



XXII Encontro Nacional de Pesquisa em Ciência da Informação – XXII ENANCIB

ISSN 2177-3688

GT-11 – Informação & Saúde

PROVENIÊNCIA DE DADOS EM SISTEMAS DE INFORMAÇÃO EM SAÚDE

DATA PROVENANCE IN HEALTH INFORMATION SYSTEMS

Márcio José Sembay. UFSC

Douglas Dyllon Jeronimo de Macedo. UFSC

Alexandre Augusto Gimenes Marquez Filho. UFSC

Modalidade: Resumo Expandido

Resumo: Para que a proveniência de dados em Sistemas de Informação em Saúde tenha êxito, padrões de interoperabilidade como *Health Level 7* e *Fast Healthcare Interoperability Resources* são essenciais. Assim, este estudo em andamento, visa propor um esforço computacional utilizando-se desses padrões para o desenvolvimento de um método de gerenciamento de dados de proveniência em Sistemas de Informação em Saúde a ser aplicado em um Hospital Universitário. A metodologia de pesquisa deste estudo é aplicada em um estudo de caso, com abordagem quantitativa e qualitativa. Com a proposta do método em andamento, destaca-se contribuições significativas com relação aos processos de gerenciamento e rastreamento dos dados de proveniência de saúde, além da manutenção da segurança desses dados e da relação custo-benefício obtido para a instituição de saúde.

Palavras-Chave: Proveniência de Dados. Sistemas de Informação em Saúde. Informação em Saúde.

Abstract: For data provenance in Health Information Systems to be successful, interoperability standards such as Health Level 7 and Fast Healthcare Interoperability Resources are essential. Thus, this ongoing study aims to propose a computational effort using these standards for the development of a management method of provenance data in Health Information Systems to be applied in a University Hospital. The research methodology of this study is applied in a case study, with a quantitative and qualitative approach. With the proposed method in progress, significant contributions are highlighted in relation to the processes of management and tracking of health provenance data, in addition to maintaining the security of these data and the cost-benefit ratio obtained for the health institution.

Keywords: Data Provenance. Health Information Systems. Health Information.



1 INTRODUÇÃO

Os Sistemas de Informação em Saúde (SIS) são definidos por Haux (2006, p. 270) 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 relacionado 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)¹ e *Fast Healthcare Interoperability Resources* (FHIR)² baseado no modelo de proveniência da *World Wide Web Consortium* (W3C) mais conhecido como W3C PROV³, são essenciais.

O trabalho proposto compõe parte de uma pesquisa acadêmica para a obtenção do grau de doutor em Ciência da Informação, possuindo caráter aplicado em um estudo de caso, com abordagem quantitativa e qualitativa e em desenvolvimento. A questão que norteia esse resumo expandido é a seguinte: 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? Assim, o presente trabalho tem por objetivo propor um esforço computacional para o desenvolvimento de um método de gerenciamento de dados de

¹ <https://www.hl7.org/>

² <http://hl7.org/fhir/>

³ <https://www.w3.org/TR/prov-primer/>



proveniência em SIS, que seja capaz de apontar em quais aspectos são necessários os processos de gerenciamento desses dados, o qual corresponde a um dos objetivos específicos da tese.

A proposta do método utiliza-se como base de estratégia computacional, o HL7 FHIR baseado em W3C PROV, tendo como cenário de aplicação o Sistema Integrado de Telemedicina e Telessaúde (STT)⁴ da Universidade Federal de Santa Catarina (UFSC) no que compete ao telediagnóstico por imagens médicas.

