

HL7 FHIR BASEADO EM W3C PROV PARA ALCANÇAR A PROVENIÊNCIA DE DADOS EM SISTEMAS DE INFORMAÇÃO EM SAÚDE

HL7 FHIR based on W3C PROV to achieve Data Provenance in Health Information Systems

**Márcio José Sembay¹, Douglas Dyllon Jeronimo de Macedo²,
Alexandre Augusto Gimenes Marquez Filho³**

(1) Universidade Federal de Santa Catarina (UFSC), marcio.sembay@posgrad.ufsc.br

(2) Universidade Federal de Santa Catarina (UFSC), douglas.macedo@ufsc.br

(3) Universidade Federal de Santa Catarina (UFSC), alexandre.agmf@gmail.com

Resumo:

A proveniência de dados e a interoperabilidade são requisitos fundamentais para os Sistemas de Informação em Saúde (SIS) para garantir não apenas o intercâmbio efetivo e eficiente dos dados de saúde, mas também, para definir fontes de dados confiáveis. Esforços de vários anos em padrões internacionais como *Health Level 7 (HL7)* e *Fast Healthcare Interoperability Resources (FHIR)* com base no modelo de proveniência de dados do *World Wide Web Consortium (W3C)*, mais conhecido como W3C PROV, confirmam esses requisitos. O objetivo desta pesquisa bibliográfica e exploratória, com abordagem qualitativa, é analisar e caracterizar o uso do padrão HL7 FHIR baseado em W3C PROV, possibilitando observar o alcance da proveniência de dados em SIS. As análises realizadas neste estudo basearam-se na literatura, possibilitando caracterizar a proveniência de dados no uso do HL7 FHIR com base no W3C PROV, além de destacar a importante intersecção desses elementos no cenário dos SIS. Assim, este estudo considera o HL7 FHIR baseado no W3C PROV, como um dos padrões internacionais que mais contribuem para alcançar a proveniência de dados em SIS.

Palavras-chave: Proveniência de Dados; W3C PROV; Sistemas de Informação em Saúde; Interoperabilidade; HL7 FHIR.

Abstract:

Data provenance and interoperability are fundamental requirements for Health Information Systems (HIS) to ensure not only the effective and efficient exchange of health data, but also to define reliable data sources. Multi-year efforts at international standards such as Health Level 7 (HL7) and Fast Healthcare Interoperability Resources (FHIR) based on the World Wide Web Consortium (W3C) data provenance model, better known as the W3C PROV, confirm these requirements. The objective of this bibliographical and exploratory research with a qualitative approach is to analyze and characterize the use of the HL7 FHIR standard based on W3C PROV, making it possible to observe the reach of the data provenance in HIS. The analyzes carried out in this study were based on the literature, making it possible to characterize the data provenance in the use of the HL7 FHIR based on the W3C PROV, in addition to highlighting the important intersection of these elements in the HIS scenario. Thus, this study considers the HL7 FHIR based on the W3C PROV, as one of the international standards that most contribute to achieving data provenance in HIS.

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

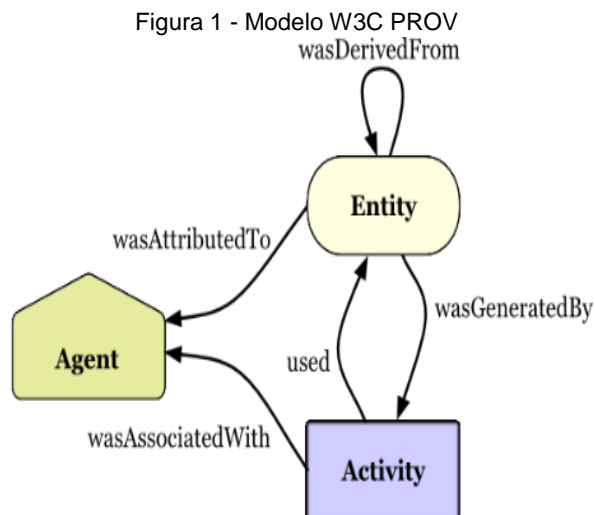
1. Introdução

Os Sistemas de Informação em Saúde (SIS) podem ser considerados como um Sistema de Informação (SI) que compõe a coleta, o processamento, a comunicação e o uso de informações fundamentais para aperfeiçoar a eficiência dos serviços de saúde (WORLD HEALTH ORGANIZATION, 2004). Sembay e Macedo (2022) ressaltam que os SIS têm a capacidade de gerar e armazenar grandes volumes de dados, tornando-se um desafio atual.

