funcionamento nos 6 módulos e tipos de componentes conforme apresentado na Figura 2.

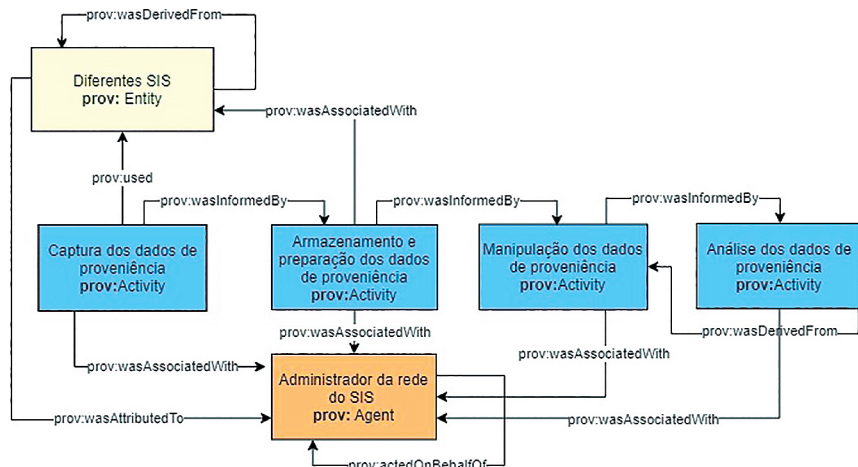


Figura 2. Modelagem do método proposto baseado no modelo PROV-DM.

Fonte: Dados da Pesquisa, 2020.

1. Diferentes SIS (Entity): são selecionadas as fontes de dados de diferentes SIS, como, por exemplo, Sistema de Informação Hospitalar (*Hospital Information System - HIS*), Sistema de Informação Clínica (*Clinical Information Systems - CIS*), Sistema de Arquivamento de Imagens (*Picture Archiving and Communication System - PACS*), Sistema de Informação de Radiologia (*Radiology Information System - RIS*), Prontuário Eletrônico do Paciente (PEP), entre outros sistemas que o agente administrador de rede tenha acesso às bases de dados que venham a ser necessárias.
2. Administrador de rede do SIS (Agent): acessa as diferentes bases de dados heterogêneas das várias entidades de saúde no intuito de realizar a captura dos dados de proveniência.
3. Captura dos dados de proveniência (Activity): são inicializados os processos de captura de dados nos bancos de dados descentralizados dos diferentes SIS e sistemas integrados por meio de comandos computacionais que venham ser necessários dependendo da demanda

informacional. Ainda, nesses processos de captura de dados de proveniência é observada a existência de padrões internacionais de interoperabilidade entre os sistemas integrados para que ocorra a melhor captura possível dos dados de proveniência.

4. Armazenamento e preparação dos dados de proveniência (*Activity*): é criado um espelhamento dos dados de proveniência obtidos dos diferentes SIS para um repositório de dados de proveniência local, utilizando recursos computacionais que venham ser necessários para a preparação desses dados.
5. Manipulação dos dados de proveniência (*Activity*): são manipulados os dados de proveniência com o uso de ferramentas computacionais, aplicações de algoritmos, bibliotecas e funções de linguagens de programação de alto nível para manipulação dos dados, ou seja, processos computacionais conforme a necessidade encontrada.
6. Análise dos dados de proveniência (*Activity*): nesse módulo é realizada uma interface de acesso aos dados de proveniência por meio de uma aplicação computacional para que seja possível a visualização dos dados e realização de análises, consultas e auditorias em relação aos dados de proveniência já manipulados e armazenados no repositório de dados de proveniência local. As análises podem gerar gráficos, tabelas (dados estatísticos, gráficos de volume de dados, entre outros) para contribuir em tomadas de decisão necessárias. Consequentemente, com as análises realizadas no repositório de dados de proveniência local, oportuniza consultas utilizando estratégias computacionais que podem ser realizadas por demanda, podendo-se, assim, descobrir a origem de um determinado dado, como, por exemplo, dados de origem de um exame de um paciente; e a auditoria nos dados de proveniência pode ser determinada pela instituição quando houver a necessidade de fiscalização de alguma situação informacional.

