



UNIVERSIDADE FEDERAL DE SANTA CATARINA
CENTRO DE CIÊNCIAS DA EDUCAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO

Márcio José Sembay

PROV-Health: método para gerenciamento de dados de proveniência em sistemas de
informação em saúde

Florianópolis

2023

Márcio José Sembay

**PROV-Health: método para gerenciamento de dados de proveniência em sistemas de
informação em saúde**

Tese submetida ao Programa de Pós-Graduação em
Ciência da Informação da Universidade Federal de Santa
Catarina como requisito parcial para a obtenção do título
de Doutor em Ciência da Informação.

Orientador: Prof. Dr. Douglas Dyllon Jeronimo de
Macedo.

Florianópolis

2023

Ficha de identificação da obra elaborada pelo autor,
através do Programa de Geração Automática da Biblioteca Universitária da UFSC.

Sembay, Márcio José
PROV-Health : método para gerenciamento de dados de
proveniência em sistemas de informação em saúde / Márcio
José Sembay ; orientador, Douglas Dyllon Jeronimo de
Macedo, 2023.
311 p.

Tese (doutorado) - Universidade Federal de Santa
Catarina, Centro de Ciências da Educação, Programa de Pós
Graduação em Ciência da Informação, Florianópolis, 2023.

Inclui referências.

1. Ciência da Informação. 2. Proveniência de dados. 3.
Interoperabilidade de dados. 4. Sistemas de informação em
saúde. 5. W3C PROV. I. Macedo, Douglas Dyllon Jeronimo de.
II. Universidade Federal de Santa Catarina. Programa de Pós
Graduação em Ciência da Informação. III. Título.

Márcio José Sembay

PROV-Health: método para gerenciamento de dados de proveniência em sistemas de informação em saúde

O presente trabalho em nível de Doutorado foi avaliado e aprovado, em 5 de julho de 2023, pela banca examinadora composta pelos seguintes membros:

Prof. Moisés Lima Dutra, Dr.
Universidade Federal de Santa Catarina

Prof.^a Regina Maria Maciel Braga, Dr.^a
Universidade Federal de Juiz de Fora

Prof. Mario Antonio Ribeiro Dantas, Dr.
Universidade Federal de Juiz de Fora

Certificamos que esta é a versão original e final do trabalho de conclusão que foi julgado adequado para obtenção do título de Doutor em Ciência da Informação.

Insira neste espaço a
assinatura digital

Coordenação do Programa de Pós-Graduação

Insira neste espaço a
assinatura digital

Prof. Douglas Dyllon Jeronimo de Macedo, Dr.
Orientador

Florianópolis, 2023.

AGRADECIMENTOS

Desafio tão grande quanto escrever esta tese, foi utilizar apenas uma página para agradecer a todos que fizeram parte desta minha trajetória de doutorado.

Início os agradecimentos primeiramente a Deus, o qual me escutou no silêncio das minhas falhas e na inquietude da minha mente nos momentos difíceis, me dando a sabedoria necessária para finalizar uma etapa tão importante da minha vida.

À minha esposa Tayse e aos meus filhos Sebastian e Coraline, que mesmo sabendo das dificuldades encontradas na minha jornada e na busca do meu sonho, esperaram por mim. Obrigado... amo vocês.

Agradeço à minha mãe Antonieta por suas orações e pelas palavras de coragem para que eu pudesse enfrentar meus obstáculos da minha caminhada nos estudos. Ao meu pai Manuel (*in memoriam*), que mesmo não estando aqui, sempre me incentivou.

Meus sinceros agradecimentos ao meu orientador Douglas Dyllon Jeronimo de Macedo, por toda confiança e compreensão, que de forma paciente e sábia, soube guiar-me para que eu pudesse chegar até aqui. Tornou-se um exemplo de profissional e de pessoa para a minha vida. Muito obrigado.

Fica meu agradecimento ao professor Dr. Adilson Luiz Pinto, com o qual iniciei meus estudos de doutorado em sua disciplina. Agradeço por toda a atenção e motivação.

Aos membros da banca, professores, Dra. Regina Maria Maciel Braga, Dr. Mario Antonio Ribeiro Dantas e Dr. Moisés Lima Dutra pela participação e por todos os apontamentos de melhoria que foram apresentados para esta pesquisa.

À direção, coordenação, a todos os professores e funcionários do PGCIN pela dedicação e apoio dispensado em todos os momentos de necessidade.

Aos colegas de turma pelos momentos de estudo e muita aprendizagem nas disciplinas que fizemos juntos, em especial às colegas Gislaine Parra Freund, Heloisa Costa e Tatiane Rossi, as quais, gentilmente sempre me ajudaram muito.

Ao meu amigo, Alexandre Augusto Gimenes Marques Filho, pelas inúmeras reuniões de estudo, me ajudando a decifrar diversas dificuldades no meu caminho de pesquisa. Sou extremamente grato por isso.

Ao Dr. Sérgio Fernando Torres de Freitas, diretor do STT/RCTM, por ter permitido que a experimentação desta tese fosse realizada nessa instituição.

Agradeço a todos que, embora não mencionados, direta ou indiretamente contribuíram para a realização desta tese. Muito obrigado!

“A persistência é o menor caminho do êxito”. (Charles Chaplin)

RESUMO

A Ciência da Informação e a Ciência da Computação, compactuam aspectos interdisciplinares no que diz respeito à Proveniência de Dados entre diferentes Sistemas de Informação em Saúde (SIS). Os SIS armazenam grandes volumes de dados em seus repositórios descentralizados, tornando um desafio o gerenciamento e a interoperabilidade desses dados. Assim, a Proveniência de Dados associada a tecnologias computacionais contribui para a rastreabilidade dos dados de origem nos SIS, a fim de identificar erros e transformações nos dados e atribuí-los às fontes, sendo útil para tomadas de decisões nos mais variados cenários de saúde. Nesta tese apresenta-se um método para gerenciamento de dados de proveniência em SIS com base no instanciamento do modelo da World Wide Web Consortium Provenance (W3C PROV) em conformidade com os aspectos de interoperabilidade entre diferentes bases de dados de saúde descentralizadas. Para alcançar os objetivos definidos para esta tese, assume-se uma metodologia de pesquisa aplicada em um estudo de caso, caracterizada como experimental e exploratória, com abordagem quantitativa e qualitativa. Quanto aos procedimentos técnicos no desenvolvimento e validação do método, foram realizados experimentos em um cenário de saúde real com dados reais, os quais possibilitaram identificar além de benefícios, desafios e ameaças durante sua experimentação. O método foi avaliado via formulário eletrônico enviado a especialistas da área, os quais destacaram pontos importantes para reflexão na estrutura do método. Conclui-se que ao final dos experimentos e das etapas de análises dos resultados, o método proposto mostrou-se significativo no que se refere ao gerenciamento de dados de proveniência em diferentes SIS, sendo possível adaptá-lo a qualquer cenário de saúde.

Palavras-chave: proveniência de dados; gerenciamento de dados de proveniência; sistemas de informação em saúde; interoperabilidade de dados; W3C PROV.

ABSTRACT

Information Science and Computer Science share interdisciplinary aspects when it comes to Data Provenance between different Health Information Systems (HIS). HIS store large volumes of data in their decentralized repositories, making the management and interoperability of this data a challenge. Thus, Data Provenance associated with computer technologies contributes to the traceability of source data in HIS, in order to identify errors and transformations in the data and attribute them to the sources, being useful for decision-making in the most varied health scenarios. This thesis presents a method for managing provenance data in HIS based on instancing the World Wide Web Consortium Provenance (W3C PROV) model in accordance with interoperability aspects between different decentralized health databases. In order to achieve the objectives defined for this thesis, an applied research methodology was used in a case study, characterized as experimental and exploratory, with a quantitative and qualitative approach. As for the technical procedures for developing and validating the method, experiments were carried out in a real health scenario with real data, which made it possible to identify not only benefits, but also challenges and threats during experimentation. The method was evaluated via an electronic form sent to experts in the field, who highlighted important points for reflection on the structure of the method. The conclusion is that at the end of the experiments and the stages of analysis of the results, the proposed method proved to be significant in terms of managing provenance data in different HIS and can be adapted to any healthcare scenario.

Keywords: data provenance; provenance data management; health information systems; data interoperability; W3C PROV.

RESUMEN

La Ciencias de la Información y la Ciencias de la Computación, comparten aspectos interdisciplinarios cuando se trata de la Procedencia de los Datos entre distintos Sistemas de Información en Salud (SIS). Los SIS almacenan grandes volúmenes de datos en sus repositorios descentralizados, lo que convierte la gestión e interoperabilidad de estos datos en todo un reto. Por ello, la Procedencia de Datos asociada a las tecnologías informáticas contribuye a la trazabilidad de los datos fuente en los SIS, con el fin de identificar errores y transformaciones en los datos y atribuirlos a las fuentes, siendo útil para la toma de decisiones en los más variados escenarios de salud. Esta tesis presenta un método de gestión de datos de procedencia en SIS basado en la instanciación del modelo World Wide Web Consortium Provenance (W3C PROV) de acuerdo con aspectos de interoperabilidad entre diferentes bases de datos de salud descentralizadas. Para alcanzar los objetivos definidos para esta tesis, se utilizó una metodología de investigación aplicada en un estudio de caso, caracterizado como experimental y exploratorio, con un enfoque cuantitativo y cualitativo. En cuanto a los procedimientos técnicos para desarrollar y validar el método, se realizaron experimentos en un escenario real de salud con datos reales, lo que permitió identificar beneficios, desafíos y amenazas durante el experimento. El método se evaluó mediante un formulario electrónico enviado a expertos en la materia, que destacaron puntos importantes para la reflexión sobre la estructura del método. La conclusión es que al final de los experimentos y de las fases de análisis de los resultados, el método propuesto demostró ser significativo en términos de gestión de datos de procedencia en diferentes SIS, y puede adaptarse a cualquier escenario de salud.

Palabras clave: procedencia de los datos; Gestión de datos de procedencia; sistemas de información en salud; interoperabilidad de datos; W3C PROV.

LISTA DE FIGURAS

| | |
|---|-----|
| Figura 1 – Organização da família PROV | 48 |
| Figura 2 – Entidades do modelo PROV | 49 |
| Figura 3 – Fluxo do telediagnóstico no estado de Santa Catarina..... | 67 |
| Figura 4 – Número de estudos por etapa da fase de condução da SLR..... | 82 |
| Figura 5 – Processo de execução <i>Snowballing (backward and forward)</i> | 86 |
| Figura 6 – Nome do método apresentado | 119 |
| Figura 7 – Instanciamento do modelo W3C PROV para concepção do PROV-Health..... | 122 |
| Figura 8 – Instrumentalização genérica dos serviços de saúde dos SIS para uso no PROV-Health | 127 |
| Figura 9 – Arquitetura genérica proposta para o PROV-Health | 131 |
| Figura 10 – Camada de Agentes (A) do PROV-Health..... | 134 |
| Figura 11 – Camada de Agentes (B) do PROV-Health..... | 135 |
| Figura 12 – Camada de Agentes (C) do PROV-Health..... | 136 |
| Figura 13 – Camada de Agentes (D) do PROV-Health..... | 137 |
| Figura 14 – Camada de Agentes (E) do PROV-Health..... | 137 |
| Figura 15 – Camada de Entidades (A) do PROV-Health..... | 138 |
| Figura 16 – Camada de Atividades (A) do PROV-Health | 140 |
| Figura 17 – Camada de Atividades (B) do PROV-Health..... | 142 |
| Figura 18 – Camada de Atividades (C) do PROV-Health genérico..... | 144 |
| Figura 19 – Arquitetura de abrangência para a experimentação do PROV-Health no STT/RCTM | 151 |
| Figura 20 – Consumidores e clientes do PROV-Health no STT/RCTM | 153 |
| Figura 21 – Tela principal do Forge FHIR para a criação de uma nova definição de estrutura de profile | 159 |
| Figura 22 – Instrumentalização específica dos serviços de saúde no STT/RCTM (ampliada) para uso no PROV-Health | 169 |
| Figura 23 – Instrumentalização específica dos serviços de saúde (reduzida) para uso no PROV- Health | 171 |
| Figura 24 – Execução simplificada do PROV-Health no STT/RCTM | 176 |
| Figura 25 – Arquitetura específica do PROV-Health para experimentação no STT/RCTM. | 178 |
| Figura 26 – Camada de Agentes (A) para aplicação do PROV-Health no STT/RCTM..... | 180 |
| Figura 27 – Camada de Agentes (B) do PROV-Health no STT/RCTM | 181 |

| | |
|---|-----|
| Figura 28 - Camada de Agentes (C) do PROV-Health no STT/RCTM..... | 182 |
| Figura 29 - Camada de Agentes (D) do PROV-Health no STT/RCTM..... | 182 |
| Figura 30 – Camada de Agentes (E) do PROV-Health no STT/RCTM | 183 |
| Figura 31 – Camada de Entidades (A) do PROV-Health no STT/RCTM | 184 |
| Figura 32 – Camada de Entidades (B) do PROV-Health no STT/RCTM..... | 184 |
| Figura 33 – Camada de Entidades (C) do PROV-Health no STT/RCTM..... | 184 |
| Figura 34 – Camada de Atividades (A) do PROV-Health para experimentação no STT/RCTM | 186 |
| Figura 35 – Estrutura de diretórios com aplicações de coletas parametrizadas para execução no HGCR..... | 188 |
| Figura 36 – Execução contínua de processos Python para coleta de exames agendados do HGCR | 188 |
| Figura 37 – Utilização da IDE Visual Studio Code para o PROV-Health no STT/RCTM ... | 190 |
| Figura 38 – Serviço Códigos UFSC para gerenciamento de código do PROV-Health no STT/RCTM..... | 191 |
| Figura 39 – STT on FHIR onde estão alocados os profiles utilizados no PROV-Health..... | 192 |
| Figura 40 – Camada de Atividades (B) do PROV-Health para experimentação no STT/RCTM | 201 |
| Figura 41 – Validações do funcionamento do FHIR utilizando Insomnia | 201 |
| Figura 42 – Camada de Atividades (C) do PROV-Health aplicada ao STT/RCTM..... | 206 |
| Figura 43 – Total de dados coletados por experimento da Camada de Atividades (A) | 211 |
| Figura 44 – Total da coleta de dados de proveniência por experimento no HIJG | 212 |
| Figura 45 – Total da coleta de dados de proveniência por experimento no HGCR | 213 |
| Figura 46 – Total da coleta de dados de proveniência por experimento no HRSJ..... | 214 |
| Figura 47 – Total das integrações no Experimento 1 por hospital | 216 |
| Figura 48 – Total das integrações do Experimento 2 por hospital | 217 |
| Figura 49 – Total das integrações do Experimento 3 por hospital | 218 |

LISTA DE GRÁFICOS

| | |
|--|-----|
| Gráfico 1 – Relação do número de estudos encontrados de acordo com o ano e tipo de publicação..... | 102 |
| Gráfico 2 – Percentuais referentes aos tipos de publicações..... | 102 |
| Gráfico 3 – Percentuais referentes à frequência dos tipos de SIS..... | 103 |
| Gráfico 4 – Quantidade de dados coletados por experimento e por hospital..... | 223 |
| Gráfico 5 – Quantidade de dados integrados por hospital no experimento 1..... | 227 |
| Gráfico 6 – Quantidade de dados integrados por hospital no experimento 2..... | 228 |
| Gráfico 7 – Quantidade de dados integrados por hospital no experimento 3..... | 229 |
| Gráfico 8 – Idade dos participantes..... | 245 |
| Gráfico 9 – Função profissional dos participantes..... | 245 |

LISTA DE QUADROS

| | |
|--|-----|
| Quadro 1 – Especificações da família PROV..... | 47 |
| Quadro 2 – Classes do modelo PROV..... | 49 |
| Quadro 3 – Relações do modelo PROV | 49 |
| Quadro 4 – Ferramentas para interoperabilidade, integração e armazenamento de dados em SIS | 56 |
| Quadro 5 – Demandas do PEP e contribuições da ciência da informação | 72 |
| Quadro 6 – Questões gerais e questões específicas de pesquisa | 76 |
| Quadro 7 – Critérios de inclusão | 78 |
| Quadro 8 – Critérios de exclusão | 78 |
| Quadro 9 – Critérios de qualidade..... | 79 |
| Quadro 10 – Seções dos estudos relacionados | 79 |
| Quadro 11 – Escala de Likert-3..... | 80 |
| Quadro 12 – Escala para cada critério de qualidade..... | 80 |
| Quadro 13 – Artigos encontrados em bases de dados eletrônicas internacionais | 83 |
| Quadro 14 – Lista final dos 17 estudos selecionados..... | 95 |
| Quadro 15 – Cenários/aplicações, métodos, técnicas, modelos e metodologias..... | 97 |
| Quadro 16 – Principais vantagens do gerenciamento dos dados de proveniência dos SIS | 105 |
| Quadro 17 – Tecnologias que contribuem para o gerenciamento dos dados de proveniência dos SIS | 106 |
| Quadro 18 – Categorias que contribuem para o gerenciamento dos dados de proveniência em SIS | 108 |
| Quadro 19 – Resumo das características metodológicas..... | 113 |
| Quadro 20 – Resumo dos aspectos metodológicos | 115 |
| Quadro 21 – Descrição das relações do instanciamento do modelo W3C PROV para concepção do PROV-Health | 123 |
| Quadro 22 – Tecnologias computacionais selecionadas para o PROV-Health no STT/RCTM | 154 |
| Quadro 23 – Identificação dos campos de agendamentos de exames do HIS..... | 160 |
| Quadro 24 – Estrutura de dados para utilização do Profile SolicitacaoHospitalar no PROV- Health | 162 |
| Quadro 25 – Dados de proveniência relevantes relacionadas a Pacientes | 164 |
| Quadro 26 – Estrutura de dados para utilização do Profile EstudoDICOM..... | 165 |

| | |
|---|-----|
| Quadro 27 – Identificação dos campos LaudoExameHospitalar no STT/RCTM..... | 166 |
| Quadro 28 – Estrutura de dados para utilização do <i>Profile LaudoExameHospitalar</i> no STT/RCTM..... | 167 |
| Quadro 29 – Metodologia da experimentação para coleta de dados de proveniência no STT/RCTM..... | 173 |
| Quadro 30 – Metodologia realizada para a integração dos dados de proveniência no STT/RCTM | 175 |
| Quadro 31 – Parâmetros necessários a execução das coletas | 187 |
| Quadro 32 – Estrutura de desenvolvimento utilizado para o PROV-Health no STT/RCTM. | 190 |
| Quadro 33 – Coleta de dados Agendados..... | 193 |
| Quadro 34 – Mapeamento do Profile de SolicitacaoHospitalar | 194 |
| Quadro 35 – Mapeamento do sexo do paciente..... | 194 |
| Quadro 36 – Mapeamento do médico solicitante | 194 |
| Quadro 37 – Mapeamento da data do exame | 195 |
| Quadro 38 – Mapeamento da modalidade DICOM..... | 195 |
| Quadro 39 – Geração do documento JSON conforme o profile SolicitacaoHospitalar | 195 |
| Quadro 40 – Coleta de dados Estudos DICOM..... | 196 |
| Quadro 41 – Consulta SQL realizada para coleta de Estudo DICOM | 196 |
| Quadro 42 – Transformações do nome do paciente | 197 |
| Quadro 43 – Transformação da data de nascimento..... | 197 |
| Quadro 44 – Data do estudo DICOM..... | 198 |
| Quadro 45 – Geração do documento JSON conforme profile EstudoDICOM..... | 198 |
| Quadro 46 – Código em Python para coleta dos dados de laudos de exames hospitalares.... | 199 |
| Quadro 47 – Modelo parametrizado para “laudos_template.json.j2” | 199 |
| Quadro 48 – Configuração dos serviços para execução do HAPI FHIR no STT/RCTM..... | 203 |
| Quadro 49 – Processo Python para os processos de verificação das solicitações hospitalares integradas com estudos DICOM..... | 204 |
| Quadro 50 – Processo Python para os processos de verificação das solicitações hospitalares integradas com laudos publicados | 205 |
| Quadro 51 – Processo Python para os processos de verificação dos laudos publicados com estudos DICOM..... | 205 |
| Quadro 52 – Coleta dos recursos FHIR armazenados no PROV-Health | 207 |
| Quadro 53 – processamento de um item obtido a partir da consulta realizada no FHIR | 207 |
| Quadro 54 – Campos das tabelas utilizadas pelo PROV-Health-BI | 208 |

| | |
|--|-----|
| Quadro 55 – Execução da coleta se fez a partir de um script desenvolvido em Bash..... | 208 |
| Quadro 56 – Agendamento é realizado diariamente por meio do serviço de “cronab” do Linux | 209 |
| Quadro 57 – Serviço Grafana adicionado na Stack do PROV-Health | 209 |
| Quadro 58 – Código realizado no PostgreSQL para a totalidade dos dados por experimento | 210 |
| Quadro 59 – Código realizado no PostgreSQL para a totalidade dos experimentos coletados do HIJG | 212 |
| Quadro 60 – Código realizado no PostgreSQL para a totalidade dos experimentos coletados do HGCR..... | 212 |
| Quadro 61 – Código realizado no PostgreSQL para a totalidade dos experimentos coletados do HRSJ..... | 213 |
| Quadro 62– Códigos realizados no PostgreSQL para as integrações do experimento 1 por hospital | 215 |
| Quadro 63 – Códigos realizados no PostgreSQL para as integrações do experimento 2 por hospital | 216 |
| Quadro 64 – Códigos realizados no PostgreSQL para as integrações do experimento 3 por hospital | 217 |
| Quadro 65 – Identificação da existência das principais características da proveniência de dados no PROV-Health experimentado no STT/RCTM (Experimentos de Coletas)..... | 226 |
| Quadro 66 – Identificação da existência das principais características da proveniência de dados no PROV-Health experimentado no STT/RCTM (Experimentos de integrações) | 232 |
| Quadro 67 – Identificação da existência das principais características da proveniência de dados no PROV-Health experimentado no STT/RCTM (Visualização das coletas) | 235 |
| Quadro 68 – Identificação da existência das principais características da proveniência de dados no PROV-Health experimentado no STT/RCTM (Visualização das integrações) | 236 |
| Quadro 69 – Respostas para a questão 1 da avaliação do PROV-Health..... | 246 |
| Quadro 70 – Respostas para a questão 2 da avaliação do PROV-Health..... | 247 |
| Quadro 71 – Respostas para a questão 3 da avaliação do PROV-Health..... | 249 |
| Quadro 72 – Respostas para a questão 4 da avaliação do PROV-Health..... | 250 |
| Quadro 73 – Respostas para a questão 5 da avaliação do PROV-Health..... | 251 |
| Quadro 74 – Respostas para a questão 6 da avaliação do PROV-Health..... | 252 |
| Quadro 75 – Respostas para a questão 7 da avaliação do PROV-Health..... | 253 |

LISTA DE TABELAS

| | |
|---|-----|
| Tabela 1 – Níveis de qualidade dos estudos | 96 |
| Tabela 2 – Extrato geral das análises dos experimentos por meio do PROV-Health no STT/RCTM..... | 238 |

LISTA DE ABREVIATURAS E SIGLAS

| | |
|---------|---|
| ABP | Aprendizado Baseado em Problemas |
| ABRAHUE | Associação Brasileira de Hospitais Universitários |
| APS | Atenção Primária à Saúde |
| CC | Ciência da Computação |
| CFM | Conselho Federal de Medicina |
| CI | Ciência da Informação |
| CLARA | Cooperação Latino-Americana de Redes Avançadas |
| CMU | <i>Couverture Maladie Universelle</i> |
| CNPq | Conselho Nacional de Desenvolvimento Científico e Tecnológico |
| DeMAC | Departamento de Média e Alta Complexidade |
| DICOM | <i>Digital Imaging and Communications in Medicine</i> |
| DNA | <i>Deoxyribonucleic acid</i> |
| ETL | <i>Extract Transform Load</i> |
| FHIR | <i>Fast Healthcare Interoperability Resources</i> |
| FINEP | Financiadora de Estudos e Projetos |
| GOe | <i>Global Observatory for eHealth</i> |
| GROUP | <i>Cyclops Group</i> |
| HC | Hospital de Clínicas |
| HIS | <i>Hospital Information System</i> |
| HL7 | <i>Health Level Seven</i> |
| IHE | <i>Integrating the Healthcare Enterprise</i> |
| MS | Ministério da Saúde |
| OCDE | Organização de Cooperação e de Desenvolvimento Econômico |
| OMS | Organização Mundial da Saúde |
| ONU | Organização das Nações Unidas |
| OPM | <i>Open Provenance Model</i> |
| PACS | <i>Picture Archiving and Communication System</i> |
| PEP | Prontuário Eletrônico do Paciente |
| PEP | Prontuário Eletrônico do Paciente |
| PGCIN | Programa de Pós-Graduação em Ciência da Informação |
| POR | <i>Proof of Retrievability</i> |

| | |
|-----------|---|
| PROV-DM | <i>Provenance Data Model</i> |
| QTL | <i>Quantitative Trait Loci</i> |
| RAS | Redes de Atenção à Saúde |
| RCTM | Rede Catarinense de Telemedicina |
| REDECOMEP | Redes Comunitárias Metropolitanas de Educação e Pesquisa |
| RES | Registro Eletrônico de Saúde |
| RIS | <i>Radiology Information System</i> |
| RNP | Rede Nacional de Ensino e Pesquisas |
| RSL | Revisão sistemática de Literatura |
| RUTE | Rede Universitária de Telemedicina |
| SES/MG | Secretaria Estadual de Saúde de Minas Gerais |
| SGWfCs | Sistemas Gerenciadores de <i>Workflows</i> Científicos |
| SIGs | <i>Special Interest Groups</i> |
| SMS/BH | Secretaria Municipal de Saúde de Belo Horizonte |
| SPICE | <i>Setting, Perspective, Intervention, Comparison, Evaluation</i> |
| STT/SC | Sistema Integrado de Telemedicina e Telessaúde |
| SUS | Sistema Único de Saúde |
| TI | Tecnologia da Informação |
| TIC | Tecnologias de Informação e Comunicação |
| UFMG | Universidade Federal de Minas Gerais |
| UFSC | Universidade Federal de Santa Catarina |
| VOIP | <i>Voice Over Internet Protocol</i> |
| W3C PROV | <i>World Wide Web Consortium Provenance</i> |
| WHO | <i>Global Observatory for eHealth</i> |

SUMÁRIO

| | | |
|----------------|--|-----------|
| 1 | INTRODUÇÃO | 23 |
| 1.1 | CONTEXTUALIZAÇÃO | 26 |
| 1.2 | MOTIVAÇÃO E JUSTIFICATIVA | 27 |
| 1.3 | PROBLEMAS DE PESQUISA | 28 |
| 1.4 | QUESTÃO DE PESQUISA E HIPÓTESE..... | 30 |
| 1.5 | OBJETIVOS..... | 30 |
| 1.5.1 | Objetivo Geral | 30 |
| 1.5.2 | Objetivos Específicos | 30 |
| 1.6 | INEDITISMO DA TESE | 31 |
| 1.7 | DELIMITAÇÃO DO ESCOPO | 32 |
| 1.8 | CONTRIBUIÇÕES DA PESQUISA | 33 |
| 1.9 | ADERÊNCIA DO TEMA AO PGCIN E À ÁREA DE CIÊNCIA DA INFORMAÇÃO..... | 35 |
| 1.10 | ESTRUTURA DA TESE | 37 |
| 2 | REFERENCIAL TEÓRICO | 40 |
| 2.1 | CIÊNCIA DA INFORMAÇÃO | 40 |
| 2.1.1 | Conceito, histórico e epistemologia | 40 |
| 2.1.2 | Relações interdisciplinares: Ciência da Informação e Ciência da Computação | 41 |
| 2.1.3 | Ciência da informação e saúde | 42 |
| 2.2 | PROVENIÊNCIA DE DADOS | 43 |
| 2.2.1 | Componentes e funcionamento | 44 |
| 2.2.2 | Principais modelos | 46 |
| 2.2.3 | Aplicações gerais | 50 |
| <i>2.2.3.1</i> | <i>Aplicações em saúde</i> | <i>51</i> |
| 2.3 | SISTEMAS DE INFORMAÇÃO EM SAÚDE | 52 |
| 2.3.1 | Principais tipos | 54 |
| 2.3.2 | Interoperabilidade, integração e armazenamento de dados | 56 |
| 2.3.3 | HL7 FHIR baseado em W3C PROV para garantir a Proveniência de dados em Sistemas de Informação em Saúde | 59 |
| 2.4 | TELEMEDICINA | 60 |
| 2.4.1 | Funcionamento e aplicações | 61 |
| 2.4.2 | Telemedicina no Brasil | 63 |

| | | |
|----------------|--|------------|
| 2.4.2.1 | <i>Telemedicina em Santa Catarina</i> | 66 |
| 2.5 | PRONTUÁRIO ELETRÔNICO DO PACIENTE (PEP) | 68 |
| 2.5.1 | Estrutura do PEP | 69 |
| 2.5.2 | Aspectos Éticos e Legais do PEP | 70 |
| 2.5.3 | Contribuições da Ciência da Informação para o PEP | 71 |
| 2.6 | CONSIDERAÇÕES DA SEÇÃO | 73 |
| 3 | ESTADO DA ARTE | 74 |
| 3.1 | REVISÃO SISTEMÁTICA DE LITERATURA..... | 74 |
| 3.1.1 | Contextualização e questões de pesquisa | 75 |
| 3.1.2 | Escolha das bases de dados e termos para estratégia de busca | 77 |
| 3.1.3 | Critérios para seleção dos estudos primários | 77 |
| 3.1.4 | Avaliação da qualidade da seleção dos estudos primários | 78 |
| 3.1.5 | Extração e Síntese dos Dados | 80 |
| 3.1.6 | Identificação dos estudos primários | 81 |
| 3.1.7 | Síntese dos estudos relevantes | 81 |
| 3.1.8 | Condução da RSL | 81 |
| 3.1.9 | Utilização da técnica de Snowballing (backward and forward) | 84 |
| <i>3.1.9.1</i> | <i>Execução do Snowballing (backward and forward)</i> | <i>85</i> |
| 3.2 | ESTUDOS SELECIONADOS..... | 87 |
| 3.2.1 | Resumo dos estudos selecionados no conjunto inicial da RSL | 87 |
| 3.2.2 | Resumo dos estudos selecionados no Snowballing (backward and forward) | 93 |
| 3.2.3 | Considerações sobre os estudos selecionados | 94 |
| 3.3 | RESULTADOS DA RSL..... | 95 |
| 3.3.1 | Respostas para as Questões de Pesquisa da RSL | 97 |
| 3.3.2 | Principais categorias identificadas em relação ao gerenciamento de dados de proveniência em SIS | 108 |
| 3.4 | LIÇÕES APRENDIDAS..... | 110 |
| 3.5 | CONSIDERAÇÕES DA SEÇÃO | 111 |
| 4 | ASPECTOS METODOLÓGICOS | 112 |
| 4.1 | CARACTERIZAÇÃO DA PESQUISA..... | 112 |
| 4.2 | UNIVERSO, POPULAÇÃO E AMOSTRA..... | 114 |
| 4.3 | PROCEDIMENTOS METODOLÓGICOS | 114 |
| 4.4 | COLETA E ANÁLISE DOS DADOS | 116 |
| 5 | PROV-HEALTH | 119 |

| | | |
|--------------|---|------------|
| 5.1 | METODOLOGIA DE CONCEPÇÃO..... | 119 |
| 5.2 | INSTANCIANDO O MODELO W3C PROV EM UM CONTEXTO GENÉRICO DE SIS PARA A CONCEPÇÃO DO PROV-HEALTH..... | 121 |
| 5.3 | PROPOSTA DE INSTRUMENTALIZAÇÃO GENÉRICA DOS SERVIÇOS DE SAÚDE PARA USO NO PROV-HEALTH | 125 |
| 5.4 | ARQUITETURA GENÉRICA DO PROV-HEALTH..... | 129 |
| 5.4.1 | Camada de Agentes | 133 |
| 5.4.2 | Camada de Entidades | 138 |
| 5.4.3 | Camada de Atividades | 139 |
| 5.4.3.1 | <i>Módulo de captura, coleta e preparação dos dados de proveniência em SIS</i> | <i>139</i> |
| 5.4.3.2 | <i>Módulo de organização, controle e armazenamento dos dados de proveniência dos SIS.....</i> | <i>141</i> |
| 5.4.3.3 | <i>Módulo de visualização, consulta e análise dos dados de proveniência em SIS</i> | <i>143</i> |
| 5.4.4 | Verificação dos componentes e relações da arquitetura genérica do PROV-Health com base no W3C PROV | 146 |
| 5.5 | CONSIDERAÇÕES DA SEÇÃO | 146 |
| 6 | EXPERIMENTAÇÃO DO PROV-HEALTH | 148 |
| 6.1 | CONTEXTUALIZAÇÃO DO AMBIENTE DE EXPERIMENTAÇÃO | 148 |
| 6.2 | ARQUITETURA DO STT/RCTM PARA O AMBIENTE DE EXPERIMENTAÇÃO..... | 150 |
| 6.3 | DEFINIÇÃO DOS CONSUMIDORES E CLIENTES DO PROV-HEALTH..... | 153 |
| 6.4 | TECNOLOGIAS COMPUTACIONAIS UTILIZADAS PARA COMPOR A ESTRATÉGIA COMPUTACIONAL DO PROV-HEALTH..... | 154 |
| 6.5 | DEFINIÇÃO DOS PROFILES FHIR PARA EXPERIMENTAÇÃO DO PROV-HEALTH NO STT/RCTM..... | 158 |
| 6.5.1 | Definição do profile <i>SolicitacaoHospitalar</i> | 160 |
| 6.5.2 | Definição do profile <i>EstudoDICOM</i> | 163 |
| 6.5.3 | Definição do profile <i>LaudoExameHospitalar</i> | 165 |
| 6.6 | INSTRUMENTALIZAÇÃO ESPECÍFICA DOS SERVIÇOS DE SAÚDE PARA USO NA EXPERIMENTAÇÃO DO PROV-HEALTH..... | 167 |
| 6.6.1 | Abrangência da instrumentalização específica dos serviços de saúde (ampliada)..... | 167 |
| 6.6.2 | Instrumentalização específica dos serviços de saúde (reduzida)..... | 171 |
| 6.7 | METODOLOGIA DE EXPERIMENTAÇÃO DO PROV-HEALTH NO | |

| | | |
|--------------|--|------------|
| | STT/RCTM..... | 172 |
| 6.8 | ARQUITETURA ESPECÍFICA PARA EXPERIMENTAÇÃO DO PROV-HEALTH..... | 177 |
| 6.8.1 | Aplicação das Camadas de Agentes..... | 179 |
| 6.8.2 | Aplicação das Camadas de Entidades..... | 183 |
| 6.8.3 | Aplicação das Camadas de Atividades..... | 185 |
| 6.8.3.1 | <i>Camada de Atividades (A) – Módulo (A): captura, coleta e preparação dos dados de proveniência do STT/RCTM.....</i> | <i>185</i> |
| 6.8.3.1.1 | Experimento 1: Coleta de dados Agendados (<i>Profile SolicitacaoHospitalar</i>)..... | 193 |
| 6.8.3.1.2 | Experimento 2: Coleta de dados Estudos DICOM (<i>Profile EstudoDICOM</i>)..... | 196 |
| 6.8.3.1.3 | Experimento 3: Coleta de dados Laudos (<i>Profile: LaudoExameHospitalar</i>)..... | 198 |
| 6.8.4.1 | <i>Camada de Atividades (B) – Módulo (B): organização, controle e armazenamento dos dados de proveniência do STT/RCTM.....</i> | <i>200</i> |
| 6.8.4.1.1 | Experimento 1: Integração de solicitações hospitalares com estudos DICOM..... | 203 |
| 6.8.4.1.2 | Experimento 2: Integração de solicitações hospitalares com laudos publicados | 204 |
| 6.8.4.1.3 | Experimento 3: Integração de laudos publicados com estudos DICOM..... | 205 |
| 6.8.5.1 | <i>Camada de Atividades (C) – Módulo (C): visualização, consulta e análise dos dados de proveniência do STT/RCTM.....</i> | <i>206</i> |
| 6.8.5.1.1 | Visualização dos dados de proveniência dos experimentos da Camada de Atividades (A)..... | 210 |
| 6.8.5.1.2 | Visualização dos dados de proveniência dos experimentos da Camada de Atividades (B)..... | 215 |
| 6.9 | DESAFIOS E AMEAÇAS ENCONTRADOS NA IMPLEMENTAÇÃO E EXPERIMENTAÇÃO DO PROV-HEALTH NO STT/RCTM..... | 219 |
| 6.10 | CONSIDERAÇÕES DA SEÇÃO..... | 220 |
| 7 | ANÁLISES E AVALIAÇÕES DA EXPERIMENTAÇÃO DO PROV-HEALTH..... | 222 |
| 7.1 | ANÁLISES DOS EXPERIMENTOS DE COLETAS NO PROV-HEALTH..... | 222 |
| 7.2 | ANÁLISES DOS EXPERIMENTOS DE INTEGRAÇÕES DO PROV-HEALTH..... | 227 |
| 7.3 | ANÁLISES DAS VISUALIZAÇÕES DOS EXPERIMENTOS: COLETAS E INTEGRAÇÕES..... | 233 |
| 7.4 | DISCUSSÕES FINAIS DAS ANÁLISES DOS EXPERIMENTOS DE COLETAS E INTEGRAÇÕES..... | 237 |
| 7.5 | AVALIAÇÃO DO PROV-HEALTH..... | 242 |

| | | |
|-------|---|-----|
| 7.5.1 | Objetivos da avaliação | 242 |
| 7.5.2 | Planejamento e preparação da avaliação | 243 |
| 7.5.3 | Aplicação da avaliação | 244 |
| 7.5.4 | Apresentação e análise dos resultados | 244 |
| 7.6 | CONSIDERAÇÕES DA SEÇÃO | 254 |
| 8 | CONCLUSÕES E TRABALHOS FUTUROS | 256 |
| 8.1 | TRABALHOS FUTUROS | 262 |
| 8.2 | ARTIGOS PUBLICADOS | 264 |
| 8.2.1 | Artigos completos publicados em periódicos nacionais e internacionais | 264 |
| 8.2.2 | Artigos completos publicados em anais de congressos | 265 |
| 8.2.3 | Resumos publicados em anais de congressos | 265 |
| 8.2.4 | Artigos publicados em capítulos de livros | 265 |
| 8.2.5 | Artigos submetidos e em análise | 266 |
| | REFERÊNCIAS | 268 |
| | APÊNDICE A – FOLHA DE ROSTO PARA PESQUISA ENVOLVENDO SERES HUMANOS | 287 |
| | APÊNDICE B – PROJETO DE PESQUISA ENVOLVENDO SERES HUMANOS | 288 |
| | APÊNDICE C – PARECER CONSUBSTANCIADO DO CEP | 294 |
| | APÊNDICE D – DECLARAÇÃO (INSTITUIÇÃO PARA COLETA DE DADOS) | 301 |
| | APÊNDICE E – TERMO DE COMPROMISSO PARA UTILIZAÇÃO DE DADOS (TCUD) | 302 |
| | APÊNDICE F – TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO (TCLE) | 304 |
| | APÊNDICE G – ROTEIRO PARA ENTREVISTA SEMIESTRUTURADA | 307 |
| | APÊNDICE H – SOLICITAÇÃO DE REGISTRO DO PROV-HEALTH NO INPI | 309 |
| | APÊNDICE I – REGISTRO DO PROV-HEALTH NO INPI | 310 |

1 INTRODUÇÃO

Em uma era em que os grandes volumes de dados se transformam em coleções exponenciais de informações, observa-se este processo como alicerce para a construção de conhecimento nos mais variados setores relacionados às atividades humanas, sociais e econômicas. Nesse sentido, os dados alimentam o crescimento de diferentes tipos de tecnologias que personalizam experiências para o avanço na busca por conhecimento.

A palavra ‘dados’ surgiu na metade do século XVII e vem do plural latim *data*, que significa uma única peça de informação. O *Oxford English Dictionary Online* define a palavra dados como fatos e/ou estatísticas reunidas para referência ou análise de algo, as quantidades (informação estatística), os caracteres e os símbolos de operações lógicas executadas por um computador (*software*). Também significa o registro armazenado, preservado e transmitido dessas operações sob forma de sinal elétrico (digital) que possa ser gravado em qualquer tipo de mídia magnética, óptica ou mecânica (*hardware*). A palavra dados se aplica a uma série de objetos que são originários de ambientes digitais, como textos, números, imagens, vídeos, áudios, *softwares*, algoritmos, equações, animações, modelos e simulações (NATIONAL SCIENCE BOARD, 2011; KOLTAY, 2017; OXFORD ENGLISH DICTIONARY, 2015).

Assim, a Proveniência de Dados (origem dos dados) é uma preocupação constante nos diversos setores que atuam com dados. Nesse contexto, segundo Moreau e Groth (2013, p. 4): “A Proveniência de Dados é definida como um registro que descreve as pessoas, instituições, entidades e atividades envolvidas na produção, influência ou entrega de um dado ou coisa”. Em complemento, a Proveniência de Dados diz respeito ao histórico dos dados, desde sua concepção, transformação e destino. Em diversos cenários os conceitos de Proveniência de Dados são utilizados, pois um dos principais objetivos é a identificação e o rastreamento dos dados em suas fontes originais no intuito de auxiliar nas tomadas de decisão.

Com a evolução dos processos de coleta, armazenamento e processamento dos dados, consequentemente as estruturas se transformam em informação. Para Silva e Ribeiro (2012) os autores propõem um conceito de informação no qual é um conjunto estruturado de representações mentais e emocionais codificadas (sinais e símbolos) que são modeladas pela interação social. A informação é algo capaz de ser registrado em qualquer formato de suporte (físico ou digital) por conseguinte, comunicada de uma maneira assíncrona e multidirecional de forma interativa.

Portanto, após conceituar dado e informação, a palavra-chave para a concretização de resultados de pesquisas e formulação de hipóteses, vem ao encontro da tecnologia,

representando o mecanismo de trabalho desses dois conceitos já citados (dado e informação) de forma aplicada ao desenvolvimento de novas soluções aos problemas humanos. Nesse sentido, Cupani (2013, p. 18), define a tecnologia como uma forma específica de conhecimento “[...] que implica maneiras específicas de conhecer o mundo material, maneiras que incorporam o saber científico, mas que possuem igualmente características próprias”. Esse fator ressalta o uso da tecnologia como metodologia racionalizada, algo que gera e responde a questões relativas ao que a tecnologia é realmente. Pode ser um método, uma técnica ou o próprio conhecimento materializado em processos de informação, habilidades e/ou procedimentos (KLINE, 2006).

Nesse sentido, a informação é fundamental e essencial nos processos tecnológicos, e observa-se que no campo da Ciência da Informação, é uma das áreas do conhecimento que tem como principal objeto de estudo a informação em seus diversos aspectos. Ressalta-se, que a Ciência da Informação também tem preocupações que são contempladas nos estudos da Ciência da Computação, como é o caso da Proveniência de Dados e a transformação desses dados em informação, além da preocupação constante com o crescimento exponencial nos mais variados setores.

Nesse sentido, a preocupação com o crescimento exponencial da informação no que diz respeito ao seu armazenamento, coloca a tecnologia como fundamento da relação entre Ciência da Informação e Ciência da Computação. Nessa relação reside a aplicação de tecnologias computadorizadas no armazenamento e recuperação da informação, assim como nos produtos, serviços e redes associadas (SARACEVIC, 1996).