Nesse sentido, as abordagens da proveniência de dados contribuem nesse quesito para responder perguntas como: “por quê?”, “como?”, “onde?”, “quando?”, “por

quem?” e “para quê?” os dados foram produzidos nos SIS, contribuindo para a rastreabilidade e gerenciamento dos dados de proveniência nos SIS (SEMBAY; MACEDO, 2022). A proveniência de dados identifica transformações pelas quais os dados passaram ao longo do tempo, e no contexto da saúde vive um cenário de pesquisa crescente, principalmente nas tecnologias aplicadas nessa área que vêm obtendo resultados expressivos (SEMBAY; MACEDO; DUTRA, 2020a; SEMBAY; MACEDO; DUTRA, 2020b). No estudo de Freund, Sembay e Macedo (2019), os autores consideram que a proveniência de dados pode ser considerada como um

requisito importante para estabelecer confiabilidade e prover segurança em SI. Dessa forma, na Figura 1, destacamos o modelo de proveniência de dados PROV criado em 2013. (GIL; MILES, 2013).



Fonte: Gil e Miles (2013).

Conforme a Figura 1, existem três tipos diferentes de elementos no W3C PROV (GIL; MILES, 2013): *Entity* (entidades), *Activity* (atividades) e *Agent* (agente). Uma entidade é um tipo físico, digital, conceitual ou outro tipo de coisa com alguns aspectos fixos. Uma atividade é algo que age sobre ou com entidades; pode incluir consumir, processar, transformar, modificar, realocar, usar ou gerar entidades. Finalmente, um agente é algo que carrega alguma forma de responsabilidade por uma atividade que ocorre, pela existência de uma entidade ou pela atividade de outro agente. Entre as entidades, atividades e agentes, pode haver diferentes tipos de relações como apresentado na Figura 1.

Nesse sentido, é importante ressaltar que o W3C PROV aplicado em SIS, pode contribuir para que os dados de saúde possam ser interoperáveis. Assim, destacamos a interoperabilidade de software no domínio da saúde utilizando um padrão consistente como *Health Level Seven (HL7)* e *Fast Healthcare Interoperability Resources (FHIR)*. O HL7 é uma organização de desenvolvimento de padrões para troca, integração, compartilhamento e recuperação de informações de saúde. O FHIR criado pelo HL7, é um padrão que descreve

formatos e elementos de dados para troca de registros eletrônicos de saúde interoperáveis (HL7 FHIR, 2022a). Porém, Landgrebe e Smith (2011) afirmam que o HL7 FHIR pode ter diversos desafios de interoperabilidade. Nesse sentido, Margueri *et al* (2020) ressaltam as bases de proveniência de recurso do HL7 FHIR com o W3C PROV, possibilitando além de compartilhar atributos de proveniência, sanar problemas de interoperabilidade semântica.

Assim, para que todas as relações de proveniência de dados do HL7 FHIR baseadas no W3C PROV sejam realizadas, é importante destacar o padrão HL7 FHIR 5Ws que incluem as seguintes questões: *Who?* (Quem?), *What?* (O quê?), *When?* (Quando?), *Where?* (Onde?), *Why?* (Por quê?). Daí o nome deste padrão (5Ws) (HL7 FHIR, 2022b). Esse padrão também inclui muitos recursos, possibilitando mapear a proveniência com base no W3C PROV (HL7 FHIR, 2022c).

Portanto, este estudo tenta responder a seguinte questão: Qual a importância do uso do padrão HL7 FHIR com base no W3C PROV para alcançar a proveniência de dados em cenários de SIS?

2. Objetivo

Este estudo tem como objetivo analisar e caracterizar o uso do padrão HL7 FHIR baseado em W3C PROV, possibilitando observar o alcance da proveniência de dados em SIS.

3. Procedimentos Metodológicos

No presente estudo realizamos uma pesquisa de caráter bibliográfico e exploratório de natureza qualitativa acerca da temática do uso do HL7 FHIR, com base no W3C PROV para alcançar a proveniência de dados em SIS. Adiante, os métodos deste estudo são descritos em três passos para a realização das análises.