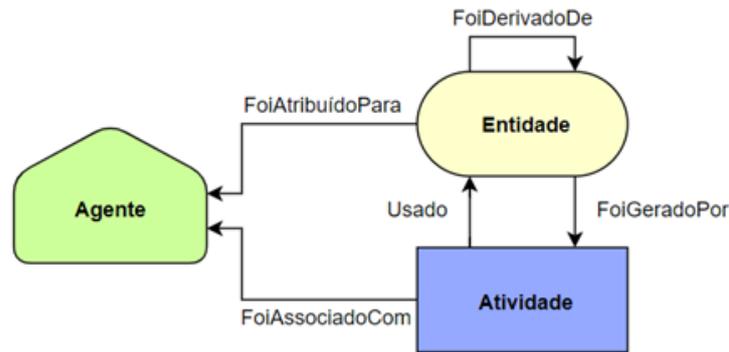
2 PROVENIÊNCIA DE DADOS

A proveniência de dados, também chamada de 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 (BUNEMAN; KHANNA; TAN, 2001). Nesse sentido, a proveniência de dados pode ser considerada um requisito importante para estabelecer confiabilidade e prover segurança em sistemas computacionais de informação. Concentra-se na identificação das fontes de dados e nas transformações pelas quais passou ao longo do tempo. Está relacionada a diferentes cenários de aplicação, sendo o de saúde foco deste estudo (FREUND; SEMBAY; MACEDO, 2019; SEMBAY; MACEDO; DUTRA, 2020).

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). A proveniência de dados apresenta vários modelos na literatura, mas para esse estudo em específico, a Figura 1 apresenta o modelo de proveniência de dados W3C PROV traduzido de Gil e Miles (2013).

⁴ <https://site.telemedicina.ufsc.br/>

Figura 1 – Modelo de proveniência de dados W3C PROV.



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

Conforme apresentado na Figura 1, o modelo de proveniência de dados W3C PROV surge 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 (GIL; MILES, 2013): a) (entidade), que pode ser algo físico, digital ou conceitual, real ou imaginário; b) (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) (agente) sendo alguma forma de responsabilidade por uma atividade que está ocorrendo, pela existência de uma entidade ou pela atividade de outro agente. Entre as entidades, atividades e agentes, pode haver vários tipos diferentes de relações como mostrado na Figura 1 (GIL; MILES, 2013).

3 SISTEMAS DE INFORMAÇÃO EM SAÚDE

Os SIS podem ser conceituados como um Sistema de Informação (SI) que integra 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, é importante ressaltar 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 desafios 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 (WHO-GLOBAL OBSERVATORY FOR eHEALTH, 2010).

Dessa forma, em um contexto brasileiro, agregada aos avanços e a expansão do acesso às tecnologias de informação e informática que facilitaram e ampliaram o uso de dados gerados pelos setores da saúde, expandiu o interesse pela utilização dos SIS de base nacional, como é o caso do Sistema Único de Saúde (SUS) (VICTORA; AQUINO; LEAL, MONTEIRO; BARROS; SZWARCOWALD, 2011).

Entretanto, uma preocupação a destacar é a utilização de padrões internacionais de interoperabilidade de dados entre SIS. Existem diversos padrões, mas para este estudo destacamos o HL7 FHIR. Assim, a interoperabilidade de software no domínio da saúde pode ser realizada utilizando padrões consistentes como HL7, uma organização de desenvolvimento de padrões para troca, integração, compartilhamento e recuperação de informações de saúde (OLARONKE; SORIYAN; GAMBO; OLALEKE, 2013). Nesse sentido, o FHIR criado pelo HL7, é outro padrão importante que descreve formatos e elementos de dados e uma interface de programação de aplicativo para troca de registros eletrônicos de saúde interoperável (HONG; MORRIS; SEO, 2017). Portanto, o HL7 FHIR é um padrão que define recursos, incluindo definições de conteúdo, arquitetura, modelos e paradigmas de intercâmbio de informações em saúde (SILVA; SLOANE; COOPER, 2020).

4 INTEROPERABILIDADE E PROVENIÊNCIA DE DADOS EM SISTEMAS DE INFORMAÇÃO EM SAÚDE: HL7 FHIR BASEADO EM W3C PROV

Importante 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, 2022). Assim, a proveniência de dados no HL7 FHIR é baseada na especificação W3C PROV onde 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, apoiam o



W3C PROV (MARGHERI; MASI; MILADI; SASSONE; ROSENZWEIG, 2020). O W3C PROV, é a representação padrão de fato para informações de proveniência interoperáveis adotada pelo HL7 FHIR (PRASSER; KOHLBACHER; MANSMANN; BAUER; KUHN, 2018).