Sendo assim, trazer os conceitos de dado, informação e tecnologia se faz necessário para melhor compreender a Ciência da Informação nos diversos contextos de pesquisa que envolve áreas da Ciência da Computação, como é o caso desta tese.

Dessa forma, a informação ganha estruturas que são vistas como algo tangível e como parte dos processos de um Sistema de Informação (SI) (BUCKLAND, 1991). Os SI constituem um conjunto de pessoas, tecnologias e negócios, que interconectados, desempenham e executam processos que são possíveis de análise para tomadas de decisões nos mais variados setores. Os SI se alimentam de dados que são transformados em informação, gerando processos informacionais para a produção de conhecimento (O’BRIEN, 2004).

Nessa perspectiva, destaca-se os Sistemas de Informação em Saúde (SIS), os quais são estruturas tecnológicas compostas por uma gama de Tecnologias de Informação e Comunicação (TIC) capazes de serem alimentados com dados por intermédio humano ou por outros sistemas (WORLD HEALTH ORGANIZATION, 2004). Esses sistemas transformam dados em

informação que dinamizam os processos que servem como base para tomadas de decisões essenciais nos mais variados setores de saúde. A dependência de profissionais de saúde para atuarem nos processos de inserção, coleta, classificação, armazenamento, análise, divulgação e recuperação dos dados, também é essencial para que esses sistemas tenham um bom funcionamento nas organizações de saúde.

Apesar dos desafios técnicos de implantação e operacionalidade das tecnologias envolvidas nos SIS, a disponibilidade dos dados dentro desses sistemas é de extrema importância. Para Abouzahr e Boerma (2005) a tomada de decisões em saúde depende criticamente da disponibilidade oportuna de dados sólidos gerados pelos SIS, para que de fato, seja possível analisar e divulgar esses dados posteriormente.

Assim, para elucidar o tema de pesquisa desta tese, é importante justificar que, a adoção do termo “método”, teve base no conceito de Wiggins e McTighe (2005), onde os autores definem como conjuntos usados para realizar tarefas, abordagens sistemáticas para alcançar um resultado ou objetivo específico e oferece uma descrição de uma forma coesa e (científica) consistente da abordagem que leva ao resultado/objetivo desejado. Os métodos convencionais compreendem a observação/exploração empírica, a teorização e a simulação (HEY; TANSLEY; TOLLE, 2009). Dessa forma, os pressupostos teóricos foram suficientes para a adoção do termo método em relação ao que se pretende nesta tese.

Nesse contexto, esta tese apresenta um método que contempla o gerenciamento de dados de proveniência em SIS de forma que esses dados possam ser centralizados dentro de uma rede interna de saúde para auxiliar nas atividades de captura, preparação, organização, armazenamento, visualização, consulta e análise de dados de proveniência advindos de bases de dados descentralizadas. Importante destacar também, que a centralização dos dados de proveniência através do método aplicado nesta tese serviu como amostragem para as análises realizadas, garantindo que essas análises pudessem apresentar a situação atual, possibilitando ainda compreender os fluxos informacionais analisados por diferentes fontes descentralizadas por meio de uma amostragem centralizada.

Pretende-se, com esta pesquisa, que o método seja utilizado por instituições de saúde que necessitam gerenciar dados de proveniência e que contribua com os processos de tomada de decisões estratégicas com base na Proveniência dos Dados, oportunizando assim a melhoria contínua dos tratamentos e cuidados em saúde. Pretende-se também, que o método possibilite que diferentes profissionais de Tecnologia da Informação (TI) e pesquisadores da comunidade acadêmica deem continuidade a estratégia computacional voltada a Proveniência de Dados desenvolvida até aqui, sanando a problemática apresentada em estudos futuros.

1.1 CONTEXTUALIZAÇÃO

Os SI e as TIC, nos deram uma nova e inovadora forma de comunicação, como viver no espaço cibernético, mensagens instantâneas, comunicar-se com alguém em qualquer lugar, entre outras formas de comunicação que estão mudando não apenas o estilo de vida, mas também o modo como acontecem os negócios em todos os setores (WHO-GLOBAL OBSERVATORY FOR eHEALTH, 2008).

Nesse sentido, realçando a importância das TIC voltadas aos seres humanos, entende-se que essas tecnologias proporcionam a construção de um dos principais cenários tecnológicos voltados à saúde humana, como é o caso dos SIS, definido 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 SIS integrados é a Telemedicina, que utiliza uma gama de equipamentos médicos e sistemas integrados possibilitando gerar diagnósticos e tratamentos independente da distância geográfica, gerando benefícios para os pacientes como, por exemplo, viagens reduzidas ou acesso mais rápido aos conhecimentos adequados (HÅKANSSON; GAVELIN, 2000; WHITTEN, COLLINS, 1997). A Telemedicina gera grandes quantidades de dados, e segundo Bakker *et al.* (2007), o setor da saúde é caracterizado por uma crescente produção de dados relacionados às demandas de atendimentos aos pacientes.

Nessa perspectiva, em consonância com o tema desta tese, no uso da Proveniência de Dados, é importante ressaltar que, o termo “proveniência” de dados diz respeito à origem ou procedência dos dados, e pode também estar relacionado à auditoria, triagem, linhagem e origem do dado (DAVIDSON; FREIRE, 2008). A proveniência pode ser aplicada em ambientes digitais, armazenando os dados em bancos de dados e registrando seus históricos e processos manuais em que podem ser registrados os estados de cada processo manualmente. Ela também auxilia 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 que?”. Nesse sentido, o principal fundamento da Proveniência de Dados é recolher os indícios quanto ao tempo, lugar e, se for o caso, à pessoa responsável pela criação, produção, descoberta ou inserção do dado (BRAGA; BANON, 2008, TAN, 2007; HILLS, 2015; SIMMHAN; PLALE; GANNON, 2005, 2006).

Nesse processo, a Proveniência de Dados desempenha um papel importante no rastreamento dos dados armazenados em diversos tipos de repositórios de dados, apresentando

a validade, a autenticidade e a integridade dos fluxos gerados por esses dados. A configuração da Proveniência de Dados nos vários sistemas de informação, depende de métodos computacionais que devem ser transformados para que haja sucesso nos mais diferenciados sistemas (MOREAU; FOSTER, 2006; BUNEMAN; KHANNA; CHIEW, 2001; BUNEMAN; TAN, 2007).

Nessa perspectiva, ressalta-se a existência dos grandes volumes de dados que são coletados e armazenados de forma descentralizada pelos SIS, os quais precisam ser gerenciados em seus ciclos de armazenamento a partir de suas fontes de origem. Assim, é importante destacar que, são requisitos contribuidores em SIS a Proveniência de Dados e a transparência de uma versão dos dados para compreensão dos fluxos informacionais existentes nesses sistemas.

Nesse contexto, salienta-se a importância da interoperabilidade em conjunto com a Proveniência de Dados, sendo requisitos fundamentais para os SIS, não só para garantir a troca eficaz e eficiente de dados de saúde, mas também para definir fontes de dados confiáveis. Para isso, é importante e natural que os dados de saúde estejam sempre em formato digital, para que todo processo de interoperabilidade possa ser exercido com eficácia em cada etapa entre diferentes SIS. Para Braunstein (2018) e Kilintzis *et al.* (2019) a gestão de dados de saúde é um domínio com várias propostas de soluções e conhecimentos acumulados ao longo de anos de investigação. O uso de padrões de interoperabilidade, por exemplo, é um pré-requisito em quase todos os sistemas médicos.

1.2 MOTIVAÇÃO E JUSTIFICATIVA

A motivação e justificativa social para o desenvolvimento desta tese se dá pela importância do tema citado para as mais variadas áreas, inclusive a Medicina. O Conselho Federal de Medicina (CFM), por meio da Resolução nº 1.821/2007, autoriza as instituições de saúde a utilizarem o SIS para realizar os processos informacionais necessários de armazenamento e manipulação de dados de pacientes no âmbito da saúde, disponibilizando esses dados quando solicitados pelo paciente. Nesse sentido, as instituições de saúde buscam por dados que os guiem para tomadas de decisão no intuito de melhorar a saúde da população e os cuidados de saúde em geral, no contexto de utilizar SIS que contribuam para esse processo.

Em relação à motivação e justificativa científica, com foco voltado para os SIS, a Proveniência de Dados contribui para a interpretação e o entendimento de resultados de um fluxo de dados, pois examinando a sequência de passos que levaram ao resultado, pode-se

compreender o raciocínio utilizado em todo fluxo, sendo possível verificar os dados de entrada, em alguns casos, reproduzir o resultado. Esses processos aplicados em SIS são primordiais para a tomada de decisões e a geração de conhecimento nesses cenários. Logo, esta tese pode proporcionar discussões sobre a Proveniência de Dados em SIS, que de fato, poderão ser ampliadas em outros contextos.

Já em relação à motivação e justificativa tecnológica, ressalta-se que a ciência se apoia cada vez mais nos dados, principalmente em tecnologias que possibilitam a eficiência na coleta, processamento, análise e armazenamento em infraestruturas computacionais, principalmente em contextos de saúde, especificamente em SIS. De fato, esse cenário gerou diversos desafios científicos em gerenciar grandes quantidades de dados em diferentes formatos em cenários de saúde descentralizados. Essa perspectiva apresenta-se nos conceitos de Proveniência de Dados em relação aos SIS, em que os dados capturados se transformam em informações relevantes para tomadas de decisão em cenários de saúde.

Portanto, a relevância do tema proposto nesta tese, pode contribuir para a compreensão do funcionamento dos processos de interoperabilidade que envolvem os SIS em um contexto de saúde pública no Brasil e em países que utilizam SIS para o gerenciamento de seus dados. Assim, importante destacar que o crescimento exponencial de armazenamento dos dados de saúde no mundo é um problema crítico, principalmente no que diz respeito ao rastreamento da proveniência desses dados, e a proposta desta tese tem sua relevância pautada neste quesito. Por fim, no campo da Ciência da informação interdisciplinarmente com o campo da Ciência da Computação, esta tese contribui para que diferentes pesquisadores dessas duas grandes áreas, aproximem pesquisas sobre o tema em âmbito nacional e internacional, com intuito de construir saberes que oportunizem melhorias para o gerenciamento de dados de proveniência em SIS.

1.3 PROBLEMAS DE PESQUISA

Problemas relacionados aos processos de gestão de dados em SIS, podem gerar a perda desses dados, além de comprometer a privacidade, segurança e a disponibilidade, afetando tomadas decisões em situações que necessitam urgência. Nesse contexto, considerando o grande volume de dados produzidos diariamente por SIS, a variedade de fontes descentralizadas e os diferentes formatos em que os dados de saúde são encontrados, têm gerado uma problemática desafiadora em relação à interoperabilidade e o gerenciamento dos dados entre diferentes cenários de saúde.

Sendo assim, com a expansão e popularização dos SIS, são necessárias estratégias

computacionais (algoritmos) e ferramentas computacionais cada vez mais interoperáveis para os processos de gestão dos dados nesses sistemas, que em sua maioria são de fontes descentralizadas. Nesse sentido, a interoperabilidade é um problema em toda a área da saúde, pois diferentes formatos de dados de saúde geram processos complexos de padronização clínica, criando enormes barreiras para oferecer transferências de dados entre diferentes SIS, consequentemente dificultando experiências aprimoradas e resultados de atendimento aos pacientes.

Assim, é importante ressaltar que existem muitos problemas de gerenciamento de dados na área da saúde, principalmente em SIS que trabalham com insumos de quantidades avassaladoras de dados fragmentados, tornando-se um grande desafio gerenciar toda o fluxo de dados que compõem esses sistemas devido à complexidade de todos os processos que envolvem os dados. No contexto de saúde, os dados podem ser armazenados em vários locais de forma descentralizada, apresentando complexas estruturas computacionais necessárias para seu gerenciamento, principalmente no que compete à proveniência desses dados.

Nesse entendimento, ainda é importante ressaltar que, das poucas iniciativas no desenvolvimento de métodos para o gerenciamento de dados de proveniência voltados para SIS descritos na literatura, principalmente com base no uso de modelos de Proveniência de Dados, de fato, possuem uma estrutura matemática para cada caso. Ou seja, a falta de métodos que sejam capazes de rastrear e gerenciar os dados de proveniência em SIS fora do âmbito dos gerenciadores nativos, que ainda se encontram em um estágio incipiente, constitui em traçar todos os processos, estruturas e tecnologias a serem empregadas no desenvolvimento de um método para esse fim. Isso, de fato, exige um estudo aprofundado sobre o ambiente de aplicação e suas respectivas funcionalidades para garantir a adaptação ideal, constituindo um desafio no método que se propõe nesta tese.

Por fim, considerando a ausência de um método que apresente abordagens para o gerenciamento de dados de proveniência em diferentes SIS, a relação desta tese com a problemática aqui apresentada, é promover ao longo dessa pesquisa, discussões que possibilitem compreender e discutir os esforços computacionais já existentes nos fluxos informacionais dos SIS, para com base nesses pressupostos propor um esforço computacional para o desenvolvimento de um método que contribua para resolver, parte dos problemas aqui apresentados, ou sua totalidade.

1.4 QUESTÃO DE PESQUISA E HIPÓTESE

Considerando os problemas de pesquisa apresentados e no intuito de nortear a investigação e o desenvolvimento desta tese, foi definida a seguinte questão geral para esta pesquisa: 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, dada a complexidade do gerenciamento dos dados de proveniência em SIS, foi definida uma hipótese a ser validada com a execução do método proposto nesta tese. Segundo Marconi e Lakatos (2003), as hipóteses são respostas possíveis e provisórias em relação às questões de pesquisa, tornam-se também instrumentos importantes como guias na tarefa de investigação. Nesse sentido, para esta tese foi definida a seguinte hipótese: **H1** – É possível afirmar que um método para gerenciamento de dados de proveniência em SIS, constituído de processos de interoperabilidade em consonância com as principais características da proveniência de dados, pode ser ajustado para diferentes contextos informacionais de saúde, contribuindo para a descrição da proveniência, das ações e dos envolvidos na construção e alteração desses dados.

1.5 OBJETIVOS

Nas próximas subseções são apresentados o objetivo geral e os objetivos específicos que norteiam esta tese.

1.5.1 Objetivo Geral

Propor um método de gerenciamento de dados de proveniência em SIS que possa contribuir para os processos de rastreamento e interoperabilidade de dados de saúde em fontes descentralizadas.

1.5.2 Objetivos Específicos

Aqui são descritos os objetivos específicos que correspondem aos resultados que esta tese pretende alcançar e contribuem para o alcance do objetivo geral.

- a) Apresentar os principais pressupostos necessários para o gerenciamento de dados de proveniência em SIS;
- b) Identificar métodos, técnicas, modelos ou metodologias computacionais que apoiam o gerenciamento dos dados de proveniência em SIS;
- c) Elaborar um método para gerenciamento de dados de proveniência em SIS com base no instanciamento do modelo W3C PROV;
- d) Aplicar o método em um cenário de saúde real para o estudo de caso;
- e) Validar o método com especialistas em gerenciamento de dados em SIS.

1.6 INEDITISMO DA TESE

Com intuito de comprovar a originalidade do método e o ineditismo desta tese, apresentando trabalhos relacionados com o tema, foi elaborada uma Revisão Sistemática da Literatura (RSL). Primeiramente, foram realizadas buscas nas principais bases de dados nacionais e internacionais relacionadas à Ciência da Informação, como: Base de Dados em Ciência da Informação (BRAPCI), Repositório Institucional Digital do IBICT (RIDDI), *Scientific Electronic Library Online (SciELO)*, *Information Science & Technology Abstracts (ISTA)*, *Library, Information Science & Technology Abstracts (LISTA)*, *Library and Information Science Abstracts (LISA)*, *Digital Library of Information Science & Technology (DLIST)* e *E-prints in Library and Information Science (E-LIS)*, porém, não foram encontrados trabalhos relacionados ao tema desta tese.

Nesse sentido, optou-se por cinco bases de dados internacionais, que teriam relações interdisciplinares com a Ciência da Informação e a Ciência da Computação, para de fato, tentar encontrar trabalhos relacionados ao tema desta tese, quais sejam: *ACM Digital Library*, *IEEEExplore*, *Science Direct*, *Scopus*, *Springer Link* e *Web of Science*. As buscas foram realizadas por meio da utilização de palavras-chave nos mecanismos de busca das bases de dados, sempre adequando as *strings* aos mecanismos de cada base. Foram utilizados os seguintes termos na *string* de busca: *Data Provenance*, *Health*, *Telemedicine*, *e-Health*, *m-Health* e *Healthcare*, os quais retornaram o maior número de documentos que condiziam com a temática. O recorte de tempo para as buscas nas bases de dados foi entre 2010 e 2020, por se tratar do período que mais retornou documentos, durante os testes.

Após realizar os processos que integram a RSL, foi possível observar a escassez de trabalhos relacionados à temática nos resultados encontrados, pois os eventos importantes que relacionam a Proveniência de Dados não são indexados em sua totalidade nas bases de dados

pesquisadas nesta tese. Nesse contexto, ainda foi possível refletir que diversos trabalhos versam sobre a importância da Proveniência de Dados nos mais variados aspectos, porém, não foram encontradas na literatura publicações que enfatizam métodos para o gerenciamento dos dados de proveniência em SIS especificamente com as estratégias propostas nesta tese. Foram encontradas apenas estruturas similares com a proposta aqui em questão, o que de fato, faz desta tese ter seu caráter de ineditismo, oportunizando novas pesquisas sobre o tema no campo da Ciência da Informação e da Ciência da Computação em conjunto com às outras áreas do conhecimento.

1.7 DELIMITAÇÃO DO ESCOPO

Primeiramente, é importante ressaltar que a pesquisa desta tese perpassa o campo da Ciência da Informação no quesito interdisciplinar com a Ciência da Computação. A informação como objeto de estudo entre essas duas ciências, compactua com o tema desta tese, utilizando-se de aparatos tecnológicos para compreender melhor a problemática desta pesquisa. Ambas as ciências compactuam com a área de saúde, especificamente os SIS, procurando realizar um esforço teórico-prático para uma solução que servirá de base para pesquisas futuras sobre a temática aqui abordada.

À vista disso, o método aqui proposto não tem a intenção de explorar questões relacionadas com a interação do usuário, tampouco com a estrutura de design dos sistemas, mas, sim, visa contribuir para o gerenciamento dos dados de proveniência em SIS para posterior avaliação dos responsáveis pelo gerenciamento dos dados em SIS. Além disso, para o método aqui proposto, foi desenvolvido um protótipo que aborda o gerenciamento dos dados de proveniência em SIS sem o objetivo de desenvolver uma aplicação *web* completa em sua versão final. Ou seja, o desenvolvimento do protótipo tem a função de provar a hipótese proposta nesta tese (H1), e responder aos objetivos traçados, não contemplando uma visão de implementação em larga escala devido ao tempo que seria necessário para este processo. Dito isso, outro ponto a ressaltar, é em relação à infraestrutura de rede, a qual, não se teve a intenção de tratar nesta tese, pois pressupõe-se a compatibilidade da estratégia computacional aqui desenvolvida (protótipo) para apresentação do método proposto para este estudo.

Portanto, devido ao grande volume de dados nos SIS, o método aqui proposto por meio do protótipo desenvolvido não tem a pretensão de estabelecer padrões de qualidade de sistemas já existentes, ou seja, sistemas gerenciadores nativos, e nem de demonstrar falhas, se houver. Outro ponto a ser destacado, é que esta tese não tem a intenção de avaliar os gerenciadores de

dados atuais em funcionamento nos SIS, mas, sim, ir ao encontro de uma contribuição para sanar os problemas já descritos no desenvolvimento de um método que realize o gerenciamento de dados de proveniência em SIS para os diferentes setores de saúde.

Por fim, o que se espera, é que quando os engenheiros de software para esses cenários de saúde se depararem com os problemas de pesquisa mencionados nesta tese, possam utilizar o método como uma das alternativas para estabelecer a melhoria dos fluxos informacionais em relação ao gerenciamento e rastreamento de dados de proveniência em repositórios de dados descentralizados de diferentes SIS.

1.8 CONTRIBUIÇÕES DA PESQUISA

A priori, a temática desta tese no campo da Ciência da Informação, vem ao encontro de contribuir para a interdisciplinaridade em relação a outras áreas, como: Ciência da Computação e Ciências da Saúde. Além da interdisciplinaridade entre essas ciências, conseqüentemente o desenvolvimento de novos saberes, favorece novas formas de aproximação no contexto social, científico e tecnológico em relação à pesquisa desta tese.

Quanto às contribuições desta pesquisa no âmbito social, estas são voltadas no intuito de contribuir para garantia do gerenciamento dos dados de proveniência em SIS que necessitam ser rastreados em ambientes descentralizados, para a transformação desses dados em informação necessária a tomadas de decisão que geram conhecimento nesses cenários de saúde. Assim, profissionais de TI juntamente com os gestores de instituições de saúde, terão a oportunidade de ampliar suas visões com a proveniência dos dados de saúde advindos de diferentes SIS, contribuindo assim, para uma gestão mais focada na geração de benefícios à saúde em geral. Em relação às contribuições sociais desta tese no campo da Ciência da Informação, vale ressaltar que esta ciência possui um papel social importante quando associada às questões de saúde humana na relação com a tecnologia, buscando compreender as transformações existentes da informação em ambientes de saúde.

Em relação às contribuições no âmbito científico, o método aqui proposto é uma abstração da realidade onde acredita-se ter capturado características marcantes do processo de gerenciamento dos dados de proveniência em SIS. Com isso, a comunidade acadêmica e pesquisadores da área, podem se beneficiar com a pesquisa, pois a mesma traz à tona estudos teóricos e práticos relacionados na literatura internacional por meio de uma RSL, oportunizando ampliações do tema. Indo além, o tema desta tese permite a pesquisadores desenvolver novas tecnologias, possibilitando novas observações sobre os possíveis métodos de gerenciamento de

dados de proveniência em SIS, que, por sua vez, permitem ampliar a construção de novos conhecimentos científico-teóricos-práticos, que, em seguida, vão inspirar outras pesquisas. Dessa forma, os resultados desta pesquisa são disponibilizados no campo da Ciência da Informação, possibilitando a pesquisadores de diferentes áreas que compactuam com essa ciência, a imergir no escopo desta pesquisa ampliando a construção e propagação dos novos conhecimentos, estimulando a busca por novas respostas para os resultados encontrados.

Já em relação às contribuições no âmbito tecnológico, foram utilizadas ferramentas tecnológicas de interoperabilidade de dados, banco de dados e linguagens de programação que proporcionaram o desenvolvimento do método, com intuito de contribuir no gerenciamento de dados de proveniência de grandes volumes de dados gerados pelos SIS. Dessa forma, entende-se que esta pesquisa de tese pode proporcionar discussões sobre a Proveniência de Dados em SIS como forma de contribuição nos diversos processos gerenciais dos dados gerados por estes sistemas, sendo de significativa importância o controle da origem dos dados para tomadas de decisões dentro desses cenários de saúde. Nesse âmbito, com método apresentado nesta tese, busca uma solução tecnológica para os problemas já identificados. Assim, a temática desta tese, relacionada à proveniência de dados no campo da Ciência da Informação, busca garantir no uso de tecnologias no campo da Ciência da Computação, que os dados capturados em repositórios de dados descentralizados em SIS tenham suas origens detalhadas, sendo considerados confiáveis para o gerenciamento de novos fluxos de informação e de conhecimento no campo das Ciências da Saúde.

Nesse sentido, destaca-se a contribuição desta tese em relação à temática aqui apresentada, por se tratar de um assunto pouco explorado na literatura voltada a SIS, como poderá ser conferido por meio do referencial teórico e da Revisão Sistemática de Literatura (RSL) elaborados nesta pesquisa. De fato, com base nos pressupostos teóricos, a serem descritos nas próximas seções, evidencia-se contribuir para o aprimoramento de estudos relacionados com a problemática voltada ao gerenciamento de dados de proveniência em SIS, inclusive contribuindo de forma interdisciplinar dentro do campo da Ciência da Informação, a qual traz subsídios para compreender os fluxos informacionais que trata a Proveniência de Dados no campo da Ciência da Computação. Ambas as ciências se entrelaçam na pesquisa desta tese e promovem contribuições significativas para o campo da saúde.

Por fim, os SIS são um domínio em constante evolução, sendo encontrados desafios em projetar e desenvolver tecnologias que beneficiem os processos e serviços de saúde desses sistemas. Assim, no que permeia o campo dos profissionais da Ciência da Informação e da Ciência da Computação envolvidos nos processos tecnológicos de saúde, oferece um espaço de

reflexão na construção de novos conhecimentos técnicos perante as perspectivas críticas relacionadas aos cuidados de saúde, influenciando o desenvolvimento de métodos que contribuem para a evolução dos serviços de saúde em SIS, como o método proposto nesta tese.

1.9 ADERÊNCIA DO TEMA AO PGCIN E À ÁREA DE CIÊNCIA DA INFORMAÇÃO

O Programa de Pós-graduação em Ciência da Informação (PGCIN¹), da Universidade Federal de Santa Catarina (UFSC), *Campus* de Florianópolis, foi criado no ano de 2003, e visa a formação de profissionais altamente qualificados e capazes para pesquisas de aspectos teóricos, metodológicos e práticos. Essas pesquisas estão relacionadas à produção, comunicação e representação do conhecimento e da informação. A investigação dos processos, ambientes, serviços, produtos e sistemas de gestão da informação e do conhecimento também fazem parte dos pressupostos do PGCIN/UFSC (UNIVERSIDADE FEDERAL DE SANTA CATARINA, 2022a).

Pesquisas voltadas ao desenvolvimento epistemológico, teórico e prático são promovidas desde a criação do PGCIN/UFSC na área de concentração, Informação e Sociedade, com ênfase no estudo dos fenômenos relacionados aos processos centrais de gerenciamento, seleção, processamento, organização, recuperação, uso e disseminação da informação, visando principalmente abordar, explorar, compreender e explicar esses fenômenos (UNIVERSIDADE FEDERAL DE SANTA CATARINA, 2022b).

Por conseguinte, o tema desta tese se insere na linha de pesquisa “Dados, Inteligência e Tecnologia”, parte do PGCIN/UFSC, a qual tem seus pressupostos em

Investigar os processos, ambientes, serviços, produtos e sistemas envolvendo dados, informação e conhecimento, por meio de abordagens interdisciplinares sobre coleta, gerenciamento, produção, armazenamento, compartilhamento, transmissão, recuperação, acesso, segurança, visualização, interconexão e avaliação de dados e informações existentes nos mais diversos contextos, tendo em vista a sustentabilidade das organizações. Como suporte, estuda, aplica e desenvolve técnicas, métodos e ferramentas para a criação de tecnologias inteligentes, prospectivas e inovadoras (UNIVERSIDADE FEDERAL DE SANTA CATARINA, 2022b).

Dessa maneira, a área de concentração em Informação e Sociedade do PGCIN/UFSC a qual esta tese está alinhada, possibilita estudar, compreender a concepção, o desenvolvimento e implantação de soluções voltadas ao tema desta tese.

Feito o alinhamento com a área de concentração e a linha de pesquisa do PGCIN/UFSC,

¹ <https://pgcin.ufsc.br/>

ainda há uma ligação direta nos processos de armazenamento e recuperação da informação, segurança da informação, informação em saúde, SIS e visualização da informação, sendo subáreas compreendidas para o tema desta tese.

Importante ressaltar também, que a interação mais elaborada entre os saberes da Ciência da Computação e da Ciência da Informação, ultrapassa uma abordagem que vem se tornando usual, a “via de mão dupla”. Dessa forma, compreende-se que a Ciência da Computação pode contribuir com a Ciência da Informação oferecendo mecanismos automatizados para o armazenamento e recuperação da informação, e que de fato, inversamente, a Ciência da Informação pode contribuir com a Ciência da Computação no sentido de propor modelos de representação da informação que oportunizem um arranjo lógico mais elaborado sobre o qual os sistemas computacionais possam agir (CAFEZEIRO; COSTA; KUBRUSLY, 2016).

Sendo assim, dentro do campo da Ciência da Informação, especificamente em relação ao tema desta tese, observa-se evidências relacionadas à Proveniência de Dados, a qual é oriunda da Ciência da Computação e transita entre a Ciência da Informação. Ressalta-se que o termo “Proveniência” está contemplado no “Dicionário de Biblioteconomia e Arquivologia” descrito pelos pesquisadores Murilo Bastos da Cunha e Cordélia Robalinho de Oliveira Cavalcanti, publicado em 2008.

p. da proveniência *principle of provenance, respect des fonds arq* princípio arquivístico fundamental, segundo o qual os documentos ou os arquivos originários de uma instituição, de uma corporação, de uma família ou de uma pessoa não devem ser incorporados a documentos ou arquivos de outras procedências; inclui, às vezes, o princípio do respeito à ordem original; princípio do respeito aos fundos, respeito aos fundos. <=> integridade dos fundos, princípio do respeito à estrutura arquivística, princípio do respeito à ordem original, proveniência, proveniência territorial. (CUNHA; CAVALCANTI, 2008, p. 291).

A Ciência da Informação absorve o termo de Proveniência de Dados em diversos cenários de aplicação, pois os dados geram informação e a informação gera conhecimento para tomada de decisões em diferentes cenários informacionais por meio de tecnologias empregadas.

Desde a origem da Ciência da Informação, muitos processos da Ciência da Computação foram incorporados por essa ciência em seus respectivos tempos, seja de forma direta ou indiretamente. As tecnologias de informação compactuando com as definições de proveniência, foram utilizadas para o desenvolvimento de pesquisas na descrição das origens dos dados, contribuindo assim para o desenvolvimento das duas ciências: Ciência da Informação e Ciência da Computação (FREUND; SEMBAY; MACEDO, 2019).

Dessa forma, os SIS que compactuam com a Proveniência de dados no tema desta tese, também tem destaque no campo da Ciência da Informação, especialmente sob a perspectiva

interdisciplinar. Dito isto, é importante ressaltar que estudos entre áreas que envolvem a saúde e a Ciência da Informação são possíveis, pois conforme Ribeiro (2009), a Ciência da Informação se assume como uma ciência social, e a medicina e saúde ao tratar dos indivíduos, trabalha no nível do entendimento dos contextos sociais, além do que trabalha diretamente com informação, seja de natureza clínica, científica, informação de caráter administrativo, entre outras informações diferentes, caracterizando objeto de estudo direto da Ciência da Informação.

Ainda, é possível afirmar que a Ciência da Informação é um cenário ideal para os pressupostos desta tese, pois trata-se de uma ciência interdisciplinar que assume diversas disciplinas do conhecimento tecnológico, e que, de fato, contribui para geração de novas pesquisas científicas também para a área de SIS.

Por fim, a proposta do tema desta tese em relação ao método proposto dentro do PGCIN da Universidade Federal de Santa Catarina (UFSC), compactuada a preocupação com a Proveniência de Dados no contexto de SIS, considerando os campos de pesquisa em Ciência da Informação e sua interdisciplinaridade com a Ciência da Computação. Nesse ínterim, considerando o grande volume de dados que necessitam ser gerenciados dentro dos SIS, a Ciência da Informação contribui para os quesitos de análise, classificação e visualização da informação em saúde, oportunizando a compreensão dos fluxos informacionais dentro desses sistemas, e a Ciência da Computação, contribui na disponibilização dos recursos necessários para a criação da estratégia computacional do método proposto.

1.10 ESTRUTURA DA TESE

Esta tese está dividida em oito seções. Na primeira seção encontra-se a Introdução, que descreve uma contextualização geral sobre o tema e as justificativas. Na sequência são definidos os problemas de pesquisa, a questão de pesquisa e hipótese, o objetivo geral e os objetivos específicos. Logo após, foram descritos o ineditismo, delimitações, contribuições e aderência do tema ao PGCIN e a Ciência da Informação; e, ainda, a estrutura do documento.

Na segunda seção é contemplado o referencial teórico, no qual são elencadas as bases teóricas relativas ao tema desta tese, contendo os seguintes assuntos: Ciência da Informação, Proveniência de Dados, Sistemas de Informação e Saúde, Telemedicina, Prontuário Eletrônico do Paciente, fechando a seção com as considerações finais da seção.

Na terceira seção é apresentada a Revisão Sistemática de Literatura com complemento da técnica de Snowballing (*backward* and *forward*) de forma detalhada, contendo, o protocolo utilizado, o resumo dos estudos selecionados e resultados. Esta seção fecha com as lições

aprendidas em consideração aos estudos encontrados e evidencia originalidade e ineditismo desta tese nas considerações da seção.

Na quarta seção são abordados os aspectos metodológicos relacionados ao desenvolvimento desta tese, especificando sua caracterização, universo, população, amostra e em seguida detalhando os procedimentos metodológicos que se pretende implementar para o alcance dos objetivos juntamente com o detalhamento sobre as premissas de coleta e análise dos dados.

Na quinta seção é apresentado o método “PROV-Health”² para o gerenciamento de dados de proveniência em SIS de forma genérica, detalhando em seguida a sua metodologia de concepção, o instanciamento realizado com base no modelo de proveniência de Dados W3C PROV e a proposta de instrumentalização dos serviços de saúde a serem utilizadas no método. Logo após, a arquitetura genérica do método é apresentada com base na constituição de suas camadas (Agentes, Entidades e Atividades), sendo que na camada de Atividades, três módulos são detalhados. Por fim, a verificação dos componentes e relações da arquitetura genérica do método com base no W3C PROV foi realizada a fim de afirmar sua concepção, finalizando com as considerações da seção.

Na sexta seção é apresentada a experimentação realizada por meio do PROV-Health, onde foi contextualizado o ambiente de estudo de caso, o STT/RCTM, no contexto de sua arquitetura para experimentação. Logo em seguida, foram definidos os consumidores e clientes do PROV-Health e as tecnologias empregadas na estratégia computacional, ambos no STT/RCTM. Logo após, foram definidos os perfis FHIR para os mapeamentos de dados necessários, juntamente com a instrumentalização específica dos serviços de saúde para o STT/RCTM. Em seguida, a metodologia de experimentação do PROV-Health foi descrita. Na sequência, e arquitetura específica para experimentação do PROV-Health, detalhando sua estratégia computacional por camadas e módulos, bem como, os experimentos realizados, foram apresentados. Em seguida, foram descritos os principais desafios e ameaças encontradas no processo de experimentação do PROV-Health no STT/RCTM. Por fim, finaliza-se com as considerações da seção.

Na sétima seção são apresentadas as análises referentes aos experimentos realizados por meio do PROV-Health no STT/RCTM. Em seguida, é apresentada a avaliação sob a ótica dos especialistas em relação ao PROV-Health. Por fim, finaliza-se com as considerações da seção.

² Programa de Computador desenvolvido pelo autor e profissional da equipe técnica do STT/RCTM, juntamente com o orientador desta tese, com registro no Instituto Nacional da Propriedade Industrial (INPI), sob o número BR 51 2023 001481 8, expedido em 06 de junho de 2023. (Ver pedido no APÊNDICE H e registro no APÊNDICE I).

Na oitava seção são apresentadas as conclusões e trabalhos futuros, seguido dos artigos publicados na jornada de doutoramento e publicações submetidas para análise. E, por fim, são listadas as referências utilizadas nesta tese, além dos apêndices.

2 REFERENCIAL TEÓRICO

Nesta seção é apresentado o referencial teórico, em que são tratados os seguintes assuntos: Ciência da Informação — área a qual esta tese está inserida, seus conceitos, relações interdisciplinares e o contexto na saúde; Proveniência de Dados — tema-chave desta tese, relacionando seus componentes e funcionamento, principais modelos, aplicações gerais e na saúde; SIS — sistemas que geram grandes volumes de dados de proveniência, apresentando seus principais tipos, interoperabilidade, integração, armazenamento e a base entre W3C PROV e HL7 nos processos de Proveniência de Dados em SIS; Telemedicina — SIS ideal para uso em estudo de caso no que se refere ao gerenciamento dos dados de proveniência, apresentando seu funcionamento e aplicações, seu foco no Brasil e em Santa Catarina; Por fim, Prontuário Eletrônico do Paciente — um dos SIS mais utilizados no Brasil, sendo apresentada a sua estrutura, aspectos éticos e legais e as contribuições da Ciência da Informação nesse sistema.

2.1 CIÊNCIA DA INFORMAÇÃO

Esta seção apresenta um breve conceito, histórico e epistemologia da Ciência da Informação. Logo após, são apresentados pressupostos de interdisciplinaridade entre a Ciência da Informação e a Ciência da Computação, as quais, são duas ciências que se entrelaçam na proposta desta tese. E, por fim, a Ciência da Informação em saúde, destacando a importância desta ciência em relação ao tema desta tese.

2.1.1 Conceito, histórico e epistemologia

Para Capurro e Hjørland (2007), o foco da Ciência da Informação implica na abordagem tanto sociológica quanto epistemológica, voltada para as fases que Borko (1968) aponta como: geração, coleta, organização, interpretação, armazenamento, recuperação, disseminação, transformação e uso da informação.

Ao avaliar os procedimentos que implicam a origem da Ciência da Informação, destaca-se as observações histórico-sociais da disciplina na proposição de elementos que foram essenciais para a sua constituição em seu campo científico no vislumbrar da sua configuração identitária, primeiro na Inglaterra e depois nos Estados Unidos, nas primeiras décadas do século XX (ARAÚJO, 2011; FEATHER; STURGES, 2003; FREIRE; SILVA, 2012).

Nesse sentido, a Ciência da Informação tem seu próprio estatuto científico, como ciência

social, sendo que, por sua natureza interdisciplinar apresenta interfaces com a Matemática, Lógica, Linguística, Psicologia, Ciência da Computação, Engenharia da Produção, Artes Gráficas, Comunicação, Biblioteconomia, Administração e outros campos científicos semelhantes (BORKO, 1968; FOSKETT, 1980; PINHEIRO, 1999).

A Ciência da Informação com elementos de ciência pura e aplicada, investiga seu objeto (a informação) no desenvolvimento de serviços e produtos. Com relação à abordagem interdisciplinar desta ciência, os estudos epistemológicos buscam mapear as relações entre a Ciência da Informação e outras áreas científicas (BORKO, 1968; FOSKETT, 1980; PINHEIRO, 1999).

2.1.2 Relações interdisciplinares: Ciência da Informação e Ciência da Computação

É importante destacar as relações interdisciplinares da Ciência da Informação e da Ciência da Computação nesta tese, pois trata-se de uma proposta com elementos tecnológicos que compactuam entre as duas ciências.

Nessa perspectiva, as relações interdisciplinares da Ciência da Informação com a Ciência da Computação, conforme Machlup e Mansfield (1983), não existem de tal forma numa completa união, mas sim, numa coabitação, ou seja, na coexistência das duas disciplinas forma um guarda-chuva curricular.

Dessa forma, vale ressaltar que a preocupação com a origem dos dados e informações é uma constante na Ciência da Informação e também é contemplada nos estudos de Ciência da Computação no que se refere à proveniência dos dados, apontando, assim, mais um quesito da interdisciplinaridade entre as ciências por compartilharem as mesmas preocupações, porém, de diferentes pontos de vista e com diferentes aparatos científicos e tecnológicos (FREUND; SEMBAY; MACEDO, 2019). Conforme aponta Alvarenga (2001, p. 12), “mudam-se os meios, sofisticam-se os instrumentos e técnicas e surgem nomes novos para designar coisas que já se faziam no passado”. Entretanto, a essência das coisas permanece. Desse modo, pode-se observar que nos princípios de evolução da tecnologia, o uso de Proveniência de Dados não é um termo novo, porém pode estar sendo contemplado e pensado com outro contexto. Esta afirmação, combinada com o apontamento de Borko (1968), ao aludir sobre a origem da Ciência da Informação, evidencia a relação interdisciplinar existente entre essa ciência, a partir das premissas da Ciência da Computação.

Em relação às áreas interdisciplinares mencionadas por Borko (1968), Capurro (2003) afirma que duas raízes da Ciência da Informação se destacam: uma é a Biblioteconomia, em

termos mais gerais, o estudo dos problemas relacionados com a transmissão de mensagens e está ligada a todos os aspectos sociais e culturais próprios do mundo humano; e a outra raiz, a Ciência da Computação, que se refere ao impacto nos processos de produção, coleta, organização, interpretação, armazenagem, recuperação, disseminação, transformação e uso da informação.

Nessa concepção, além da interdisciplinaridade da Ciência da Informação, aponta a participação ativa e deliberada na evolução da sociedade da informação, que a apresenta como uma ciência ativa nos processos de construção de conhecimento na sociedade atual, além da preocupação com a informação registrada, as tecnologias e os serviços que facilitem a sua gestão e utilização (SARACEVIC, 1996; SARACEVIC, 2010).

2.1.3 Ciência da informação e saúde

A relação da Ciência da Informação e saúde vai além dos acervos documentais e científicos pela produção de informações por parte dos profissionais da área de saúde, pois o ciclo da informação no que compete ao seu uso, abre perspectivas para o estudo do comportamento informacional traçada pela colaboração entre os membros das duas áreas (NEVES; BRAZ, 2018).

Sobre isso, Ribeiro (2009) aponta que no campo da saúde, todas as questões de produção, preservação e uso da informação, seja no contexto de serviços, seja no contexto de sistemas e organizações, depende da percepção e explicação à luz da fundamentação da Ciência da Informação.

Levando em conta a temática desta tese, Ribeiro (2009) aponta que a Ciência da Informação e saúde podem compactuar no que compete a análise de sistemas de informação e sua otimização em termos de funcionamento. Nesses sistemas de informação, leva-se em conta desde estudos orgânico-funcionais, análise e representação dos fluxos de informação, aplicação de normas para representação da informação, estudos de avaliação de informação com vista à sua preservação a longo tempo, implicando isso a implementação de modelos e esquemas de metainformação.

Portanto, a relação da Ciência da Informação com a saúde, gera competência em informação em saúde, que para Fong (2018), é compreendida como a capacidade de obter, processar e compreender informações básicas de saúde e serviços necessários para a tomada de decisões apropriadas no campo da saúde. Ressalta ainda que, a tecnologia da informação, na maioria das vezes, encaixa nesses movimentos de inter-relação entre Ciência da Informação e

a saúde, “[...] a fim de contribuir para o alcance da competência em informação que se espera da população assistida e das organizações de saúde que assistem tal população” (BRITO *et al.*, 2020, p. 184).

2.2 PROVENIÊNCIA DE DADOS

Para Moreau (2006), existem mais de uma centena de formas diferentes de utilização do termo proveniência, o que faz necessária uma clara e ampla definição do conceito de proveniência. O dicionário inglês Oxford, fornece a seguinte definição para o termo proveniência (provenance): *(i) the fact of coming from some particular source or quarter; origin, derivation. (ii) the history or pedigree of a work of art, manuscript, rare book etc.* (OXFORD ENGLISH DICTIONARY, 2015). A definição pelo dicionário em inglês deixa clara a adoção do termo na construção textual desta tese, pontuando essa definição para a descrição de um dado, ou seja, a adoção da Proveniência de Dados como foco de pesquisa no contexto dos Sistemas de Informação em Saúde.