O primeiro passo é caracterizar a proveniência de dados em relação ao HL7 FHIR com base no W3C PROV. Para isso, comparamos as questões base do padrão HL7 FHIR 5Ws (HL7 FHIR, 2022b) responsável pelo mapeamento dos recursos de proveniência em SIS, juntamente com as questões base da proveniência de dados

definidas por Buneman, Khanna e Wang-Chiew (2001). Logo depois, são caracterizadas as aplicações da proveniência de dados de parte da taxonomia definida por Simmhan *et al.* (2005): i) qualidade dos dados (baseado nos dados originais e nas suas derivações); ii) trilha de auditoria (detectando erros na geração de dados); iii) receitas de replicação (repetição da derivação de dados e sua atualização); iv) atribuição (direitos autorais e propriedade de dados); e, v) informativo (interpretação de dados). Essas aplicações são caracterizadas com o HL7 FHIR baseado no W3C PROV de acordo com Margheri *et al.* (2020) e Kohlbacher *et al.* (2018), destacando sua importância de aplicação em SIS.

O segundo passo é demonstrar por meio da literatura científica a importância do uso do HL7 FHIR baseado no W3C PROV em SIS. Para validar esse passo, selecionamos cinco artigos na literatura científica de pesquisas recentes que envolvessem soluções dirigidas para alcançar a proveniência de dados em SIS por meio do HL7 FHIR baseado em W3C PROV, observando as importantes visões consolidadas nas pesquisas nesta área. Assim, deste corpus científico foram possíveis realizar análises para apresentar os principais termos e tecnologias identificadas no uso do HL7 FHIR baseado no W3C PROV em diferentes tipos de SIS.

Por fim, no terceiro passo, apresentamos a proveniência de dados como intersecção entre o HL7 FHIR com base no W3C PROV e HL7 FHIR 5Ws, ambos como fatores de importância e necessidade em relação à interoperabilidade existente nos cenários dos SIS.

4. Análises e Resultados

As análises realizadas basearam-se na metodologia descrita neste estudo e seguem descritas nas próximas subseções.

4.1 Caracterizando a proveniência de dados em HL7 FHIR com base no W3C PROV para SIS

A Tabela 1 apresenta as questões do padrão HL7 FHIR 5Ws definidos pelo HL7 FHIR (2022b) com as questões de

proveniência de dados definidas por Buneman, Khanna e Wang-Chiew (2001).

Tabela 1 - Comparando as questões base - HL7 FHIR 5Ws com a proveniência de dados em SIS

Padrão HL7 FHIR 5Ws	Proveniência de Dados
Who? (Quem?)	Who? (Quem?)
What? (O que?)	How? (Como?)
When? (Quando?)	When? (Quando?)
Where? (Onde?)	Where? (Onde?)
Why? (Por quê?)	Why? (Por quê?)

Fonte: dados da pesquisa (2022).

Pode-se observar na Tabela 1 que praticamente todas as questões comparadas são iguais, confirmando a necessidade da existência da proveniência de dados para que haja interoperabilidade de dados realizada no HL7 FHIR com base no W3C PROV, que de fato, para o domínio dos SIS é significativo. Outro ponto importante observado na Tabela 1, é que o padrão HL7 FHIR 5Ws possibilita o mapeamento de recursos de proveniência por se basear nas premissas do W3C PROV e conseqüentemente contribui para a segurança usada no rastreamento, status, histórico e autoria de dados para diferentes SIS. Já na Tabela 2 é apresentada a caracterização das aplicações da proveniência de dados definidas por Simmhan *et al.* (2005) em relação ao HL7 FHIR baseado no W3C PROV conforme indicado por Margheri *et al.* (2020) e Kohlbacher *et al.* (2018).

Tabela 2 - Caracterizando aplicações da proveniência de dados com HL7 FHIR baseado no W3C PROV para SIS

Aplicações da Proveniência de Dados	HL7 FHIR baseado em W3C PROV
Qualidade dos Dados	Estima a qualidade e a confiabilidade dos dados com base nas transformações de origem e dados.
Trilha de Auditoria	Determina o uso de recursos e detecta erros nos dados de geração. Evento de Auditoria de Recursos.
Receitas de replicação	Permite a replicação e derivação de dados através de informações detalhadas de procedência.
Atribuição	Isso possibilita determinar a propriedade e os direitos sobre os dados.
Informativo	O <i>Representational State Transfer (RESTful)</i> , uma arquitetura regrada de software

	mantém o controle das informações de proveniência relacionadas a recursos.
--	--

Fonte: dados da pesquisa 2022.