5 METODOLOGIA

Este estudo, em andamento, é pautado em uma pesquisa de natureza aplicada, que de acordo com Gerhardt e Silveira (2009, p. 35), “tem como objetivo gerar conhecimentos para aplicação prática, dirigidos à solução de problemas específicos”. O teor de aplicação deste estudo se refere à proposta do método para o gerenciamento de dados de proveniência em SIS a ser aplicado no STT/UFSC por meio de uma estratégia computacional em andamento.

Com relação à abordagem, classifica-se este estudo como quantitativo e qualitativo. A pesquisa quantitativa, segundo Knechtel (2014) é baseada no teste de uma teoria e composta por variáveis quantificadas em números, as quais são analisadas de modo estatístico. Neste estudo, o caráter quantitativo está atrelado à quantidade de dados de proveniência que o método irá coletar e gerenciar. Já em relação à pesquisa qualitativa, considera-se que a análise dos dados seja realizada de forma indutiva, em que o pesquisador desenvolve conceitos, ideias e entendimentos a partir de padrões encontrados nos dados (RENEKER, 1993). Nesse sentido, é importante ressaltar que a própria proveniência de dados qualifica os dados.

Por sua vez, em relação ao estudo de caso, segundo Yin (2001) é o método que visa compreender fenômenos sociais complexos, preservando as características holísticas e significativas dos eventos da vida real. O estudo de caso será aplicado no STT/UFSC no que compete ao telediagnóstico por imagens médicas em relação aos dados de proveniência gerados nos SIS descentralizados. O método de gerenciamento de dados de proveniência em SIS proposto para aplicação no STT/UFSC, utilizará em sua base computacional o HL7 FHIR baseado em W3C PROV entre outras tecnologias que ainda estão sendo definidas para a implementação final.

É importante ressaltar que, este estudo em andamento, já conta com a aprovação do Comitê de Ética em Pesquisa com Seres Humanos da Universidade Federal de Santa Catarina (CEPSH-UFSC) no que tange aos seguintes documentos: a) declaração que apresenta os efeitos legais do consentimento do responsável que autoriza o local (instituição) para a coleta dos dados; b) Termo de Compromisso para Utilização de Dados (TCUD); c) Termo de



Consentimento Livre e Esclarecido (TCLE); e, d) roteiro para a entrevista semiestruturada. Os pesquisadores envolvidos neste estudo se comprometem a zelar pela confidencialidade dos dados e privacidade dos participantes, de acordo com a Resolução CNS 466/2012 (BRASIL, 2012). Declaram também conhecer e cumprir os requisitos da Lei Geral de Proteção de Dados, nº 13.709, de 14 de agosto de 2018 (BRASIL, 2018). Isso se deve ao fato de que haverá tratamento de dados pessoais e dados pessoais sensíveis, necessários para a execução deste estudo.

O universo deste estudo se volta ao cenário do STT/UFSC, no qual se pretende aplicar o método. Com relação à população selecionada, serão utilizados os dados digitais e manuais inseridos nos SIS integrados no STT/UFSC. A amostra será realizada com base no repositório de dados de proveniência local (método proposto), o qual configurar-se-á em um banco de dados local que receberá dados do STT/UFSC por meio de estratégias computacionais a serem criadas dentro da mesma rede, sem acesso externo ou por pessoas não autorizadas.

A coleta de dados se dará por intermédio do método proposto que acessará os bancos de dados dos SIS integrados ao STT/UFSC. Dessa forma, os dados de proveniência serão armazenados pelo método para posteriormente serem processados e manipulados para geração de análises e consultas de forma detalhada, a fim de gerar conhecimento para uso em tomadas de decisão no cenário do STT/UFSC.

Para a análise e interpretação dos dados, serão consideradas a qualidade e a quantidade dos dados coletados pela estratégia computacional (método), aplicado nos SIS descentralizados e integrados ao STT/UFSC. Inserido ao método, no quesito de quantidade, ferramentas computacionais irão gerar gráficos relacionados aos dados analisados por especificidades, sendo que, no quesito de qualidade, a própria proveniência de dados qualificará os dados.