A Proveniência de Dados possui várias definições, porém, nesta tese foram selecionadas aquelas que mais se aproximam e possibilitam o alinhamento com os aspectos da Ciência da Informação no uso do termo escrito em português. Segundo Buneman, Khanna e Chiew (2001), a Proveniência de Dados é a documentação complementar de um determinado dado que contém a descrição de “como”, “quando”, “onde” e “por que” ele foi obtido e “quem” o obteve. Simmhan, Plale e Gannon (2005) exemplificam que ao comprar uma obra de arte é importante saber a sua procedência, desde a sua criação, incluindo todos os antigos donos, ou seja, essa informação será essencial para se estabelecer o valor da obra de arte. De igual forma a Proveniência de Dados possibilita garantir a qualidade e a veracidade dos dados. Freire *et al.* (2008) ressalta que a Proveniência de Dados pode ser definida como informação que auxilia a determinar a derivação histórica do produto de dado, a partir de suas fontes de origem, sendo considerado um componente essencial para permitir reprodutibilidade do resultado, compartilhamento e reuso de conhecimento pela comunidade científica.

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 sentido, a Proveniência de Dados identifica transformações pelas quais os dados inseridos nos SI passaram ao longo do tempo (DAVIDSON; FREIRE, 2008; SEMBAY; MACEDO; DUTRA, 2020a).

Nesse contexto, é importante ressaltar que, a Proveniência de Dados possui

características particulares que justificam sua aplicação em diferentes cenários. Assim, conforme Simmhan, Plale e Gannon (2005), os vários cenários demandam muitos esforços de pesquisa e a criação de uma taxonomia para categorizar esses esforços, contribui para que sistemas de proveniência possam ser construídos para operar de diversas formas, com diferentes características e operações. Dessa forma, parte da taxonomia definida pelos autores, apresenta as principais características da Proveniência de Dados a diferentes cenários de aplicações, sendo essas características:

- a) Qualidade dos dados: a linhagem pode ser usada para estimar a qualidade e confiabilidade dos dados com base nos dados originais e nas suas transformações, sendo possível oferecer declarações de prova na derivação dos dados;
- b) Trilhas de auditoria: a proveniência pode ser usada para rastrear a trilha de auditoria determinar o uso e detectar erros na geração de dados;
- c) Receitas de replicação: informações detalhadas de proveniência permitem a repetição da derivação de dados, ajudam a manter a atualização dos dados
- d) Atribuição: o pedigree dos dados pode estabelecer direitos autorais e propriedade de dados, possibilitar citações e determinar a responsabilidade na ocorrência de dados incorretos; e,
- e) Informacional: o uso genérico da linhagem dos dados é consultar com base em metadados para a descoberta de dados. Possibilita fornecer um contexto para interpretar os dados.

Dito isso, nesta tese, a Proveniência de Dados contribui para o entendimento interdisciplinar entre a Ciência da Informação e a Ciência da Computação, factível para contribuição de novos estudos no que compete ao tema aqui em questão no contexto entre essas duas ciências.

2.2.1 Componentes e funcionamento

Em relação aos componentes da Proveniência de Dados, um dos mais importantes é a causalidade. Nesse componente é guardada a descrição do processo (ou sequência dos passos) que, junto com dados de entrada e seus respectivos parâmetros, levam à criação de um conjunto de dados final. Assim, as dependências dos processos são usadas para documentar sua criação, bem como auxiliar na reprodução e validação desse processo. A causalidade pode ser inferida tanto para a proveniência prospectiva como para a retrospectiva (FREIRE *et al.*, 2008).

Em relação à proveniência prospectiva e retrospectiva, explica Freire *et al.* (2008), que

a forma prospectiva captura os passos que devem ser seguidos para a geração de um dado produto (por exemplo, conjunto de processos, um *script*). A forma de captura de proveniência retrospectiva captura os passos executados por uma tarefa computacional assim como a informação sobre o ambiente utilizado para derivar um dado produto específico, ou seja, um *log* detalhado sobre a execução da tarefa.

Ressalta-se ainda a proveniência evolutiva, que embora não seja definida formalmente na literatura, ela está presente em diversas abordagens (PRABHUNE *et al.*, 2018). Assim, a proveniência evolutiva reflete as alterações feitas na proveniência prospectiva e retrospectiva executadas no fluxo de trabalho (ou seja, o histórico de evolução), mantendo todas as alterações aplicadas ao longo do seu ciclo de vida (KOOP, FREIRE, 2014).

Outro componente-chave para a proveniência, são os dados advindos dos processos de anotação pelo usuário, ou seja, dados que não são capturados automaticamente. Com isso, possuem diferentes níveis de granularidade e podem estar associados a diferentes componentes da proveniência prospectiva como da retrospectiva. Esse tipo de registro se torna muito importante, pois contém informações sobre decisões tomadas e observações feitas pelo usuário (FREIRE *et al.*, 2008).

Em relação ao funcionamento da Proveniência de Dados, em geral, o que acontece é: os dados, em um banco de dados, são movidos para outro banco de dados quando consultas ou programas são executados; e, assim, é feita uma descrição especificando as relações entre as fontes envolvidas (TAN, 2007).

Importante ressaltar, que em uma pesquisa em banco de dados, a proveniência auxilia no processo de rastreamento e sinalização da origem dos dados, como também sua movimentação entre diferentes fontes de dados (SIMMHAN; PLALE; GANNON, 2005).

Nesse contexto, destaca-se o funcionamento da Proveniência de Dados voltados a experimentos científicos que necessitam de mecanismos de captura de proveniência para acessar os detalhes relevantes de uma tarefa computacional, registrando seus passos, informações de execução e anotações de usuários (FREIRE *et al.*, 2008). Destaca-se, ainda, que seu funcionamento na grande maioria das situações depende também dos *workflows* científicos que capturam os dados por intermédio de ferramentas computacionais apresentando em seu ciclo de vida os passos executados em tal procedimento ou tarefa. Como exemplo cita-se a captura de dados de sensores de diferentes tipos de funcionalidades, passíveis de tratamento de proveniência.

Outro ponto de destaque, é que nem sempre o funcionamento da Proveniência de Dados está diretamente ligado a bancos de dados, vai além dessa tecnologia, ou seja, dados manuais

também compõem a estrutura de proveniência, como por exemplo, os dados de uma experimentação científica com anotações de forma manual que podem ser repassados para um repositório de dados em específico para servir de base para novos experimentos. Enfim, podem ser explorados de diferentes formatos que sejam necessários procedimentos de tratamento de sua proveniência para tomadas de decisões.

2.2.2 Principais modelos

As iniciativas para a representação da proveniência por meio de recursos informacionais em um âmbito geral tiveram início com as discussões da construção do *Open Provenance Model* (OPM) em 2006, no primeiro *International Provenance and Annotation Workshop* (IPAW) (MOREAU, 2006). A proposta do OPM foi definir um modelo de dados que seja aberto do ponto de vista da interoperabilidade, mas também com relação à comunidade de seus colaboradores, revisores e usuários (MOREAU *et al.*, 2011). O modelo OPM procura demonstrar a relação causal entre eventos que afetam objetos (digitais ou não) e descreve essa relação através de um grafo acíclico direcionado (OPEN PROVENANCE MODEL, 2010).

Dessa forma, os pesquisadores que estudavam o OPM, juntamente com o grupo de trabalho de proveniência da *World Wide Web* (W3C), avançaram suas pesquisas para um novo modelo denominado PROV. Para Groth e Moreau (2013) a família de documentos PROV define um modelo, serializações e outras definições necessárias de apoio correspondente que permitem o intercâmbio de informações de proveniência em ambientes heterogêneos, como é o caso da Web. A família de documentos PROV conta com quatro recomendações, The PROV Data Model (PROV-DM), The PROV Ontology (PROV-O), The Provenance Notation (PROV-N) e Constraints of the PROV Data Model (PROV-CONSTRAINTS).

Ainda foram publicadas algumas notas e recomendações que auxiliam no mapeamento e nas informações sobre o modelo PROV, com intuito de progressão para domínios mais avançados por parte de usuários e desenvolvedores. Dessa forma, para elucidar o PROV, os documentos foram classificados com recomendações para três perfis de público, onde o primeiro grupo é destinado ao público geral de usuários no entendimento e no uso de aplicativos que suportam o PROV; o segundo grupo é para desenvolvedores que queiram criar ou desenvolver aplicativos que criem e consumam proveniência; e o terceiro grupo é constituído pelo processo avançado em relação a validadores, novas serializações ou outros sistemas baseados no PROV (GROTH; MOREAU 2013).

O Quadro 1 apresenta as especificações da família PROV no quesito relacionado ao

público-alvo do documento, tipo (nota ou recomendação) e documento (breve definição e função do documento). Importante ressaltar que o PROV-OVERVIEW é considerado uma “Nota”, e é a própria visão geral da família de documentos PROV, sendo, portanto, considerado 12 documentos propostos no total pela W3C.

Quadro 1 – Especificações da família PROV

| Parte | Público | Tipo | Documento |
|-------|---------------|--------------|---|
| 1 | Usuário | Nota | PROV-PRIMER é o ponto de início para o PROV, oferece uma introdução ao modelo de dados de proveniência. É aqui que você deve começar e, para muitos, pode ser o único documento necessário. |
| 2 | Desenvolvedor | Recomendação | O PROV-O define uma ontologia <i>Web Ontology Language (OWL)</i> ³ 2 leve para o modelo de dados de proveniência. Isso é destinado à comunidade <i>Linked Data</i> ⁴ e <i>Semantic Web</i> ⁵ . |
| 3 | Desenvolvedor | Nota | O PROV-XML define um esquema XML ⁶ para o modelo de dados de proveniência. Isso é destinado a desenvolvedores que precisam de uma serialização XML nativa do modelo de dados PROV. |
| 4 | Avançado | Recomendação | O PROV-DM define um modelo de dados conceitual para a proveniência, incluindo diagramas UML. PROV-O, PROV-XML e PROV-N são serializações deste modelo conceitual. |
| 5 | Avançado | Recomendação | O PROV-N define uma notação legível para o modelo de proveniência. Isso é usado para fornecer exemplos dentro do modelo conceitual, bem como usado na definição de PROV-CONSTRAINTS. |
| 6 | Avançado | Recomendação | PROV-CONSTRAINTS define um conjunto de restrições no modelo de dados PROV que especifica uma noção de proveniência válida. Ele é especificamente destinado aos implementadores de validadores. |
| 7 | Desenvolvedor | Nota | O PROV-AQ define como usar mecanismos baseados na Web para localizar e recuperar informações de proveniência. |
| 8 | Desenvolvedor | Nota | O PROV-DC define um mapeamento entre o <i>Dublin Core</i> ⁷ e o PROV-O. |
| 9 | Desenvolvedor | Nota | PROV-DICIONARY define construções para expressar a proveniência de estruturas de dados no estilo de dicionário. |
| 10 | Avançado | Nota | O PROV-SEM define uma especificação declarativa em termos de lógica de primeira ordem do modelo de dados PROV. |
| 11 | Avançado | Nota | O PROV-LINKS define extensões para o PROV para permitir a vinculação de informações de proveniência em pacotes de descrições de proveniência. |

Fonte: Adaptado de Groth e Moreau (2013)

³ <https://www.w3.org/OWL/>

⁴ <https://www.w3.org/standards/semanticweb/data>

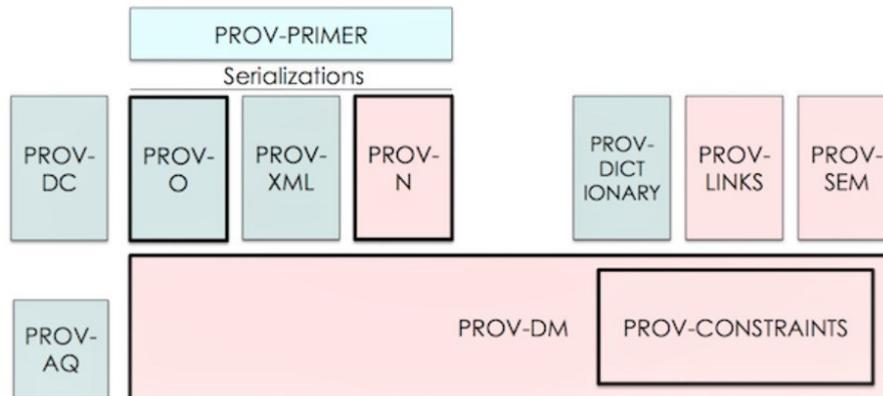
⁵ https://www.w3.org/standards/semanticweb/#w3c_overview

⁶ https://www.w3schools.com/xml/xml_what.asp

⁷ <https://www.dublincore.org/>

Levando em consideração os documentos descritos no Quadro 1, uma melhor visualização pode ser observada na Figura 1, a qual apresenta a organização da família PROV em uma estrutura de fluxograma.

Figura 1 – Organização da família PROV

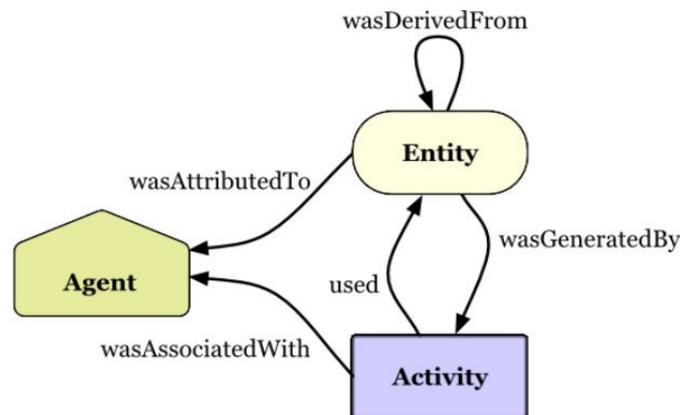


Fonte: Groth e Moreau (2013).

Observa-se na Figura 1, a organização do PROV e como os documentos dependem uns dos outros. O esquema de cores corresponde ao roteiro da organização do PROV, onde o documento em azul claro (PROV-PRIMER) é para o público geral, os documentos em rosa correspondem ao público avançado e os documentos em azul mais escuro são destinados aos desenvolvedores. Em seu núcleo está o (PROV-DM), que define um vocabulário comum usado para descrever a proveniência como base para todos os documentos. Isso é instanciado por várias serializações usadas por implementações para a troca de proveniência. O documento que apresenta um contorno em preto destaca-se como recomendação da W3C.

Nesse sentido, o PROV-PRIMER apresenta conceitos fundamentais para entender os demais documentos da família PROV, servindo como uma introdução e guia para o modelo de dados PROV, promovendo o intercâmbio interoperável da proveniência em ambientes heterogêneos como a Web (GIL; MILES, 2013). Para Gil e Miles (2013) a estrutura central do PROV baseia-se na definição das entidades, atividades e agentes que estão envolvidos na produção de um dado ou de uma coisa e, como eles estão relacionados, definindo quatro tipos de propriedade: *wasGeneratedBy*, *wasAssociatedBy*, *wasAttributedTo* e *used*. O diagrama apresentado na Figura 2, fornece uma visão geral da estrutura dos registros PROV.

Figura 2 – Entidades do modelo PROV



Fonte: Gil e Miles (2013).

A seguir, nos Quadros 2 e 3, são explicitadas as classes e relações do modelo W3C PROV para que se tenha um entendimento de suas ligações.

Quadro 2 – Classes do modelo PROV

| Classes | Descrição |
|-----------------|--|
| Entity | A classe “Entity” é utilizada para representar qualquer tipo de entidade, como por exemplo, física, conceitual ou digital. |
| Agent | A classe “Agent” se refere aos agentes que podem influenciar na proveniência representada, como por exemplo, um <i>software</i> que realiza uma ação, ou um usuário. |
| Activity | A classe “Activity” se relaciona às atividades, como por exemplo, processos ou ações. |

Fonte: Adaptado de Gil e Miles (2013).

Quadro 3 – Relações do modelo PROV

| Propriedades | Descrição |
|--------------------------|---|
| wasAttributedTo | Relaciona um agente a uma entidade, a qual ele foi atribuído |
| wasAssociatedWith | Relaciona um agente a uma atividade, a qual ele é responsável |
| used | Relaciona uma entidade a uma atividade que a usou |
| wasGeneratedBy | Relaciona uma atividade a uma entidade pela qual foi gerada |
| wasDerivedFrom | Relaciona uma entidade a outra pela qual a mesma foi gerada |

Fonte: Adaptado de Gil e Miles (2013)

Para Gil e Miles (2013) uma entidade pode ser algo físico, digital, conceitual ou outro tipo de coisa com algum aspecto de entidade, sendo real ou imaginária. Já uma atividade é algo que ocorre durante um período e age sobre ou com entidades, podendo incluir o consumo, processamento, transformação, modificação, realocação, uso ou geração de entidades. E finalmente, um agente é algo que carrega alguma forma de responsabilidade por uma atividade que está ocorrendo, pela existência de uma entidade, ou para a atividade de outro agente. O

Grupo de Trabalho de Proveniência da W3C sugeriu que as formas padrão para cada tipo de nó em um gráfico de proveniência são: entidade – elipse, atividade – retângulo, agente – pentágono.

Nessa perspectiva, como base para o desenvolvimento do método adotado nesta tese, foram utilizados como base teórica os documentos: PROV-OVERVIEW e PROV-PRIMER. O PROV-OVERVIEW foi utilizado para a compreensão da família PROV, e o PROV-PRIMER foi utilizado como base para modelar e apresentar os principais conceitos relacionados à proveniência. Assim, foi abordado o W3C PROV como base para as discussões e construção do método desta tese, por se tratar de um modelo de destaque na literatura internacional em relação à sua aplicação em diversos domínios, inclusive nos domínios de saúde no uso do HL7 FHIR que possui recursos básicos de proveniência disponíveis em sua estrutura⁸.

Na seção 3 desta tese, intitulada “Estado da Arte”, é abordada a RSL, onde são citados alguns estudos aplicados nos cenários de saúde, especificamente em SIS, utilizando o W3C PROV para auxiliar no gerenciamento dos dados de proveniência, justificando assim, a prospecção do seu uso nesta tese. Dito isto, é importante ressaltar também, que o modelo W3C PROV é o padrão atualmente adotado para captura de proveniência nos mais variados cenários, o que de fato, para o cenário escolhido desta tese, a RSL comprova seu uso e benefícios em SIS. Outro ponto importante a ser destacado pela escolha do W3C PROV nesta tese, é que os padrões de interoperabilidade HL7 FHIR estão em consonância com o W3C PROV⁹, tornando mais clara a ideia do seu uso e instanciamento para propor o método desta tese. Isso de fato, possibilita a criação da estrutura computacional para o gerenciamento dos dados de proveniência em SIS, contemplando o objetivo geral aqui proposto nesta tese.

Por fim, existem outros modelos de Proveniência de Dados na literatura, porém, os principais modelos ou extensões existentes são baseadas nos documentos da família PROV, os quais são referências internacionais regidas pela W3C. Dito isto, esta tese baseia-se no cerne dos modelos de Proveniência de Dados, o W3C PROV.

2.2.3 Aplicações gerais

Em relação às aplicações da Proveniência de Dados, a mesma pode ser aplicada em diversas áreas, nas quais informações detalhadas sobre a origem dos dados são necessárias. Esse tipo de informação é muitas vezes referido como origem de dados ou linhagem de dados. A

⁸ <https://build.fhir.org/ig/HL7/US-Core/basic-provenance.html>

⁹ <https://build.fhir.org/provenance-mappings.html>

proveniência de um item de dados inclui informações sobre os processos e itens de dados de origem que levam à sua criação e representação atual (GLAVIC; DITTRICH, 2007).

Assim, a utilidade e aplicação da proveniência está diretamente relacionada à granularidade dos dados coletados (SIMMHAN; PLALE; GANNON, 2005). Granularidade é o nível de detalhamento dos dados capturados e pode ser classificada em dois tipos: grão fino (quando os dados de proveniência são capturados com muitos detalhes) e grão grosso (quando os dados de proveniência são capturados com poucos detalhes (BUNEMAN; TAN, 2007). Esse tipo de informação é importante porque muitas vezes o resultado completo não está disponível e apenas uma parte do resultado pode ser de suma importância, além de ser também muito mais simples.

Para se obter os benefícios advindos a partir das aplicações das informações de proveniência, torna-se fundamental que tais dados sejam capturados, modelados e armazenados para posterior utilização. O gerenciamento dos dados de proveniência é uma questão em aberto que tem merecido tratamento da comunidade científica (FREIRE *et al.*, 2008; MARGHERI *et al.*, 2020; RAHMAN *et al.*, 2020; HARDIN; KOTZ, 2021; MICHALA *et al.*, 2022).

2.2.3.1 Aplicações em saúde

A Proveniência de Dados está relacionada a diferentes cenários de aplicação, como é o caso da saúde, foco desta tese. O uso de dados de proveniência no contexto da saúde vive um cenário crescente de pesquisa com base nos mais variados cenários de saúde, e as tecnologias aplicadas nessa área vêm obtendo resultados expressivos (SEMBAY; MACEDO; DUTRA, 2020b).

Existem várias aplicações da Proveniência de Dados nos cenários de saúde, porém, serão citados apenas alguns exemplos para elucidar a complexidade das estruturas de tratamento dos dados de proveniência em setores de saúde específicos.

Em Álvarez *et al.* (2006), a aplicação da proveniência se deu no contexto da administração e distribuição de transplante de órgãos. O trabalho descreve o desenvolvimento de uma arquitetura orientada a serviços usando proveniência em sistemas médicos para auxiliar no processo de decisão de um transplante.

Outra iniciativa que trabalha com Proveniência de Dados na saúde foi descrita em Li *et al.* (2008). O projeto atua em parceria com o Centro para a Modelagem de Imunidade Pulmonar (CMPI), criado com a associação das Universidades de Pittsburgh, *Carnegie Mellon e de Michigan*. Esse trabalho foi responsável por projetar e desenvolver uma plataforma de

distribuição de dados, o DataXS, na qual dados experimentais, análises, e modelos são distribuídos para os projetos participantes. Esse projeto utiliza a proveniência para manter o registro de onde o dado se originou, e não as etapas pelas quais os dados foram processados.

Já no estudo de Margueri *et al.* (2020), os autores relatam a importância da criação e troca de registros eletrônicos de saúde dos pacientes, que se desenvolveram significativamente na última década. Os registros dos pacientes são, no entanto, distribuídos em silos de dados em várias instalações de saúde, apresentando desafios clínicos que podem colocar em risco a segurança dos pacientes. A plataforma alavanca dois pilares: o uso de padrões internacionais tais como Integrating the Healthcare Enterprise (IHE), Health Level Seven para recursos de interoperabilidade internacional (HL7) e Fast Healthcare Interoperability Resources (FHIR) para alcançar interoperabilidade e o uso de um processo de criação de proveniência. Os autores utilizam o modelo de proveniência de dados W3C PROV.

Em um recente estudo de Werder *et al.* (2022), os autores relatam preocupações com a proveniência de dados relacionadas a aplicações de recomendações de Inteligência Artificial (IA) na área da saúde, em que é destacado o uso de técnicas de proveniência embutidas em sistemas de registros eletrônicos de saúde, para prever a sepse (uma condição potencialmente fatal na qual a resposta do corpo a uma infecção danifica seus próprios tecidos).

Por fim, outras aplicações da Proveniência de Dados em saúde, especificamente em SIS, podem ser mais bem compreendidas na seção de “Estado da Arte” desta tese, onde é descrita uma Revisão Sistemática de Literatura (RSL).

2.3 SISTEMAS DE INFORMAÇÃO EM SAÚDE

Promover a qualidade e o acesso aos serviços é uma preocupação crescente aos profissionais da área, governos, consumidores, ou seja, o uso de SI nesse ambiente possui grande potencial “para melhorar a relação custo-eficácia, qualidade e acessibilidade aos cuidados em saúde” (CHIASSON; DAVIDSON, 2004, p. 156). A difusão da teoria em sistemas de informação para novos campos sociais (como exemplo o campo da saúde) exige que se dedique maior atenção a respeito das variações que ocorrem no contexto de aplicação da teoria (AVGEROU, 2000).

As Tecnologias de Informação e Comunicação (TIC) na área da Saúde têm sido utilizadas em Sistemas de Informação Hospitalar, Prontuário Eletrônico e Telemedicina, tendo um enorme potencial para propulsionar melhorias que envolvem ganhos de qualidade e acessibilidade no que tange aos cuidados em saúde e a melhoria considerável do custo-benefício

(SOUZA-JUNIOR; CUNHA; DORNELAS, 2009). Os SIS se diferenciam dos Sistemas de Informações Hospitalares, que são caracterizados como uma instância de um SIS, por apenas gerir as informações de unidades hospitalares. Assim, os SIS também podem ser conceituados como um SI que integra a coleta, o processamento, a comunicação e o uso de informações fundamentais para aperfeiçoar a eficácia e a eficiência dos serviços de saúde, por meio de melhor gestão em todos os níveis do setor de saúde. Esse tipo de sistema produz informações relevantes e de qualidade para apoiar a gestão e o planejamento de programas em saúde (WORLD HEALTH ORGANIZATION, 2004; HAUX, 2006). Para Sembay e Macedo (2022) os SIS estão sendo implementados em todos os aspectos da saúde, desde a administração aos sistemas de apoio às decisões clínicas, gerando grandes volumes de dados em seus repositórios descentralizados.

Nesse sentido, no cenário da reforma do setor da saúde e do processo de descentralização da prestação de serviços na área é cada vez mais necessária a gerência dos serviços o mais próximo possível do nível da prestação de serviços para o surgimento de novas demandas informacionais estabelecendo modificações importantes nas estruturas dos SIS (WORLD HEALTH ORGANIZATION, 2008).

No contexto brasileiro, destaca-se o uso das TIC na esfera pública de saúde, em que, nas últimas duas décadas, podem ser observados grandes avanços no processo de implantação dos (SIS), com a ampliação da cobertura, do acesso e da utilização dos bancos de dados nacionais sobre nascimentos, óbitos, doenças e agravos, atenção primária, imunização, hospitalização, procedimentos ambulatoriais, estabelecimentos de saúde, orçamentos públicos, assim como informações demográficas, sociais e econômicas, dentre outras (MOTA; CARVALHO, 2003; MINISTÉRIO DA SAÚDE, 2006).

Outro fator importante a ser mencionado em relação aos SIS, é o que diz respeito à interoperabilidade entre os sistemas, que pode ser definida como a habilidade de dois ou mais sistemas de informação interoperar seus dados, de acordo com um padrão definido compartilhando informações dentro de um cenário de saúde. Distintos recursos baseados em TIC introduzidas ao longo do tempo necessitam estabelecer conexões tanto para prover serviços quanto para fornecer informações, independentemente de onde estejam os dados e as pessoas solicitantes, como é o caso da Telemedicina (SOUZA-JUNIOR; CUNHA; DORNELAS, 2009; SANTOS; REINHARD, 2011).

2.3.1 Principais tipos

Existem diferentes tipos de SIS, e para o escopo desta tese, serão apresentados os que mais se destacam em relação ao gerenciamento de dados de proveniência em um contexto mundial. Esses SIS são:

- a) **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).
- b) **Registro de Saúde Pessoal/Personal Health Record (PHR):** esses registros são frequentemente criados e monitorados pelos próprios pacientes. Eles podem ser baseados em desktop, baseados na internet por meio de dispositivos móveis (por exemplo, localizados inteiramente em um celular ou dispositivo de armazenamento portátil) (LIU; SHIH; HAYES, 2011).
- c) **Sistemas de Monitoramento de Saúde/Health Monitoring System (HMS):** esses sistemas monitoram a saúde em um campo potencial de aplicação de sensores vestíveis. Esses sensores vestíveis e ambientais medem dados relacionados à saúde em ambientes cotidianos de usuários ou pacientes (KORHONEN; PARKKA; VAN GILS, 2003).
- d) **Sistema de Informação Hospitalar/Hospital Information System (HIS):** Nesse contexto, 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 *et al.*, 2000).
- e) **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). Segundo Honeyman (1999), um Sistema de Informação em Radiologia é tipicamente um sistema especificamente desenvolvido para departamentos de radiologia, para receber interpretações e preparar listas de pacientes. Este sistema pode prover relatórios históricos dos radiologistas e frequentemente transmite o relatório final para o HIS.
- f) **Sistema de informação de laboratório/Laboratory Information System (LIS):** traduzido como sistema de informática laboratorial e descrito como sendo um sistema

composto por um ou mais softwares, ligados direta ou indiretamente, que visam gerenciar informações dentro de um laboratório de análises clínicas, sejam estas de cunho técnico, operacional, administrativo, gerencial ou de uma mescla destes (BLICK, 1997).

- g) **Sistema de Comunicação e Armazenamento de Imagens/Picture Archiving and Communication System (PACS):** consiste em subsistemas integrados através de redes de computadores para aquisição, armazenamento e visualização de imagens e dados. 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, *gateway*¹⁰ *Digital Imaging and Communications in Medicine (DICOM)*¹¹, controle de armazenamento e estações de visualização.
- h) **Sistema de Informações de Pesquisa Clínica/Clinical Research Information System (CRIS):** software projetado para apoiar pesquisas clínicas que possam reduzir os custos de estudos de pesquisa. A CRIS apoia o atendimento clínico, coleta de dados para pesquisa e suporte às operações hospitalares (NADKARNI, 2019).
- i) **Sistema de Aprendizagem de Saúde/ Learning Healthcare System (LHS):** esse sistema tem a capacidade de compartilhar dados e aproveitar seu potencial para gerar conhecimento rapidamente e informar decisões transformadoras que contribuam para uma melhor saúde. Possui uma infraestrutura para atingir esse objetivo em escala, combinando tecnologia, processo e política (FRIEDMAN *et al.*, 2014).

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

- a) **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.

Existem outros SIS na literatura que também fazem parte dos cenários de saúde, porém,

¹⁰ Dispositivo de tradução de protocolo em hardware ou software que permite que os usuários que trabalham em uma rede possam acessar outra.

¹¹ <https://www.DICOMstandard.org/>

os sistemas aqui listados estão no escopo de estudo desta tese em relação ao método de gerenciamento de dados de proveniência que se pretende aplicar, conforme detalhado na seção 5.

2.3.2 Interoperabilidade, integração e armazenamento de dados

A necessidade de padrões de interoperabilidade entre SIS é inerente para que haja comunicação e a troca de dados de saúde, estabelecendo meios de interoperar os processos existentes entre diferentes plataformas de saúde. Destaca-se que a transformação digital dos dados de saúde é um importante ponto de necessidade para os diferentes SIS, pois a interoperabilidade nos diversos setores de saúde pode contribuir para o desenvolvimento econômico de todo ecossistema de saúde.

Para que os SIS desempenhem seu papel, eles necessitam basicamente de ferramentas computacionais que realizam e intermedeiam todo o processo de interoperabilidade, integração e armazenamento dos dados de saúde conforme destacado no Quadro 4, os quais estão em conformidade com esta tese.

Quadro 4 – Ferramentas para interoperabilidade, integração e armazenamento de dados em SIS

| Tipo | Descrição |
|---|---|
| Protocolos de Interoperabilidade | <p>DICOM - O DICOM foi desenvolvido a partir da cooperação do <i>American College of Radiology (ACR)</i> com a <i>National Electrical Manufacturers Association (NEMA)</i>. DICOM é um padrão orientado a objeto, definindo objetos de informação, serviços e classes de serviços para executar estes serviços. Cada dispositivo tem um conjunto de objetos definidos que é preparado para reconhecer o arquivo e permitir o acesso a ele e aos serviços providos, e negociar entre dois dispositivos qual tem a necessidade de transferir a imagem. O padrão DICOM é adotado pelos fabricantes de equipamentos médicos e desenvolvedores de sistemas de informática para a área da saúde como o padrão para a troca de imagens em um formato digital (HONEYMAN, 1999; MILDENBERGER; EICHELBERG; MARTIN, 2002; OOSTERWIJK, 2002). É importante ressaltar, ainda, que o padrão DICOM define o Service-Object Pair Class (SOP Class), serviço de rede que permite a transferência de informações entre dispositivos diferentes, desde que suportados os mesmos serviços e objetos (NEMA, 2007). A transferência realizada por intermédio do SOP Class é definida pela união de IOD e elementos de serviços DICOM Message Service Element (DIMSE), o qual oferece interfaces para aquisição, arquivamento e processamento por meio de serviços para armazenamento (C-STORE), consulta (C-FIND), recuperação (C-GET) e transferência (C-MOVE), não havendo serviço recente para atualização de forma intencional para reduzir possibilidade de alterações de dados (NEMA, 2007).</p> <p>HL7 FHIR - 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</p> |

| | |
|----------------------------|--|
| | <p>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). O HL7 FHIR suporta a <i>Service-oriented Architecture</i> (SOA) e <i>Representational State Transfer</i> (REST) também chamado abordagem RESTful para troca de informações. A arquitetura RESTful permite uma comunicação usando apenas o protocolo HTTP onipresente sem camadas adicionais (HL7 INTERNATIONAL INC, 2022c).</p> <p>IHE - O protocolo <i>Integration the Healthcare Enterprise</i> (IHE), que teve seu início em novembro de 1998, e é um modelo de informação de alto nível para dirigir adaptações do padrão HL7 e DICOM. A missão do IHE foi definir e estimular os fabricantes a utilizarem os padrões para que equipamentos e sistemas de informação fossem complacentes, facilitando as operações clínicas diárias (HUANG, 2004). O IHE conseqüentemente abre uma efetiva contribuição para todos os profissionais de saúde, que podem sinalizar as principais instâncias que emergem diariamente no raio de visão de suas atividades. Originalmente especificado para a radiologia, hoje, ele objetiva junto aos médicos especialistas e profissionais da tecnologia da informação criar regras para identificar e solucionar as dificuldades que impedem uma integração efetiva e funcional dos sistemas de informação em saúde. A arquitetura técnica do IHE define uma linguagem, um vocabulário e um modelo comum usando DICOM e HL7 para completar um conjunto radiológico bem definido e transações clínicas para certos serviços (BERNARDINI <i>et al.</i>, 2003; HUANG, 2004). O objetivo do IHE é proporcionar ao usuário final uma melhora no acesso às informações críticas e clínicas do paciente, armazenadas em todos os sistemas conectados em rede de um hospital, tendo como proposta facilitar a eficiência, prognosticar e integrar as funcionalidades entre os sistemas incompatíveis (BOOCHEVER, 2004).</p> |
| Integração de Dados | <p>ETL – O processo <i>Extract-Transform-Load</i> (ETL) tem sido usado para integração de dados de várias fontes ou aplicativos, possivelmente de diferentes domínios. Refere-se a um processo de armazenamento de dados que extrai dados de fontes externas, os transforma para atender às necessidades operacionais, que podem incluir verificações de qualidade e os carrega no banco de dados de destino, mais especificamente, armazenamento de dados operacionais, como por exemplo, um data warehouse. As três fases desse processo são: a) Extrair: esta é a primeira fase do processo que envolve a extração de dados de fontes de dados apropriadas. Os dados geralmente estão disponíveis em formatos de arquivo simples, como (.csv), (.xls) e (.txt), ou estão disponíveis por meio de um cliente RESTful; b) Transformar: esta fase envolve a limpeza dos dados para cumprir o esquema de destino. Algumas das atividades típicas de transformação envolvem a normalização de dados, remoção de duplicatas, verificação de violações de restrição de integridade, filtragem de dados com base em algumas expressões regulares, classificação e agrupamento de dados, aplicação de funções incorporadas quando necessário; e, c) Carga: esta fase envolve a propagação dos dados em um data warehouse que atende um Big Data (BANSAL, 2014).</p> |
| Servidor de dados | <p>HAPI FHIR - É uma implementação completa do padrão HL7 FHIR para interoperabilidade de assistência médica em Java. É uma comunidade aberta de desenvolvimento de software licenciado sob a Apache Software License 2.0 (HAPI FHIR, 2022). HAPI FHIR é um produto da Smile CDR. O HAPI FHIR é um projeto aberto de servidor JPA (Java Persistence API. A comunidade</p> |

| | |
|--|--|
| | <p>HAPI concentrou-se na criação de uma biblioteca FHIR que permite consumir e expor APIs FHIR. O projeto é rico em recursos, incluindo: o servidor, excelente documentação, uma comunidade de suporte e lançamentos frequentes (corrigindo bugs, adicionando recursos e expandindo o suporte para recursos e padrões FHIR mais recentes) (SÁNCHEZ; DEMURJIAN, BAIHAN, 2017). O servidor HAPI FHIR possibilita hospedar conjuntos de dados de saúde personalizados e retê-los por um longo período, sendo possível de realizar esse processo em um servidor público oferecido pela <i>University Health Network</i> (UHN) (HAPI FHIR, 2022). A equipe do HAPI FHIR enfatiza que a biblioteca foi construída para ser flexível acima de tudo. Alguns exemplos de cenários de uso incluem o seguinte: a) usando o analisador e o codificador HAPI FHIR para converter entre FHIR e o modelo de dados da aplicação; b) o cliente HAPI FHIR em um aplicativo para buscar ou armazenar recursos em um servidor externo; c) o uso do servidor HAPI FHIR em um aplicativo para permitir que aplicativos externos acessem ou modifiquem os dados; d) o HAPI JPA/Database Server para implantar um servidor FHIR totalmente funcional no qual se pode desenvolver aplicativos (HUSSAIN; LANGER; KOHLI, 2018).</p> |
|--|--|

Fonte: Elaborado pelo autor (2022).

Os padrões destacados no Quadro 4 são os utilizados no Brasil. Cabe também destacar, que o Ministério da Saúde (2011) publicou a Portaria nº 2.073, de 31 de agosto de 2011¹², que define, entre outros padrões para interoperabilidade em SIS, o uso do padrão DICOM para informação relativa a exames de imagem e HL7 para troca de documentação clínica (BRAGHETTO; CORDEIRO, 2014).

Nesse contexto, pode-se destacar os padrões de interoperabilidade utilizados no PEP conforme o estudo de Miranda e Pinto (2015), em que os referidos autores destacam que os principais padrões indicados para a área da saúde elencados pela portaria nº 2.073/2011 do Ministério da Saúde que regulamenta o uso de padrões de interoperabilidade e informação em saúde no âmbito do Sistema Único de Saúde (SUS), aponta um catálogo com 14 padrões, os quais são: CID10, Tabela SUS, TISS, TUSS/CBHPM, DICOM, HL7, HL7 CDA, CIAP-2, ISO 13.606-2, ISBT 128, IHE-PIX, SNOMED-CT, OpenEHR e LOINC.

Ainda, importa mencionar que a portaria nº 1.434, de 28 de maio de 2020¹³, institui o Programa Conecte SUS e altera a Portaria de Consolidação nº 1/GM/MS, de 28 de setembro de 2017, para instituir a Rede Nacional de Dados em Saúde e dispor sobre a adoção de padrões de interoperabilidade em saúde no Brasil (MINISTÉRIO DA SAÚDE, 2020).

Diante disso, é importante destacar que existem diversas ferramentas computacionais disponíveis na indústria de sistemas médicos que realizam a interoperabilidade, integração e armazenamento dos dados de saúde em SIS, porém, busca-se ferramentas computacionais como

¹² https://bvsmis.saude.gov.br/bvsmis/saudelegis/gm/2011/prt2073_31_08_2011.html

¹³ <https://www.in.gov.br/en/web/dou/-/portaria-n-1.434-de-28-de-maio-de-2020-259143327>

observadas no Quadro 4 pelo baixo custo de implantação e manutenção.

Porém, é importante ressaltar que, as ferramentas apresentadas no Quadro 2, são algumas das quais atualmente são mais utilizadas e promissoras no contexto brasileiro dos SIS de nível público, tanto pelo fato do baixo custo de manutenção e facilidade de operação e integração.

2.3.3 HL7 FHIR baseado em W3C PROV para garantir a Proveniência de dados em Sistemas de Informação em Saúde

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, 2022b). Dessa forma, esforços de vários anos em padrões internacionais como HL7 FHIR, o qual possui definições dos recursos de proveniência baseado no modelo W3C PROV, confirmam esses requisitos (SEMBAY; MACEDO, MARQUEZ FILHO, 2022c).

Nesse sentido, é importante destacar que, existem diversos padrões de interoperabilidade, mas, o padrão comumente utilizado nos ambientes de saúde, até mesmo por questões de custo é o conjunto de padrões HL7, sendo indiscutivelmente o esforço de padronização mais reconhecido no campo da informática em saúde. O seu padrão mais recente, o HL7 FHIR, é um protocolo aberto, moderno e líder em interoperabilidade, sendo, portanto, ideal para atuar como fonte de inspiração e repositório das melhores práticas digitais em um domínio de saúde. O padrão HL7 FHIR tem o potencial de solucionar vários problemas de interoperabilidade, que causam o longo impedimento da informática em atingir seu impacto na qualidade da saúde pública, na relação de custo-benefício e na segurança de dados.

Assim, uma característica significativa para esta tese é que 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, 2022a). Ressalta-se ainda, que o HL7 FHIR baseado no W3C PROV, possui consonância com o padrão HL7 FHIR 5Ws¹⁴, o qual executa os processos referentes aos recursos de proveniência¹⁵. Assim, no padrão HL7 FHIR 5Ws, todos os recursos de proveniência incluem informações de atribuição. Classicamente, essas informações incluem: Who?, What?, When?, Where?, Why?. Daí o nome deste padrão (5Ws)

¹⁴ <http://hl7.org/fhir/fivews.html>

¹⁵ <https://build.fhir.org/provenance-mappings.html>

(HL7 INTERNATIONAL INC, 2022b). Importante destacar também que o HL7 FHIR com base no modelo W3C PROV traz todas as características das relações causais.

Dessa forma, o HL7 FHIR traz como base em relação à proveniência de recurso, o W3C PROV com intuito de contribuir para a garantia da Proveniência dos Dados em SIS. De fato, isso mostra que a Proveniência de Dados é um elemento importante e estruturado no cerne do HL7 FHIR utilizado no gerenciamento dos dados em SIS.

O W3C PROV fornece meios de design e implementação para compartilhar atributos de proveniência semanticamente interoperáveis. Além disso, importantes órgãos de saúde, como *Integrating the Healthcare Enterprise (IHE)* suportam W3C PROV (MARGUERI *et al.*, 2020).

Ressalta-se ainda que o uso da Proveniência de Dados em SIS contribui para o gerenciamento e a rastreabilidade dos dados e podem contribuir também para o gerenciamento dos metadados que acompanham os dados de proveniência, o que é fundamental para o acompanhamento dos fluxos informacionais, além de assegurar que os dados atendam à legislação, às normas e aos regulamentos técnicos vigentes dentro do SIS.

2.4 TELEMEDICINA

A Telemedicina foi o primeiro termo usado na década de 1960, cujo significado estrito era o tratamento do paciente pelo médico, a distância. A partir de 1999, com a expansão da área, novas definições surgiram, com especial atenção na proposta durante o Congresso da *American Telemedicine Association em Washington DC*, Estados Unidos, onde foi mencionado o uso das TIC para compartilhar e transferir informações médicas em processos de diagnóstico, terapia e educação (NORRIS, 2002). Com o uso das TIC, houve também o aparecimento de termos a partir da Telemedicina, como Telessaúde, que se pensava ser mais “politicamente correto”, e outros termos como saúde on-line e e-saúde (WOOTTON, 2001). Porém, os termos Telemedicina/Telessaúde foram os que mais se destacaram na literatura, sendo um conjunto de serviços ligados aos cuidados com a saúde, nos casos em que a distância é um fator crítico, ampliando a assistência e também a cobertura (WHO-GLOBAL OBSERVATORY FOR eHEALTH, 2010). No entanto, a Telemedicina e Telessaúde são sinônimos, e usados de forma intercambiável. Assim, nesta tese será utilizado o termo Telemedicina em todas as aplicações.