A Tabela 2 caracteriza fatores que contribuem para os SIS nos seguintes aspectos: i) gestão de dados para profissionais de saúde; ii) controle social dos SIS, favorecendo a coleta qualificada e a interoperabilidade de informações (dados sensíveis) que apoiam o acesso à saúde, bem como a alocação de recursos de proveniência; e, iii) possibilitando a interoperabilidade de dados nos SIS, que de fato é um dos maiores desafios do setor.

Ainda, enfatizando a Tabela 2, no estudo de Sembay, Macedo e Marquez Filho (2022), os autores comparam as mesmas aplicações da proveniência de dados com a tecnologia Blockchain, também responsável por contribuir com a proveniência de dados em SIS.

4.2 Usando HL7 FHIR baseado no W3C PROV em SIS

Com base na metodologia descrita para esta pesquisa, destacamos cinco estudos encontrados na literatura, os quais possibilitaram extrair oito principais termos apresentados na Tabela 3.

Tabela 3 - Principais termos que caracterizam o uso do HL7 FHIR baseado no W3C PROV em SIS

Autor/Ano	Termos	HL7 FHIR baseado em W3C PROV
Bittins <i>et al.</i> (2021), Margheri <i>et al.</i> (2020), Massi <i>et al.</i> (2018), Kohlbacher <i>et al.</i> (2018), Mense e Blobel (2017)	Interoperabilidade	Sintática e semântica
	Segurança	<i>Hyper Text Transfer Protocol Secure (HTTPS)</i> , - <i>Secure Sockets Layer (SSL)</i>
Bittins <i>et al.</i> (2021) Massi <i>et al.</i> (2018), Margheri <i>et al.</i> (2020)	Criptografia	Além dos protocolos HL7 v2, ele suporta outras normas
Bittins <i>et al.</i> (2021) Margheri <i>et al.</i> (2020), Massi <i>et al.</i> (2018), Kohlbacher <i>et al.</i> (2018),	Privacidade	Implementar privacidade e fácil devido às <i>Application Programming Interface (APIs)</i>

Mense e Blobel (2017)		
Bittins <i>et al.</i> (2021) Kohlbacher <i>et al.</i> (2018)	Compatibilidade	Facilmente integrado com: <i>Digital Imaging and Communications in Medicine (DICOM)</i>
Massi <i>et al.</i> (2018), Kohlbacher <i>et al.</i> (2018)	Flexibilidade	Muito flexível
Bittins <i>et al.</i> (2021) Margheri <i>et al.</i> (2020), Massi <i>et al.</i> (2018), Kohlbacher <i>et al.</i> (2018), Mense e Blobel (2017)	Proveniência	Modelo W3C PROV
	Arquitetura	Com base no protocolo <i>RESTful</i>

Fonte: dados da pesquisa 2022.

Em relação aos oito termos apresentados na Tabela 3, podemos destacar que todos são discutidos no W3C PROV para que o processo de proveniência seja possível através do padrão HL7 FHIR. Indo além, a Tabela 4 apresenta três tipos de SIS: (*Electronic Health Record (EHR)*), *Personal Health Record (PHR)* e *Electronic Medical Record (EMR)* e as principais tecnologias utilizadas em conjunto de aplicações do HL7 FHIR com base no W3C PROV, sendo elas: Blockchain, *Integrating the Healthcare Enterprise (IHE)*, DICOM e *Clinical Document Architecture (CDA)*.

Tabela 4 - Tipos de SIS e tecnologias que contribuem para o uso do HL7 FHIR baseado no W3C PROV em SIS

Autor/Ano	Tipos de SIS	Principais tecnologias
Bittins <i>et al.</i> (2021)	EHR	HL7 FHIR, IHE, DICOM, CDA e Blockchain
Margheri <i>et al.</i> (2020)	EHR	HL7 FHIR, Blockchain e IHE
Massi <i>et al.</i> (2018)	EHR	HL7 FHIR, Blockchain e IHE
Kohlbacher <i>et al.</i> (2018)	EMR	HL7 FHIR, IHE, DICOM e CDA
Mense e Blobel (2017)	EHR e PHR	HL7 FHIR e CDA

Fonte: dados da pesquisa (2022).