6 RESULTADOS PRELIMINARES

No desenvolvimento do método proposto para o gerenciamento de dados de proveniência em SIS no STT/UFSC, resultados preliminares puderam ser alcançados, como:

- Compreensão da importância de gerenciar os grandes volumes de dados de proveniência nos SIS descentralizados do STT/UFSC, por parte dos especialistas em banco de dados, processo que ainda não existia;



- Redução do tempo de rastreabilidade dos dados de proveniência nos diversos sistemas integrados e descentralizados do STT/UFSC;
- Redução de custos na obtenção dos dados de diferentes repositórios interligados do STT/UFSC; e,
- Possibilidades de melhorias dos fluxos informacionais que envolvem a gestão dos dados de proveniência em SIS do STT/UFSC.

7 CONSIDERAÇÕES PARCIAIS

A proveniência de dados em SIS pode contribuir para os processos de gerenciamento e rastreamento dos dados de saúde, mantendo a segurança desses dados e contribuindo também para o custo-benefício da instituição de saúde. Os SIS geram e armazenam grandes volumes de dados em seus repositórios descentralizados, o que torna de fato, um desafio no que diz respeito ao rastreamento e ao gerenciamento desses dados. Nesse quesito, a proveniência de dados contribui para garantir diversos benefícios de rastreabilidade nas fontes de dados de origem dos SIS, sendo útil para tomadas de decisão. O método de gerenciamento de dados de proveniência em SIS já em andamento, utiliza-se como base computacional o padrão HL7 FHIR baseado em W3C PROV para garantir os benefícios da proveniência de dados especificamente no STT/UFSC que possui diversas bases de dados descentralizadas.

Assim, espera-se que esse estudo obtenha como resultado da aplicação do método no STT/UFSC, uma contribuição significativa nos quesitos de identificação, rastreamento, gerenciamento e reuso desses dados possibilitando a geração de conhecimento, contribuindo para tomadas de decisão nesse cenário. Desse modo, a expectativa é que o resultado da aplicação do método sirva como piloto para adequação para outros cenários de saúde do STT/UFSC, além de outras instituições de saúde do país.

REFERÊNCIAS

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.



BRASIL. Resolução nº 466, de 12 de dezembro de 2012. Dispõe sobre diretrizes e normas regulamentadoras de pesquisas envolvendo seres humanos. **Diário Oficial [da] República Federativa do Brasil**, Brasília, DF, 13 jun. 2013. Disponível em: <http://bit.ly/1mTMIS3>. Acesso em: 29 abr. 2022.

BUNEMAN, Peter; KHANNA, Sanjeev; TAN, Wang-Chiew. Why and where: a characterization of data provenance. *In*: VAN DEN BUSSCHE, J., VIANU, V. (ed.). **Database theory: ICDT 2001**. Lecture Notes in Computer Science, v. 1973. Berlin, Heidelberg: Springer, 2001. p. 1-15. Disponível em: https://repository.upenn.edu/cgi/viewcontent.cgi?article=1209&context=cis_papers. Acesso em: 02 maio 2022.

DAVIDSON, Susan; FREIRE, Juliana. Provenance and scientific workflows: challenges and opportunities. *In*: ACM SIGMOD INTERNATIONAL CONFERENCE ON MANAGEMENT OF DATA, p. 1345-1350, 2008. **Proceedings [...]**. Disponível em: <https://vgc.poly.edu/~juliana/pub/freire-tutorial-sigmod2008.pdf>. Acesso em: 02 maio 2022.

FREUND, Gislaïne 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/article/view/21203/23548>. Acesso em: 20 maio 2022.

GERHARDT, Tatiana. E.; SILVEIRA, Denise; T. **Métodos de pesquisa**. Porto Alegre: Ed. da UFRGS, 2009.

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: 20 maio 2022.

GROTH, Paul; MOREAU, Luc (ed.). **PROV-Overview: an overview of the PROV family of documents World Wide Web Consortium**, Monograph (Project Report), 2013. Disponível em: <https://eprints.soton.ac.uk/356854/>. Acesso em: 15 abr. 2022.