A Telemedicina pode ser definida como uma ciência aberta e em constante evolução, pois incorpora novos avanços em tecnologia e responde e se adapta às mudanças e necessidades de saúde e contextos das sociedades. Alguns distinguem a Telemedicina à prestação de serviços

por intermédio somente de médicos, e Telessaúde por serviços prestados por profissionais de saúde em geral, incluindo enfermeiros, farmacêuticos e outros (WHO-GLOBAL OBSERVATORY FOR eHEALTH, 2010). Nesse sentido, atualmente, a Telemedicina assume o uso de informação médica por meio da comunicação eletrônica para a saúde do paciente e da educação dos profissionais da saúde (KIM, 2004).

A Telemedicina também pode ser conceituada como um processo ou cadeia de atendimento mais amplo, ou seja, no uso de aparatos tecnológicos no intuito de melhorar essa cadeia e, assim, melhorar a qualidade e a eficiência dos cuidados de saúde. A Telemedicina também pode ser vista como o rápido acesso à experiência (médica) por meio de tecnologias de telecomunicação e informação, não importando onde estejam localizados o paciente ou a informação (ROINE; OHINMAA; HAILEY, 2001; CRAIG; PATTERSON, 2005). Dessa maneira, a Telemedicina, aparece como um meio de reduzir os problemas de desigualdade social no quesito de saúde, pois torna possível o atendimento sem a necessidade de o profissional de saúde e o paciente estarem no mesmo local simultaneamente, rompendo as barreiras geográficas e os custos (WHO-GLOBAL OBSERVATORY FOR eHEALTH, 2010).

Para a Organização Mundial da Saúde, a Telemedicina deve ser adotada tanto em países pobres, em desenvolvimento, como em países ricos, industrializados. Sua missão é entendida como a entrega de serviços de saúde — cuja distância é fator crítico — por todos os profissionais de saúde, usando TIC para troca de informações. Essa troca corresponde a informações válidas para diagnósticos, tratamento e prevenção de doenças e lesões, pesquisa e avaliação, e para educação continuada de provedores de cuidados de saúde, todos com interesse de promover a saúde dos indivíduos e suas comunidades (WHO-GLOBAL OBSERVATORY FOR eHEALTH, 2010).

Entretanto, é importante ressaltar, que os sistemas de Telemedicina têm se popularizado exponencialmente, ou seja, com a maturidade das redes de comunicação ao redor do mundo e o poder computacional que atualmente pode-se adquirir na utilização desses sistemas. Esses dois adventos, somados ao fato da modernidade das linguagens de programação e à alta tecnologia aplicada em equipamentos médico hospitalares, resultaram em uma oportunidade de sucesso para implantação dos sistemas de Telemedicina (MACEDO, 2014).

2.4.1 Funcionamento e aplicações

A Telemedicina tem seu funcionamento em dois tipos básicos, de acordo com o momento das informações transmitidas e a interação entre os indivíduos envolvidos, que são:

assíncrona e síncrona (CRAIG; PATTERSON, 2005; WILSON; MAEDER, 2015).

Os sistemas síncronos ou em tempo real, suportam interação imediata. A principal vantagem de uma abordagem síncrona é a eficiência obtida ao se ter a oportunidade de refinar detalhes pertinentes ao episódio de cuidado, durante a sessão, buscando informações ou dados adicionais e, em muitos casos, fornecendo uma decisão clínica ou aconselhamento dentro da sessão. A videoconferência típica, entre médico e paciente, é um exemplo de um modelo síncrono.

Os sistemas assíncronos desacoplam os componentes da interação, para que ocorram, em momentos diferentes, na conveniência dos participantes. Esse modelo ainda permite que ocorram interações com vários componentes, mas, o efeito da realização das tarefas em tempo diferente, pode adicionar a sobrecarga de atualização necessária, no contexto das solicitações de Telemedicina. São exemplos de sistemas assíncronos a Telepatologia e Telerradiologia (CRAIG; PATTERSON, 2005; WILSON; MAEDER, 2015).

Em relação às classificações da Telemedicina, armazenar e encaminhar informações de forma assíncrona, envolve a troca de dados pré-gravados entre dois ou mais indivíduos em momentos diferentes. Por exemplo, o paciente ou profissional de saúde encaminhador envia uma descrição por e-mail de um caso médico a um especialista que posteriormente envia uma opinião sobre o diagnóstico e o tratamento ideal (RAO; LOMBARDI, 2009). Por outro lado, a Telemedicina em tempo real ou síncrona exige que os indivíduos envolvidos estejam presentes simultaneamente para troca imediata de informações, como no caso da videoconferência (RAO; LOMBARDI, 2009).

Sendo assim, na Telemedicina síncrona e assíncrona, informações relevantes podem ser transmitidas em uma variedade de mídias, como texto, áudio, vídeo ou imagem estática. Essas duas abordagens básicas da Telemedicina são aplicadas a uma ampla gama de serviços em ambientes diversos, incluindo os serviços que são os mais favoráveis para o emprego da Telemedicina: Teledermatologia, Telepatologia e Telerradiologia (CURRELL, 2000; WOOTTON; MENZIES; FERGUSON, 2009).

No que se refere aos pressupostos anteriores, outro ponto importante para que as aplicações em Telemedicina tenham sucesso, são as várias questões éticas e legais quanto à prática da mesma, tais como o controle da prática médica “virtual” e a responsabilidade pelas informações divulgadas na Internet. Destaca-se que o paciente deve consentir formalmente com o uso e o envio de informações, e estar orientado sobre as limitações do método, em comparação às consultas presenciais (MALLET, 2003). A segurança e a privacidade são outros aspectos bastante discutidos; nos Estados Unidos e na Europa já existem legislações específicas para

esse fim (DEMIRIS, 2003). Nesse sentido, as aplicações da Telemedicina são muito amplas, pois abrangem desde o primeiro contato entre o médico e o paciente, a formulação diagnóstica, o tratamento clínico e até intervenções cirúrgicas com o auxílio de qualquer meio de comunicação distantes fisicamente. Assim, esse método é capaz de conectar centros de referência com unidades de atendimento da periferia, proporcionando assistência em dúvidas diagnósticas, condutas e ensino (HERSH *et al.*, 2001; VUCKOVIC *et al.*, 2003).

Nesse contexto, as aplicações da Telemedicina podem ser divididas em dois grandes setores: Teleassistência e Teleducação. Dentro desses setores, podem ser aplicadas diversas áreas da medicina, como a dermatologia, cardiologia, traumatologia, emergência, patologia, cirurgia, psiquiatria, radiologia, entre outras. Outras situações que também são favoráveis ao seu emprego são associadas à localização dos recursos, como em áreas rurais, prisões, presença de obstáculos geográficos (ilhas, montanhas, geleiras, desertos), situações de guerras ou desastres naturais, asilos, estações espaciais, entre outros (KIM, 2004).

Em relação à Teleassistência, um exemplo simples é o acompanhamento de pacientes a distância por meio da monitoração e aconselhamento via telefone, como é o caso de abordagens empregadas com sucesso no manejo e na melhor aderência ao tratamento da hipertensão arterial sistêmica, diabetes mellitus e na monitoração de arritmias cardíacas. Ainda na aplicação da Teleassistência, é possível a inclusão de consultas por vídeo em tempo real, e sistemas de armazenamento e envio de imagens (*store-and-forward*) na busca de informações e de segunda opinião especializada (ROINE; OHINMAA; HAILEY, 2001; KIM, 2004).

Já no contexto da Teleducação, os recursos podem ser utilizados para educação médica continuada, através de teleconferências e de tutoriais. O Aprendizado Baseado em Problemas (ABP) é um método educacional que utiliza casos clínicos como contexto para estimular os alunos a desenvolverem habilidades na solução de problemas e na aquisição do conhecimento. Vários modelos de ABP podem ser elaborados por meio da revisão da literatura, síntese das evidências disponíveis, definição de metas de aprendizagem e questões relacionadas aos casos (DEMIRIS, 2003).

2.4.2 Telemedicina no Brasil

O Brasil é um país que oferece oportunidades ímpares para o desenvolvimento e as aplicações dos sistemas de Telemedicina. Sua grande extensão territorial, milhares de locais isolados e de difícil acesso, distribuição extremamente desigual de recursos, médicos com boa qualidade, entre outros aspectos que vêm desafiando a efetivação do direito à saúde —

universal, integral e equânime — permitem prever a existência de um grande potencial de expansão da Telemedicina no País. Os esforços dos governos estaduais e federal na implementação da Telemedicina consubstanciam essa perspectiva (SABBATINI, 2012).

No Brasil, a Universidade Federal de Minas Gerais (UFMG), em 2001, iniciou o desenvolvimento do serviço de Telessaúde dentro do Hospital de Clínicas (HC/UFMG). Por meio de diversos projetos desenvolvidos e implantados em parceria com a Secretaria Municipal de Saúde de Belo Horizonte (SMS/BH), Secretaria Estadual de Saúde de Minas Gerais (SES/MG) e Ministério da Saúde (MS), a UFMG estabeleceu e consolidou o Centro de Telessaúde do HC/UFMG. O primeiro projeto de Telemedicina para apoio à rede pública de assistência em saúde foi implantado em 2004 e recebeu o nome de BH Telessaúde. A partir disso, o Centro de Telessaúde do HC/UFMG, em parceria com as três instâncias governamentais, expandiu a Telemedicina para todo o estado de Minas Gerais (ALKMIM *et al.*, 2012).

Após a iniciativa da UFMG no Brasil, ocorreram diversas iniciativas no processo de desenvolvimento da Telemedicina. No que se refere a políticas governamentais, o Programa Telessaúde Brasil, a Rede Nacional de Ensino e Pesquisas (RNP) e a Rede Universitária de Telemedicina (RUTE) são as mais importantes (MALDONADO; MARQUES; CRUZ, 2016). Ainda, ressalta-se que, em 2016 foi disponibilizado um guia de referência sobre a Telemedicina no Brasil intitulado de “Agenda Tecnológica Setorial ATS – Complexo Industrial da Saúde TELEMEDICINA PANORAMA TECNOLÓGICO” (ASSOCIAÇÃO BRASILEIRA PARA O DESENVOLVIMENTO DA INDÚSTRIA, 2016).

O Programa Telessaúde Brasil, como componente da Estratégia e-Saúde (Saúde Digital) para o Brasil, tem como finalidade a expansão e melhoria da rede de serviços de saúde, sobretudo da Atenção Primária à Saúde (APS), e sua interação com os demais níveis de atenção fortalecendo as Redes de Atenção à Saúde (RAS) do Sistema único de Saúde (SUS) (TELESSAÚDE, 2019). Esse projeto iniciou em 2007, como Projeto Piloto em apoio à Atenção Básica envolvendo nove Núcleos de Telessaúde localizados em universidades nos estados do Amazonas, Ceará, Pernambuco, Goiás, Minas Gerais, Rio de Janeiro, São Paulo, Santa Catarina e Rio Grande do Sul.

Após a publicação do Decreto nº 9795, de 17 maio de 2019, o Ministério da Saúde, por meio do Departamento de Saúde Digital, estabelecerá as Diretrizes para a Telessaúde no Brasil, no âmbito do SUS (TELESSAÚDE, 2019), que são: transpor barreiras socioeconômicas, culturais e, sobretudo, geográficas, para que os serviços e as informações em saúde cheguem a toda população; maior satisfação do usuário, maior qualidade do cuidado e menor custo para o

SUS; atender aos princípios básicos de qualidade dos cuidados de saúde: segura, oportuna, efetiva, eficiente, equitativa e centrada no paciente; reduzir filas de espera; reduzir tempo para atendimentos ou diagnósticos especializados; e, evitar os deslocamentos desnecessários de pacientes e profissionais de saúde.

Os campos de atuação do Programa Telessaúde no Brasil estão voltados aos seguintes segmentos (TELESSAÚDE, 2019): Inovação em saúde digital e Telessaúde; Teleconsultoria; Telediagnóstico; Telemonitoramento; Telerregulação e Teleducação.

- a) Inovação em saúde digital e Telessaúde: é transversal às iniciativas de Telessaúde e busca nas Tecnologias de Informação e Comunicação (TIC), no quesito de explorar novas ideias para a resolução de problemas crônicos, de difícil solução pelos métodos usuais e devem partir de necessidades em saúde da população;
- b) Teleconsultoria: consultoria registrada e realizada entre trabalhadores, profissionais e gestores da área de saúde, por meio de instrumentos de telecomunicação bidirecional, com o fim de esclarecer dúvidas sobre procedimentos clínicos, ações de saúde e questões relativas ao processo de trabalho em saúde, podendo ser em tempo real ou por meio de mensagens off-line;
- c) Telediagnóstico: consiste em serviço autônomo que utiliza as TIC para a realização de serviços de Apoio ao Diagnóstico, como a avaliação de exames a distância, facilitando o acesso a serviços especializados. Busca reduzir o tempo de diagnóstico possibilitando tratamento para complicações previsíveis por meio do diagnóstico precoce;
- d) Telemonitoramento: monitoramento a distância de parâmetros de saúde e/ou doença de pacientes por meio das TIC. O monitoramento pode incluir a coleta de dados clínicos, a transmissão, o processamento e o manejo por um profissional de saúde utilizando sistema eletrônico;
- e) Telerregulação: conjunto de ações em sistemas de regulação com intuito de equacionar respostas adequadas às demandas existentes, promovendo acesso e equidade aos serviços, possibilitando a assistência à saúde. Inclui também a avaliação e o planejamento das ações, fornecendo à gestão uma inteligência reguladora operacional. A telerregulação visa fortalecer o atendimento na Atenção Primária em Saúde, permitindo qualificar e reduzir as filas de espera no atendimento especializado;
- f) Teleducação: disponibilização de objetos de aprendizagem interativos sobre temas relacionados à saúde, ministrados a distância por meio de TIC, com foco na aprendizagem no trabalho, que por sua vez, ocorre transversalmente em seus campos de atuação.

Além disso, é importante destacar a Rede Nacional de Ensino e Pesquisas (RNP), uma plataforma digital para educação, pesquisa e inovação no Brasil que contribui para trazer a Internet para o País, e hoje é considerada um sistema que beneficia milhões de alunos, professores e pesquisadores. Fazem parte do ecossistema da RNP: Instituições de educação superior e pesquisa; Agências de fomento à pesquisa; Museus e instituições culturais; Empresas inovadoras; Estabelecimentos de saúde com ensino e pesquisa; e Ambientes promotores de inovação (parques e polos tecnológicos) (REDE NACIONAL DE ENSINO E PESQUISAS, 2019).

Outra rede importante a ser destacada, é a Rede Universitária de Telemedicina (RUTE), a qual é uma iniciativa do Ministério da Ciência e Tecnologia, apoiada pela Financiadora de Estudos e Projetos (Finep) e pela Associação Brasileira de Hospitais Universitários (ABRAHUE) e coordenada pela Rede Nacional de Ensino e Pesquisa (RNP), que visa apoiar o aprimoramento de projetos em Telemedicina já existentes e incentivar o surgimento de futuros trabalhos interinstitucionais. (REDE UNIVERSITÁRIA DE TELEMEDICINA, 2019).

2.4.2.1 Telemedicina em Santa Catarina

No estado de Santa Catarina, no sul do Brasil, a principal iniciativa é o Sistema Integrado de Telemedicina e Telessaúde (STT/SC), integrante da Rede Catarinense de Telemedicina (RCTM), situado na capital do estado (Florianópolis). Essa infraestrutura foi aplicada inicialmente em 1999, como um projeto de aplicação de algumas tecnologias que estavam naquele momento sendo desenvolvidas pelo Cyclops Group (GROUP, 2013), da UFSC. Essas tecnologias estavam sendo desenvolvidas em parceria com instituições da Alemanha e ainda contavam com financiamento do CNPq e da Fundação Volkswagen (Wolfsburg/German) (MAIA; WANGENHEIM; NOBRE, 2006; WALLAUER *et al.*, 2008; WANGENHEIM *et al.*, 2009; MACEDO, 2014).

No ano de 2002, o projeto piloto dessa rede buscava interligar pequenas clínicas médicas dos municípios de Florianópolis e São José. Já em 2004, com o sucesso da aplicação da tecnologia nessas clínicas, o Governo do Estado de Santa Catarina, juntamente com o Projeto Cyclops, iniciou a implementação de um piloto em nível estadual, para ser implementado no primeiro semestre de 2005. O piloto tinha o intuito de demonstrar sua aplicação para implantação no Sistema Único de Saúde (SUS). Já em 2012, o STT/RCTM estava presente em 287 dos 293 municípios catarinenses, com uma base composta por 2.118.709 exames, sendo que mais de 193 municípios já estavam com a capacidade de enviar exames a distância. Os

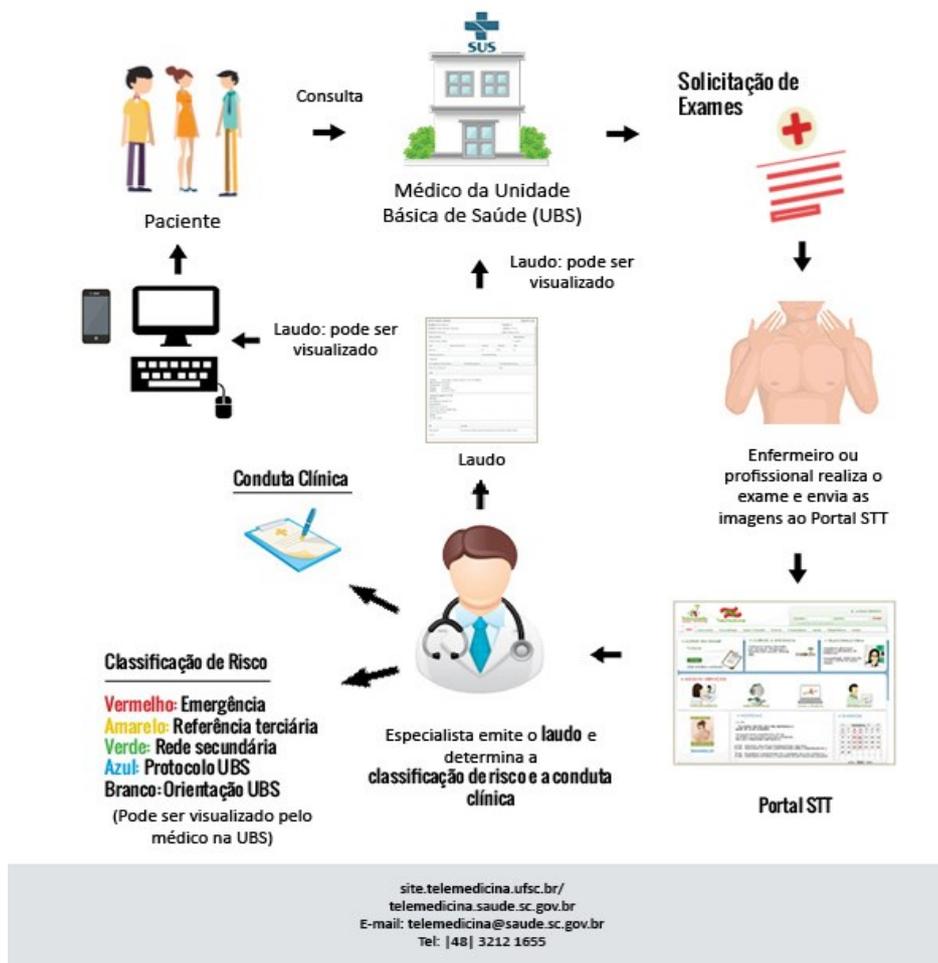
exames são provenientes, em sua maioria, de imagens médicas e, isso significa um volume de metadados exorbitante (MACEDO, 2014).

Exemplificando melhor o sistema de Telemedicina de Santa Catarina, na capital do estado, Florianópolis, destaca-se o sistema de telediagnóstico para exames dermatológicos, que possui diversos exames realizados através do STT/RCTM (TELEMEDICINA – SANTA CATARINA, 2016). A Figura 3 apresenta a estrutura do funcionamento do fluxo de Teledermatologia em um modelo catarinense de telediagnóstico.

Figura 3 – Fluxo do telediagnóstico no estado de Santa Catarina

Modelo Catarinense de Telediagnóstico

O fluxo apresenta as etapas fundamentais que compõem o atendimento e realização do Telediagnóstico em Santa Catarina



Fonte: Telemedicina – Santa Catarina (2016).

A Figura 3 apresenta um dos principais resultados desde a adesão ao STT/SC: a redução do tempo de espera para o atendimento dermatológico. De acordo com o Departamento de

Média e Alta Complexidade (DeMAC) da Secretaria Municipal de Saúde de Florianópolis, o tempo de espera para uma consulta com especialista antes da Telemedicina era em torno de um ano. Atualmente, a espera é de uma a duas semanas para o exame dermatológico. O laudo do especialista, realizado via STT/SC, é emitido na mesma semana em que o exame é feito. A espera para que o médico de família avalie o resultado e realize a conduta clínica ou encaminhe o paciente para consulta presencial com um dermatologista, depende de cada Centro de Saúde (TELEMEDICINA – SANTA CATARINA, 2016).

2.5 PRONTUÁRIO ELETRÔNICO DO PACIENTE (PEP)

Os registros apontam a Clínica Mayo, nos Estados Unidos, como sendo a pioneira em 1907, ao adotar um registro individual e cronológico de informações do paciente, criando um prontuário focado no paciente. Essa iniciativa estabeleceu em 1920 um conjunto mínimo de informações que deviam compor os registros, estes dados foram tão importantes que constituem a estrutura e apresentação do Prontuário Eletrônico do Paciente (PEP) (MASSAD MARIN; AZEVEDO, 2003; MOTA; BABETTO, 2004).

Os primeiros registros em sistema informatizado do PEP são datados da década de 1960, inicialmente para facilitar a intercomunicação hospitalar, culminando posteriormente no armazenamento das informações do paciente. A partir do primeiro congresso realizado nos Estados Unidos em 1972, organizado pelo Research and Development National Center for Health Services e o National Center for Health Statistics, foram estabelecidos os critérios mínimos para registros em prontuários médicos e, conseqüentemente, o surgimento dos primeiros PEP (COSTA, 2001).

Independentemente se em suporte tradicional ou eletrônico, ao estudar o Prontuário Eletrônico do Paciente, constata-se que ele possui várias denominações: Prontuário do Paciente, Prontuário Médico, Registro do Paciente, PEP, sendo esta última denominação a que será adotada nesta tese. No que diz respeito ao PEP, segundo França e Gaunt (1994, p. 194), é “uma memória escrita das informações clínicas, biológicas, diagnósticos e terapêuticas de uma pessoa, às vezes individual e coletiva, constantemente atualizado”. O PEP visa substituir o processo manual pela via eletrônica, de forma a torná-lo descentralizado e independente do esforço de localização física. O acesso aos dados do paciente por meio eletrônico, em um único local, proporciona o suporte automatizado à decisão e facilita a comunicação no âmbito do registro médico, tornando o registro em papel desvantajoso (WILCOX *et al.*, 2005).

O sistema de PEP foi criado para que médicos e enfermeiros recordassem de forma

sistemática dos fatos e eventos clínicos ocorridos em um indivíduo, de forma que os demais profissionais da saúde envolvidos no processo de atenção pudessem ter acesso a essas informações. Assim, ele é o mais relevante veículo de comunicação entre os membros de uma equipe de saúde responsável pelo atendimento (MASSAD; MARIN; AZEVEDO, 2003).

Por sua vez, o Conselho Federal de Medicina (CFM), no Artigo 1º da Resolução de nº 1.638/2002, define o PEP como sendo um documento único constituído por:

um conjunto de informações, sinais e imagens registradas, geradas a partir de fatos, acontecimentos e situações sobre a saúde do paciente e a assistência a ele prestada, de caráter legal, sigiloso e científico, utilizado para possibilitar a comunicação entre membros da equipe multiprofissional e a continuidade da assistência prestada ao indivíduo (BRASIL, 2002a).

A Resolução nº 1.639, de 10 de julho de 2002, aprova as normas técnicas para o uso de sistemas informatizados para a guarda e manuseio do prontuário médico, dispondo sobre a guarda permanente dos prontuários armazenados eletronicamente e estabelecendo critérios para certificação dos sistemas de informação (BRASIL, 2002b). Nesse sentido, além de ser uma ferramenta de trabalho, o PEP fortalece o atendimento integral e ajuda a estabelecer as necessidades do paciente como sujeito singular; além disso, o PEP é considerado não apenas uma ferramenta de registro, mas um facilitador da comunicação e integração dos vários saberes, que auxilia a superar inúmeras dicotomias fundamentadas no modelo biopsicossocial e no fazer interdisciplinar (MOERSCHBERGER; CRUZ; LANGARO, 2017).

2.5.1 Estrutura do PEP

Segundo Bieger e Glock (1985), com base na taxonomia das diferentes categorias de informação, o PEP é um documento multimídia, uma vez que em sua estrutura são encontrados textos verbais e não verbais, como sinais, imagens (exame radiológico), sons (batidas do coração), código de barras e símbolos (laudos e exames). De maneira geral, pode-se dizer que, como qualquer documento, o PEP, também é constituído pela estrutura física e lógica. A primeira pode ser entendida como sendo a categoria de tópicos concernentes ao paciente, convênio e internação, e aqueles atinentes à gestão de saída das organizações de saúde (PINTO, 2006).

A estrutura física do PEP está relacionada a:

- a) Dados relativos ao paciente, convênio e internação: ficha clínica, dados referentes aos exames clínicos iniciais e complementares solicitados e seus respectivos resultados, diagnóstico(s) definitivo(s), tratamento(s) efetuado(s), evolução diária do paciente,

dados referentes à nutrição, ao Serviço Social, à assistência psicológica, prescrição sobre as doses de medicamentos a serem utilizadas, avaliação pré-operatória, recuperação pós-operatória, relatório cirúrgico, controle de antimicrobiano, boletim anestésico, sistema de controle clássico e resumo de alta da sala de recuperação.

- b) Dados respeitantes à gestão da saída do paciente: prestação de contas referente à hospitalização, relatório de alta, receituário contendo orientações e prescrições a serem seguidas pelo paciente, consumo do centro cirúrgico e notificação de óbito. (PINTO, 2006).

Portanto, o PEP é um documento caracterizado, de um lado, pela sua unicidade — é único para cada paciente —, e de outro, pela pluralidade de autores que produzem seus textos e pela variedade de tipos de informações nele registradas, sendo também polifônico por sua própria natureza, escrito por vários autores: médicos, auxiliares, enfermeiros, assistentes sociais, psicólogos, nutricionistas, fonoaudiólogos, farmacêuticos bioquímicos, entre outros. É um documento constituído de informações objetivas — identificação do paciente, prescrições, cuidados etc. e informações subjetivas — compreensão dos sintomas das doenças, queixas dos pacientes e interpretação dos exames (PINTO, 2006).

Nesse contexto, ressalta-se que os padrões DICOM, HL7 e IHE mencionados no estudo de Miranda e Pinto (2015) em uso no PEP, estão em conformidade com o método desta tese, comprovando seu uso e aplicabilidade no contexto dos SIS no Brasil, com destaque para o padrão DICOM, sendo um dos padrões mais conhecidos e utilizados.

2.5.2 Aspectos Éticos e Legais do PEP

Os aspectos éticos do PEP estão diretamente relacionados aos profissionais que recebem, registram, manipulam, digitam, armazenam e processam informações relacionadas ao paciente. Este profissional é responsável pela guarda, integridade, sigilo e privacidade destes dados, que são de propriedade do paciente (MASSAD; MARIN; AZEVEDO, 2003). No contexto mundial, desde 1990, as Nações Unidas, por intermédio do seu Alto Comissariado, exigem respeito aos princípios de confidencialidade, não discriminação em relação aos dados pessoais, segurança dos arquivos, de declarações, de legitimidade das informações (noções dos benefícios ou de serviços). Por exemplo, questões éticas não devem constar nos prontuários, salvo se realmente forem necessárias. Por outro lado, a Organização de Cooperação e de Desenvolvimento Econômico (OCDE) recomenda aos seus Estados Membros que limitem a coleta dos dados apenas aos que forem considerados úteis e sem constrangimentos aos

indivíduos. No contexto europeu, desde 24 de novembro de 1995, foram estabelecidas ações para harmonizar as normas. Na França, a Lei de nº 78-17, de 06 de janeiro de 1978, relativa à informática, aos arquivos e liberdades, diz respeito apenas às questões de saúde. A Lei nº 94-548, de 01 de julho 1994, é mais específica em relação aos dados nominativos concernentes à pesquisa médica, complementa a Lei de 1978. A Lei nº 99-461, de 27 de julho de 1999, denominada *Couverture Maladie Universelle (CMU)*, também trata dos PEPs (PINTO, 2006).

Os profissionais e instituições de saúde são responsáveis pela privacidade, ou seja, pela custódia das informações contidas no PEP, sendo obrigados a não dar ciência sem consentimento prévio do paciente com exceção nos casos de necessidade de cuidado assistencial ou dever legal. A instituição de saúde tem a responsabilidade de programar, aprimorar, estabelecer normas e rotinas que garantam o controle e identificação do acesso dos usuários às informações do PEP (MOTTA, 2003). No Brasil, o Conselho Federal de Medicina (CFM) está empenhado no contexto legal e ético do prontuário do paciente, tendo já aprovado a Resolução CFM nº 1.331/89, que trata da temporalidade do PEP, e as portarias de nº 1.638/2002 e nº 1.639/2002 que normatiza o uso de sistemas informatizados, a guarda e o manuseio de prontuários. Portanto, todos os aspectos legais do PEP vêm ao encontro da autenticidade, integridade, confidencialidade/privacidade, auditoria, assinatura eletrônica e guarda de documentos (PINTO, 2006).

2.5.3 Contribuições da Ciência da Informação para o PEP

Ao estudar o PEP observa-se que as informações contidas em sua estrutura partem de um pressuposto ligado à Ciência da Informação em relação aos seus processos informacionais. Assim, a Ciência da Informação pode contribuir em diversos aspectos informacionais, evidenciando um melhor gerenciamento, juntamente com as TIC, das informações registradas.

Para Galvão e Ricarte (2011), apresentam as demandas informacionais e tecnológicas associadas aos seus processos (criação, comunicação, identificação, seleção, aquisição, organização, recuperação, armazenagem, preservação, análise e síntese, e avaliação) do prontuário eletrônico do paciente, bem como as contribuições necessárias da Ciência da Informação e de seus profissionais para atendê-las, conforme pode ser observado no Quadro 5.

Quadro 5 – Demandas do PEP e contribuições da ciência da informação

| Processo | Demanda | Contribuição necessária |
|---|--|--|
| Criação da informação | Informações e dados que garantam a integralidade e continuidade da assistência em saúde no contexto institucional, regional e nacional | Projetar, validar e implementar metodologias que permitam a interoperabilidade sintática e semântica entre sistemas de prontuários, considerando normas existentes |
| Comunicação da informação | Registro no prontuário empregando terminologias que possam ser compreendidas pela equipe multiprofissional presente e futura | Planejar, traduzir, harmonizar e manter terminologias e classificações clínicas, especialmente em língua portuguesa |
| Identificação da informação | Identificação unívoca de pacientes, profissionais da saúde, medicamentos e procedimentos | Projetar, validar e implementar metodologias para registros e identificação de pessoas. Associar instrumentos de apoio terminológico ao prontuário |
| | Anonimização de dados e informações do prontuário, visando gestão, ensino e pesquisa | Projetar, validar e implementar metodologias para anonimizar dados e informações |
| Seleção da informação | Seleção de dados e informações que compõem o prontuário frente às mudanças tecnológicas, científicas, e de políticas em saúde | Estabelecer políticas e critérios de seleção e atualização das informações a serem registradas no prontuário |
| Aquisição da informação | Variação de tipos de registro e formatos (textual, visual, sonoro e suas combinações), em cada área e especialidade profissional | Projetar, validar e implementar metodologias para a aquisição e integração de dados e informações nos formatos mais adequados Capacitar os diferentes profissionais da saúde para o adequado registro no prontuário |
| | Recepção de dados e informações oriundos de contextos externos à instituição (dispositivos móveis e outras instituições) | Projetar, validar e implementar metodologias relacionadas ao fluxo, compartilhamento e intercâmbio seguro de dados e informações |
| Organização e recuperação da informação | Organização e apresentação flexível de dados e informações | Projetar, validar e implementar metodologias de organização e apresentação de dados e informações que atendam os diferentes profissionais da saúde |
| | Localização de dados e informações precisos para assistência ao paciente, gestão, pesquisa e ensino em saúde | Projetar, validar e implementar metodologias de navegação e mecanismos de busca adequados às necessidades informacionais, respeitados os limites éticos e legais |
| Armazenagem e preservação da informação | Controle de situações ambientais ou tecnológicas inadequadas que levem à perda de dados | Projetar, validar e implementar metodologias para o armazenamento seguro de dados e informações |
| | Guarda permanente dos dados e informações de prontuários | Projetar, validar e implementar metodologias de preservação digital |

| | | |
|---------------------------------|---|---|
| | eletrônicos | de longo prazo para os dados e informações |
| Análise e síntese da informação | Necessidades de informações sumarizadas sobre um paciente no momento da assistência em saúde | Projetar, validar e implementar metodologias para a síntese automática de dados e informações do prontuário de paciente orientada a diferentes perfis de profissionais da saúde |
| | Necessidades de informações sumarizadas para gestão e estabelecimento de políticas públicas em saúde baseadas em evidências | Projetar, validar e implementar metodologias para a síntese automática de dados e informações de conjuntos de prontuários de pacientes |
| | | Projetar, validar e implementar metodologias para a criação de novos produtos informacionais, respeitados os limites éticos e legais |
| Avaliação da informação | Controle da qualidade da informação registrada em prontuários | Projetar, validar e implementar metodologias para o registro de dados e informações corretos, sem prejudicar a atuação do profissional da saúde |
| | Uso de informação em conformidade com legislação vigente e códigos de ética profissional | Projetar, validar e implementar metodologias para verificação do registro e do acesso a dados e informações referentes às diferentes funções exercidas pelos profissionais da saúde |

Fonte: Galvão e Ricarte (2011).

No Quadro 5, vale ressaltar que o aprofundamento do PEP enquanto objeto de estudo pela Ciência da Informação constitui campo fértil para teste de hipóteses, tendo em vista a ampliação, aperfeiçoamento ou mesmo revisão de teorias e metodologias (BOOTH, 2003). Por fim, em relação aos estudos do PEP, a Ciência da Informação pode contribuir para o desenvolvimento de novas pesquisas ligadas ao tema nas diferentes áreas do conhecimento, e de forma interdisciplinar juntamente com a Ciência da Computação e as Ciências da Saúde.

2.6 CONSIDERAÇÕES DA SEÇÃO

Nesta seção foram vistos assuntos relacionados ao referencial teórico desta tese, em que foram apresentados estudos para embasar o método que está sendo proposto. O referencial teórico mostra a importância da temática em questão, no que diz respeito aos cenários de saúde ideais para a motivação do desenvolvimento e concretização desta pesquisa.

3 ESTADO DA ARTE

Nesta seção é discutido o estado da arte sobre o tema de pesquisa desta tese com intuito de verificar e evidenciar o ineditismo e a originalidade do método proposto, além de averiguar a existência de trabalhos relacionados ou similares. Para a realização desse processo, são apresentados o protocolo da RSL proposto por Kitchenham (2004) e a execução da técnica de *Snowballing (backward and forward)* proposta por Wohlin (2014) para cobrir os estudos que não foram identificados anteriormente na execução da RSL. Ainda, serão tratados os estudos selecionados ao tema e os trabalhos realizados com intuito de experimentação para a construção deste método. Posteriormente, foi realizada a análise dos estudos selecionados em relação à pesquisa desenvolvida apresentando as lições aprendidas. Por fim, são apresentadas as considerações finais da seção.

3.1 REVISÃO SISTEMÁTICA DE LITERATURA

Com o propósito de identificar o ineditismo e a originalidade desta tese, foi realizada uma Revisão Sistemática de Literatura (RSL) para identificar estudos relacionados ou similares ao tema aqui apresentado. A RSL é um tipo de investigação científica que tem por objetivo reunir, avaliar criticamente e conduzir uma síntese dos resultados de múltiplos estudos primários. Esse tipo de revisão também objetiva responder a uma pergunta claramente formulada, utilizando métodos sistemáticos e explícitos para identificar, selecionar e avaliar as pesquisas relevantes, coletar e analisar dados de estudos incluídos na revisão, sendo que os métodos estatísticos podem ou não ser usados para analisar e sumarizar os resultados dos estudos incluídos (COOK; MULROW; HAYNES, 1997; CLARKE; HORTON, 2001; LINDE; WILLICH, 2003; KITCHENHAM, 2004; KITCHENHAM; CHARTERS, 2007; PETTICREW; ROBERTS, 2006).

Uma RSL é necessária para uma pesquisa científica, pois dentre os inúmeros estudos publicados sobre determinado tema, aqueles com superior qualidade metodológica, podem ser utilizados nas mais variadas práticas. Além disso, como resultados conflitantes frequentemente emergem de estudos diferentes abordando a mesma questão e estudos individuais raramente possuem poder estatístico suficiente para fornecer respostas definitivas, as revisões sistemáticas têm grande importância como ferramenta de investigação científica para a tomada de decisão, com custos bem menores do que os exigidos para a realização de estudos em larga escala (COOK; MULROW; HAYNES, 1997; DIXON; HAMEED; SUTHERLAND, 2005).

Destaca-se, ainda, que esse tipo de revisão, disponibiliza um resumo das evidências relacionadas à uma estratégia de intervenção específica, mediante a aplicação de métodos explícitos e sistematizados de busca, apreciação crítica e síntese da informação selecionada. É definida por um protocolo que estabelece as etapas dos procedimentos a serem realizados durante a revisão. Os procedimentos metodológicos, presentes no protocolo, permitem tanto avaliar o estado atual dos conhecimentos da área, como manter a atualização de pesquisas em bases avançadas (COOK; MULROW; HAYNES, 1997; CLARKE; HORTON, 2001; LINDE; WILLICH, 2003; KITCHENHAM, 2004; KITCHENHAM; CHARTERS, 2007; PETTICREW; ROBERTS, 2006).

Nessa perspectiva, a RSL aqui desenvolvida para esta tese, foi baseada no trabalho original de Kitchenham (2004), o qual define as atividades para o desenvolvimento de um estudo em engenharia de software, sendo adaptado para a temática em questão. A metodologia da RSL descrita por Kitchenham (2004) é reconhecida internacionalmente, sendo empregada nesta tese para responder às questões de pesquisa.

3.1.1 Contextualização e questões de pesquisa

De fato, não há padronização para métodos, técnicas, modelos ou metodologias para o gerenciamento da Proveniência de Dados em SIS, o que motivou o desenvolvimento de vários tipos de abordagens. Essas abordagens podem assumir diferentes aplicações, nas mais variadas situações e cenários que envolvem SIS. Nesse sentido, a fim de buscar um alinhamento com o tema desta tese, foi realizada uma RSL na tentativa de identificar diferentes métodos, técnicas, modelos ou metodologias para o gerenciamento da proveniência de dados em SIS, sendo uma contribuição para a área do conhecimento relacionada à temática em questão.

A RSL aqui apresentada poderá contribuir, também, para construir novas hipóteses, evidências e sintetizar resultados que auxiliarão novas pesquisas envolvidas com a Proveniência de Dados aplicada em cenários de SIS. Por isso, um dos processos essenciais de qualquer RSL é a seleção de questões de pesquisa (KITCHENHAM, 2004; PETTICREW; ROBERTS, 2006; KITCHENHAM; CHARTERS, 2007). Sendo assim, foram separadas as questões em dois grupos: a) questões gerais de pesquisa (QG) e questões específicas de pesquisa (QE).

Nesse sentido, antes é importante ressaltar que, para a formulação das questões gerais desta RSL foi utilizado o mnemônico SPICE, proposto por Booth (2004). O mnemônico – SPICE, compreende: *Setting – where?* (Cenário - Onde? Ex.: Em que contexto você está abordando a questão?), *Perspective – for whom?* (Perspectiva - Para quem? Ex.: Quem são os

participantes?), *Intervention – what?* (Intervenção – O que? Ex.: O que está sendo feito?), *Comparison – compared with what?* (Comparação – comparado com o que? Ex.: Quais as suas alternativas?), *Evaluation – with what result?* (Avaliação - com qual resultado? Ex.: Com qual (que) resultado? Como você medirá se a intervenção foi bem-sucedida?).

As fases do mnemônico SPICE, adaptadas para este projeto são:

- a) Cenário: Tipos de SIS;
- b) Perspectiva: Profissionais e usuários do SIS;
- c) Intervenção: armazenamento, disponibilidade, rastreabilidade, confidencialidade, integridade, autenticidade e auditabilidade;
- d) Comparação: além de métodos, técnicas, modelos ou metodologias para o gerenciamento de dados de proveniência, alternativas que podem ser comparadas, como as tecnologias utilizadas nos SIS;
- e) Avaliação: resultados apresentados utilizando métodos, técnicas, modelos, metodologias e tecnologias utilizadas para gerenciar dados de proveniência em SIS.

Vale ressaltar que essas cinco fases foram estabelecidas com vistas a garantir a qualidade dos artigos a serem revisados por meio das QG formuladas. O Quadro 6 apresenta as QG formuladas pelo mnemônico SPICE e as respectivas QE para esta RSL.

Quadro 6 – Questões gerais e questões específicas de pesquisa

| Questão | Tipo de questão |
|---|---|
| Questões Gerais de pesquisa | |
| QG01 | Quais são os diferentes métodos, técnicas, modelos e metodologias utilizadas para o gerenciamento de dados de proveniência em SIS? |
| QG02 | Quais são os desafios em relação aos diferentes métodos, técnicas, modelos e metodologias identificados a respeito do gerenciamento de dados de proveniência em SIS? |
| Questões Específicas de pesquisa | |
| QE01 | Levando em conta os aspectos mais representativos quanto ao gerenciamento de dados de proveniência em SIS, como esses sistemas abordam os diferentes métodos, técnicas, modelos e metodologias? |
| QE02 | Quais são as principais vantagens da aplicação de diferentes métodos, técnicas, modelos ou metodologias para o gerenciamento de dados de proveniência em SIS? |
| QE03 | Quais são as principais tecnologias identificadas nos diferentes métodos, técnicas, modelos ou metodologias que contribuíram para o gerenciamento de dados de proveniência em SIS? |

Fonte: Elaborado pelo autor (2020).

3.1.2 Escolha das bases de dados e termos para estratégia de busca

As pesquisas foram realizadas por meio da configuração de VOIP da Universidade Federal de Santa Catarina (UFSC), que proporciona o acesso às bases de dados pela Internet em qualquer local onde o pesquisador esteja. Para isso, a escolha das bases de dados internacionais utilizadas, partiram do pressuposto de maior aderência em relação às pesquisas relacionadas à temática, em fontes bibliográficas da área de estudo, sendo que as bases utilizadas foram: *ACM Digital Library*; *IEEEExplore*; *Science Direct*; *Springer Link*; *Scopus* e, *Web of Science*.

O intervalo utilizado para as buscas nas bases de dados escolhidas foi de 2010 a 2020, sendo estipulado o ano de 2020 como limite por motivos da pandemia de Covid-19, que em 11 de março de 2020 foi caracterizada pela OMS como uma pandemia mundial, impedindo muitos estudos de serem realizados. Nesse sentido, esse intervalo de tempo foi estipulado a partir dos pré-testes realizados nas bases de dados que apresentaram mais resultados após o ano de 2010. Os termos pesquisados foram adequados conforme as configurações fornecidas pelas bases de dados, sendo que a pesquisa foi limitada somente ao título, abstract e as palavras-chave dos artigos nas respectivas buscas nas bases de dados.