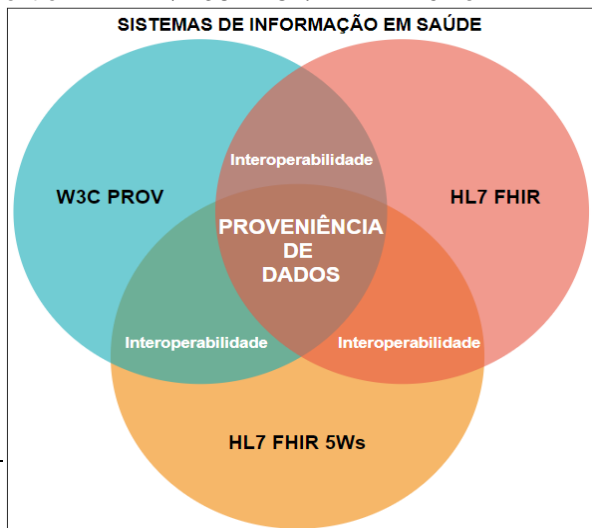
Observa-se na Tabela 4 que o EHR se destaca nos estudos de Bittins *et al.* (2021), Margheri *et al.* (2020), Massi *et al.* (2018) e

Mense e Blobel (2017), por se tratar do SIS mais popular e mais utilizado em vários países. Logo após, o EMR que garante o registro médico protegido apresentado no estudo de Kohlbacher *et al.* (2018) e o PHR, relacionado a dispositivos móveis, apresentado no estudo de Mense e Blobel (2017), são considerados essenciais pelos autores. É importante ressaltar que, tanto no EHR, PHR e EMR os dados podem ser armazenados de forma distribuída, possibilitando a proveniência de dados em saúde. A Tabela 4 também destaca o uso das seguintes tecnologias que auxiliam no alcance da proveniência de dados em SIS: Blockchain nos estudos de Bittins *et al.* (2021), Margheri *et al.* (2020) e Massi *et al.* (2018) no quesito de interoperabilidade e imutabilidade dos dados; IHE nos estudos de Bittins *et al.* (2021), Margheri *et al.* (2020), Massi *et al.* (2018) e Kohlbacher *et al.* (2018) permitindo a melhoria da interoperabilidade; e por fim, CDA encontrado nos estudos de Bittins *et al.* (2021), Kohlbacher *et al.* (2018) e Mense e Blobel (2017) e DICOM encontrado nos estudos de Bittins *et al.* (2021) e Kohlbacher *et al.* (2018) relacionados à alta sensibilidade dos dados de saúde.

4.3 Proveniência de dados como intersecção entre o HL7 FHIR com base no W3C PROV e HL7 FHIR 5Ws no cenário de SIS

A Figura 2 apresenta a proveniência de dados como intersecção entre HL7 FHIR/W3C PROV/HL7 FHIR 5Ws para mapear a proveniência de recursos nos cenários de SIS.

Figura 2 - Proveniência de dados como intersecção entre HL7 FHIR/W3C PROV/HL7 FHIR 5Ws



Fonte: Elaborada pelo autor.

Na Figura 2, a proveniência de dados, mostra-se como o principal insumo e benefício de trabalho do HL7 FHIR baseado no W3C PROV, oportunizando a geração de conhecimento e contribuindo para a tomada de decisão concisa nos setores de saúde.

5. Considerações Finais

Os resultados obtidos das análises realizadas vão ao encontro do objetivo proposto nesta pesquisa e têm o potencial de contribuir com futuros estudos que visem melhorar as práticas associadas ao uso do HL7 FHIR baseado em W3C PROV, que apesar de enfrentar desafios de interoperabilidade, muitas vezes únicos para cada instituição, continua sendo aperfeiçoado para contribuir no alcance da proveniência de dados em SIS. A importância do uso do padrão HL7 FHIR com base no W3C PROV foi evidenciada neste recente e dinâmico campo de estudos que se forma na intersecção da proveniência de dados no cenário de SIS. Considera-se ainda que as contribuições identificadas neste estudo são representativas do potencial impacto exercido nas práticas científicas da Ciência da Computação e da Ciência da Informação, prospectando trabalhos futuros que visem ampliar as reflexões e contribuições aqui descritas.