Ainda conforme a Figura 2, a modelagem apresentada do método aqui proposto foi instanciada na perspectiva das ligações entre os componentes do modelo PROV-DM, sendo possível observar a organização das relações do método, que podem ser descritas da seguinte maneira:

- (i) *actedOnBehalfOf*: o administrador da rede do SIS (*Agent*) tem suas autoridades e responsabilidades no gerenciamento dos SIS em relação aos processos informacionais.
- (ii) *wasAssociatedWith*: as atividades para captura de dados de proveniência (*Activity*), armazenamento e preparação dos dados de proveniência (*Activity*), manipulação dos dados de proveniência (*Activity*) e análise dos dados de proveniência (*Activity*) são atividades associadas

com o administrador da rede do SIS (*Agent*) para total acesso aos dados na realização das soluções e tomadas de decisões necessárias para obter os dados de proveniência. O armazenamento e a preparação dos dados de proveniência (*Activity*) também estão associados aos diferentes SIS (*Entity*), por se tratar de diversos tipos de armazenamento de dados.

- (iii) *wasAttributeTo*: os diferentes SIS (*Entity*) estão atribuídos ao administrador da rede do SIS (*Agent*), o qual tem autorização de acesso aos dados conforme a demanda informacional;
- (iv) *wasDerivedFrom*: como os diferentes SIS (*Entity*) podem ser derivados de outros sistemas integrados e descentralizados em seu contexto geral, esta relação fica entre essa entidade. As análises dos dados de proveniência (*Activity*) são derivadas da manipulação dos dados de proveniência (*Activity*), por se tratar de trocas constantes de dados para chegar a um resultado satisfatório.
- (v) *wasInformedBy*: as atividades para captura de dados de proveniência (*Activity*), armazenamento e preparação dos dados de proveniência (*Activity*) e a manipulação dos dados de proveniência (*Activity*) trocam informações entre si para completar os processos necessários, chegando ao último processo de análise dos dados de proveniência (*Activity*) para promover soluções informacionais sobre a origem e rastreabilidade dos dados de proveniência nos diferentes SIS.
- (vi) *Used*: no módulo de captura de dados de proveniência (*Activity*) são usados os diferentes SIS (*Entity*) para gerar os processos necessários para obtenção dos dados de proveniência dos diferentes SIS.

Portanto, por meio do instanciamento do modelo de proveniência de dados PROV-DM, é possível justificar que essas instâncias foram necessárias para a proposta do método deste estudo no intuito do gerenciamento dos dados de SIS.

4. CONSIDERAÇÕES FINAIS

Este estudo foi realizado com base no instanciamento do modelo de proveniência de dados genérico PROV-DM recomendado pela W3C. Foi possível propor um método para o gerenciamento de dados de proveniência em SIS, sendo aplicável para diferentes contextos de saúde.

A estratégia computacional que envolverá o gerenciamento dos dados de proveniência em SIS, assim como a execução dos módulos apresentados no método aqui proposto, será realizada na continuidade deste estudo, por meio de um piloto para validação do método a fim de responder à pergunta levantada neste estudo. Futuros trabalhos incluem a utilização

de tecnologias *blockchain* aplicadas em SIS com intuito de contribuir para a interoperabilidade, rastreabilidade, imutabilidade e segurança dos dados de proveniência, bem como testar o método para garantir um fluxo informacional em ambientes computacionais ligados a uma rede de saúde descentralizada. E, também, posteriormente, adicionar outros modelos de proveniência de dados indicados pela W3C e tecnologias de *web* semântica ao método aqui proposto com intuito de incorporar outras funcionalidades. Posteriormente, os resultados e a discussão da validação e aplicação do método serão apresentados em novos estudos.

5. CONFLITOS DE INTERESSE

Os autores declaram não haver conflito de interesses.

6. DECLARAÇÃO DE CONTRIBUIÇÃO

Conceitualização, Investigação, Metodologia, Visualização, Redação – rascunho original: Márcio José Sembay

Administração do projeto, Supervisão: Douglas Dyllon Jeronimo de Macedo

Validação, Redação – revisão e edição: Márcio José Sembay, Douglas Dyllon Jeronimo de Macedo

7. DECLARAÇÃO DE CONSENTIMENTO DE DADOS

Nenhum dado foi gerado durante o desenvolvimento deste estudo. 