O objetivo da estratégia de pesquisa estava focado em concatenar os termos “*Data Provenance*” e “*Health*” para identificar diferentes métodos, técnicas, modelos ou metodologias de Proveniência de Dados, considerando os diferentes tipos de SIS e diferentes cenários de saúde. De acordo com o trabalho de Kitchenham (2004), foram selecionadas variantes e sinônimos relacionados ao tema da pesquisa para resultados mais precisos que associavam o tema em questão com o estudo de caso desta tese, como: “*Telemedicine*”, “*e-Health*”, “*m-Health*” e “*Healthcare*”. Finalmente, compactuando os termos de foco da pesquisa juntamente com as variantes e sinônimos associados, foram utilizados os operadores lógicos booleanos (AND, OR) para formar a seguinte *string* de busca: *Data Provenance AND (Health OR Telemedicine OR e-Health OR m-Health OR Healthcare)*. Não foram realizadas variações desses termos, tendo em vista que na aplicação dos pré-testes foram observadas a irrelevância dessas variações nos resultados encontrados.

3.1.3 Critérios para seleção dos estudos primários

Foram analisados artigos somente no idioma inglês, publicados em *Journals* ou em *Conferences*, devido à relevância desses na área de Ciência da Computação. Outros documentos

como teses, dissertações, livros, entre outros, não foram selecionados. Nesse sentido, para realizar esta etapa de seleção dos estudos primários, um conjunto de critérios de inclusão (Quadro 7) e exclusão (Quadro 8) foram definidos.

Quadro 7 – Critérios de inclusão

| ID | Critérios de Inclusão |
|------|---|
| CI01 | Estudos publicados em artigos de <i>Journal e Conference</i> . |
| CI02 | Estudos que indicam primária ou secundariamente abordagens experimentais ou teóricas, onde exemplos de aplicações ou descrições de experimentos de casos reais são apresentados utilizando diferentes métodos, técnicas, modelos ou metodologias para o gerenciamento de dados de proveniência, considerando os tipos de HIS. |

Fonte: Elaborado pelo autor (2020).

Quadro 8 – Critérios de exclusão

| ID | Critérios de Exclusão |
|------|---|
| CE01 | Estudos duplicados nas bases pesquisadas. |
| CE02 | Estudos que não estejam no idioma inglês. |
| CE03 | Estudos com menos de seis páginas. |
| CE04 | Estudos que não se referiam a nenhum tipo de método, técnica, modelo ou metodologia para o gerenciamento de dados de proveniência em SIS. |
| CE05 | Estudos que não atendam às questões de pesquisa. |

Fonte: Elaborado pelo autor (2020).

3.1.4 Avaliação da qualidade da seleção dos estudos primários

A qualidade de um artigo pode ser mensurada por sua relevância e pelo valor científico de seu conteúdo, considerado como um critério de exclusão a ser aplicado durante a condução da pesquisa (PETERSEN; VAKKALANKA; KUZNIARZ, 2015). Para a avaliação da qualidade dos estudos primários desta tese, foram levantados critérios a fim de verificar se os artigos são estudos relevantes, pois conforme descrito por Kitchenham (2004) esses procedimentos são necessários para avaliar a qualidade dos trabalhos selecionados. A avaliação da qualidade dos estudos primários neste projeto consiste nos artigos selecionados, considerando o propósito de pesquisa, contextualização, revisão da literatura, trabalhos relacionados, metodologia, resultados, e a conclusão, de acordo com os objetivos e indicação de estudos futuros. Assim, durante a análise dos estudos primários e a coleta de resultados, os critérios formulados no Quadro 9 foram aplicados permitindo um processo adicional mais amplo de validação dos estudos.

Quadro 9 – Critérios de qualidade

| ID | Critérios de qualidade |
|------|--|
| CQ01 | Os objetivos do estudo devem estar alinhados com a gestão de dados de proveniência em SIS. |
| CQ02 | O estudo deve apresentar teorias ou aplicações para o gerenciamento de dados de proveniência em contextos de SIS. |
| CQ03 | O estudo deve apresentar a proposta ou experimento sobre os métodos, técnicas, modelos ou metodologias para o gerenciamento de dados de proveniência em SIS. |
| CQ04 | O estudo deve apresentar resultados relevantes para o uso da teoria ou aplicações de gerenciamento de dados de proveniência em SIS. |
| CQ05 | O estudo deve apresentar a conclusão vinculada aos objetivos da pesquisa. |

Fonte: Elaborado pelo autor (2020).

Nessa perspectiva, a avaliação da qualidade pode servir como recomendação de estudos para futuras pesquisas, fornecendo informações a respeito da qualidade das informações de cada estudo avaliado (PETERSEN; VAKKALANKA; KUZNIARZ, 2015). Desse modo, os estudos primários identificados foram lidos em sua totalidade e, então, foram avaliados utilizando os critérios de qualidade, os quais descreviam de forma abrangente o suficiente para cobrir o escopo dos estudos a serem considerados, sendo possível buscar respostas às perguntas criadas para realização desta RSL, apresentadas no Quadro 10.

Quadro 10 – Seções dos estudos relacionados

| Capítulo | Descrição | Questões de Pesquisa |
|-----------------|--|-------------------------------|
| Título | Título do estudo específico | QG01, QG02 |
| Resumo | Resumo do estudo | QG01, QG02 |
| Palavra-chave | Palavras do conteúdo do texto | Todas as questões de pesquisa |
| Introdução | Problema a ser resolvido | Todas as questões de pesquisa |
| Desenvolvimento | Conceitos relacionados com a proposta | Todas as questões de pesquisa |
| Métodos | Metodologia científica | Todas as questões de pesquisa |
| Resultados | Resultado das avaliações | Todas as questões de pesquisa |
| Discussão | Dados quantificados em comparação com a literatura | QE01, QE02, QE03 |
| Conclusão | Descobertas relacionadas a objetivos e hipóteses | QE01, QE02, QE03 |

Fonte: Elaborado pelo autor (2020).

Ainda, para avaliar o grau de adequação aos critérios de qualidade, foi adotada a estratégia de avaliação proposta por Likert (1932), permitindo respostas gradativas de 0 (discordo totalmente) a 2 (concordo totalmente), como apresentado no Quadro 11.

Quadro 11 – Escala de Likert-3

| Escala de Likert-3 | |
|-------------------------|--|
| Concordo totalmente (2) | Deve ser concedido no caso em que o trabalho apresente no texto os critérios que atendam totalmente a questão. |
| Neutro (1) | Deve ser concedido no caso em que o trabalho não deixe claro se atende ou não a questão; |
| Discordo totalmente (0) | Deve ser concedido no caso em que não existe nada no trabalho que atenda aos critérios da questão. |

Fonte: Adaptado de Likert (1932).

Para auxiliar a avaliação, a escala de Likert-3 foi adaptada para cada critério de qualidade proposto nesta RSL, como se pode observar no Quadro 12.

Quadro 12 – Escala para cada critério de qualidade

| Critério | Escala |
|----------|---|
| CQ01 | 2 - Define os objetivos do estudo com o contexto da pesquisa. 1 - Define os objetivos, mas não deixa claro o contexto da pesquisa. 0 - Não define os objetivos. |
| CQ02 | 2 - As teorias e aplicações são apresentadas no estudo. 1 - Apenas uma das características é informada. 0 - O estudo não relaciona a teoria com a prática. |
| CQ03 | 2 - O estudo apresenta a proposta e o experimento. 1 - O estudo apresenta apenas uma das características. 0 - O estudo não apresenta nenhuma das características. |
| CQ04 | 2 - Os resultados estão totalmente aderentes ao objetivo do estudo. 1 - Os resultados são parcialmente aderentes ao objetivo do estudo. 0 - Não é alcançado nenhum resultado. |
| CQ05 | 2 - As conclusões são claramente apresentadas. 1 - As conclusões não são claras. 0 - Não são apresentadas conclusões. |

Fonte: Elaborado pelo autor (2020).

Após essas etapas, foram extraídos os dados e realizada a sintetização desses dados para obtenção de uma visão mais ampla sobre a temática aqui proposta nesta tese.

3.1.5 Extração e Síntese dos Dados

A extração e sintetização dos dados dos estudos recuperados foram realizados de forma estruturada, a partir da exportação dos documentos para o *Mendeley* para a eliminação de estudos duplicados. Para melhor visualização dos dados foi gerada uma planilha eletrônica contendo (base de dados, tipo de estudo, ano, idioma, autores, título, resumo, palavras-chave, fonte, DOI e link para acesso ao estudo). Para melhor obter os dados nos estudos resultantes, eles foram sintetizados com base nas seções designadas para responder às questões gerais e

específicas desta tese, seguindo os passos de avaliação dos estudos selecionados, exposto anteriormente no Quadro 10.

3.1.6 Identificação dos estudos primários

As bibliotecas eletrônicas já mencionadas no protocolo desta RSL para recuperar estudos primários compõem as metas para cobrir *Journals* ou *Conferences* essenciais dentro da ciência da computação voltados ao uso de SIS. Portanto, os resultados da pesquisa ainda tiveram que passar pelos filtros dos processos desta RSL e pela fase de síntese dos estudos relevantes.

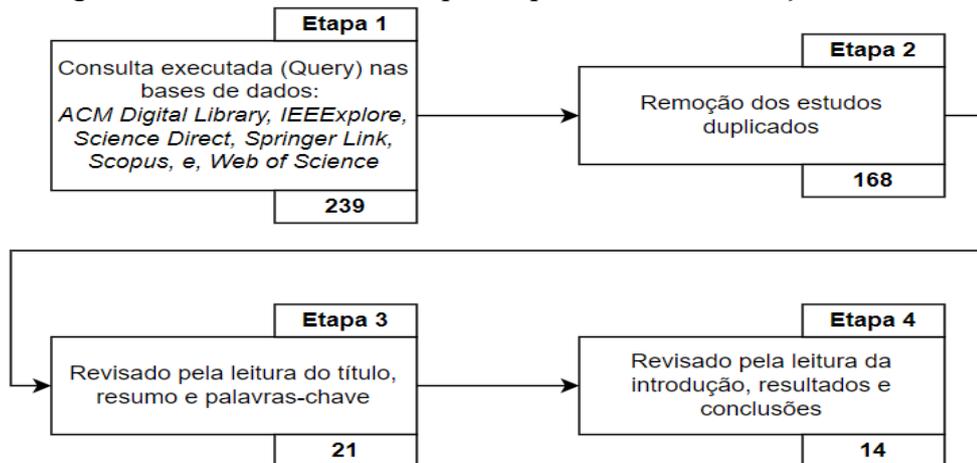
3.1.7 Síntese dos estudos relevantes

O processo de síntese dos estudos relevantes para essa RSL, seguiu alguns passos com o objetivo de filtrar aqueles que tinham alguma relação com o método proposto nesta tese. Após os primeiros processos, descritos anteriormente, foi executado o primeiro filtro de leitura do (título, resumo e palavras-chave) juntamente com a execução dos critérios de inclusão e exclusão. Em seguida, foi executado o segundo filtro de leitura da (introdução, resultados e conclusão), sendo ainda possível descartar estudos semelhantes (resumos e ou extensões) dos estudos originais. Dessa forma, foi possível selecionar somente os artigos que atenderam aos critérios de seleção anteriormente especificados e possíveis de responderem as questões desta RSL.

3.1.8 Condução da RSL

Os estudos primários foram identificados e utilizados para responder às perguntas já definidas. A Figura 4 mostra o processo de seleção de estudos primários em cada etapa da RSL.

Figura 4 – Número de estudos por etapa da fase de condução da SLR



Fonte: o próprio autor (2021)

Observamos na Figura 4, que muitos estudos duplicados foram encontrados. Isso acontece porque as bases de dados digitais frequentemente indexam estudos primários de outras bases. Diferentes fatores podem justificar a quantidade de estudos que retornou em cada banco de dados. Essa quantidade pode ser justificada por fatores, como: a ordem de execução das buscas; a quantidade total de estudos na fonte; e a relevância da fonte para a questão de pesquisa. Para entender melhor a Figura 4, na etapa 1, a consulta foi executada nas bases de dados selecionadas entre 25, 26 e 27 de junho de 2021. O intervalo de busca é de 10 anos (2010 a 2020), retornando um total de 239 estudos. Destes 239 estudos recuperados, 11 foram retirados da *ACM Digital Library*, 50 da *IEEEExplore*, 3 da *Science Direct*, 66 da *Scopus*, 59 da *Springer Link* e 50 da *Web of Science*. Ressalta-se que as buscas realizadas nessas bases de dados não retornaram estudos indexados nos principais eventos de Proveniência de Dados, como o Provenance Week (<http://provenanceweek.org/>) que engloba eventos relacionados em torno do tema da proveniência, tais quais: International Provenance & Annotation Workshop (IPAW) – (<http://ipaw.info/>) e International Workshop on the Theory and Practice of Provenance (TaPP) – (<https://www.usenix.org/conferences/byname/186>) entre outros eventos voltados especificamente aos estudos de Proveniência de Dados. Isso é um fato comum, pois nem todos os eventos são indexados nas bases de dados científicas, estando os documentos relativos a esses eventos restritos a buscas somente nas instituições que os organizam.

Portanto, para melhor visualizar os resultados dessas buscas, no Quadro 13 são apresentadas as datas das buscas, a *string* de busca (estruturadas para utilização em cada base) e a quantidade de artigos recuperados.

Quadro 13 – Artigos encontrados em bases de dados eletrônicas internacionais

| Bases de Dados | Data da busca | Estratégia de busca | Quantidade de Artigos |
|-----------------------|----------------------|--|------------------------------|
| ACM Digital Library | 25/06/2021 | Searched The ACM Guide to Computing Literature [Abstract: "data provenance"] AND [[Abstract: health] OR [Abstract: telemedicine] OR [Abstract: "e-health"] OR [Abstract: "m-health"] OR [Abstract: healthcare]] AND [Publication Date: (01/01/2010 TO 12/31/2020)] | 11 |
| IEEEExplore | 25/06/2021 | "(("Abstract": "Data Provenance") AND ("Abstract": Health OR "Abstract": Telemedicine OR "Abstract": "e-Health" OR "Abstract": "m-Health" OR "Abstract": Healthcare)) Filters Applied: 2010 - 2020" | 50 |
| Science Direct | 25/06/2021 | Year:2010-2020 Title, abstract, keywords: "Data Provenance" AND (Health OR Telemedicine OR "e-Health" OR "m-Health" OR Healthcare) Article type: Review article OR Research articles OR Conference abstracts | 3 |
| Scopus | 25/06/2021 | (TITLE-ABS-KEY ({Data Provenance}) AND TITLE-ABS-KEY (health OR telemedicine OR {e-Health} OR {m-Health} OR healthcare)) AND PUBYEAR > 2009 AND PUBYEAR < 2021 AND (LIMIT-TO (DOCTYPE , "cp") OR LIMIT-TO (DOCTYPE , "ar")) | 66 |
| Springer Link | 27/06/2021 | ""data AND provenance"" AND (Health OR Telemedicine OR "e-Health" OR "m-Health" OR Healthcare)' within 2010 - 2020 Com filtros de tipo de conteúdo: article, conference paper" | 59 |
| Web of Science | 27/06/2021 | "(de Principal Coleção do Web of Science) TÓPICO: ("Data Provenance") AND TÓPICO: (Health OR Telemedicine OR "e-Health" OR "m-Health" OR Healthcare) Tempo estipulado: 2010-2020. Índices: SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI. " | 50 |
| TOTAL | | | 239 |

Fonte: Elaborado pelo autor (2020).

Como resultado da etapa 2 foram encontrados 71 estudos duplicados, restando 168 estudos para o filtro da etapa 3. Assim, na etapa 3, foi realizado o primeiro filtro (leitura do título, resumo e palavras-chave), em que foram descartados 147 estudos por não atenderem aos critérios de inclusão desta RSL, restando 21 estudos para a execução do segundo filtro na etapa

4. É importante ressaltar que, ao ler os resumos dos estudos, observou-se que eles possuíam características relacionadas ao gerenciamento dos dados de proveniência em SIS. Na Etapa 4, foi realizado o segundo filtro (leitura da introdução, resultados e conclusões) dos 21 estudos restantes. Assim, fortes relações com o gerenciamento dos dados de proveniência dos SIS foram cuidadosamente observadas nesses estudos. No entanto, sete estudos foram descartados por serem muito semelhantes (resumos e/ou extensões). Por fim, 14 estudos foram selecionados para leitura completa, pois atenderam a todos os critérios de seleção especificados nas 4 etapas.

Para a avaliação da qualidade dos 14 estudos primários selecionados para compor esta RSL, pode-se afirmar que todos os estudos foram avaliados seguindo todos os critérios de qualidade já mencionados.

Ainda, é importante ressaltar que, o processo de exclusão que resultou em 225 estudos descartados dos 239 recuperados nas bases de dados selecionadas, não estavam relacionados ao gerenciamento dos dados de proveniência em SIS. Embora a Proveniência de Dados possa ter sido mencionada em seus resumos como um dos casos de uso, não foi o foco da pesquisa do autor. Esses estudos só mencionaram a Proveniência de Dados em uma das subseções como área potencial de aplicação em saúde, sem contribuir para novas ideias aplicadas ao gerenciamento dos dados de proveniência em SIS.

3.1.9 Utilização da técnica de Snowballing (backward and forward)

De acordo com Kitchenham e Charters (2007), as RSL devem ser executadas estritamente seguindo uma estratégia de pesquisa predefinida. Essa estratégia de busca deve ser imparcial e deve permitir que a integridade da pesquisa seja avaliada. Para Kitchenham e Charters (2007), os autores argumentam que as buscas iniciais por estudos podem ser realizadas por meio de diversas bibliotecas digitais e indicam que outras pesquisas complementares devem ser empregadas (por exemplo, pesquisas manuais em periódicos). Um exemplo de um procedimento manual frequentemente usado além da RSL é o *Snowballing (backward and forward)*. Esta é uma estratégia de busca em que partindo de um conjunto inicial de artigos, visita-se referências dele (*backward snowballing*) e para ele (*forward snowballing*). Esse processo é repetido até que não existam mais referências a serem adicionadas (KITCHENHAM; CHARTERS, 2007; JALALI; WOHLIN, 2012; WOHLIN, 2014; BADAMPUDI; WOHLIN; PETERSEN, 2015). Por essa razão, foi identificado um conjunto inicial, definido como o ponto de partida. Este conjunto inicial é uma coleção de estudos já selecionados para compor o mapeamento sistemático, a partir do qual suas referências e citações serão verificadas (JALALI;

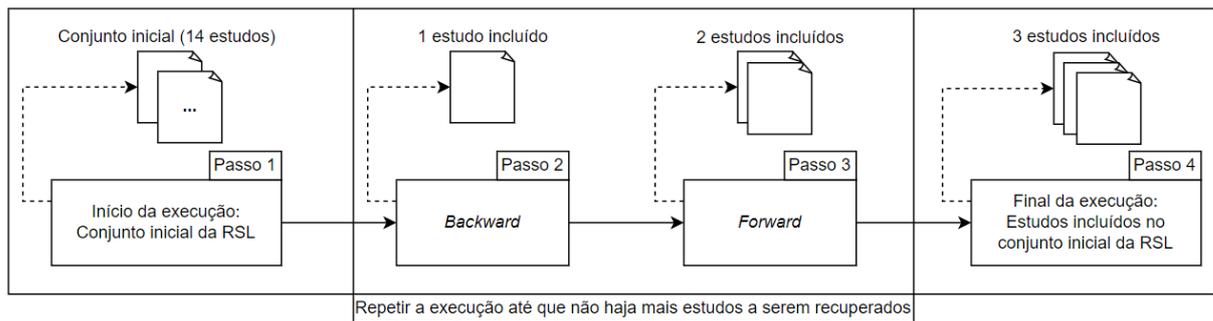
WOHLIN, 2012; WOHLIN, 2014). Nesse conjunto, foram incluídos apenas os estudos que farão parte da análise final. O próximo passo é iniciar a primeira iteração, conduzindo *Snowballing (backward and forward)*. Após a execução dos processos de *backward* e *forward*, os documentos recuperados são adicionados ao total do conjunto inicial que foi avaliado no início do processo (WOHLIN, 2014). As iterações são definidas por (WOHLIN, 2014) da seguinte forma:

- a) *Snowballing (backward)*: utiliza-se a lista de referência de cada estudo para encontrar novos trabalhos a serem incluídos. Ao verificar a lista de referências, exclua de acordo com critérios básicos de exclusão, como ano de publicação, linguagem escrita ou tipo de publicação. O próximo passo é excluir os estudos já encontrados antes, e então os outros são candidatos à inclusão, e depois ler as outras informações e partes com maior relevância;
- b) *Snowballing (forward)*: utiliza-se a lista de citações das obras incluídas. O *Google Scholar* pode visualizar as citações que cada artigo tem. Cada citação é analisada a partir de uma visão geral, e se informações como título e resumo forem suficientes, o artigo pode ser incluído na lista para leitura posterior.

3.1.9.1 Execução do *Snowballing (backward and forward)*

Um segundo passo na RSL foi realizado, executando então a técnica de *Snowballing (backward and forward)* com base em Wohlin (2014) para cobrir estudos primários que não foram identificados anteriormente. Nesse processo, os 14 estudos primários selecionados na RSL foram utilizados como conjunto inicial de entrada para execução do *Snowballing (backward and forward)*. É importante ressaltar que, os critérios de inclusão e exclusão utilizados na execução da técnica de *Snowballing* foram os mesmos utilizados na RSL, apresentados anteriormente no Quadro 5 no Quadro 6 desta tese. Porém, é importante destacar uma ressalva: no critério de inclusão (CI01) no Quadro 5, considerou-se, além de estudos publicados em artigos de *Journal* e *Conference*, artigos de relatórios técnicos (*Technical Report*) e artigos de e-book para maior amplitude no uso da execução da técnica de *Snowballing*. Na execução da técnica de *Snowballing*, foram analisadas as referências, e os estudos primários que atendem aos interesses desta RSL, sendo adicionados ao escopo estudado. A Figura 5 mostra os passos utilizados na técnica de *Snowballing* aplicada nesta RSL.

Figura 5 – Processo de execução *Snowballing (backward and forward)*



Fonte: Elaborado pelo autor (2021).

Conforme apresentado na Figura 5, o processo de execução do *Snowballing (backward and forward)* foi realizado em quatro passos e são descritos a seguir:

- a) Passo 1: o conjunto inicial de 14 estudos selecionados na RSL foi avaliado para iniciar o processo de execução do *Snowballing (backward and forward)*;
- b) Passo 2: no processo de execução do *Snowballing (backward)*, foram verificados os anos de publicação dos artigos na lista de referência para ver se eles atendiam aos critérios previamente definidos. Logo depois, foram realizadas quatro verificações seguindo as recomendações de Wohlin (2014): 1) título verificado na lista de referências; 2) localização de referência verificada; 3) leitura do resumo do estudo referenciado, e 4) verificação das referências completas do estudo referenciado. Três iterações foram realizadas através de uma busca manual. Na primeira iteração, foi avaliado o conjunto inicial proposto com as referências listadas pela RSL. Na segunda iteração, 73 estudos foram indicados para possível inclusão. No entanto, na terceira iteração, apenas um dos 73 estudos teve sua relevância associada ao tema desta tese, e 72 foram excluídos. Portanto, o processo de execução do *Snowballing (backward)* resultou em apenas um novo trabalho de inclusão ao conjunto inicial, e o processo foi concluído;
- c) Passo 3: no processo de execução do *Snowballing (forward)*, foi utilizado o *Google Scholar* como um mecanismo de busca de citação. De acordo com Wohlin (2014), o *Google Scholar* evita um viés na pesquisa, pois indexa as principais bases de pesquisa entre outras bases internacionais renomadas. Verificou-se o ano de publicação dos artigos para verificação dos critérios previamente definidos. Logo depois, foram realizadas quatro verificações seguindo as recomendações de Wohlin (2014): 1) verificou-se o título do

estudo citado; 2) foi realizada a leitura do resumo do estudo; 3) leitura do local de citação realizada; e 4) verificou-se a citação completa do estudo. Assim, foram realizadas três iterações de pesquisa manual para avaliação dos estudos na lista de citações. Na primeira iteração, foram avaliadas citações do conjunto inicial apresentado pela RSL. Na segunda iteração, foram avaliados 37 artigos relacionados ao tema para possível inclusão. Na terceira iteração, observou-se que nos estudos encontrados, as iterações tendem a deixar o tema inicial cada vez mais disperso, e não foram encontradas fontes mais relevantes. Assim, foram avaliados os 37 estudos, sendo observado que dois desses estudos tiveram relevância associada ao tema desta tese. Portanto, 2 estudos foram incluídos no processo de execução do *Snowballing (forward)*, e 35 estudos foram excluídos. Nenhum novo estudo foi encontrado, e o processo de execução do *Snowballing (forward)* foi concluído;

- d) Passo 4: Finalmente, três estudos foram adicionados ao conjunto inicial fornecido pela RSL.

Nessa fase, foram incluídos três estudos por meio do processo de execução de *Snowballing (backward and forward)*, permitindo aprofundar no tema apresentado nesta tese.

3.2 ESTUDOS SELECIONADOS

As próximas subseções descrevem os 14 estudos encontrados no conjunto inicial da RSL e os três estudos encontrados no processo de *Snowballing (backward and forward)*. Assim, os resumos dos 17 estudos selecionados no total, são descritos nas subseções a seguir.

3.2.1 Resumo dos estudos selecionados no conjunto inicial da RSL

No estudo *MediAlly: A Provenance-Aware Remote Health Monitoring Middleware* (CHOWDHURY; FALCHUK; MISRA, 2010) foi proposto o MediAlly, uma arquitetura de *middleware* para apoiar o monitoramento remoto de longo prazo por meio de um smartphone, utilizando técnicas para a coleta sob demanda de proveniência contextual para registrar os dados em um repositório de PHR. O modelo utilizado para a arquitetura de *middleware* foi baseado no modelo híbrido de Proveniência de Dados denominado *Time-Value Centric* (TVC) (Billount, 2007). Além disso, foi adotado um paradigma chamado *Activity Triggered Deep Monitoring* (ATDM), onde os episódios de coleta de dados são acionados apenas quando o assunto é

determinado como tendo um contexto específico. Além de TVC e ATDM, os autores utilizam um algoritmo com técnicas de Proveniência de Dados para middleware. Como resultado, este estudo demonstrou que o protótipo desenvolvido teve boa viabilidade em SIS. A relação desse estudo com o método desta tese, está na utilização do fluxo que descreve a proveniência em torno do processo de coleta de dados.

Em Kovalchuk *et al.* (2013), o estudo intitulado *Provenance-aware pervasive computing in clinical applications*, os autores propõem uma arquitetura de um CRIS, genérico e inteligente em hospitais no uso de técnicas de coleta de dados de proveniência com base nas especificações da W3C PROV. Os autores utilizaram técnicas de coleta de dados de proveniência baseadas nas especificações PROV, PROV-O e PROV-DM. Como resultado, a solução genérica e escalonável apresentada pelos autores, permite um hospital coletar dados de proveniência para validar seus dados clínicos. A relação desse estudo com o método desta tese, está na utilização das especificações do modelo W3C PROV, o qual tem especificações muito semelhantes aqui mencionadas. Ainda, para a validação do método utilizado nesta tese, o cenário de utilização também se assemelha muito ao SIS.

No estudo intitulado *Fuzzy Reasoning of Accident Provenance in Pervasive Healthcare Monitoring Systems* (WANG; HU, 2013), os autores propõem uma metodologia baseada em *Backward Fuzzy Time Reasoning Net* (BFTRN), combinando dados de proveniência semântica e tempo difuso para um HMS. Nesse sentido, a metodologia utilizada não é especificamente de Proveniência de Dados, mas utiliza técnicas inseridas em algoritmos de Proveniência de Dados que apoiam o modelo final criado pelos autores, denominado *Framework of provenance-aware life-critical monitoring system*. Como resultado, o modelo se mostra eficiente quando aplicado em ambientes abrangentes de monitoramento de saúde. A relação desse estudo com o método desta tese, está associada com a forma de contribuição para gerenciamento dos dados de proveniência em SIS.

Em Prasad *et al.* (2013), no estudo intitulado *Provenance framework for mHealth*, os autores propõem uma estrutura de *middleware* para EHR, utilizando técnicas de Proveniência de Dados inseridas no algoritmo criado para dispositivos em *mHealth*, com intuito de coletar e compartilhar metadados de proveniência e ajudar o consumidor de dados a verificar se certas propriedades de proveniência são satisfeitas no recebimento dos dados de saúde de fora do hospital, podendo, assim, compartilhar com outros sistemas de saúde. Como complemento, esse estudo descreve as técnicas utilizadas no modelo de programação para a estrutura proposta pelos autores, descrevendo as regras a serem implementadas para fornecer antecipadamente a coleta de dados de proveniência para aplicativos *mHealth*. Como resultado, o estudo

demonstrou que o protótipo desenvolvido teve boa viabilidade em testes para implantação em aplicativos *mHealth* para SIS. A relação desse estudo com o método desta tese, está associada à utilização do repositório desenvolvido para o armazenamento descentralizado dos dados de proveniência.

No estudo *Mobile-Based Medical Data Accessibility in mHealth* (LOMOTÉY; DETERS, 2014), os autores propõem uma arquitetura de um sistema em nuvem para *mHealth*, utilizando técnicas de Proveniência de Dados incorporadas em um *middleware* denominado *Med App* para auxiliar profissionais de saúde a terem acesso remoto aos dados do EHR. Como resultados, na avaliação do sistema já mencionado, fica claro que a disseminação dos dados médicos pode ser alcançada de forma eficiente e segura. A relação desse estudo com o método aqui proposto nesta tese, está associada à estrutura de construção do armazenamento dos dados de proveniência, e por se tratar da aplicação em um SIS.

Em Seneviratne e Kagal (2014), no estudo intitulado *Enabling Privacy Through Transparency*, os autores apresentam um sistema denominado *Privacy Enabling Transparent Systems* (PETS), que usa a metodologia *Provenance Tracking Network* (PTN) de Proveniência de Dados em sua arquitetura, que é uma rede global aberta confiável de servidores de mesmo nível para a *web*. O sistema apresentado pelos autores (PETS) juntamente com a metodologia de Proveniência de Dados (PTN), permite que os consumidores de dados sejam transparentes em relação ao uso de dados e determinam se há violações de privacidade nesse processo. O estudo é aplicado a um sistema baseado na Web para dados em EHR, entre sistemas hospitalares para envio e recebimento de dados de pacientes para posterior tratamento e análise. Os autores enfatizam o uso do *Health Insurance Portability and Accountability Act* (HIPAA) na EHR nos Estados Unidos. A HIPAA é um conjunto de normas que as organizações de saúde dos Estados Unidos devem cumprir para proteger as informações de saúde contra crimes cibernéticos (ANNAS, 2003; SENEVIRATNE; KAGAL, 2014). Como resultado, as avaliações demonstraram que a arquitetura é promissora no que diz respeito a proteger a privacidade de vazamentos não intencionais de dados confidenciais de saúde. A relação desse estudo com esta tese, está associada ao cenário de aplicação e do repositório de armazenamento dos dados de proveniência, ambos semelhantes a proposta desta tese.

Em Ramesh *et al.* (2015), no estudo intitulado *Insight: Semantic Provenance and Analysis Platform for Multi-center Neurology Healthcare Research*, os autores propõem um sistema chamado Insight, com base em tecnologia da Web Semântica, para apoiar a análise secundária em grande escala de dados de saúde para pesquisa clínica em neurologia em dados no EHR. O sistema utiliza especificações do W3C PROV, especificamente no modelo de

Proveniência de Dados PROV-O e PROV-DM. Como resultado, o estudo aponta o sistema Insight na avaliação de desempenho dos dados e módulos de transformação, carregamento e desempenho de consulta, muito eficazes aplicados em plataformas de saúde. A relação desse estudo com o método nesta tese, está associada ao uso das especificações do W3C PROV e ao cenário de aplicação em saúde mencionado pelos autores.

No estudo denominado *A method of electronic health data quality assessment: Enabling data provenance* (SUN; LU; GU, 2017), os autores propõem um método de coleta de dados para um sistema chinês de PHR em parceria com a *Shanghai Disabled Federation*, com base no modelo de Proveniência de Dados OPM para validar e qualificar os dados eletrônicos de saúde na plataforma utilizada por eles. Como resultado, o protótipo construído teve eficiência na avaliação da qualidade dos dados de saúde utilizando o modelo de Proveniência de Dados mencionado. A relação deste estudo com o método proposto para esta tese, está no uso do OPM, que além de muitas particularidades com o modelo W3C PROV aplicado ao método aqui empregado, está associado à aplicação em SIS.

Em Curcin *et al.* (2017), no estudo intitulado *Templates as a method for implementing data provenance in decision support systems*, os autores relatam sobre o uso de LHS, como, por exemplo, o projeto EU FP7 TRANSFoRm (DELANEY *et al.*, 2015), uma ferramenta de decisão diagnóstica que promove inúmeras práticas de ponta de suporte à decisão clínica. Como complemento, inclui padrões de usabilidade precisamente definidos, integração com um EHR, permitindo recomendações no ponto de atendimento como parte do fluxo de trabalho do médico e uma proveniência *back-end* que captura dados de proveniência sobre os aspectos computacionais, à tarefa de diagnóstico. Os autores mencionam os modelos da W3C PROV, como o PROV-O, PROV-N, além do OPM, são utilizados em sua proposta. Como resultado, a aplicação dos modelos de Proveniência de Dados foi eficaz para a construção do projeto de um LHS, como no caso do EU FP7 TRANSFoRm. Isso significa, que um projeto dessa proporção em utilização de modelos de proveniência é promissor para os SIS. A relação desse estudo com o método desta tese, está no uso dos modelos de Proveniência de Dados de especificação do W3C PROV e o cenário de aplicação em SIS.

No estudo intitulado *Visualizing the provenance of personal data using comics* (SCHREIBER; STRUMINKSI, 2018), os autores propõem uma técnica de visualização de dados de proveniência pessoais de saúde em uma rede de nuvem segura para uso em sistemas de saúde, especificamente em PHR. Essa técnica segue a mesma lógica de histórias em quadrinhos, em que cada tira do quadrinho representa uma determinada atividade, tais como: inserir dados usando um aplicativo de smartphone, armazenar ou recuperar dados em um

serviço em nuvem ou gerar um diagrama a partir dos dados. As histórias em quadrinhos são geradas automaticamente usando gráficos de proveniência registrados. Os autores desenvolveram um aplicativo web denominado PROV-Comics para uso em PHR baseado no PROV, PROV-O, PROV-DM e PROV-N de forma mista. Como resultado, a aplicação dos modelos de Proveniência de Dados se mostrou eficaz ao auxiliar o aplicativo PROV-Comics de forma a apresentar melhor entendimento dos dados de proveniência. A relação desse estudo com o método proposto nesta tese, se refere ao uso dos modelos de proveniência mencionados pelos autores, os quais têm muitas particularidades com o modelo W3C PROV aplicado nesta tese e por sua aplicação ser direcionada à SIS.

Em Gong, Lin e Li (2019), no estudo intitulado *Research on Personal Health Data Provenance and Right Confirmation with Smart Contract*, com o objetivo de resolver o fenômeno de vazamento, abuso e aquisição ilegal de dados de saúde pessoal, os autores propõem um modelo de Proveniência de Dados denominado PROV-Chain, baseado em tecnologias blockchain e no modelo OPM. O PROV-Chain está focado em aplicações em PHR para uso de dispositivos em um cenário de IoT. Como resultados, as avaliações do modelo não só garantem a rastreabilidade de dados pessoais de saúde, mas também identificam o direito de uso pessoal desses, o que garante ainda mais a segurança. A relação desse estudo com o método nesta tese, se refere à técnica para a construção do algoritmo estabelecida para a coleta dos dados de proveniência e pelo cenário de aplicação em saúde.

No estudo intitulado *Prov-IoT: A Security-Aware IoT Provenance Model* (JAIGIRDAR; RUDOLPH; BAIN, 2020), os autores discutem a importância de adicionar metadados de segurança a um gráfico de Proveniência de Dados. Eles propuseram um novo modelo de Proveniência de Dados para IoT, o Prov-IoT para uso em PHR. Esse modelo documenta o histórico de registros de dados considerando o processamento e a agregação de dados, juntamente com metadados de segurança para permitir uma base de confiança nos dados. Com base nos modelos PROV-DM e OPM, o modelo Prov-IoT foi desenvolvido e aplicado em um cenário IoT para PHR para demonstrar o impacto do modelo proposto. Como resultado, o Prov-IoT, na maioria dos casos, não constrói um gráfico de proveniência com reconhecimento de segurança e deixa a avaliação para o usuário final. Em vez disso, divide a documentação de procedência e valida metadados de segurança em etapas. Por outro lado, o usuário final recebe apenas os resultados desta validação juntamente com os segmentos relevantes do gráfico de origem significativo. A relação desse estudo com o método desta tese, está no uso dos modelos de Proveniência de Dados PROV-DM e OPM, os quais têm muitas particularidades com o método abordado nesta tese no uso do W3C PROV, e por sua aplicação ser direcionada a um

cenário de saúde.

No estudo intitulado *Secure and Provenance Enhanced Internet of Health Things Framework: A Blockchain Managed Federated Learning Approach* (RAHMAN *et al.*, 2020), os autores ressaltam que os avanços recentes na Internet of Health Things (IoHT) deram início à ampla adoção de dispositivos IoT em relação ao gerenciamento diário de saúde. Para que os dados da IoHT sejam aceitáveis pelas partes interessadas, os aplicativos que incorporam a IoHT devem ter uma provisão para a Proveniência de Dados, além da precisão, segurança, integridade e qualidade dos dados. Nesse sentido, os autores apresentam um modelo de aprendizagem federada híbrida em que os contratos inteligentes de blockchain gerenciam o plano de treinamento, oportunizando confiança e autenticação de nós dos participantes federados. Para oferecer suporte à total privacidade e anonimato dos dados IoHT, o modelo foi testado com vários aplicativos de aprendizado, projetados para ensaios clínicos com pacientes de COVID-19 em um PHR. Como resultados, o modelo no uso de aplicativos, demonstrou um forte potencial para uma adoção mais ampla do gerenciamento de saúde baseado em IoHT em PHR de maneira segura através de um algoritmo de segurança e privacidade que pudesse ser usado de forma particular nos dispositivos de tecnologia IoHT que exigiam uma garantia de privacidade para dados de saúde de propriedade privada durante o treinamento colaborativo. A relação desse estudo com o método desta tese, se refere à técnica para a construção do algoritmo estabelecida para a coleta dos dados de proveniência e pelo cenário de aplicação em saúde.

No estudo intitulado *Decentralised provenance for healthcare data* (MARGHERI *et al.*, 2020), os autores propõem uma plataforma para gerenciamento do rastreamento de dados de proveniência de registros eletrônicos de saúde com base na tecnologia Blockchain. A plataforma utiliza-se de padrões internacionais de interoperabilidade de dados de saúde como: IHE, HL7 e FHIR para realizar a interoperabilidade e o processo de criação de Proveniência de Dados. A plataforma também utiliza um *Smart Contract* implementado dentro do *Hyperledger Fabric Blockchain*, que gerencia a proveniência de acordo com o modelo W3C PROV para documentos médicos em formatos padronizados, por exemplo (um documento CDA, um recurso FHIR, um estudo DICOM etc.). A plataforma descentralizada proposta pelos autores foi integrada com o motor SpiritEHR, um sistema de saúde de nível empresarial. Como resultado, os autores afirmam que usando uma solução de Blockchain baseada em nuvem, é possível observar que a sobrecarga adicionada ao tempo de processamento de leitura e gravação de dados médicos, é da ordem de milissegundos. Além disso, a integração de um proxy utilizado no nível de troca de mensagens em sistemas EHR permite o uso transparente de dados de proveniência em vários cenários de saúde, auxiliando nas tomadas de decisões, rastreabilidade

dos dados e auditoria com consentimento do paciente. A relação desse estudo com o método proposto nesta tese, se refere à técnica para construção do algoritmo estabelecido para a coleta dos dados de proveniência, os padrões internacionais de interoperabilidade utilizados, o modelo de proveniência utilizado W3C PROV e o cenário de aplicação em SIS. Esse trabalho, de fato, é o que mais se assemelha ao estudo desta tese.

3.2.2 Resumo dos estudos selecionados no Snowballing (backward and forward)

Um único estudo foi encontrado no *Snowballing (backward)*, intitulado *Integrating blockchain for data sharing and collaboration in mobile healthcare applications* (LIANG *et al.*, 2017). Os autores propuseram uma solução inovadora de compartilhamento de dados de saúde centrada no usuário baseada em Blockchain controlada pelo usuário móvel para compartilhamento de dados em nuvem em um PHR. Eles usaram técnicas de coleta de dados de proveniência baseadas em algoritmos usando tecnologias Blockchain. Como resultado, a solução apresentava um algoritmo capaz de gerenciar dados compartilhados de um *mHealth*, preservando a integridade e a privacidade. A solução também pode ser estendida para acomodar o uso de dados de saúde para fins de pesquisa. A relação desse estudo com o método proposta nesta tese, se refere à técnica para construção do algoritmo estabelecido para a coleta dos dados de proveniência, levando em consideração também o cenário de aplicação em SIS.

Por outro lado, no *Snowballing (forward)*, dois estudos foram encontrados. No primeiro estudo intitulado *Practical Extension of Provenance to Healthcare Data Based on the W3C PROV Standard* (KOCK-SCHOPPENHAUER *et al.*, 2018) os autores propuseram uma estratégia para capturar e gerenciar dados interoperáveis de proveniência necessários para os SIS. Avaliaram dados de saúde que são reutilizados em um contexto de pesquisa. O modelo W3C PROV foi utilizado para gerenciar dados de proveniência em um HIS. Como resultado, a estratégia foi implementada em forma de protótipo para demonstrar a viabilidade da abordagem, disponibilizando e utilizando dados de proveniência em um HIS. A relação desse estudo com o método proposto nesta tese, se refere ao uso do modelo de proveniência W3C PROV, o mesmo aplicado nesta tese e por sua aplicação ser direcionada à SIS.

No segundo estudo encontrado no *Snowballing* intitulado *Using PROV and Blockchain to Achieve Health Data Provenance* (MASSI *et al.*, 2018) os autores propõem resolver o problema do gerenciamento de dados de proveniência de saúde em EHR baseados em tecnologias de Blockchain descentralizadas com base no modelo W3C PROV. Como resultado, a proposta apresentada pelos autores baseou-se nas normas internacionais existentes que

garantem a interoperabilidade entre os dados de saúde. O uso de sistemas abertos e tecnologia Blockchain com base no modelo W3C PROV promoveu a segurança e a imutabilidade dos registros de saúde no EHR. A relação desse estudo com o método proposto nesta tese, se refere ao uso do modelo de proveniência W3C PROV, sendo similar ao método proposto nesta tese e por sua aplicação ser direcionada à SIS.