6. Referências

BITTINS, Soeren *et al.* Healthcare data management by using blockchain technology. *In: Applications of blockchain in healthcare*. Singapore: Springer, 2021. p. 1-27.

BUNEMAN, Peter; KHANNA, Sanjeev; WANG-CHIEW, Tan. **Why and where:** a characterization of data provenance. *In: INTERNATIONAL CONFERENCE ON DATABASE THEORY*. Berlin, Heidelberg: Springer, 2001. p. 316-330.

FREUND, Gislaine P.; SEMBAY, Márcio J.; MACEDO, Douglas D. J. de. Proveniência de Dados e Segurança da Informação: relações interdisciplinares no domínio da ciência da informação. **Revista Ibero-Americana de Ciência da Informação**, [s. l.], v. 12, n. 3, p.

807-825, 14 set. 2019. Disponível em: <https://periodicos.unb.br/index.php/RICI/articloe/view/21203/23548>. Acesso em: 20 maio 2022.

GIL, Yolanda; MILES, Simon. **PROV Model Primer**. In: W3C, 2013 April 30. W3C Working Group Note. Disponível em: <https://www.w3.org/TR/prov-primer/>. Acesso em: 02 ago. 2022.

HL7 FHIR. **Welcome to FHIR**. 2022a. Disponível em: Disponível em: <https://hl7.org/fhir/>. Acesso em: 02 ago. 2022.

HL7 FHIR. **5Ws Pattern**. Pattern FiveWs: Content. 2022b. Disponível em: <https://build.fhir.org/fivews.html>. Acesso em: 02 ago. 2022.

HL7 FHIR. **Provenance**. 2022c. Disponível em: <https://www.hl7.org/fhir/provenance.html>. Acesso em: 05 ago. 2022.

KOHLBACHER, Oliver *et al.* **Data Integration for Future Medicine (DIFUTURE): an architectural and methodological overview**. 2018. DOI <https://doi.org/10.3414/ME17-02-0022>

LANDGREBE, Jobst; SMITH, Barry. The HL7 approach to semantic interoperability. In: **Proceedings of the 2nd International Conference on Biomedical Ontology**. CEUR, vol. 833. pp. 139-146, 2011.

MARGHERI, Andrea *et al.* Decentralised provenance for healthcare data. **International Journal of Medical Informatics**, v. 141, p. 1-21, set. 2020. Elsevier BV. DOI <http://dx.doi.org/10.1016/j.ijmedinf.2020.104197>.

MASSI, Massimiliano *et al.* **Using PROV and blockchain to achieve health data provenance**. 2018. Disponível em: https://eprints.soton.ac.uk/421292/1/PROV_BC_Healthcare.pdf. Acesso em: 02 ago. 2022.

MENSE, Alexander; BLOBEL, Bernd. HL7 standards and components to support implementation of the European general data

protection regulation. **European Journal for Biomedical Informatics**, v. 13, n. 1, p. 27-33, 2017.

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**, v. 2, p. 192-201, Tallinn, Estonia: ColNes Publishing, 2022. DOI: 10.47909/anis.978-9916-9760-3-6.101.

SEMBAY, Márcio José; MACEDO, Douglas Dyllon Jeronimo de; MARQUEZ FILHO, Alexandre Augusto Gimenes. Identification of the relationships between Data Provenance and Blockchain as a contributing factor for Health Information Systems. In: EAI INTERNATIONAL CONFERENCE ON DATA AND INFORMATION IN ONLINE ENVIRONMENTS, 3., 2022, Florianópolis. **Proceedings** [...]. Florianópolis: DIONE 2022, p. 189-203.

SEMBAY, Márcio José; MACEDO, Douglas Dyllon Jeronimo de; DUTRA, Moisés Lima. L. A proposed approach for provenance data gathering. **Mobile Networks & Applications**, p. 1-13, 2020a. DOI 10.1007/s11036-020-01648-7

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: **Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering**. 1. ed. [S. l.]: Springer International Publishing, 2020b. v. 1, p. 89-102. DOI 10.1007/978-3-030-50072-6_8

SIMMHAN, Yogesh L. *et al.* A survey of data provenance techniques. **Computer Science Department**, Indiana University, Bloomington IN, v. 47405, p. 69, 2005.

WORLD HEALTH ORGANIZATION *et al.* **Developing health management information systems: a practical guide for developing countries**. 2004.