8. REFERÊNCIAS

- BELHAJJAME, K., B'FAR, R., CHENEY, J., COPPENS, S., CRESSWELL, S., GIL, Y., GROTH, P., KLYNE, G., LEBE, T., MCCUSKER, J., MILES, S., MYERS, J., & SAHOO, S. (2013, April 30). *PROV-DM: The PROV Data Model. W3C Recommendation*. <https://www.w3.org/TR/prov-dm/>
- CHIASSON, M. W., & DAVIDSON, E. (2004). Pushing the contextual envelope: developing and diffusing is theory for health information systems research. *Information And Organization*, 14(3), 155-188. <http://dx.doi.org/10.1016/j.infoandorg.2004.02.001>
- DAVIDSON, S. B., & FREIRE, J. (2008) Provenance and scientific workflows: challenges and opportunities. In *ACM Sigmod International Conference On Management Of Data*, Vancouver. <https://vgc.poly.edu/~juliana/pub/freire-tutorial-sigmod2008.pdf>

- FERREIRA, J. E. DE S. M., OLIVEIRA, L. R. DE, MARQUES, W. S., LIMA, T. S. DE, BARBOSA, E. DA S., CASTRO, R. R., & GUIMARÃES, J. M. X. (2020). Sistemas de Informação em Saúde no apoio à gestão da Atenção Primária à Saúde: revisão integrativa. *Revista Eletrônica de Comunicação, Informação e Inovação em Saúde*, 14(4), 970-982. <http://dx.doi.org/10.29397/reciis.v14i4.1923>
- FREIRE, J., KOOP, D., SANTOS, E., & SILVA, C. T. (2008). Provenance for computational tasks: a survey. *Journal Computing in Science and Engineering*, 10(3), 11-21. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.147.3801&rep=rep1&type=pdf>
- FREUND, G. P., SEMBAY, M. J., & MACEDO, D. D. J. (2019). Proveniência de Dados e Segurança da Informação: relações interdisciplinares no domínio da ciência da informação. *Revista Ibero-Americana de Ciência da Informação*, 12(3), 807-825. <https://doi.org/10.26512/rici.v12.n3.2019.21203>
- GARCIA, P. T., & REIS, R. S. (2016). *Gestão Pública em Saúde: sistemas de informação de apoio à gestão em saúde*. UFMA.
- GROTH, P., & MOREAU, L. (2013). PROV-Overview: an overview of the prov family of documents. <https://www.w3.org/TR/prov-overview/>
- MOREAU, L., GROTH, P., CHENEY, J., LEBE, T., & MILES, S. (2015). The rationale of PROV. *Journal of Web Semantics*, 35, 235-257. <https://doi.org/10.1016/j.websem.2015.04.001>
- SEMBAY, M. J., MACEDO, D. J., & DUTRA, M. L. A. (2020a) Method for collecting provenance data: a case study in a Brazilian hemotherapy center. In *Lecture Notes of the Institute for Computer Sciences, Social InformaTIC and Telecommunications Engineering*. 1. ed. (pp. 89-102). Springer International Publishing. Doi 10.1007/978-3-030-50072-6_8
- SEMBAY, M. J., MACEDO, D. J., & DUTRA, M. L. (2020b). A proposed approach for provenance data gathering. In *Mobile Networks & Applications*, (pp. 1-13). <https://doi.org/10.1007/s11036-020-01648-7>
- SIMMHAN, Y., PLAILE, B., & GANNON, D. (2006). A framework for collecting provenance in data-centric scientific workflows. *Proceedings International Conference On Web Services, ICWS'06*, Chicago, EUA, Doi 10.1109/ICWS.2006.5
- TAN, W. C. (2008). *Provenance in databases: past, current, and future*. IEEE Data Eng. Bull, 30(4), 3-12, 2008.
- W3C (2013). PROV-DM: The PROV Data Model. W3C. <http://www.w3.org/TR/prov-dm/>