3.2.3 Considerações sobre os estudos selecionados

Dos 17 estudos encontrados nesta RSL, todos tem particularidades e utilizam diferentes modelos de proveniência para tentar solucionar os problemas encontrados. Porém, em relação à problemática desta tese, nenhum dos estudos encontrados nesta RSL apresenta uma estrutura 100% semelhante ao que foi desenvolvido, nem tão pouco o cenário específico de execução. Ainda, tais trabalhos não apresentam propostas que se assemelham à pesquisa desta tese, tampouco contemplam abordagens metodológicas que estruturam a Proveniência de Dados de forma genérica para uso em aplicações específicas. Apesar do modelo W3C PROV ter sido mencionado em alguns estudos, nenhum deles apresentou o instanciamento desse modelo para um domínio de saúde para criação de um método capaz de gerenciar dados de proveniência para um domínio específico de SIS. Com isso, ressalta-se que poucas são as iniciativas de desenvolvimento de métodos para o gerenciamento de dados de proveniência voltados para SIS descritas na literatura, tornando esta tese com um diferencial relevante nesse contexto.

O método desenvolvido nesta tese, vem ao encontro estender o modelo W3C PROV direcionando-o à construção de um método que seja capaz de contribuir para o gerenciamento de dados de proveniência em um domínio desafiador, seja pelos processos de interoperabilidade e manipulação do grande volume de dados existentes nos SIS.

Ressalta-se que, atualmente, a medicina avança exponencialmente, principalmente no que diz respeito aos SIS, e esta tese traz uma proposta inovadora ainda não encontrada na literatura, como foi descrita na RSL realizada. Portanto, a logística e arquitetura aqui apresentadas em relação ao desenvolvimento do método, diferenciam-se dos estudos encontrados na RSL, o que de fato pode contribuir e forma significativa para a prática médica dentro dos SIS.

Porém, vale ressaltar que desafios no gerenciamento dos grandes volumes de dados nos SIS onde são processados e compartilhadas informações sigilosas de pacientes ainda vão continuar a existir, sendo que o método proposto nesta tese é genérico e pode ser ajustado para contribuir na solução de problemas específicos de cada SIS.

Por fim, ainda é importante ressaltar que, em relação à problemática desta tese e sua temática em geral, não foram encontrados estudos dentro da Ciência da Informação na execução da RSL, sendo considerado outro ponto de destaque que impulsiona esta tese dentro da interdisciplinaridade entre a Ciência da Informação e a Ciência da Computação a ser um estudo que contribui com os estudos no contexto do Brasil.

3.3 RESULTADOS DA RSL

A partir desta seção, os 17 estudos selecionados farão parte da base de análise para gerar os resultados desta RSL. Nesse sentido, para melhor visualização dos estudos selecionados, o Quadro 14 apresenta os autores, ano de publicação, título do estudo e os tipos de publicações, sendo esses dois últimos descritos em inglês. Importante ressaltar que, para todos os resultados apresentados nesta seção, muitos termos foram mantidos em inglês por se tratar de estudos internacionais.

Quadro 14 – Lista final dos 17 estudos selecionados

| N. | Autores | Ano | Título | Tipo de publicação |
|----|----------------------------------|------|--|--------------------|
| 1 | Margheri <i>et al.</i> | 2020 | Decentralised provenance for healthcare data | Journal Article |
| 2 | Rahman <i>et al.</i> | 2020 | Secure and Provenance Enhanced Internet of Health Things Framework: A Blockchain Managed Federated Learning Approach | Journal Article |
| 3 | Jaigirdar, Rudolph e Bain | 2020 | Prov-IoT: A Security-Aware IoT Provenance Model | Conference Paper |
| 4 | Gong, Lin e Li | 2019 | Research on Personal Health Data Provenance and Right Confirmation with Smart Contract | Conference Paper |
| 5 | Kock-Schoppenhauer <i>et al.</i> | 2018 | Practical Extension of Provenance to Healthcare Data Based on the W3C PROV Standard | Article e-book |
| 6 | Massi <i>et al.</i> | 2018 | Using PROV and Blockchain to Achieve Health Data Provenance | Technical Report |
| 7 | Schreiber e Struminski | 2018 | Visualizing the provenance of personal data using comics | Journal Article |
| 8 | Liang <i>et al.</i> | 2017 | Integrating blockchain for data sharing and collaboration in mobile healthcare applications. | Conference Paper |
| 9 | Curcin <i>et al.</i> | 2017 | Templates as a method for implementing data provenance in decision support systems | Journal Article |

| | | | | |
|----|----------------------------|------|---|------------------|
| 10 | Sun, Lu e Gu | 2017 | A method of electronic health data quality assessment: Enabling data provenance | Conference Paper |
| 11 | Ramesh <i>et al.</i> | 2015 | Insight: Semantic provenance and analysis platform for multi-center neurology healthcare research | Conference Paper |
| 12 | Seneviratne e Kagal | 2014 | Enabling privacy through transparency | Conference Paper |
| 13 | Lomotey e Deters | 2014 | Mobile-Based Medical Data Accessibility in mHealth | Conference Paper |
| 14 | Prasad <i>et al.</i> | 2013 | Provenance framework for mHealth | Conference Paper |
| 15 | Wang e Hu. | 2013 | Fuzzy reasoning of accident provenance in pervasive healthcare monitoring systems | Journal Article |
| 16 | Kovalchuk <i>et al.</i> | 2013 | Provenance-aware pervasive computing in clinical applications | Conference Paper |
| 17 | Chowdhury, Falchuk e Misra | 2010 | MediAlly: A provenance-aware remote health monitoring middleware | Conference Paper |

Fonte: Elaborado pelo autor (2021).

Com base no Quadro 14, em relação à avaliação da qualidade dos 17 estudos selecionados nesta RSL, percebeu-se que todos os estudos avaliados assumiram e por unanimidade a escala “2” para os cinco critérios de qualidade (QC01, QC02, QC03, QC04, QC05) com base (Quadros 9 e 10). Após a escala de Likert-3 ser adaptada para cada critério de qualidade apresentada no Quadro 9, foram analisados os níveis de qualidade, conforme proposto por Beecham *et al.* (2008). Os níveis de qualidade podem ser observados na Tabela 1.

Tabela 1 – Níveis de qualidade dos estudos

| Baixa < 26% | Média 26%-45% | Bom 46%-65% | Muito Bom 66%-85% | Excelente >86% | TOTAL |
|----------------|------------------|----------------|----------------------|-------------------|-------|
| 0 | 0 | 0 | 5 | 12 | 17 |
| ~0% | ~0% | ~0% | ~29,4% | ~70,6% | 100% |

Fonte: Adaptado de Beecham *et al.* (2008).

Assim, com base na Tabela 1, foi possível observar que cinco estudos (29,4%) se apresentaram como “muito bons”, satisfazendo os critérios de qualidade de forma positiva. Entretanto, 12 estudos (70,6%) foram considerados “excelentes”, o que mostra que a seleção de estudos nesta RSL teve um alto nível de estudos com qualidade significativa para o tema relacionado ao gerenciamento dos dados de proveniência dos SIS para a tese aqui em questão. Os percentuais positivos apresentados e analisados pela escala Likert-3 apresentada no Quadro 11, mostraram-se favoráveis à qualificação dos estudos selecionados, contribuindo para a

valorização do conteúdo apresentado pelos autores dos estudos.

Isso mostra que a qualidade dos estudos descritos nesta tese, tanto nos estudos resultantes da RSL quanto nos estudos resultantes da técnica de *Snowballing (backward and forward)*, contribuem para pesquisas significativas sobre o tema aqui em questão, sendo possível obter resultados de boa qualidade também. Porém, realizando uma análise mais ampla para sintetizar esses 17 estudos resultantes, foram respondidas as questões de pesquisa desta RSL nas próximas subseções.

3.3.1 Respostas para as Questões de Pesquisa da RSL

Nesta subseção são apresentadas as respostas tanto da questão geral, quanto das questões específicas desta RSL.

QG01 - Quais são os diferentes métodos, técnicas, modelos e metodologias utilizados para o gerenciamento de dados de proveniência em SIS?

Para responder a essa questão, no Quadro 15 são sintetizadas algumas informações como: autores, anos, tipos de SIS/cenário de saúde, método/técnica/modelo/metodologia para gerenciamento da Proveniência de Dados no contexto de sistemas de informação em saúde.

Quadro 15 – Cenários/aplicações, métodos, técnicas, modelos e metodologias

| N. | Autores | Ano | Tipos de SIS- Cenários de Saúde | Método/técnica/modelo/metodologia |
|----|----------------------------------|------|---------------------------------------|--|
| 1 | Margheri <i>et al.</i> | 2020 | EHR-Healthcare | PROV e proveniência de dados incorporada ao (blockchain) |
| 2 | Rahman <i>et al.</i> | 2020 | PHR-IoHT | Proveniência de dados incorporada ao (blockchain) |
| 3 | Jaigirdar, Rudolph e Bain | 2020 | PHR-IoHT | PROV-IoT baseado no PROV-DM e OPM |
| 4 | Gong, Lin e Li | 2019 | PHR / IoT | PROV-Chain com base no OPM e proveniência de dados incorporada ao (blockchain) |
| 5 | Kock-Schoppenhauer <i>et al.</i> | 2018 | HIS-Healthcare | PROV |
| 6 | Massi <i>et al.</i> | 2018 | EHR-Healthcare | PROV e proveniência de dados incorporada ao (blockchain) |
| 7 | Schreiber e Struminski | 2018 | PHR-Health Monitoring | PROV-Comics baseado nos modelos PROV, PROV-DM, PROV-O, |

| | | | | |
|----|----------------------------|------|-----------------------|---|
| | | | | PROV-N |
| 8 | Liang <i>et al.</i> | 2017 | PHR- mHealth | Proveniência de dados incorporada ao (blockchain) |
| 9 | Curcin <i>et al.</i> | 2017 | LHS-Healthcare | PROV, PROV-O, PROV-N, OPM |
| 10 | Sun, Lu e Gu | 2017 | PHR-Healthcare | OPM |
| 11 | Ramesh <i>et al.</i> | 2015 | EHR-Healthcare | PROV, PROV-O, PROV-DM |
| 12 | Seneviratne e Kagal | 2014 | EHR-Healthcare | PTN baseado no modelo PROV-O |
| 13 | Lomotey e Deters | 2014 | EHR-mHealth | Algoritmo com técnicas de proveniência de dados para middleware |
| 14 | Prasad <i>et al.</i> | 2013 | EHR-mHealth | Algoritmo com técnicas de proveniência de dados para middleware |
| 15 | Wang e Hu. | 2013 | HMS-Health Monitoring | BFTRN |
| 16 | Kovalchuk <i>et al.</i> | 2013 | CRIS-Healthcare | PROV, PROV-O, PROV-DM |
| 17 | Chowdhury, Falchuk e Misra | 2010 | PHR-Health Monitoring | TVC, ATDM e algoritmo com técnicas de proveniência de dados para middleware |

Fonte: Elaborado pelo autor (2021).

Pode-se observar no Quadro 15, que diferentes métodos, técnicas, modelos e metodologias de gerenciamento de dados de proveniência foram apresentados nos estudos selecionados. Destaca-se os modelos de Proveniência de Dados recomendados pela W3C implementados por meio de ferramentas computacionais em diferentes tipos de SIS, os quais aparecem com mais frequência nos estudos selecionados, sendo eles: PROV, PROV-O, OPM, PROV-DM e PROV-N. Dessa forma, observou-se que a maioria dos estudos (sete), os autores (KOVALCHUK *et al.*, 2013; RAMESH *et al.*, 2015; CURCIN *et al.*, 2017; SCHREIBER; STRUMINKSI, 2018; KOCK-SCHOPPENHAUER *et al.*, 2018; MASSI *et al.*, 2018; MARGHERI *et al.*, 2020) utilizam o modelo PROV combinados com outros modelos, aplicados em (EHR, PHR, LHS, CRIS e HIS). Nesse sentido, evidencia-se que o modelo PROV é, de fato, um dos modelos que mais frequentemente contribuem para o gerenciamento dos dados de proveniência em SIS. Esse fato, concretiza ainda mais o uso do modelo PROV na construção do método desta tese, alicerçado na literatura comprovada nesta RSL, apontando, dessa forma, tendências de seu uso especificamente em SIS. Já para o modelo PROV-O, os cinco estudos (KOVALCHUK *et al.*, 2013; SENEVIRATNE; KAGAL, 2014; RAMESH *et al.*, 2015; CURCIN *et al.*, 2017; SCHREIBER; STRUMINKSI, 2018) utilizam esse modelo de forma combinada com outros modelos apresentando aplicações em (EHR, PHR, LHS e CRIS). No que compete ao uso do modelo OPM, os quatro estudos (SUN; LU; GU, 2017; CURCIN *et*

al., 2017, GONG; LIN; LI, 2019; JAIGIRDAR; RUDOLPH; BAIN, 2020) apresentam aplicações desse modelo em (PHR e LHS). Já o modelo PROV-DM, quatro estudos (KOVALCHUK *et al.*, 2013; RAMESH *et al.*, 2015; SCHREIBER; STRUMINKSI, 2018; JAIGIRDAR; RUDOLPH; BAIN, 2020) apresentam aplicações em (HER, PHR e CRIS). E para o uso do modelo PROV-N, dois estudos (CURCIN *et al.*, 2017; SCHREIBER; STRUMINKSI, 2018) apresentam aplicações em (PHR e LHS). Importante destacar que, esses modelos aparecem com mais frequência por serem específicos de aplicações de Proveniência de Dados recomendados pelo W3C, capazes de descrever as entidades e processos envolvidos na produção de um recurso. Nesse sentido, pode-se evidenciar que, a utilização dos modelos de proveniência pode contribuir significativamente para criação de métodos para coletar dados de proveniência em saúde com qualidade em seus processos de análises, desde a origem dos dados conforme apresentados nos estudos encontrados na literatura.

Ainda com base no Quadro 16, os estudos também apresentaram diferentes modelos de Proveniência de Dados indicados pelo W3C, combinados ou não, com diferentes métodos, técnicas, metodologias e tecnologias que possibilitaram o gerenciamento de dados de proveniência em SIS.

Quatro estudos apresentaram o gerenciamento dos dados de proveniência combinando com tecnologias Blockchain (LIANG *et al.*, 2017; GONG; LIN; LI, 2019; RAHMAN *et al.*, 2020; MARGHERI *et al.*, 2020). O uso combinado dessas duas tecnologias promove, além da segurança dos dados, uma forma de garantir a proveniência de dados por meio da imutabilidade dos dados. Assim, o uso de tecnologias blockchain para a imutabilidade de dados e o uso de contratos inteligentes como ferramenta tecnológica que permite a criação de contratos auto executados que não podem ser perdidos ou adulterados, são vistos nestes estudos. No estudo de (LIANG *et al.*, 2017), os autores usam a tecnologia blockchain para preservar a integridade dos dados de saúde, validando e recuperando permanentemente esses dados no banco de dados de PHR em nuvem na rede blockchain. O estudo (GONG; LIN; LI, 2019) apresenta um modelo de Proveniência de Dados chamado PROV-Chain, baseado em tecnologias blockchain e no modelo OPM para uso em um PHR. Já no estudo de (RAHMAN *et al.*, 2020), os autores apresentam a Proveniência de Dados incorporada ao Blockchain em um PHR. Para o estudo de (MARGHERI *et al.*, 2020), os autores apresentam o gerenciamento dos dados de proveniência incorporados ao Blockchain juntamente com o modelo PROV em um EHR. Portanto, é importante ressaltar que, a gestão dos dados de proveniência mediada pelas tecnologias blockchain é uma aliada para a segurança de dados em cenários de SIS. Dessa forma, consequentemente, promovem o rastreamento seguro de dados e garantem a imutabilidade

desses dados nos SIS. Outros dois outros estudos se destacaram por apresentar algoritmos com técnicas de aplicação de Proveniência de Dados utilizando tecnologias de middleware (PRASAD *et al.*, 2013; LOMOTEY; DETERS, 2014) com aplicações em EHR e PHR. O Middleware permite compartilhar os metadados de origem com o consumidor de dados, ou seja, qualquer informação contextual que possa atestar a autenticidade e exatidão dos dados e auxiliar na interpretação desses dados. Já no estudo dos autores (CHOWDHURY; FALCHUK; MISRA, 2010), além de utilizar o algoritmo com técnicas de aplicação de Proveniência de Dados utilizando tecnologia middleware, combinou o uso do modelo de Proveniência de Dados híbrido denominado TVC e ATDM obtendo sucesso para aplicações em PHR. Para no estudo de (JAIGIRDAR; RUDOLPH; BAIN, 2020), os autores apresentam um modelo de Proveniência de Dados denominado (PROV-IoT), baseado nos modelos PROV-DM e OPM para aplicação em PHR. Esse modelo documenta o histórico de registros de dados permitindo uma base de confiança nos dados de origem. Assim, no modelo PROV-Comics é apresentado no estudo (SCHREIBER; STRUMINKSI, 2018), baseado nos modelos PROV, PROV-DM, PROV-O e PROV-N para aplicação em PHR. PROV-Comics tem como objetivo visualizar dados de fontes de saúde pessoal através de quadrinhos. Já em relação ao PTN apresentado no estudo de (SENEVIRATNE; KAGAL, 2014) é baseado no modelo PROV-O. O PTN é uma metodologia de rede de confiança global aberta para servidores web aplicado em EHR. No que se refere à Proveniência de Dados no modelo BFTRN apresentada no estudo (WANG; HU, 2013) é aplicada juntamente com um algoritmo de raciocínio retroativo motivado pela teoria da automação do tempo com base nesse modelo. Dessa forma, ele lista alarmes críticos aplicados em um HMS. A procedência dos dados nesse modelo está associada aos dados capturados em relação aos alarmes críticos emitidos pelo sistema de monitoramento da saúde.

Portanto, é possível observar no Quadro 16, que o gerenciamento de dados de proveniência em SIS, não se limita ao uso de métodos, técnicas, modelos e metodologias específicas, mas sim, ao uso de combinações híbridas de diferentes tecnologias computacionais para alcançar o sucesso esperado.

QG02 - Quais são os desafios em relação aos diferentes métodos, técnicas, modelos e metodologias identificados a respeito do gerenciamento de dados de proveniência em SIS?

Após a leitura dos 17 estudos selecionados, foi possível observar que os autores recomendam atenção a questões relacionadas a inconsistências, vazamentos e segurança de dados que podem ocorrer nos SIS, mesmo sem intervenção do usuário. Essas questões

apontadas pelos autores são consideradas por eles como desafios ainda pendentes para tornar os dados de proveniência mais seguros e confiáveis nos SIS. Conseqüentemente, isso ajuda a garantir o armazenamento seguro, publicação e preservação desses dados.

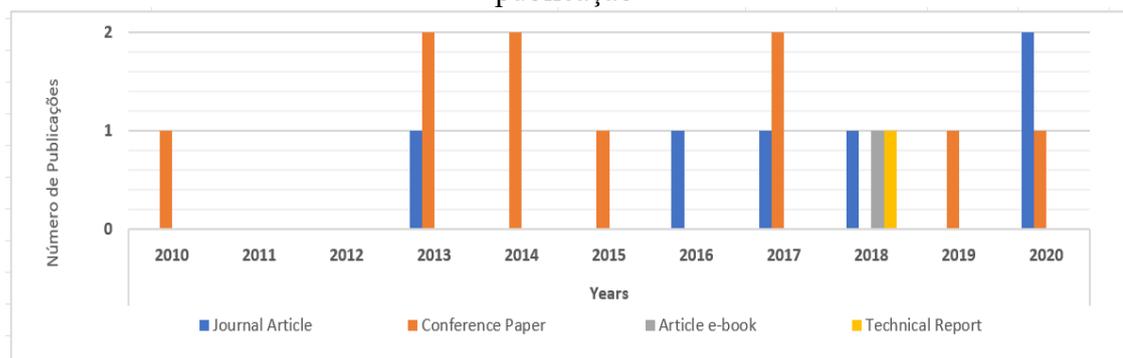
Foi observado também, que não há uma estrutura comum que possa garantir total segurança quanto ao gerenciamento de dados de proveniência em SIS. Um exemplo, é a criação e troca de registros pessoais de saúde distribuídos em múltiplos SIS que apresentam desafios técnicos e clínicos que podem colocar em risco a segurança do paciente. Porém, os atuais sistemas de compartilhamento de informações de saúde garantem a interoperabilidade dos prontuários dos pacientes nas instituições de saúde. No entanto, eles têm limites para apresentar aos profissionais de saúde o contexto clínico dos dados de prontuários médicos. Portanto, existem diversos desafios técnicos para a implementação da segurança dos dados de proveniência dos SIS conforme apresentado na literatura.

De fato, esses foram os principais fatores observados nos 17 estudos selecionados que podem promover o sucesso ou a falha da qualidade dos dados, seja no fornecimento de metadados ou em questões de interoperabilidade, privacidade e confidencialidade desses dados nos SIS.

QE01 - Levando em conta os aspectos mais representativos quanto ao gerenciamento de dados de proveniência em SIS, como esses sistemas abordam os diferentes métodos, técnicas, modelos e metodologias?

Para responder essa questão, pode-se observar *a priori*, o período de abordagem apresentado pelos estudos no quesito de uso sobre os diferentes métodos, modelos, técnicas e metodologias para o gerenciamento dos dados de proveniência dos SIS apresentado no Gráfico 1.

Gráfico 1 – Relação do número de estudos encontrados de acordo com o ano e tipo de publicação

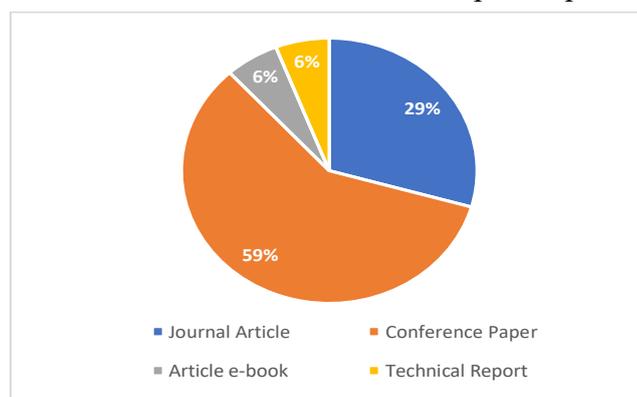


Fonte: Elaborado pelo autor (2021).

Observa-se inicialmente no Gráfico 1, que o número de publicações a cada ano varia de 1, 2 ou 3 estudos, sendo 2013, 2017, 2018 e 2020 os anos mais produtivos. Evidencia-se que isso pode estar ligado ao grande avanço das conferências internacionais no campo da ciência da computação sobre aspectos relacionados às TIC em aplicações de gerenciamento de dados de proveniência em SIS.

Nesse sentido, dos estudos selecionados mais da metade são publicados em conferências. Logo, para afirmar esse fato, no Gráfico 2 observa-se os tipos de publicações associadas a *Journal Article*, *Conference Paper*, *Article e-book* e *Technical Report*.

Gráfico 2 – Percentuais referentes aos tipos de publicações



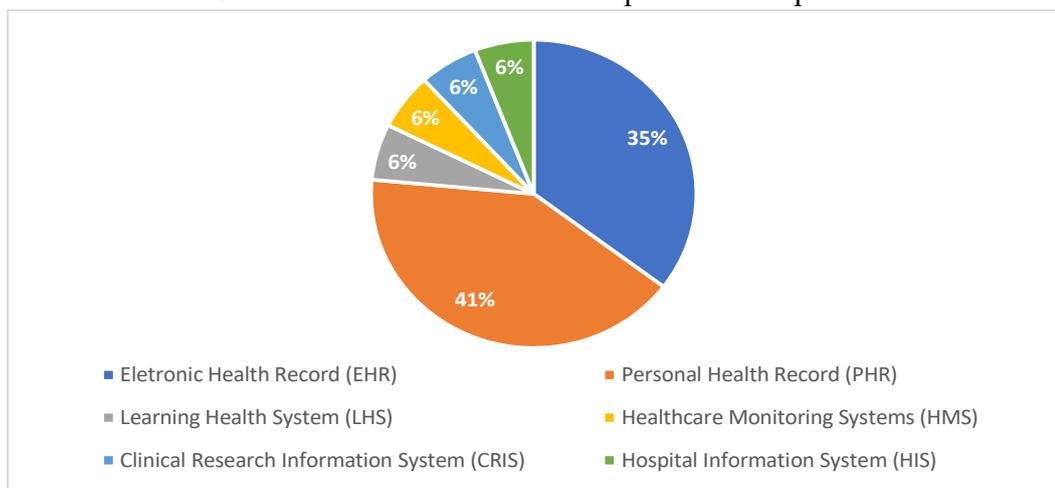
Fonte: Elaborado pelo autor (2021).

O Gráfico 2 mostra que a maioria dos estudos (59%) foi publicada em *Conference Paper*, enquanto (29%) dos estudos foram publicados em *Journal Article*. Também foram observadas publicações de *Article e-book* (6%) e *Technical Report* (6%). Na área da saúde, isso acontece porque os eventos científicos (conferências), são fontes de informações relevantes para o compartilhamento de conhecimento, permitindo a disseminação de evidências científicas, boas práticas e novas tecnologias que contribuem para o aprimoramento e

atualização da pesquisa no setor. Além disso, o maior número de publicações na área da computação está ligado a estudos empíricos publicados em conferências internacionais, que geralmente têm aplicações computacionais práticas demonstradas em estudos de caso. São temas de grande interesse para o uso dessas aplicações práticas por empresas de tecnologia nos mais variados cenários de saúde. Isso indica um dos fatores que contribuíram para o gerenciamento dos dados de proveniência em relação aos aspectos representativos e técnicos que despertaram interesse no uso dos SIS.

Também foram observados os tipos de SIS que gerenciam os dados de proveniência por meio de métodos, técnicas, modelos, metodologias e tecnologias de ferramentas computacionais. Os estudos selecionados mostram que, a partir de 2010, grandes volumes de dados começaram a ser gerados nos SIS. No Gráfico 3, são apresentados os percentuais referentes à frequência dos principais tipos de SIS encontrados nos estudos selecionados, os quais são: EHR, PHR, LHS, HMS, CRIS e HIS.

Gráfico 3 – Percentuais referentes à frequência dos tipos de SIS



Fonte: Elaborado pelo autor (2021).

Como apresentado no Gráfico 3, as maiores frequências de aplicação do gerenciamento de dados de proveniência dos SIS estão presentes em EHR e PHR, ambos com (41%) e (35%), respectivamente. Isso se deve ao fato de serem esses SIS (EHR e PHR) os mais difundidos entre as instituições de saúde atualmente em nível mundial. Em seguida aparece o LHS (6%), HMS (6%), CRIS (6%) e HIS (6%). Considerando a leitura dos 17 estudos selecionados, esses SIS foram considerados os mais aplicáveis em relação ao gerenciamento de dados de proveniência em diferentes cenários de saúde nos diferentes contextos encontrados. É importante ressaltar que os tipos de SIS mostrados no Gráfico 3 possuem características diferentes, conforme já

detalhado nas seções anteriores. Além disso, com base nas informações do Gráfico 3, os autores que mencionaram os tipos de SIS em seus estudos em relação ao gerenciamento de dados de proveniência são: EHR (PRASAD *et al.*, 2013; LOMOTEY; DETERS, 2014; SENEVIRATNE; KAGAL, 2014; RAMESH *et al.*, 2015; MASSI *et al.*, 2018; MARGHERI *et al.*, 2020), PHR (CHOWDHURY; FALCHUK; MISRA, 2010; SUN; LU; GU, 2017; LIANG *et al.*, 2017; SCHREIBER; STRUMINKSI, 2018; GONG; LIN; LI, 2019; JAIGIRDAR; RUDOLPH; BAIN, 2020; RAHMAN *et al.*, 2020), LHS (CURCIN *et al.*, 2017), HMS (WANG; HU, 2013), CRIS (KOVALCHUK *et al.*, 2013) e HIS (KOCK-SCHOPPENHAUER *et al.*, 2018).

Outro ponto importante a ser enfatizado, é que foi possível observar nos estudos selecionados o crescimento exponencial a partir de 2010 em relação aos grandes volumes de dados gerados nos SIS, o que de fato, é uma tendência global conforme comprovada por meio da literatura. Observou-se também, que foi necessário o uso de diferentes métodos, técnicas, modelos e metodologias combinadas com tecnologias computacionais para gerenciar os dados de proveniência em diferentes SIS. Com isso, uma das tendências observadas, foi o uso de ferramentas computacionais para o gerenciamento dos dados de proveniência dos SIS considerando cenários remotos de monitoramento de saúde. Dispositivos móveis ou vestíveis aliados aos conceitos de IoT empregados na área da saúde representam alguns desses cenários. Isso mostra uma vertente evolutiva global na prestação de serviços de saúde em diferentes ambientes de saúde e em diferentes SIS para os próximos anos. Portanto, pode-se evidenciar que o gerenciamento de dados de proveniência dos SIS utiliza e combina diferentes ferramentas computacionais, criando uma estratégia híbrida para atender às necessidades de gerenciar grandes volumes de dados que contribuem para as tomadas de decisões nesses sistemas.

QE02 - Quais são as principais vantagens da aplicação de diferentes métodos, técnicas, modelos ou metodologias para o gerenciamento de dados de proveniência em SIS?

Em todos os artigos selecionados, foram descritas de modo geral no Quadro 16, todas as principais vantagens observadas em relação ao gerenciamento dos dados de proveniência em SIS.

Quadro 16 – Principais vantagens do gerenciamento dos dados de proveniência dos SIS

| Principais vantagens do gerenciamento dos dados de proveniência em SIS | Descrição das vantagens |
|--|--|
| | Integridade mantida dos objetos digitais em termos de onde vieram, como foram obtidos, como chegaram ao seu estado atual, quem ou o que agiu sobre eles, gerando assim uma maior fonte de confiança; |
| | Vantagens associadas à auditabilidade, transparência, avaliação, disponibilidade, qualidade e correção dos dados, gerando confiança na interoperabilidade entre SIS; |
| | Aumento na visibilidade da fonte de dados de saúde e suas transformações em seu ciclo de vida; |
| | Aumento da confiança das instituições de saúde na autenticidade e confidencialidade dos dados compartilhados; |
| | Melhor compreensão das fontes de grandes volumes de dados de saúde; |
| | Possibilidades de utilização de dados eletrônicos originais de saúde; |
| | Recuperação de novas fontes de dados de saúde; |
| | Importância do gerenciamento dos dados de proveniência para a tomada de decisões importantes nesses sistemas. |

Fonte: Elaborado pelo autor (2021).

Porém, com base nas vantagens descritas no Quadro 16, é importante destacar que o gerenciamento dos dados de proveniência em SIS, contribui para que os profissionais de saúde possam conhecer melhor as fontes originais dos dados de saúde os quais estão sendo manipulados, evitando transtornos e comprometimento da saúde dos pacientes. Portanto, conforme a literatura, as vantagens de gerenciar dados de proveniência em SIS, apresenta benefícios positivos que vão além do rastreamento dos dados de origem e, conseqüentemente, oportuniza um melhor atendimento aos pacientes nas instituições de saúde.

QE03 - Quais são as principais tecnologias identificadas nos diferentes métodos, técnicas, modelos ou metodologias que contribuíram para o gerenciamento de dados de proveniência em SIS?

Existem uma série de tecnologias que podem ser direcionadas e combinadas para o gerenciamento dos dados de proveniência dos SIS conforme apresentado nos estudos selecionados. As principais tecnologias identificadas e os respectivos autores que as relacionam em seus estudos, foram descritos no Quadro 17.

Quadro 17 – Tecnologias que contribuem para o gerenciamento dos dados de proveniência dos SIS

| Tecnologias | Autores |
|--|---|
| O uso de ETL, ferramentas de software cuja função é extrair dados de diferentes sistemas, transformar esses dados de acordo com as regras de negócios e finalmente carregar os dados. | Ramesh <i>et al.</i> (2015) Curcin <i>et al.</i> (2017) |
| Dispositivos móveis para coleta de dados de saúde (smartphone, tablets, notebooks e sensores). | Kovalchuk <i>et al.</i> (2013) Wang; Hu (2013) Prasad <i>et al.</i> (2013) Lomotey e Deters (2014) Schreiber e Struminski (2018) Rahman <i>et al.</i> (2020) |
| Monitoramento de sinais vitais à distância por meio de Personal Digital Assistants (PDA) | Chowdhury, Falchuk e Misra (2010) |
| Uso de Extensible Markup Language (XML) recomendada pela W3C para gerar linguagens de marcação para necessidades especiais. | Chowdhury, Falchuk e Misra (2010) Lomotey e Deters (2014) Kock-Schoppenhauer <i>et al.</i> (2018) Massi <i>et al.</i> (2018) Jaigirdar, Rudolph e Bain (2020) |
| Uso de Ontology Web Language (OWL), a qual é uma linguagem para definir e instanciar ontologias na World Wide Web. Resource Description Framework (RDF) a qual é planejada como um modelo de dados para metadados foi apresentada nos estudos. | Kovalchuk <i>et al.</i> (2013) Ramesh <i>et al.</i> (2015) Curcin <i>et al.</i> (2017) Kock-Schoppenhauer <i>et al.</i> (2018) |
| Utilização de SPARQL, o qual é um acrônimo recursivo de SPARQL Protocol and RDF Query Language. Destina-se a uma linguagem padronizada para a consulta de grafos RDF. | Kovalchuk <i>et al.</i> (2013) Curcin <i>et al.</i> (2017) Liang <i>et al.</i> (2017) |
| Estruturas de Cloud Computing para armazenamento de dados de saúde e uso da capacidade de computação como forma de garantir a quantidade de recursos necessários para que as operações ocorram sem erros ou gargalos. | Chowdhury, Falchuk e Misra (2010) Lomotey e Deters (2014) Schreiber e Struminski (2018) Gong, Lin and Li (2019) Jaigirdar, Rudolph e Bain (2020) Rahman <i>et al.</i> (2020) Margheri <i>et al.</i> (2020) |
| Redes privadas para monitoramento de dados de pacientes, as quais contribuem para a segurança e o não vazamento de dados monitorados de pacientes. | Chowdhury, Falchuk e Misra (2010) Kovalchuk <i>et al.</i> (2013) Wang; Hu (2013) Lomotey e Deters (2014) Seneviratne and Kagal (2014) Sun, Lu and Gu (2017) Liang <i>et al.</i> (2017) Gong, Lin and Li (2019) Jaigirdar, Rudolph e Bain (2020) Rahman <i>et al.</i> (2020) Margheri <i>et al.</i> (2020) |
| Gerenciador de banco de dados MySQL | Chowdhury, Falchuk e Misra (2010) |

| | |
|--|---|
| Gerenciador de banco de dados Neo4j | Curcin <i>et al.</i> (2017) Schreiber e Struminski (2018) |
| Uso de Cypher query language, a qual é uma linguagem oficial de consultas do Neo4j, podendo criar, modificar e procurar os dados em uma estrutura baseada em um conjunto de grafos e seus relacionamentos. | Curcin <i>et al.</i> (2017) |
| Utilização de Digital Imaging and Communications in Medicine (DICOM), o qual é um dos principais conjuntos de normas para o tratamento, armazenamento e transmissão de informações médicas em formato eletrônico, estruturando um protocolo. Uso de Clinical Document Architecture (CDA) como um padrão de marcação baseado em XML destinado a especificar a codificação, estrutura e semântica dos documentos clínicos para troca. Utilização de Integrating the Healthcare Enterprise (IHE). IHE é uma organização internacional que envolve profissionais de saúde e representantes da indústria para juntos procurar melhores alternativas de compartilhamento de saúde. Uso de Cross Enterprise Document Sharing (XDS) sendo um sistema de padrões para catalogar e compartilhar registros de pacientes entre instituições de saúde. | Massi <i>et al.</i> (2018) Margheri <i>et al.</i> (2020) |
| Character Separated Values (CSV) sendo um arquivo de texto delimitado que usa uma vírgula para separar valores. | Schreiber e Struminski (2018) |
| Continuity of Care Documents (CCD) como padrão de marcação baseado em XML destinado a especificar a codificação, estrutura e semântica de um documento clínico de resumo do paciente para troca. | Lomotey e Deters (2014) |
| JavaScript Object Notation (JSON) o qual é um formato compacto, de padrão aberto independente, de troca de dados simples e rápida entre sistemas. | Curcin <i>et al.</i> (2017) Massi <i>et al.</i> (2018) Margheri <i>et al.</i> (2020) |
| Blockchain utilizado para garantir a imutabilidade dos dados de saúde. | Liang <i>et al.</i> (2017) Massi <i>et al.</i> (2018) Gong, Lin and Li (2019) Rahman <i>et al.</i> (2020) Margheri <i>et al.</i> (2020) |
| Middleware como forma de fornecer serviços para softwares aplicativos além daqueles disponíveis pelo sistema operacional. | Chowdhury, Falchuk e Misra (2010) Prasad <i>et al.</i> (2013) Lomotey e Deters (2014) |
| Health Insurance Portability and Accountability Act (HIPAA) é a legislação dos Estados Unidos que fornece privacidade de dados e disposições de segurança para proteger informações médicas. | Seneviratne and Kagal (2014) |
| Health Level Seven International (HL7) sendo um conjunto de normas internacionais para a representação e a transferência de dados clínicos e administrativos. E | Massi <i>et al.</i> (2018) Kock-Schoppenhauer <i>et al.</i> (2018) Margheri <i>et al.</i> (2020) |

| | |
|--|--|
| Fast Healthcare Interoperability Resources (FHIR) como padrão para descrever formatos e elementos de dados para uma interface de aplicação para troca de registros eletrônicos de saúde. | |
|--|--|

Fonte: Elaborado pelo autor (2021).

Segundo os autores do Quadro 17, essas tecnologias podem ser combinadas para gerenciar dados de proveniência em SIS. No entanto, além das vantagens já descritas na QE02, todas essas tecnologias apresentam desafios de implementação e interoperabilidade entre diferentes SIS. Assim, nem todas as tecnologias mencionadas nos estudos garantem 100% de eficiência.

De fato, com o avanço do grande volume de dados gerados em cenários de saúde, as ferramentas e estratégias tecnológicas estão sendo aprimoradas de acordo com a demanda e a necessidade de gerenciar os dados de proveniência em SIS. Isso pode depender também, não apenas das tecnologias em si, mas também de profissionais bem treinados para o uso dessas tecnologias, bem como de processos e políticas impostas pelas instituições de saúde.

3.3.2 Principais categorias identificadas em relação ao gerenciamento de dados de proveniência em SIS

Após a leitura dos artigos selecionados, observou-se que a maioria dos autores listados, direta ou indiretamente, discute sete categorias que entendemos como as principais para uma contribuição positiva em relação ao gerenciamento dos dados de proveniência dos SIS em suas propostas. São elas: armazenamento, disponibilidade, rastreabilidade, confidencialidade, integridade, autenticidade e auditabilidade. Nesse sentido, no Quadro 18 mostra-se em quais estudos essas categorias se destacam.

Quadro 18 – Categorias que contribuem para o gerenciamento dos dados de proveniência em SIS

| Estudos | Armazenamento | Disponibilidade | Rastreabilidade | Confidencialidad | Integridade | Autenticidade | Auditabilidade |
|----------------------------------|---------------|-----------------|-----------------|------------------|-------------|---------------|----------------|
| Margheri <i>et al.</i> (2020) | ✓ | ✓ | ✓ | - | ✓ | - | ✓ |
| Rahman <i>et al.</i> (2020) | ✓ | ✓ | - | ✓ | ✓ | - | ✓ |
| Jaigirdar, Rudolph e Bain (2020) | ✓ | ✓ | - | ✓ | ✓ | ✓ | - |

| | | | | | | | |
|--|---|---|---|---|---|---|---|
| Gong, Lin e Li (2019) | ✓ | - | ✓ | ✓ | ✓ | - | - |
| Kock-Schoppenhauer <i>et al</i> (2018) | ✓ | ✓ | - | - | - | - | - |
| Massi <i>et al.</i> (2018) | ✓ | ✓ | - | ✓ | ✓ | ✓ | - |
| Schreiber e Struminski (2018) | ✓ | ✓ | ✓ | - | ✓ | - | - |
| Liang <i>et al.</i> (2017) | ✓ | - | - | - | ✓ | - | - |
| Curcin <i>et al.</i> (2017) | ✓ | ✓ | ✓ | - | - | ✓ | ✓ |
| Sun, Lu e Gu (2017) | - | - | - | - | ✓ | - | - |
| Ramesh <i>et al.</i> (2015) | - | ✓ | - | - | - | - | - |
| Seneviratne e Kagal (2014) | ✓ | ✓ | | ✓ | ✓ | ✓ | ✓ |
| Lomotey e Deters (2014) | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Prasad <i>et al.</i> (2013) | ✓ | ✓ | - | ✓ | | ✓ | ✓ |
| Wang e Hu (2013) | - | ✓ | ✓ | - | ✓ | ✓ | ✓ |
| Kovalchuk <i>et al.</i> (2013) | - | ✓ | ✓ | - | - | - | ✓ |
| Chowdhury, Falchuk e Misra (2010) | ✓ | ✓ | - | ✓ | - | - | - |

Fonte: Elaborado pelo autor (2021).

As sete categorias apresentadas no Quadro 19 têm um contexto extremamente focado na contribuição dos processos de gerenciamento dos dados de proveniência em SIS. Essas categorias são explicadas abaixo:

- c) Armazenamento: dados de saúde armazenados em repositórios de dados de proveniência. O armazenamento de dados de saúde nos SIS é de grande relevância para a rastreabilidade e disponibilidade de reutilização desses dados;
- d) Disponibilidade: essencial para garantir que os dados de saúde estejam disponíveis para o bom funcionamento da assistência à saúde em andamento. Os dados de saúde disponíveis também podem contribuir para a pesquisa médica;
- e) Rastreabilidade: localização de dados de saúde em um mapa de dados. A rastreabilidade dos dados de saúde em suas diversas etapas e processos contribui para a identificação de atividades dentro de um SIS, como participantes, locais, horários entre outros;
- f) Confidencialidade: garantia de privacidade dos dados em SIS. A confidencialidade dos dados de saúde está relacionada ao acesso por pessoas expressamente autorizadas no SIS. Na verdade, é a sua proteção para impedir que pessoas não autorizadas tenham acesso;
- g) Integridade: garantir a precisão e a confiabilidade dos dados de proveniência ao longo de seu ciclo de vida. No que diz respeito à integridade dos dados de saúde, ele deve ser recuperado em sua forma original (no momento em que foi armazenado). Isso evita modificações intencionais ou acidentais não autorizadas nos SIS;
- h) Autenticidade: a certeza de que os dados de saúde vêm de fontes anunciadas e que não

foram mutados durante um processo dentro dos SIS, e;

- i) **Auditabilidade:** a auditoria dos dados de saúde nos SIS é um processo que aumenta a credibilidade dos serviços de saúde e é responsável por manter as políticas legais e internas sempre atualizadas.

Comparando as categorias apresentadas no Quadro 19 com nos respectivos estudos selecionados conforme a literatura, podemos afirmar que essas categorias fazem parte dos requisitos de segurança e uso de proveniência de dados que podem estar perfeitamente associadas às aplicações em relação ao gerenciamento de dados de proveniência em SIS, o que tornam evidências relevantes para o tema desta tese.

Portanto, observou-se ainda que, nos grandes volumes de dados gerados pelos SIS, o fato de gerenciar dados de proveniência pode contribuir significativamente para a qualidade dos serviços de saúde oferecidos pelas instituições médicas. De fato, essas categorias afetam positivamente a tomada de decisões que requerem processos de análise com base nas fontes de dados contidos nos bancos de dados dos SIS.