HÅKANSSON, Stefan; GAVELIN, Carin. What do we really know about the cost-effectiveness of telemedicine? **Journal of Telemedicine and Telecare**, 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 InformaTIC**, v. 75, p. 268-281, 2006. DOI 10.1016/j.ijmedinf.2005.08.002.

HL7 International Inc. **Provenance**. 2022. Disponível em: <https://www.hl7.org/fhir/provenance.html>. Acesso em: 20 maio 2022.

HONG, Jaeki; MORRIS, Peter; SEO, Jonghwa. Interconnected personal health record ecosystem using IoT cloud platform and HL7 FHIR. *In*: **2017 IEEE international conference on**



healthcare informatics (ICHI), IEEE, 2017. p. 362-367.

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

MARGHERI, Andrea; MASI, Massimiliano; MILADI, Abdallah; SASSONE, Vladimiro; ROSENZWEIG, Jason. Decentralised provenance for healthcare data. **International Journal Medical Informatics**, p. 104197, 2020. DOI 10.1016/j.ijmedinf.2020.104197.

MELO, Edleusa T. de; MOTA, Francisca R. L.; ARAÚJO, Nelma C. Organização da informação no Instituto Médico Legal Estácio de Lima–Maceió/AL. **Revista de Saúde Digital e Tecnologias Educacionais**, Fortaleza, v. 7, n. esp. III, p. 185-203, fev. 2022. Disponível em: https://repositorio.ufc.br/bitstream/riufc/64044/1/2022_art_etmelo.pdf. Acesso em: 20 maio 2022.

OLARONKE, Iroju; SORIYAN, Abimbola; GAMBO, Ishaya; OLALEKE, Janet. Interoperability in healthcare: benefits, challenges and resolutions. **International Journal of Innovation and Applied Studies**, v. 3, n. 1, p. 262-270, 2013. Disponível em: <http://www.ijias.issr-journals.org/abstract.php?article=IJIAS-13-090-01>. Acesso em: 18 maio 2022.

PRASSER, Fabian; KOHLBACHER, Oliver; MANSMANN, Ulrich; BAUER, Bernhard; KUHN, Klaus A. **Data Integration for Future Medicine (DIFUTURE)**: an architectural and methodological overview, 2018. Disponível em: https://www.researchgate.net/publication/326465503_Data_Integration_for_Future_Medicine_DIFUTURE. 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**, v. 63, n. 4, p. 487-507, Oct. 1993.

SEMBAY, Márcio J.; MACEDO, Douglas D. J. 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**, volume 2. Tallinn, Estonia: ColNes Publishing, 2022. DOI 10.47909/anis.978-9916-9760-3-6.101.

SEMBAY, Márcio J.; MACEDO, Douglas D. J. de; DUTRA, Moisés L. A method for collecting provenance data: a case study in a Brazilian hemotherapy center. In: **International Conference on Data and Information in Online**. Springer, Cham, 2020. p. 89-102.

SEMBAY, Márcio J.; MACEDO, Douglas D. J. de; DUTRA, Moisés L. A Proposed Approach for Provenance Data Gathering. **Mobile Networks and Applications**, v. 26, n. 1, p. 304-318, 2021.

SILVA, Ricardo J.; SLOANE, Elliot B.; COOPER, Todd. Application of HL7® FHIR for device and health information system interoperability. In: **Clinical Engineering Handbook**. Academic Press, 2020. p. 611-615.



VICTORA, Cesar G.; AQUINO, Estela M. L.; LEAL, Maria do Carmo, MONTEIRO, Carlos A.; BARROS, Fernando C.; SZWARCOWALD, Celia L. Saúde de mães e crianças no Brasil: progressos e desafios. **The Lancet**, London, p. 32-46, 2011.

WHO-GLOBAL OBSERVATORY FOR eHEALTH. Telemedicine: Opportunities and developments in Member States. **Observatory**, v. 2, p. 96, 2010. DOI 10.4258/hir.2012.18.2.153. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3402558/>. Acesso em: 21 abr. 2022.

YIN, R. K. **Estudo de caso**: planejamento e métodos. 2. ed. Porto Alegre: Bookman, 2001.