3.4 LIÇÕES APRENDIDAS

Após a análise dos artigos selecionados nesta RSL, foi extraído um conjunto de lições aprendidas sobre o processo de gerenciamento de dados de proveniência em SIS:

- a) é um processo complexo, e envolve diversas tecnologias computacionais integradas em diversos fluxos de trabalho altamente heterogêneos;
- b) existem uma série de métodos, técnicas, metodologias e modelos utilizados para realização do gerenciamento dos dados de proveniência, incluindo os modelos recomendados pelo W3C para que esse processo ocorra de forma padronizada;
- c) as normas internacionais de transmissão de dados de saúde entre SIS são essenciais para os dados de proveniência no processo de gestão dos dados;
- d) o gerenciamento dos dados de proveniência dos SIS atingiu proporções sem precedentes. Os dados dos SIS são cruciais e geram benefícios para a saúde das populações;
- e) redução de custos, segurança dos dados de saúde, centralização de dados para análise suplementar e tomada de decisão nos SIS. Conduzir decisões em SIS com dados de qualidade é uma boa abordagem para contribuir para a redução de riscos e obter resultados satisfatórios; e,
- f) a gestão do volume crescente de dados de proveniência dos SIS é fundamental para a

ciência, pois isso se traduz na geração de conhecimento para os setores de saúde.

Nesse contexto, com as lições aprendidas na execução desta RSL, foi possível compreender diferentes estruturas computacionais para o desenvolvimento do PROV-Health que podem minimizar alguns desafios apresentados nos estudos encontrados. Isso de fato, direcionou para a escolha de novas abordagens de implementação do PROV-Health, utilizando-se de ferramentas e técnicas computacionais mais abrangentes e o uso do instanciamento do modelo W3C PROV para contribuir na dinamização da coleta e no gerenciamento dos dados de proveniência no cenário de saúde referente ao estudo de caso desta tese.

3.5 CONSIDERAÇÕES DA SEÇÃO

Nesta seção foi realizada a metodologia norteadora desta tese, além de uma Revisão Sistemática de Literatura (RSL) baseada na abordagem de Kitchenham (2004) com complementação da técnica de *Snowballing (backward and forward)*, a fim de verificar o ineditismo do método aqui proposto. Foram criados processos para o desenvolvimento da RSL e questões de pesquisa para guiar a revisão. Durante o processo da RSL, foram consultadas as principais bases de conhecimento, sendo elas: *ACM Digital Library; IEEEExplore; Science Direct; Springer Link; Scopus e, Web of Science*. Além disto, foram definidos termos de busca alinhados à pergunta de pesquisa e, no processo de busca de estudos foram levantados 239 estudos que, após as etapas de aplicação dos critérios de exclusão, resultaram na seleção de 17 trabalhos alinhados ao tema. Após a leitura e análise integral dos artigos, identificou-se que nenhum deles estava totalmente alinhado com o método aqui proposto.

Por fim, é significativo destacar que, uma versão ampliada desta RSL intitulada “Provenance Data Management in Health Information Systems: a Systematic Literature Review”, foi publicada neste ano vigente (2023) em uma seção especial do *Journal of Personalized Medicine (MDPI)* “Personalized Medicine: The Future of Health Care”.

4 ASPECTOS METODOLÓGICOS

Nesta seção são apresentados os aspectos metodológicos adotados para a realização desta tese, com o intuito de responder à questão de pesquisa e atender os objetivos propostos. Primeiramente é descrita a caracterização da pesquisa, seguindo na descrição do universo, população e amostra. Logo após, os procedimentos metodológicos e a coleta e análise dos dados são descritos na sequência.

4.1 CARACTERIZAÇÃO DA PESQUISA

Em relação à sua natureza, este projeto de pesquisa classifica-se como aplicada. De acordo com Gerhardt e Silveira (2009, p. 35), “este tipo de pesquisa tem como objetivo gerar conhecimentos para aplicação prática, dirigidos à solução de problemas específicos”. O teor de aplicação desta pesquisa, se refere à proposta de um método para o gerenciamento de dados de proveniência em SIS no STT/RCTM.

Do ponto de vista dos procedimentos técnicos, esta pesquisa caracteriza-se como experimental, sendo observadas as variáveis em um estudo de caso a ser realizado no STT/RCTM. Entende-se que os experimentos a serem realizados nesta pesquisa, em relação à coleta e integração dos dados de proveniência seguindo de análises, irão colaborar para o gerenciamento de dados de proveniência dos SIS integrados no âmbito do STT/RCTM, sem exposição dos dados fora dos controles seguros da rede deste cenário de saúde, sendo apenas para aplicação e validação do método aqui proposto. Segundo Gil (2007), a pesquisa experimental consiste em determinar um objeto de estudo, selecionar as variáveis que seriam capazes de influenciá-lo, definir as formas de controle e de observação dos efeitos que a variável produz no objeto. Nesta pesquisa, os dados coletados dos SIS serão os objetos de estudo que serão gerenciados por meio do método desta tese, o qual será o controle para efeito da experimentação.

Do ponto de vista dos objetivos, esta pesquisa é exploratória. Pesquisas exploratórias são todas as que buscam explorar, investigar fatos, fenômenos ou novos conhecimentos sobre o qual ainda se tem pouca informação (TOBAR; ROMANO YALOUR, 2001). Dessa forma, nesta tese, a pesquisa exploratória está presente na RSL realizada como suporte na concepção do método desta tese. A partir da RSL realizada, o estudo exploratório permitiu mostrar que na literatura não há informação sistematizada sobre o uso de Proveniência de Dados em SIS e sobre a situação real de gerenciamento de dados de proveniência em SIS.

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. Nesta pesquisa, o caráter quantitativo está atrelado à quantidade de dados de proveniência que o método irá coletar, armazenar e gerenciar no STT/RCTM. 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, segundo Bose e Frew (2005), a própria Proveniência de Dados qualifica os dados sendo esta ação um dos benefícios da proveniência.

Contudo, optou-se para a pesquisa desta tese, o estudo de caso, o qual é especialmente indicado como estratégia, pois questões do tipo “como?” e “por quê?” são indagadas, concentrando o foco em fenômenos contemporâneos inseridos em algum contexto da vida real. A opção por uma pesquisa de estudo de caso permitirá a validação do método aqui proposto, em um cenário de saúde real com dados reais. Nesse sentido, o estudo de caso é um método que visa compreender fenômenos sociais complexos, preservando as características holísticas e significativas dos eventos da vida real (YIN, 2001). O estudo de caso a ser detalhado na seção 6 desta tese, será experimentado no STT/RCTM no que compete ao telediagnóstico por imagens médicas no âmbito da Telerradiologia em relação aos dados de proveniência gerados nos SIS descentralizados.

O Quadro 19 resume as características metodológicas e a classificação da pesquisa juntamente com os autores que as fundamentam.

Quadro 19 – Resumo das características metodológicas

| Caracterização da Pesquisa | Classificação da pesquisa | Autores/ano |
|----------------------------|---------------------------|------------------------------|
| | Aplicada | Gerhardt e Silveira (2009) |
| | Experimental | Gil (2007) |
| | Exploratória | Tobar e Romano Yalour (2001) |
| | Quantitativa | Knechtel (2014) |
| | Qualitativa | Reneker (1993) |
| | Estudo de caso | Yin (2001) |

Fonte: Elaborado pelo autor (2020).

As características metodológicas apresentadas no Quadro 20, serviram de base científica e teórica para a proposta de desenvolvimento do método para gerenciamento de dados de proveniência em SIS proposto nesta tese.

4.2 UNIVERSO, POPULAÇÃO E AMOSTRA

O universo, ou população, é o conjunto de elementos que possuem as características que serão objeto do estudo, e a amostra, ou população amostral, é uma parte do universo escolhido, selecionada a partir de um critério de representatividade (VERGARA, 1997).

Nesse sentido, o universo de pesquisa se volta ao STT/RCTM, no qual se pretende aplicar o método. Dessa forma, a população selecionada segue duas vertentes: computacional e humana. No quesito computacional, serão os dados digitais e manuais inseridos nos SIS integrados no STT/RCTM. Já no quesito humano, serão os profissionais/especialistas que atuam no gerenciamento dos dados na mesma rede.

Em relação à amostra, também seguem os mesmos quesitos da população. Nesse sentido, a amostra no quesito computacional será com base no repositório de dados de proveniência local (banco de dados local que receberá dados do STT/RCTM por meio de estratégias computacionais desenvolvidas dentro da mesma rede sem acesso externo ou por pessoas não autorizadas). Já no que compete à amostra no quesito humano, serão selecionados cinco profissionais/especialistas que atuam no STT/RCTM para responderem a um formulário eletrônico (entrevista semiestruturada) gerado a partir do *Google Forms* para validação do método proposto nesta tese. Justificando essa escolha, os cinco profissionais/especialistas validaram o método por conhecerem o fluxo computacional do STT/RCTM e as ferramentas computacionais utilizadas em sua aplicação.

Portanto, o gerenciamento dos dados de proveniência nos SIS do STT/RCTM servirá como piloto para adequação do método para outros cenários de saúde.

4.3 PROCEDIMENTOS METODOLÓGICOS

Para que o método de gerenciamento de dados em SIS proposto possa ter êxito, houve a necessidade de realizar testes piloto por meio da estratégia computacional desenvolvida (detalhado na seção 6), a fim de demonstrar os resultados para obtenção do desenho do método final. Nesse contexto, esta pesquisa caracteriza-se como um processo prático e voltado aos meios computacionais para traçar o método ideal para o cenário de saúde que se pretende executar.

O Quadro 20 resume os aspectos metodológicos realizados para alcançar as respostas, conforme os objetivos traçados, definindo as etapas de execução que compreendem desde a fundamentação teórica até os resultados esperados.

Quadro 20 – Resumo dos aspectos metodológicos

| Objetivos Específicos | Dados a serem coletados | Metodologia |
|---|--|---|
| a) Apresentar os principais pressupostos necessários para o gerenciamento de dados de proveniência em SIS; | 1. Listar os principais pressupostos que necessitam de gerenciamento de dados de proveniência em SIS. | Pesquisa bibliográfica Referencial Teórico |
| b) Identificar métodos, técnicas, modelos ou metodologias computacionais que apoiam o gerenciamento dos dados de proveniência em SIS; | 2. Listar os principais métodos, técnicas, modelos e metodologias computacionais de Proveniência de Dados existentes para adaptar estratégias concretas para o método. | RSL |
| c) Elaborar um método para gerenciamento de dados de proveniência em SIS com base no instanciamento do modelo W3C PROV; | 3. Apresentar arquiteturas e estratégias computacionais, além da modelagem de dados necessária para o ambiente de saúde ao qual se pretende aplicar o método. | Desenho das arquiteturas e estruturação dos dados conforme ferramentas de desenvolvimento computacional. |
| d) Aplicar o método em um cenário de saúde real para o estudo de caso; | 4. Apresentar as estratégias computacionais, identificando as tecnologias a serem empregadas. | Experimentações utilizando ferramentas computacionais: Simplifier ¹⁶ , Simplifier - STT/RCTM on FHIR® (R4) ¹⁷ , Forge by Firely ¹⁸ , Visual Studio Code (VS CODE) ¹⁹ , Insomnia ²⁰ , Postgresql ²¹ , SQL ²² , GitLab ²³ , Python ²⁴ , Json ²⁵ , HL7 FHIR ²⁶ , HAPI FHIR ²⁷ , DCM4CHE ²⁸ , ETL ²⁹ , Linux (Ubuntu) ³⁰ , |

¹⁶ <https://simplifier.net/>¹⁷ <https://simplifier.net/STT>¹⁸ <https://fire.ly/products/forge/>¹⁹ <https://code.visualstudio.com/>²⁰ <https://insomnia.rest/download>²¹ <https://www.postgresql.org/>²² <https://www.arcservice.com/blog/how-install-sql-server-linux-and-create-database>²³ <https://about.gitlab.com/>²⁴ <https://www.python.org/>²⁵ <https://www.json.org/json-pt.html>²⁶ <https://www.hl7.org/fhir/>²⁷ <https://hapifhir.io/>²⁸ <https://www.dcm4che.org/>²⁹ <https://www.stitchdata.com/>³⁰ <https://ubuntu.com/download/desktop>

| | | |
|---|---|--|
| | | Windows ³¹ , Bash ³² , Docker ³³ e Grafana ³⁴ . |
| e) Validar o método com especialistas em gerenciamento de dados em SIS. | 5. Perguntas fechadas e abertas de cunho técnico, operacional e gerencial relacionadas sobre o uso do método experimentado em estudo de caso. | Formulário eletrônico enviado por meio do Google Forms para os especialistas/profissionais em gerenciamento de dados em SIS. |

Fonte: Elaborado pelo autor (2020).

Outro ponto a ser considerado nos aspectos metodológicos se refere à elaboração da RSL, que irá traçar a situação teórico-tecnológica em que se encontram as aplicações dos métodos, técnicas, modelos ou metodologias computacionais nos cenários de saúde, a fim de encontrar trabalhos semelhantes à temática abordada. Os detalhes da elaboração RSL podem ser observados na seção 3, já apresentada anteriormente.

4.4 COLETA E ANÁLISE DOS DADOS

Antes de se iniciar a coleta e análise dos dados, observou-se que esse processo envolveria a participação de seres humanos e com isso fez-se necessário submeter o projeto de pesquisa desta tese para a apreciação do Comitê de Ética em Pesquisas com Seres Humanos (CEPSH) da Universidade Federal de Santa Catarina (UFSC). Após submissão do projeto de pesquisa (APÊNDICE A) com suas devidas assinaturas (APÊNDICE B), o parecer para a realização da coleta dos dados foi aceito e emitido sob o nº 5.063.111, conforme disponível no (APÊNDICE C). A declaração que apresenta os efeitos legais do consentimento do responsável que autoriza o local (instituição) para a coleta de dados dessa pesquisa pode ser observada no (APÊNDICE D). Ainda, para formalizar os aspectos éticos da coleta de dados, foram necessários o Termo de Compromisso para Utilização de Dados (TCUD) (APÊNDICE E) e o Termo de Consentimento Livre e Esclarecido (TCLE) (APÊNDICE F). Importante destacar que, os pesquisadores envolvidos nesta tese se comprometem a zelar pela confidencialidade dos dados e privacidade dos participantes, de acordo com a Resolução CNS 466/2012 (BRASIL, 2012).

Dessa forma, com os aspectos éticos formalizados, a primeira coleta de dados se dará

³¹ <https://www.microsoft.com/software-download/windows11>

³² <https://www.gnu.org/software/bash/>

³³ <https://www.docker.com/>

³⁴ <https://grafana.com/>

pelo método aqui proposto de forma prática e aplicada a ser detalhada na Seção 6, por meio de intervenção de softwares que, conforme explica Santos (2004, p. 28), “são os métodos práticos utilizados para juntar as informações necessárias à construção dos raciocínios em torno de um fato/fenômeno/problema”. O método a ser experimentado terá premissas em sua estrutura para selecionar quais dados serão capturados e posteriormente analisados. Os dados coletados por intermédio do método nos bancos de dados dos SIS descentralizados e integrados ao STT/RCTM serão armazenados em um repositório de dados de proveniência local que servirão também como amostragem dos dados de proveniência, possibilitando a realização das análises realizadas pelo método e compreender os fluxos informacionais advindos de fontes descentralizadas. Dessa forma, esses dados posteriormente serão processados e manipulados em uma ferramenta computacional de visualização de dados, a fim de gerar conhecimento para uso em tomadas de decisão no cenário de saúde em estudo. Nesse contexto, internamente, sem acesso externo, os dados serão utilizados pelos gestores e profissionais/especialistas autorizados do STT/RCTM para controle, monitoramento e rastreamento das informações técnicas especializadas para contribuir em tomadas de decisão e possíveis auditorias. Porém, é de suma importância ressaltar que os dados pessoais e dados sensíveis serão anonimizados conforme Lei Geral de Proteção de Dados Pessoais (BRASIL, 2018), pois o intuito é a aplicação e validação de forma acadêmica do método juntamente com os profissionais/especialistas atuantes no STT/RCTM.

Já a segunda forma de coleta de dados será um formulário eletrônico gerado no *Google Forms*, a qual é uma ferramenta que oferece suporte para a criação de formulários personalizados de forma simples (GOOGLE, 2020). Esse formulário eletrônico foi enviado por e-mail para cinco profissionais/especialistas que atuam no STT/RCTM, os quais validaram o método. Juntamente no formulário eletrônico foi realizada a gravação de um vídeo curto explicando os passos do método. Para Zanini (2007), existem várias vantagens associadas à utilização do formulário eletrônico quando comparado ao formulário convencional (com utilização de papel), entre elas, a facilidade na busca de dados, a utilização de armazenamento físico diminuto e distribuição fácil e rápida. Em relação ao formulário eletrônico, o roteiro da entrevista semiestruturada pode ser visualizado no (APÊNDICE G), onde são apresentadas questões de cunho técnico e operacional referentes à experimentação do método para a sua validação pelos cinco profissionais/especialistas atuantes no STT/RCTM. Após a aplicação do formulário eletrônico para descobrir a eficiência do método, as respostas serão transcritas e tabuladas para análises e interpretação dos dados.

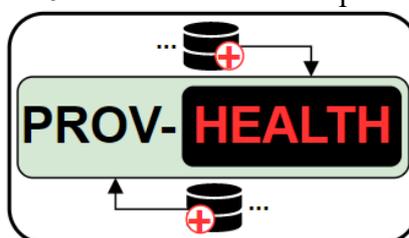
Desse modo, em relação ao contexto da análise e interpretação dos dados, foram

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 geraram gráficos relacionados aos dados coletados por especificidades, sendo que, no quesito de qualidade, a própria proveniência de dados qualificará os dados. As análises e interpretação dos dados serão descritas e detalhadas na Seção 7 desta tese.

5 PROV-HEALTH

O método a ser apresentado nesta seção foi nomeado de PROV-Health, em que “PROV” representa a palavra no inglês “provenance”, e “Health” traduzido do inglês “saúde”, de forma a representar sua tradução em um contexto geral para o português: “proveniência em saúde”, pois o PROV-Health pode trazer além dos dados de proveniência de saúde, outros pontos que sejam importantes na interpretação dos profissionais da informação que utilizarão o método. Assim, na Figura 6 é apresentada a logomarca definida para o PROV-Health.

Figura 6 – Nome do método apresentado



Fonte: Elaborado pelo autor (2023).

Portanto, nesta seção são apresentadas as etapas para a concepção do PROV-Health em um contexto genérico de SIS, onde a metodologia de concepção empregada é apresentada, seguida do instanciamento do modelo de Proveniência de Dados W3C PROV para servir de base para a construção do PROV-Health. Logo em seguida, uma proposta de instrumentalização genérica dos serviços de saúde dos SIS é apresentada para servir de parte da arquitetura do PROV-Health, descrita posteriormente de forma detalhada no que tange as suas camadas e respectivos módulos. Por fim, a verificação dos componentes e relações da arquitetura do PROV-Health com base no W3C PROV são apresentadas, para que de forma concernente, o método proposto tenha sua concepção justificada. Por fim, finaliza-se com as considerações finais da seção.

5.1 METODOLOGIA DE CONCEPÇÃO

Levando em consideração o contexto sobre o tema, com base no levantamento bibliográfico e na RSL apresentada nesta tese, as quais nortearam as discussões para a concepção do PROV-Health, algumas etapas foram necessárias para a construção dos processos.

Primeiramente foi necessário compreender o objetivo do PROV-Health com base nos

componentes ideais e necessários para a concepção do método contidos no modelo de Proveniência de Dados W3C PROV, os quais são: *Agent*, *Activity*, *Entity*. Esses componentes têm relações definidas, sendo elas: *wasAssociatedWith*, *wasAttributeTo*, *wasGeneratedBy*, *wasDerivedFrom*, *Used*. Para isso, fez-se o instanciamento do W3C PROV para o contexto genérico de SIS que fosse possível de inserir na arquitetura final do PROV-Health, 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.

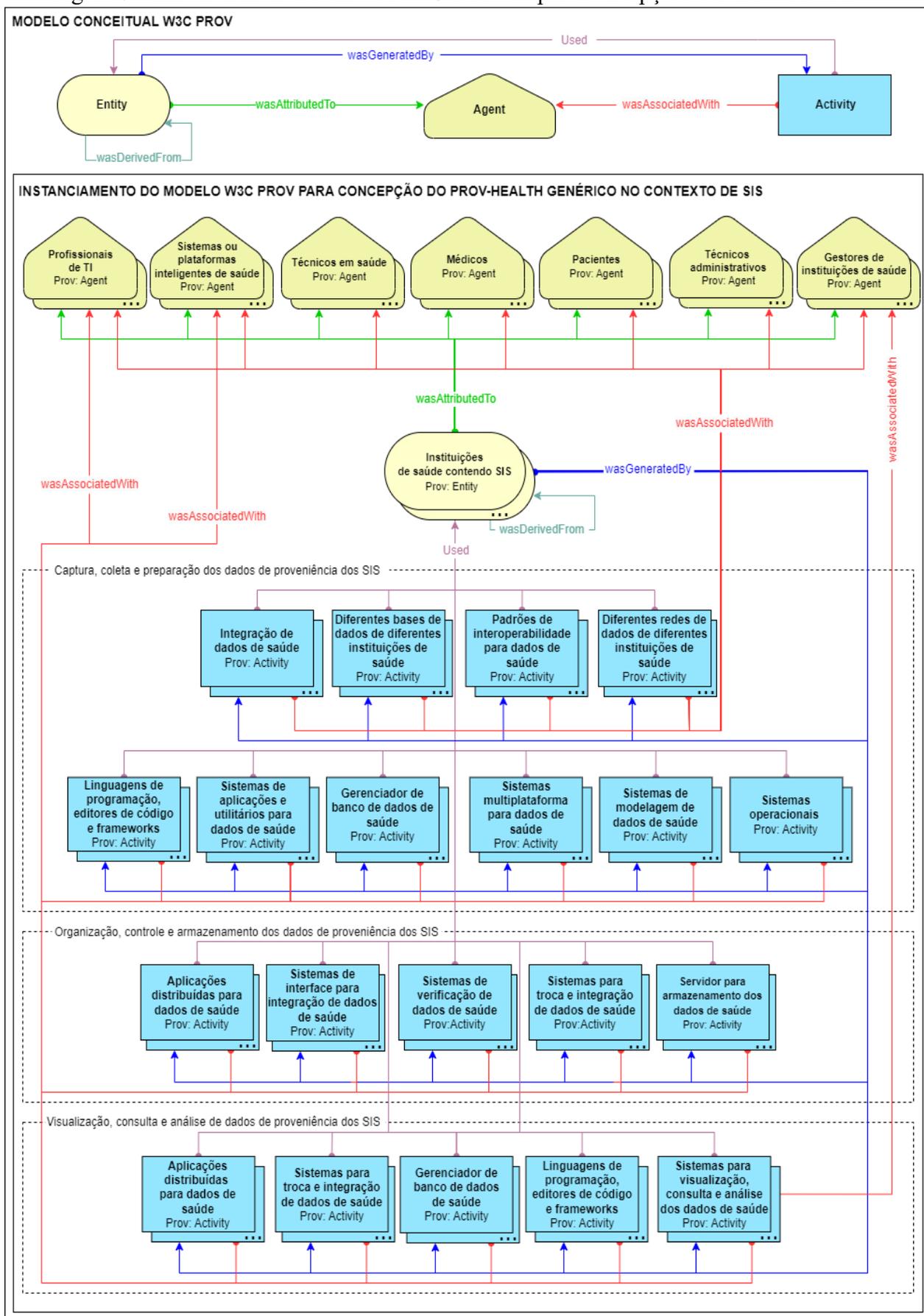
Nesse sentido, analisando os fluxos informacionais dos diferentes SIS, houve a necessidade também de propor uma instrumentalização genérica dos serviços de saúde dos SIS para auxiliar a captura, coleta e preparação dos dados de proveniência dos SIS descentralizados a serem armazenados e organizados no PROV-Health de forma centralizada, sendo possível posteriormente a realização da visualização, consulta e análise desses dados. A centralização dos dados organizada pelo PROV-Health, apesar de possibilitar um armazenamento exponencial de dados de proveniência, servirá como amostragem para análises a serem realizadas pelo PROV-Health na busca pela compreensão dos fluxos informacionais descentralizados dos diferentes SIS e nos possíveis pontos falhos existentes em toda estrutura que gera dados de saúde que sejam passíveis de manutenção.

Com isso, a arquitetura do PROV-Health com a definição detalhada de suas camadas e seus respectivos módulos foi definida, assim como, as suas relações com o modelo de Proveniência de Dados W3C PROV, o qual foi base para a sua concepção. A arquitetura relaciona os possíveis agentes, entidades e atividades necessárias em um contexto genérico de utilização em SIS. Módulos foram definidos na arquitetura do PROV-Health, como: Módulo de captura, coleta e preparação dos dados de proveniência dos SIS; Módulo de organização, controle e armazenamento dos dados de proveniência dos SIS; e, Módulo de visualização, consulta e análise de dados de proveniência dos SIS, onde a parte de visualização denominou-se “PROV-Health BI”. Cada módulo pode conter possíveis sistemas necessários para realizar as atividades geradas pelas entidades e executadas pelos agentes. Por fim, foi realizada uma verificação dos componentes e relações da arquitetura genérica do PROV-Health com base no W3C PROV.

5.2 INSTANCIANDO O MODELO W3C PROV EM UM CONTEXTO GENÉRICO DE SIS PARA A CONCEPÇÃO DO PROV-HEALTH

O modelo de Proveniência de Dados W3C PROV é considerado um modelo genérico, e foi observada a necessidade de compreender suas estruturas na perspectiva dos seus componentes (*Agent*, *Entity*, *Activity*) e as relações entre esses componentes (*wasAssociatedWith*, *wasAttributeTo*, *wasGeneratedBy*, *wasDerivedFrom*, *Used*), para realizar o instanciamento que sirva de base para a concepção do PROV-Health em um contexto genérico de SIS, com intuito de descrever os possíveis processos que podem ser adaptados para qualquer cenário de saúde mediante integração e interoperabilidade de dados entre diferentes SIS. Assim, apresenta-se na Figura 7, o instanciamento do modelo W3C PROV com seus respectivos componentes e relações para o contexto de SIS de forma genérica. Importante ressaltar que, esse instanciamento baseou-se no estudo de KOCK-SCHOPPENHAUER *et al.* (2018), o qual descreve uma extensão do modelo W3C PROV, apresentando também uma camada de instanciamento voltada para o contexto de saúde.

Figura 7 – Instanciamento do modelo W3C PROV para concepção do PROV-Health



Fonte: Elaborado pelo autor (2023).

A Figura 7 apresenta o modelo de Proveniência de Dados W3C PROV em sua estrutura conceitual e logo após é realizada o instanciamento deste modelo para o contexto de SIS para servir de base na concepção do PROV-Health em um contexto de saúde genérico. Nesse sentido, destaca-se que o PROV-Health pode ser ajustado para qualquer contexto de saúde, sendo que os componentes (*Agent*, *Entity*, *Activity*) podem sofrer alterações necessárias dependendo do contexto de saúde a ser aplicado. Assim, no PROV-Health, as ligações dos componentes (*Agent*, *Entity*, *Activity*) e as relações (*wasAssociatedWith*, *wasAttributeTo*, *wasGeneratedBy*, *wasDerivedFrom*, *Used*) entre cada componente são apresentadas na Figura 7 em relação aos fluxos utilizados para esse instanciamento como proposta para um contexto de saúde genérico. Pode-se observar também que a mesma Figura, apresenta três módulos que se sustentam das relações entre (*Agent*, *Entity*, *Activity*), 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 PROV-Health. Esses módulos também podem sofrer mudanças conforme o contexto de aplicação, assim como seus componentes podem ser ajustados. São observadas também, que algumas atividades se repetem, pois são constituintes e necessárias em diferentes módulos, a fim de executar funções computacionais compatíveis com diferentes SIS. Assim, para melhor compreensão dos fluxos da Figura 7, o Quadro 21 descreve as relações causais detalhadas entre (*Agent*, *Entity*, *Activity*).

Quadro 21 – Descrição das relações do instanciamento do modelo W3C PROV para concepção do PROV-Health

| Relação | Descrição |
|--------------------------|---|
| <i>wasAttributeTo</i> | Esta relação está voltada às instituições de saúde contendo SIS (<i>Entity</i>) que atribui a todos os agentes envolvidos direta ou indiretamente nos SIS neste instanciamento. Assim, para este instanciamento no contexto de SIS, os agentes: profissionais de TI (<i>Agent</i>), sistemas ou plataformas inteligentes de saúde (<i>Agent</i>), técnicos em saúde (<i>Agent</i>), médicos (<i>Agent</i>) e pacientes (<i>Agent</i>), técnicos administrativos (<i>Agent</i>) e Gestores de instituições de saúde (<i>Agent</i>) podem possuir atribuições de inserção, acesso e uso de dados de saúde de forma diferenciada conforme demanda. |
| <i>wasAssociatedWith</i> | Esta relação está voltada às atividades (<i>Activity</i>) e agentes (<i>Agents</i>). Cada agente pode realizar uma determinada atividade conforme sua responsabilidade e autorização. As seguintes atividades (integração de dados de saúde; diferentes bases de dados de diferentes instituições de saúde; padrões de interoperabilidade para dados de saúde e diferentes redes de dados de diferentes instituições de saúde) são associadas a todos os agentes para este instanciamento, sendo analisadas as atribuições e autorizações necessárias. Essas |

| | |
|-----------------------|---|
| | <p>associações são realizadas por se tratar do primeiro passo para a instrumentalização dos serviços de saúde em SIS, sendo observadas as responsabilidades de acesso e uso de dados de saúde de cada agente. Logo após, para que a instrumentalização de serviços de saúde em SIS possa ter êxito, e uma possível captura, coleta e preparação dos dados de proveniência dos SIS possa ser executada, as atividades (linguagens de programação, editores de código e frameworks; sistemas de aplicações e utilitários para dados de saúde; gerenciador de banco de dados de saúde; sistemas multiplataforma para dados de saúde; sistemas de modelagem de dados de saúde; e, sistemas operacionais) são associadas apenas aos agentes (profissionais de TI e sistemas ou plataformas inteligentes de saúde) por se tratar de uma responsabilidade técnica e operacional que deve ser executada por esses agentes. Assim, ainda são associadas a esses dois agentes, as atividades (aplicações distribuídas para dados de saúde; sistemas de interface para integração de dados de saúde; sistemas de verificação de dados de saúde; sistemas para troca e integração de dados de saúde; e, servidor para armazenamento dos dados de saúde) para que seja possível executar o armazenamento e organização dos dados de proveniência em SIS. E por fim, ainda são associadas a esses dois últimos agentes, as atividades (aplicações distribuídas para dados de saúde; sistemas para troca e integração de dados de saúde; gerenciador de banco de dados de saúde; linguagens de programação, editores de código e frameworks; e, sistemas para visualização, consulta e análise dos dados de saúde) com intuito de executar a visualização, consulta e análise dos dados de proveniência dos SIS que podem contribuir para tomadas de decisão voltadas aos agentes gestores de instituições de saúde.</p> |
| <i>wasGeneratedBy</i> | <p>Essa relação está voltada entre entidades (<i>Entity</i>) e atividades (<i>Activity</i>) no quesito de observar as atividades geradas pelas instituições de saúde que contêm SIS. Nesse sentido, instituições de saúde contendo SIS (<i>Entity</i>), geram todas as atividades expostas na Figura 7 e já mencionadas anteriormente na relação <i>wasAssociatedWith</i>. Em resumo, instituições de saúde contendo SIS (<i>Entity</i>) geram atividades direta ou indiretamente que contribuem para a captura, coleta, preparação, organização, controle, armazenamento, visualização, consulta e análise dos dados de proveniência dos SIS. 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.</p> |
| <i>wasDerivedFrom</i> | <p>Esta relação está voltada apenas de entidade (<i>Entity</i>) para entidade (<i>Entity</i>). Dessa forma, instituições de saúde contendo SIS (<i>Entity</i>) podem ser derivadas de outros SIS entre essas instituições, ou seja, por se tratar de um contexto descentralizado, esta relação fica entre essa entidade para com intuito de centralização dos dados de proveniência para fins de compreensão dos fluxos informacionais descentralizados existentes.</p> |
| <i>Used</i> | <p>Esta relação está voltada ao uso das atividades (<i>Activity</i>) pelas</p> |

| | |
|--|---|
| | entidades (<i>Entity</i>). Assim como descrito anteriormente na relação <i>wasGeneratedBy</i> , que entidades geram atividades, a relação <i>used</i> afirma que cada atividade tem seu uso na entidade de forma direta e indireta, ou seja, são executadas em fluxos de estratégias computacionais necessárias para contribuir para a execução da captura, coleta, preparação, armazenamento, organização, visualização, consulta e análise dos dados de proveniência dos SIS. |
|--|---|

Fonte: Elaborado pelo autor (2023).

Assim, com base nas relações (*wasAssociatedWith*, *wasAttributeTo*, *wasGeneratedBy*, *wasDerivedFrom*, *Used*) entre (*Agent*, *Entity*, *Activity*) apresentadas na Figura 7 e as descrições do Quadro 21, foi possível fundamentar a proposta do PROV-Health com intuito de preservar os fundamentos do modelo de Proveniência de Dados W3C PROV, tendo em vista, que cada agente, entidade e atividade será incorporada e descrita separadamente nas camadas necessárias para compor a arquitetura genérica do método nas próximas subseções desta seção. Importante destacar também, que as relações causais definidas na Figura 7 são baseadas no modelo W3C PROV e podem ser ajustadas conforme a necessidade e demanda do contexto de saúde a ser aplicado. Outro ponto que se evidencia na Figura 7 é que o nível de proveniência se destaca em grão grosso, ou seja, por se tratar do instanciamento do modelo W3C PROV para a concepção do PROV-Health de forma genérica; o grão fino da proveniência pode surgir conforme as necessidades de aplicações do método.

Assim, na próxima subseção é apresentada a proposta de instrumentalização genérica dos serviços de saúde dos SIS para uso na arquitetura genérica do PROV-Health posteriormente descrita, contendo a explicação de suas respectivas etapas.

5.3 PROPOSTA DE INSTRUMENTALIZAÇÃO GENÉRICA DOS SERVIÇOS DE SAÚDE PARA USO NO PROV-HEALTH

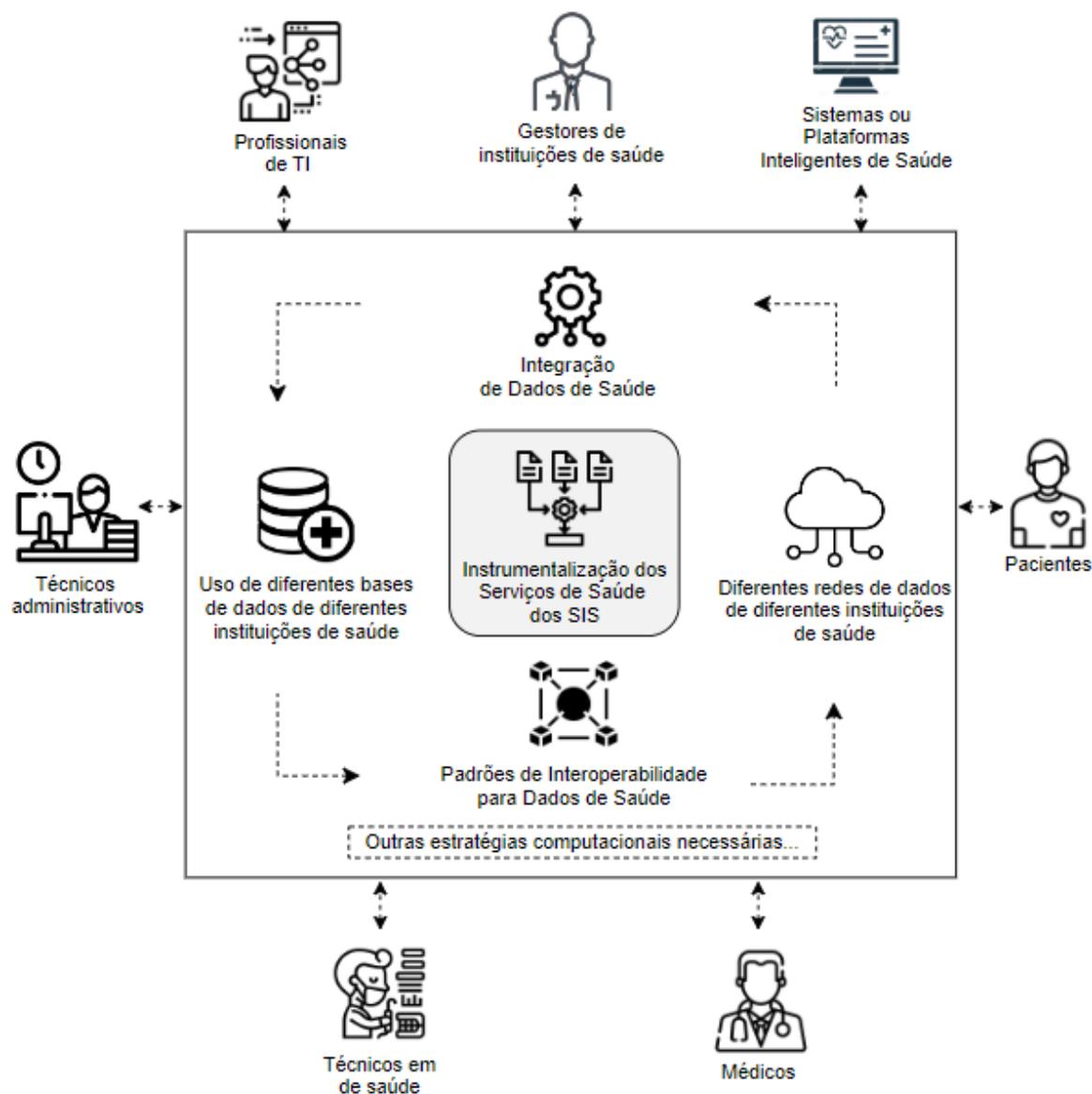
No contexto dos SIS é imprescindível manter os fluxos informacionais atualizados dentro desses sistemas em qualquer atividade desempenhada. Dito isto, preocupou-se na concepção do PROV-Health, simplificar a execução dos complexos fluxos informacionais que precisam de um controle rigoroso nos SIS, com intuito de contribuir para automatizar parte ou todos os processos gerados nesses sistemas de forma genérica. Ainda, um ponto que merece destaque está atrelado ao uso de dados de saúde que estejam em formato digital de ponta a ponta nos processos do PROV-Health, ou seja, processo que é de fato uma necessidade para que exista eficácia ao atingir os objetivos inicialmente planejados.

Isso de fato, contribui também para ampliar resultados e o engajamento dos diferentes profissionais (agentes) envolvidos em todos os processos de saúde (atividades), tornando a produtividade dos serviços prestados (entidades) nas diversas modalidades dos SIS, de forma confiável e segura.

Assim, como não existe uma instrumentalização de serviços de saúde em SIS que seja padrão na literatura, foi necessário compreender os diferentes fluxos informacionais de diferentes SIS apresentados nos estudos da RSL apresentada nesta tese, além de compreender também os principais processos e ferramentas tecnológicas mais comumente utilizadas pelos autores em seus estudos, para que de fato, servisse de base e fosse possível propor uma instrumentalização dos serviços de saúde dos SIS de forma genérica para adaptação em diferentes instituições de saúde e SIS.

Portanto, levando em consideração as diversas modalidades médicas executadas nos SIS, desde o agendamento dos exames dos pacientes até os laudos e resultados desses exames, na Figura 8 é apresentada uma proposta de instrumentalização genérica dos serviços de saúde dos SIS para compor a inicialização da concepção do PROV-Health.

Figura 8 – Instrumentalização genérica dos serviços de saúde dos SIS para uso no PROV-Health



Fonte: Elaborado pelo autor (2023).

Nota-se que na Figura 8 foram previstos os agentes identificados no instanciamento do modelo W3C PROV para a concepção do PROV-Health da Figura 7, sendo eles: Profissionais de TI, Gestores de Instituição de Saúde, Sistemas ou plataformas Inteligentes de Saúde, Técnicos Administrativos, Pacientes, Técnicos em Saúde e Médicos. Cada agente inserido no contexto da instrumentalização proposta, conta com autorizações diferenciadas em relação à inserção, acesso e uso dos dados de saúde. Profissionais de TI e sistemas ou plataformas inteligentes de saúde, podem acessar a instrumentalização para realizarem todos os processos técnicos e operacionais necessários nos fluxos de dados existentes. Já os técnicos em saúde

podem apenas inserir dados de saúde referentes aos pacientes ao fluxo da instrumentalização. Os médicos podem tanto acessar os dados, como alterá-los dependendo do tratamento de saúde do paciente. Já os técnicos administrativos, podem apenas inserir e acessar os dados de agendamento dos procedimentos médicos solicitados pelos pacientes. E, por fim, os gestores de instituições de saúde tomam decisões gerenciais conforme informações enviadas pelos técnicos administrativos, e decisões técnicas e operacionais conforme solicitações dos profissionais de TI.

Nesse contexto, conforme as autorizações dos agentes anteriormente descritas, esses estão inseridos em um fluxo de tecnologias computacionais de forma direta e indireta, fazendo parte dos processos de instrumentalização genérica dos serviços de saúde dos SIS. Lembrando que essas tecnologias são expressamente acessadas pelos Profissionais de TI ou pelos Sistemas ou Plataformas Inteligentes de Saúde, as quais contribuem para o rastreamento dos dados de proveniência dentro dos diferentes SIS. Nesse ínterim, o fluxo de tecnologias computacionais comumente encontrada servindo de base para a concepção dos processos de instrumentalização genérica dos serviços de saúde em SIS, são:

- a) **Integração de dados de saúde:** 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, certificando-se de que todos os serviços de saúde instrumentalizados possam seguir com seus processos de forma contínua e eficiente;
- b) **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 Estaduais;
- c) **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. Essas diferentes redes podem ser Federais, Municipais, Distritais ou Estaduais, sejam particulares ou públicas que possibilitem a comunicação e acesso por meio de autorizações perante as políticas e leis de segurança de proteção de dados.
- d) **Uso de diferentes bases de dados de diferentes instituições de saúde:** são 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. Profissionais de TI autorizados para manutenção dos dados de saúde nessas bases também podem ter acesso, como por exemplo os administradores de redes de dados de saúde entre outros profissionais de TI devidamente autorizados. Essas bases de dados podem ser de diferentes instituições de saúde pública, ambas com restrito acesso ou inserção de dados de saúde por meio de políticas e leis de proteção de dados vigentes.

Portanto, a instrumentalização genérica dos serviços de saúde em SIS proposta para uso no PROV-Health, 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. Salienta-se que a instrumentalização aqui apresentada, em consonância com a arquitetura proposta do PROV-Health poderá, de fato, contribuir para a captura dos dados de proveniência.

Na próxima subseção é apresentada a arquitetura do PROV-Health para um contexto genérico de SIS, sendo descritas suas camadas e seus respectivos módulos.

5.4 ARQUITETURA GENÉRICA DO PROV-HEALTH

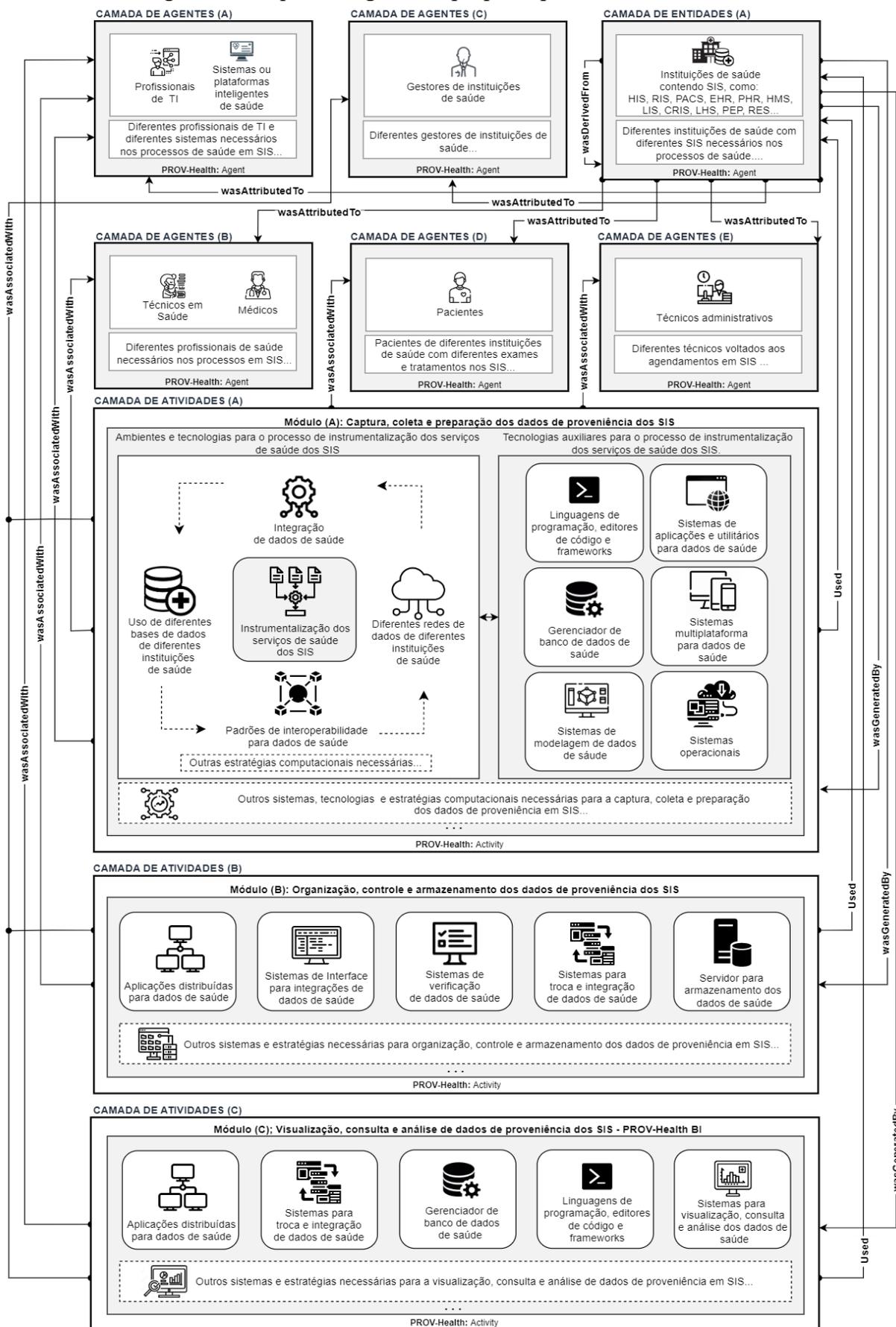
Para que a arquitetura genérica do PROV-Health pudesse ser modelada, baseou-se na revisão de literatura sobre o tema e posteriormente na RSL desta tese, considerando os estudos mais focados nos conceitos do modelo de Proveniência de Dados (W3C PROV). Além disso, houve a necessidade de considerar o grande volume de dados armazenados e compartilhados nos diferentes SIS para prever algumas tecnologias computacionais necessárias. Assim, com base no instanciamento do W3C PROV já realizada para a concepção do PROV-Health em um contexto de SIS de forma genérica conforme já apresentado na (Figura 7), foi possível modelar e propor a arquitetura do PROV-Health em um contexto genérico para diferentes SIS, que seja capaz de contribuir para o gerenciamento dos dados de proveniência desses sistemas que se encontram de forma descentralizada, servindo para adaptações em qualquer cenário de saúde. Portanto, a arquitetura genérica proposta do PROV-Health contém as seguintes especificações:

- a) Cinco camadas de agentes: Agentes (A), (B), (C), (D) e (E), sendo possível ajustar conforme a necessidade de inserção de novos agentes;
- b) Uma camada de Entidades: Entidades (A), com a possibilidade de inserção de novas entidades que sejam necessárias;

- c) Três camadas de Atividades: Atividades (A) - contendo o módulo de captura, coleta e preparação dos dados de proveniência dos SIS; Atividades (B) - contendo o módulo de organização, controle e armazenamento dos dados de proveniência dos SIS; e, Atividades (C) – contendo o módulo de visualização, consulta e análise dos dados de proveniência em SIS. É possível ajustar as estratégias e sistemas contidas nas atividades conforme necessário a demanda operacional encontrada.

Importante ressaltar que, as camadas descritas na arquitetura do PROV-Health não são destinadas a camadas de software e sim a camadas que diferenciam os processos no método proposto. Assim, a arquitetura genérica proposta do PROV-Health é apresentada na Figura 9 com suas camadas e com seus respectivos módulos.

Figura 9 – Arquitetura genérica proposta para o PROV-Health



Fonte: Elaborado pelo autor (2023).

A *priori*, o PROV-Health consiste em um método que contém em sua estrutura estratégias computacionais adaptáveis, que levam a um repositório de dados local. É importante ressaltar que um repositório de dados, consiste em uma terminologia genérica que referente a um vasto banco de dados destinado a armazenar diferentes tipos de dados, os quais são passíveis de gerenciamento para análise, distribuição e geração de relatórios. O PROV-Health caracteriza-se como um tipo de repositório de dados que armazena dados médicos, o qual pode contribuir para acelerar a tomada de decisões em diferentes instituições de saúde. Assim, o PROV-Health é um repositório de Proveniência de Dados local a ser instalado em uma rede de dados de saúde, preocupando-se com a segurança e na prevenção de possíveis vazamentos de dados. O PROV-Health pode receber diferentes formatos de dados de fontes descentralizadas de instituições de saúde que possuem diferentes SIS. Tecnicamente, o repositório de dados de proveniência local é um servidor de dados de saúde³⁵ contendo um banco de dados interno, que centraliza dados de proveniência relevantes a serem selecionados dos diferentes SIS descentralizados dentro de uma rede de dados de saúde autorizada. Isso contribui para a realização dos processos de gerenciamento, segurança e rastreabilidade dos dados de saúde, centralizando-os para a compreensão dos fluxos informacionais gerados nos diferentes SIS descentralizados.

Ressalta-se que o PROV-Health traz em sua concepção a preocupação no cumprimento dos requisitos da Lei Geral de Proteção de Dados, de nº 13.709, de 14 de agosto de 2018 (BRASIL, 2018), no que diz respeito ao tratamento de dados pessoais e dados pessoais sensíveis, sendo de suma importância ressaltar que esses dados são anonimizados conforme a necessidade, com auxílio de ferramentas de software e posteriormente poderão ser analisados por políticas internas de auditoria instituídas para cada instituição de saúde. Também, enfatiza-se que no PROV-Health genérico não são alteradas as bases de dados originais e tampouco expostos dados de saúde sem anonimização prévia, ou seja, possui uma cópia das bases de dados originais e recebe atualizações referentes às modificações ou transformações dos dados dessas bases, seguindo os conceitos de Proveniência de Dados. Nesse sentido, vale ressaltar que, em relação à anonimização dos dados de saúde, pode ser ou não aplicada na coleta desses dados, porém, deve ser obrigatoriamente aplicada caso houver exposição dos dados sensíveis fora do âmbito da rede a qual se está trabalhando. Considerando que o PROV-Health genérico realiza suas funções em ambiente interno, plataformas com controle de acesso devem existir para que não haja vazamentos de dados.

³⁵ São aplicações ou equipamentos dentro de um sistema computacional usados para armazenar e gerenciar dados de saúde de forma estruturada.

Dessa forma, para melhor compreender o PROV-Health em um contexto genérico apresentado na Figura 9, na próxima subseção serão descritas as camadas que o compõem de forma separada, assim como, a descrição de seus componentes.

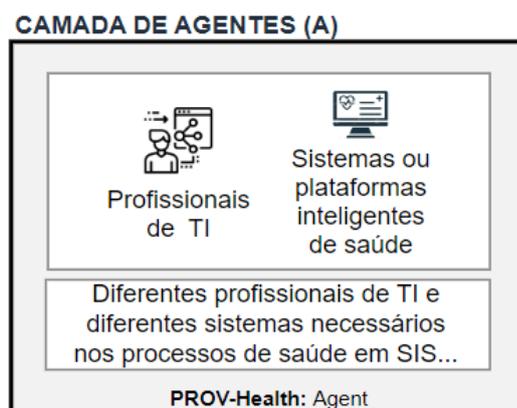
5.4.1 Camada de Agentes

Na concepção do PROV-Health, preocupou-se em manter para cada camada a concepção do modelo de Proveniência de Dados W3C PROV, estipulando agentes que fossem comuns para uso em qualquer instituição de saúde, observando os principais processos de acesso e manipulação dos dados de saúde de diferentes SIS nesse contexto.

Dessa forma, inicialmente são descritas as camadas de Agentes (A), (B), (C), (D) e (E), sendo possível incrementar mais camadas de agentes caso seja necessário. A camada de Agentes (A) é composta por profissionais de TI, os quais são agentes de Tecnologia da Informação (TI) responsáveis por gerenciar bancos de dados das instituições de saúde em redes internas ou externas por meio de softwares, sendo capazes ainda de gerir complexas estruturas computacionais no apoio a logística dos fluxos informacionais de trabalho de uma instituição, sendo neste caso, uma instituição de saúde. Esse agente direcionado aos serviços tecnológicos em SIS, é responsável também por acessar os dados de saúde internos de todas as camadas da arquitetura do PROV-Health, pois tem total acesso e autorização a toda rede de dados de saúde. Podem ser administradores de redes de dados, analistas de sistemas, engenheiros de software, entre outros profissionais de TI. Em alguns casos, esses agentes também contribuem para o desenvolvimento de novos processos computacionais necessários para que os fluxos informacionais existentes possam funcionar da melhor forma possível, e de fato, os dados possam ser gerenciados. A Camada de Agentes (A) pode ainda ser composta pelos sistemas ou plataformas inteligentes de saúde, sendo softwares autônomos considerados agentes que podem acessar os dados de saúde da mesma forma que os profissionais de TI, contendo as mesmas funções. Basta serem projetados para automação de tarefas pré-definidas por meio de algoritmos que, em muitos casos, são associados a técnicas de inteligência artificial para desempenharem funções de gerenciamento dos dados de saúde sem intervenção humana.

A Figura 10 apresenta a camada de Agentes (A) proposta no PROV-Health de forma separada.

Figura 10 – Camada de Agentes (A) do PROV-Health



Fonte: Elaborado pelo autor (2023).

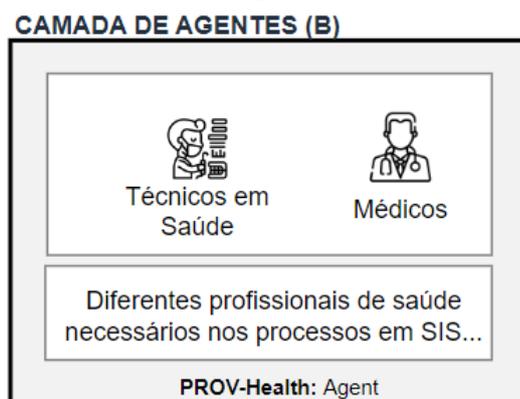
É importante ressaltar que foram destacados nesta camada de Agentes (A) os principais agentes comumente envolvidos nos processos computacionais para a execução da instrumentalização genérica dos serviços de saúde em diferentes SIS. Ressalta-se, ainda, que os profissionais de TI e os possíveis sistemas ou plataformas inteligentes de saúde que realizarão os processos computacionais relacionados aos dados de proveniência das instituições de saúde contendo SIS, deve além de cumprir os requisitos da Lei Geral de Proteção de Dados, nº 13.709, de 14 de agosto de 2018 (BRASIL, 2018), ter a devida autorização por parte dos gestores das instituições de saúde, as quais terão seus dados de saúde absorvidas de suas bases de dados originais pelo PROV-Health, para que possa dar andamento em todo processo de uso de diferentes ferramentas tecnológicas que sejam necessárias.

Já na camada de Agentes (B), ela é composta por técnicos em saúde e médicos. Os técnicos em saúde são agentes que fazem a coleta de dados sobre o estado de saúde de pacientes através de triagens, exames físicos, exames laboratoriais, anamneses entre outros. Tem acesso aos dados do SIS de maneira a complementar o prontuário do paciente com os procedimentos executados, bem como, interage com outros SIS. Esse agente pode ser, por exemplo, técnicos em radiologia, técnicos em enfermagem, técnicos em farmácia, entre outras formações técnicas em saúde. Já os médicos são agentes responsáveis por atender o paciente, encaminhar solicitações médicas, avaliar os exames realizados, realizar prescrições de medicamentos, diagnósticos e laudos, bem como, fazer o acompanhamento e tratamento do paciente. Esse agente tem acesso aos dados dos SIS e em especial, ao prontuário do paciente para tomadas de decisão, e adicionar informações complementares conforme atendimentos, procedimentos e acompanhamentos executados. Esse agente pode ter qualquer especialização médica, como por exemplo: dermatologista, cardiologista, radiologista, neurologista, entre outras especialidades médicas. Importante destacar que, esses agentes não realizam nenhum procedimento técnico

computacional no uso de ferramentas computacionais, apenas fazem parte da inserção direta ou indireta de dados de saúde no que compete as funções clínicas e operacionais de saúde para compor o ciclo de instrumentalização dos serviços de saúde em SIS.

Na Figura 11 é possível visualizar a camada de Agentes (B) proposta no PROV-Health de forma separada.

Figura 11 – Camada de Agentes (B) do PROV-Health



Fonte: Elaborado pelo autor (2023).

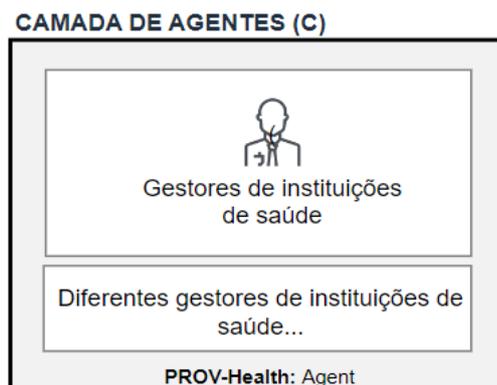
Importante ressaltar que, foram destacados nesta camada de Agentes (B), os principais agentes comumente envolvidos no ciclo da instrumentalização dos serviços de saúde em diferentes SIS no que se refere à manutenção dos dados de saúde. Esses agentes são responsáveis por quase todos os dados contidos no ecossistema dos diferentes SIS, onde diversos serviços médicos de diversas especialidades médicas, estão envolvidos com esses agentes. Esses agentes têm acesso a Camada de Atividades (A) por meio da Camada de Entidades (A). Diferentes profissionais de saúde necessários nos processos em SIS no PROV-Health podem fazer parte da instrumentalização dos serviços de saúde em SIS, como por exemplo: enfermeiros, farmacêuticos, fisioterapeutas, entre outros profissionais, podendo acessar e inserir dados de saúde mediante autorização prévia. Está contida também nesta camada, os requisitos da Lei Geral de Proteção de Dados, nº 13.709, de 14 de agosto de 2018 (BRASIL, 2018), onde a consulta e visualização dos dados de saúde devem manter o sigilo devido.

Em relação à Camada de Agentes (C), essa contém os gestores de instituições de saúde, os quais tem a função de coordenar processos de trabalho em instituições de saúde, realizar processos de planejamento estratégico, desde os requisitos de instituições colaboradoras aos recursos financeiros das instituições de saúde, além de definir fluxos de investimentos de diversos setores. Tem autorização para acessar dados de saúde e delegar funções de manutenção

desses dados aos agentes autorizados. São responsáveis por fazer com que as instituições de saúde funcionem de forma eficaz e completa, gerenciando o estado da infraestrutura física e tecnológica, além de acompanhar as atividades médicas e seus indicadores e metas pactuadas.

A Figura 12 apresenta a Camada de Agentes (C) proposta no PROV-Health de forma separada.

Figura 12 – Camada de Agentes (C) do PROV-Health



Fonte: Elaborado pelo autor (2023).

Destaca-se que a Camada de Agentes (C) pode delegar funções com base nas políticas internas de cada instituição de saúde, respeitando as responsabilidades de cada agente no contexto de SIS. Esses agentes podem ter acesso a todas as camadas da arquitetura do PROV-Health por possuírem autorização máxima para visualizar qualquer processo que seja necessário para uma tomada de decisão necessária. Ainda, nesta camada, é previsto diferentes gestores de instituições de saúde a fim de compactuar com os principais profissionais responsáveis pelo gerenciamento de todas as atividades envolvidas em uma instituição de saúde que contém diferentes SIS.

Já na Camada de Agentes (D), estão contidos os pacientes, os quais realizam diferentes exames médicos em diferentes equipamentos médicos, como por exemplo: Raio X³⁶, Eletrocardiograma³⁷, Eletroencefalograma³⁸, entre outros. Seus dados de saúde são inseridos e armazenados em diferentes SIS, tendo acesso posteriormente ao resultado dos procedimentos realizados, como relatórios de exames, diagnósticos, receituários, atestados e o laudo descrito pelo médico ao final do processo de seu exame médico. Esse agente não tem acesso aos dados internos dos SIS.

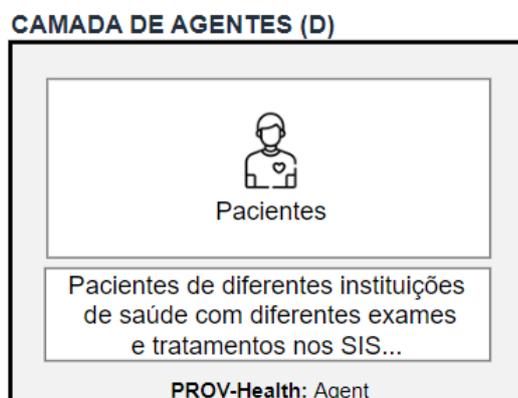
³⁶ Exame que utiliza método de imagem utilizado para produzir imagens de estruturas internas do corpo.

³⁷ Exame de avaliação da atividade elétrica do coração através de eletrodos fixados na pele.

³⁸ Exame que analisa a atividade elétrica cerebral espontânea, captada através da utilização de eletrodos.

A Figura 13 apresenta a Camada de Agentes (D) proposta no PROV-Health de forma separada.

Figura 13 – Camada de Agentes (D) do PROV-Health



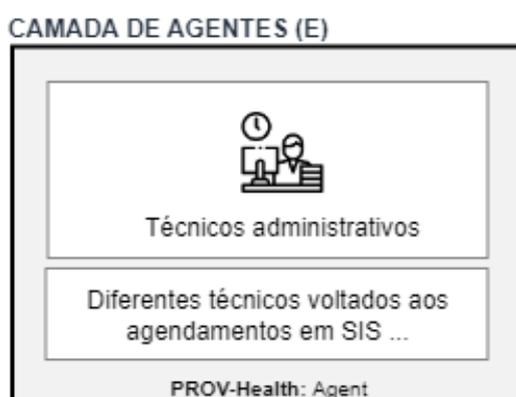
Fonte: Elaborado pelo autor (2023).

Vale destacar que os pacientes fazem parte da instrumentalização genérica dos serviços de saúde em SIS por conterem dados de saúde de seu PEP em diferentes instituições de saúde. Esses agentes possuem acessos aos dados voltados às prescrições, receituários e laudos de seus tratamentos. Importante ressaltar que esse agente só pode acessar a Camada de Atividades (A) por meio da camada de Entidades (A), onde estão inseridos seus dados de saúde.

Já em relação à Camada de Agentes (E) onde estão contidos os técnicos administrativos, os quais realizam os agendamentos necessários para todos os procedimentos clínicos voltados ao paciente.

A Figura 14 apresenta a Camada de Agentes (E) proposta no PROV-Health de forma separada.

Figura 14 – Camada de Agentes (E) do PROV-Health



Fonte: Elaborado pelo autor (2023).

Na prática, esses agentes fazem a coleta de dados para atendimento do paciente,

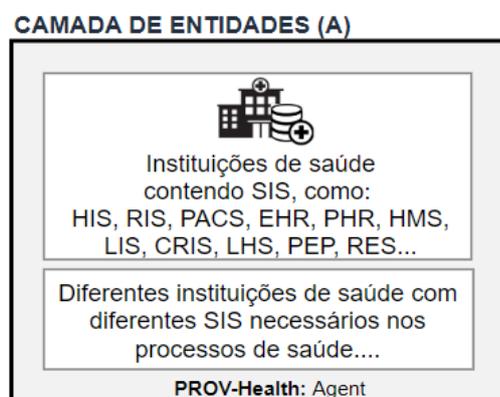
agendamentos de exames e auxilia gestores na coordenação e supervisão dos serviços de saúde, contribuindo para auxílio na compreensão das rotinas clínicas existentes. A gestão e o controle administrativo das organizações dependem de vários profissionais, e essas duas Camadas de Agentes (C) e (E) geralmente trabalham de forma conjunta nas instituições de saúde. Por fim, a Camada de Agentes (E) só tem acesso a Camada de Atividades (A) por meio da Camada de Entidades (A).

5.4.2 Camada de Entidades

Já na Camada de Entidades (A), destaca-se as instituições de saúde contendo diferentes SIS, sendo possível incrementar mais camadas de entidades que sejam necessárias. Apresenta-se os principais e mais comuns SIS voltados para os processos de instrumentalização dos serviços de saúde, sendo eles: HIS, RIS, PACS, EHR, PHR, HMS, LIS, CRIS, LHS, PEP, RES. Esses SIS consistem em um importante ecossistema de saúde de uma instituição, 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. Diferentes instituições de saúde com diferentes SIS necessários nos processos de saúde podem ser previstos para esta camada, pois para o PROV-Health, qualquer tipo de instituição de saúde e diferentes SIS, ambos que gerem dados de saúde, são considerados válidos nessa camada, independentemente de sua estrutura tecnológica.

Na Figura 15 apresenta-se a camada de entidades proposta no PROV-Health de forma separada.

Figura 15 – Camada de Entidades (A) do PROV-Health



Fonte: Elaborado pelo autor (2023).

Para a camada de Entidades (A), o PROV-Health tem a preocupação em inserir a essa camada, diferentes SIS que tenham integração entre diferentes instituições de saúde, oferecendo

soluções acessíveis e práticas quanto ao compartilhamento de dados, principalmente que gere benefícios em relação à tomada de decisão médica no auxílio direto ao paciente em suas consultas, tratamentos e exames. Desta camada são selecionados os dados que serão armazenados no servidor de dados de proveniência local do PROV-Health, 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 cumprimento dos requisitos da Lei Geral de Proteção de Dados, nº 13.709, de 14 de agosto de 2018 (BRASIL, 2018) para realização das atividades com os dados dentro das instituições de saúde. Destaca-se que a Camada de Entidades (A) pode acessar as Camadas de Agentes (A), (B), (C), (D) e (E) e gerar atividades para as Camadas de Atividades (A), (B) e (C) as quais se utilizam dos serviços de destino. A Camada de Entidades (A) ainda pode derivar suas atividades e agentes dependendo da demanda encontrada.

5.4.3 Camada de Atividades

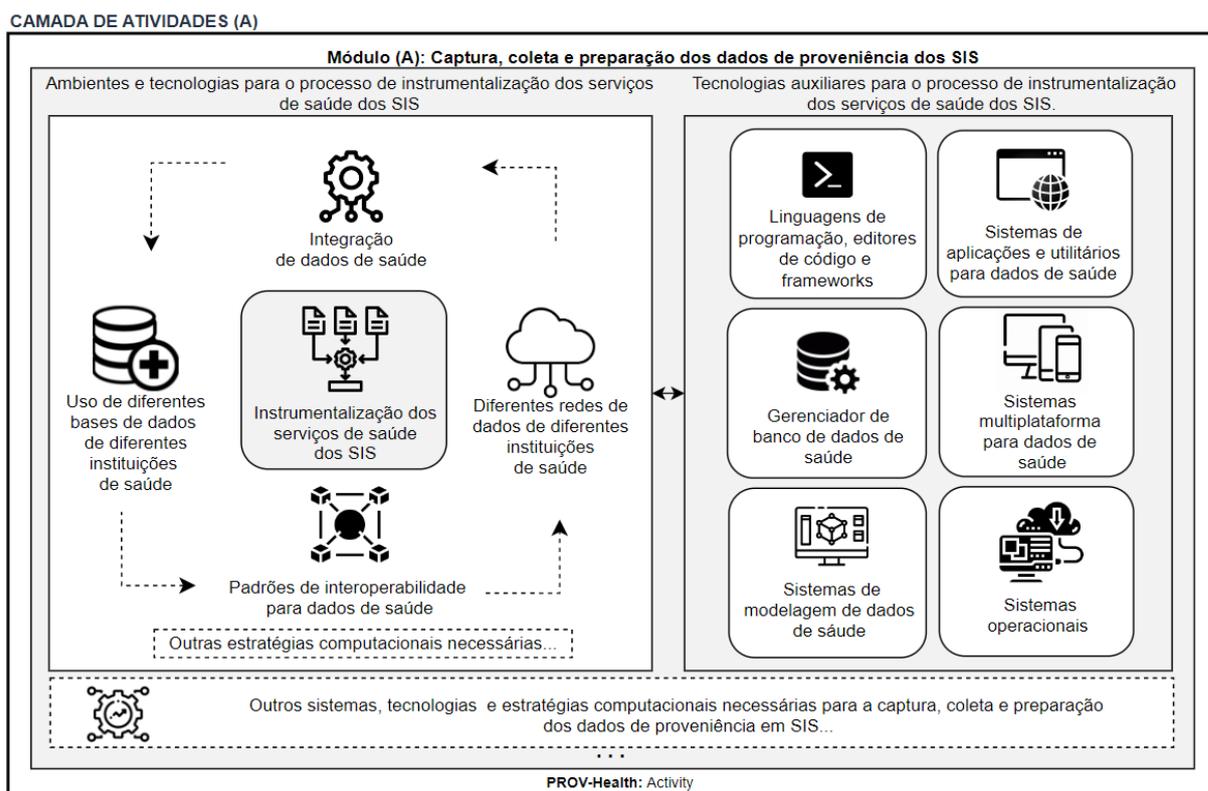
Com relação às camadas de atividades da arquitetura do PROV-Health, essas estão divididas em três: Camada Atividades (A), Camada Atividades (B) e Camada Atividades (C), sendo possível incrementar mais camadas se for necessário.

5.4.3.1 Módulo de captura, coleta e preparação dos dados de proveniência em SIS

A Camada de Atividades (A) rege o módulo de captura, coleta e preparação dos dados de proveniência em SIS, no qual, possui ambientes e tecnologias para o processo de instrumentalização dos serviços de saúde dos SIS com tecnologias auxiliares para que essa instrumentalização possa ser executada. Nessa camada, vale ressaltar, que metadados também são coletados juntamente com os dados de proveniência. Ainda nessa camada, a captura, coleta e preparação dos dados de proveniência, facilita a aplicabilidade e a interoperabilidade entre diferentes SIS.

Lembrando que na Camada de Atividades (A), a mesma mantém relações com as Camadas de Agentes (A), (B), (C), (D) e (E) já descritas anteriormente. Desta forma, para melhor compreensão da Camada de Atividades (A) da arquitetura genérica do PROV-Health, essa camada pode ser visualizada na Figura 16 de forma separada.

Figura 16 – Camada de Atividades (A) do PROV-Health



Fonte: Elaborado pelo autor (2023).

Observa-se que na Camada de Atividades (A), existe a proposta de outras tecnologias computacionais auxiliares para o processo de execução da instrumentalização genérica dos serviços de saúde dos SIS já descrita anteriormente na Figura 8, quais sejam:

- a) **Linguagens de programação, editores de códigos e frameworks:** são softwares (ambientes programáveis) para criação e edição de códigos em aplicações de algoritmos, podendo reunir uma variedade de códigos genéricos, que no caso do PROV-Health genérico podem simplificar a estruturação dos dados de saúde;
- b) **Gerenciador de banco de dados de saúde:** é uma plataforma que ajuda a organizar e armazenar os dados de saúde de forma centralizada no PROV-Health genérico, 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;
- c) **Sistemas de modelagem dos dados de saúde:** no uso de padrões de troca de informação em saúde no PROV-Health genérico, é recomendada a modelagem dos dados de saúde, utilizando-se de um sistema operacional para executar esse processo, buscando a qualidade e controle dos dados, principalmente

contribuindo para identificar os tipos de dados e seus relacionamentos. Assim, modelando os dados de saúde, recomenda-se o uso de 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;

- d) **Sistemas de aplicações e utilitários para dados de saúde:** Levando em consideração a diversidade da instrumentalização genérica dos serviços de saúde em SIS no PROV-Health, é necessário compreender através dos padrões de interoperabilidades de dados de saúde utilizados, a possibilidade de utilizar aplicações e utilitários que forneçam opções de trabalho em conjuntos de dados de saúde e imagens médicas de forma gerenciável e escalável;
- e) **Sistemas multiplataformas para dados de saúde:** a necessidade de transportar dados de saúde de diversas fontes de integração no PROV-Health genérico é indispensável. Nesse sentido, é encorajado usar 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;
- f) **Sistemas operacionais:** Para operacionalização das atividades do PROV-Health genérico, existe a necessidade de utilização de um ou mais sistemas operacionais para a execução das tarefas, tanto em interface gráfica como por programas shell³⁹ executados em uma interface de linha de comando para controlar esses sistemas.

Porém, além das tecnologias descritas, outros sistemas e estratégias computacionais necessárias para a captura, coleta e preparação dos dados de proveniência dos SIS podem ser adaptadas no contexto do PROV-Health genérico, conforme a necessidade operacional da demanda de serviços de saúde em SIS.

5.4.3.2 Módulo de organização, controle e armazenamento dos dados de proveniência dos SIS

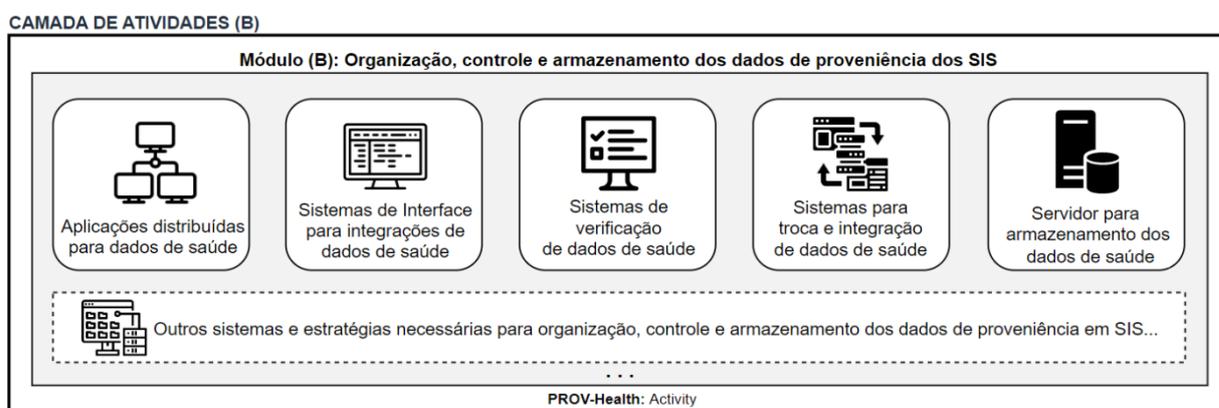
Já na Camada de Atividades (B), ela é composta pelo Módulo de organização, controle e armazenamento dos dados de proveniência dos SIS, o qual recebe os dados de proveniência

³⁹ Programa que possibilita o usuário acessar recursos do Sistema Operacional.

já capturados, coletados e preparados pela Camada de Atividade (A). Esses dados são oriundos de bases de dados originais descentralizadas de diferentes instituições de saúde com diferentes SIS. Dessa forma, são coletados os dados sem modificar a estrutura das fontes originais. Importante ressaltar que a camada de Atividades (B) é de uso exclusivo da Camada de Agentes (A), os quais tem autorização para armazenar e organizar os dados de proveniência dos SIS de forma centralizada.

Na Figura 17 pode ser visualizada de forma separada a Camada de Atividades (B) do PROV-Health genérico, para melhor compreensão.

Figura 17 – Camada de Atividades (B) do PROV-Health



Fonte: Elaborado pelo autor (2023).

Conforme observado na Camada de Atividades (B), são propostas tecnologias computacionais para apoiar todas as estratégias de do módulo de armazenamento e organização dos dados de saúde em SIS, as quais são:

- a) **Aplicações distribuídas para dados de saúde:** levando em consideração que os dados de saúde podem estar descentralizados no contexto dos SIS em diferentes redes de dados de saúde e em diferentes bancos de dados distribuídos, o PROV-Health genérico prevê a 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;
- b) **Sistemas de interface para integrações de dados de saúde:** com a quantidade de dados gerados pelos SIS, a necessidade de centralizá-los e ordená-los de forma lógica e automatizada, tornou-se um processo fundamental no PROV-Health genérico. Para isso, recomenda-se o uso de sistemas de interface para a realização da integração dos dados de saúde entre

diferentes SIS. Importante que esses sistemas sejam compatíveis com os padrões de interoperabilidade de dados de saúde, as quais facilitam e otimizam os serviços e melhoram a integração dos diferentes SIS;

- c) **Sistemas de verificação de dados de saúde:** na captura, coleta e preparação dos dados de saúde de SIS recomenda-se 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 PROV-Health;
- d) **Sistemas para troca e integração de dados de saúde:** O PROV-Health genérico prevê a utilização de padrões de interoperabilidade entre diferentes SIS, possibilitando a troca de dados de saúde nesses sistemas. Desse modo, oportuniza também, a integração entre diferentes SIS, contribuindo para que a troca de dados de saúde aconteça de forma eficiente e contínua;
- e) **Servidor para armazenamento de dados de saúde:** Levando em consideração o grande volume de dados produzidos nos setores de saúde, o PROV-Health genérico preocupa-se principalmente com o gerenciamento dos dados de proveniência dos SIS que se encontram descentralizados. Nesse sentido, como o intuito do PROV-Health genérico 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.

Entretanto, além das tecnologias computacionais mencionadas, outros sistemas e estratégias necessárias para o armazenamento e organização dos dados de proveniência dos SIS podem ser adaptadas ao PROV-Health genérico por meio de viabilização técnica entre diferentes ferramentas computacionais mediante aprovação de um profissional de TI.

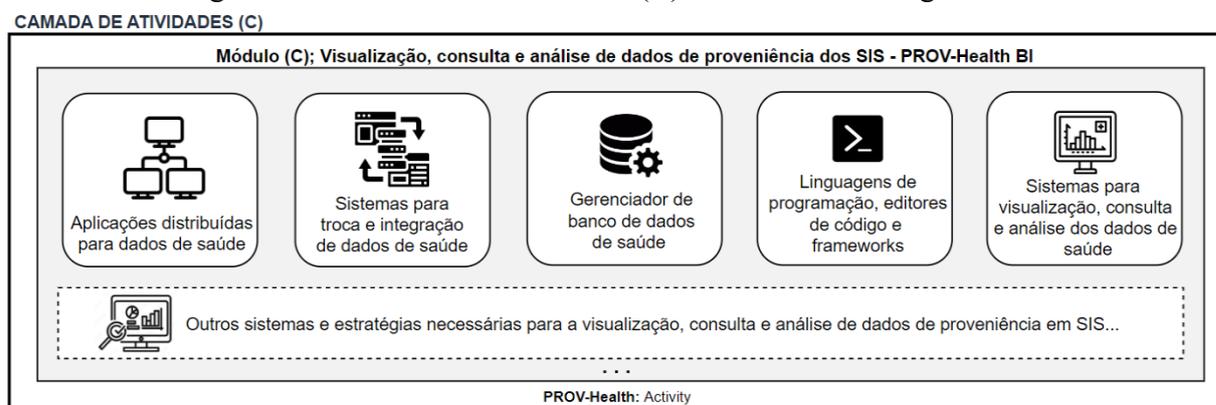
5.4.3.3 Módulo de visualização, consulta e análise dos dados de proveniência em SIS

Por fim, a última camada a ser apresentada, é a Camada de Atividades (C), a qual é composta pelo Módulo de visualização, consulta e análise dos dados de proveniência em SIS – PROV-Health BI. Essa camada recebe os dados Organizados, controlados e armazenados da Camada de Atividades (B), com intuito de possibilitar uma interface com o usuário (Camada de Agentes (A) e Camada de Agentes (C)) e, oportunizar a visualização, consulta e posteriormente análises dos dados de proveniência do PROV-Health. Essa camada tem uma

denominação de PROV-Health BI também de forma genérica, ou seja, com contextos simplificados de *Business Intelligence* (BI), os quais são absorvidos das Camadas de Atividades (A) e (B) com intuito de contribuir para apresentar os dados de proveniência contidos no PROV-Health, possibilitando oferecer suporte a gestão de negócios para acesso das Camadas de Agentes (A) e (C) para tomadas de decisões estratégicas.

Para melhor compreensão da Camada de Atividades (C) do PROV-Health genérico, pode-se visualizá-la de forma separada na Figura 18.

Figura 18 – Camada de Atividades (C) do PROV-Health genérico



Fonte: Elaborado pelo autor (2023).

Importante ressaltar que a camada de Atividades (C) é de uso exclusivo da Camada de Agentes (A) para os processos técnicos e operacionais envolvendo ferramentas computacionais, os quais têm autorização para visualizar, consultar e realizar posteriormente, possíveis análises dos dados de proveniência em SIS para geração de gráficos e relatórios destinadas a Camada de Agentes (C) para o auxílio em tomadas de decisões gerenciais no contexto dos SIS.

Ainda, na Camada de Atividades (C), ela propõe estratégias tecnológicas mediadas por tecnologias computacionais já descritas anteriormente na camada de Atividades (A) e na camada de Atividades (B), exceto a seguinte tecnologia computacional:

- j) **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 PROV-Health genérico em banco de dados para fins de análises em um contexto simplificado de ferramentas de BI em conjunto de sistemas de aplicação web multiplataforma para visualização interativa de dados de saúde. Sendo assim, esse processo contribui para a geração de tabelas, gráficos e possíveis alertas quando conectado a fontes de dados suportadas para possíveis análises necessárias no contexto de saúde.

Ressalta-se que esses sistemas de visualização, consulta e análise dos dados de saúde no

PROV-Health genérico, possibilitam gerar bases de dados apropriadas para consultas eficientes, podendo ainda possibilitar além de visualizar os dados de proveniência dos SIS por meio de tabelas e gráficos, realizar consultas em SQL⁴⁰, a qual é uma linguagem de consulta genérica suportada por diversos tipos de bancos de dados e plataformas computacionais, podendo assim, descobrir a origem de um determinado dado de proveniência em saúde, como por exemplo: qual instituição de saúde o dado pertence, dados do paciente, dados do médico que realizou o exame, entre outras consultas. Já em relação às análises, é importante destacar ainda que, a Camada de Atividades (C) é a única camada com interface para visualização geral dos dados de proveniência do PROV-Health genérico, possibilitando contribuir para a realização de um dos processos mais importantes para as Camadas de Agentes (A) e Agentes (C), o gerenciamento dos dados de proveniência dos SIS armazenados, sendo a função mais importante do PROV-Health genérico. Com a possibilidade de gerenciar os dados de proveniência advindos de diferentes SIS armazenados no PROV-Health genérico, esse processo pode contribuir para a fiscalização de uma determinada situação informacional para a compreensão das atividades, a fim de obter informações sobre falhas ou inconformidades. Além disso, obter as informações relativas a usuários, acessos, permissões, autorizações e similares para ser capaz de contribuir para a conformidade regulatória, pode ser possível nessa camada conforme a necessidade da instituição de saúde, obtendo informações sobre discrepâncias e anomalias do fluxo informacional, o que se torna essencial para identificar possíveis preocupações ou suspeitas de violações de segurança nos SIS.

Assim, destaca-se que a Camada de Agentes (A) que possui atividades relacionadas ao monitoramento e manutenção dos sistemas e banco de dados de diferentes de SIS na Camada de Atividades (C). A partir da coleta dos dados no PROV-Health genérico, podem ser relacionados a esse, dados complementares para proveniência de maneira a contribuir nas atividades desse agente. Permite gerar indicadores da eficiência do sistema, como por exemplo (falhas recorrentes que podem ser verificadas e evitadas por meio de análises no PROV-Health genérico e também na verificação de integração e troca de dados previstos entre SIS em seu fluxo informacional).

Os dados de proveniência do PROV-Health genérico, também podem ser compostos pelos dados de origens complementares que visam contemplar as atividades dos Agente (A) e Agente (C), por estes estarem mais relacionados ao funcionamento de tomadas de decisões no quesito de gerenciamento de dados.

⁴⁰ Linguagem de consulta de dados utilizada para fazer consultas, extrair informações de um banco de dados.

5.4.4 Verificação dos componentes e relações da arquitetura genérica do PROV-Health com base no W3C PROV

Com base nas premissas do modelo de Proveniência de Dados W3C PROV a observação da arquitetura genérica do PROV-Health já apresentado, pode ser compreendida da seguinte forma: as Camadas de Agentes (A), (B), (C), (D) e (E) (*PROV-Health: Agent*), tem relações definidas por (*wasAttributedTo*) com a Camada de Entidades (A) (*PROV-Health: Entity*), ou seja, todos os agentes que visualizam, acessam ou manipulam os dados de saúde, pertencem a uma entidade que mantem esses dados em seus diferentes SIS. Ainda as Camadas de Agentes (A), (B), (C), (D) e (E) (*PROV-Health: Agent*) possuem relações definidas por (*wasAssociationWith*) com as Camadas de atividades (A), (B) e (C) (*PROV-Health: Activity*), ou seja, as camadas de agentes executam atividades regidas pelas entidades. A Camada de Entidades (A) (*PROV-Health: Entity*) fica internamente na relação (*wasDerivedFrom*), ou seja, sempre derivando ou obtendo novos serviços de saúde de subsistemas dos SIS já existentes. Além disso, a Camada de Entidades (A) (*PROV-Health: Entity*) tem relações definidas por (*wasGeneratedBy*) com as Camadas de Atividades (A), (B) e (C) (*PROV-Health: Activity*), ou seja, cada entidade gera atividades necessárias aos processos de captura, coleta, preparação, armazenamento, organização, visualização, consulta e análise dos dados de proveniência dos diferentes SIS dentro do PROV-Health. Por fim, a Camada de Entidades (A) (*PROV-Health: Entity*) tem relações definidas por (*used*) com as Camadas de Atividades (A), (B) e (C) (*PROV-Health: Activity*), ou seja, caracterizando o uso dos módulos para a realização das atividades pertinentes por parte da Camada de Agentes (A), (B), (C), (D) e (E) (*PROV-Health: Agent*), respeitando os tipos de acesso e responsabilidades para com os dados no PROV-Health.

5.5 CONSIDERAÇÕES DA SEÇÃO

Nesta seção foram apresentados os elementos para concepção e desenvolvimento do PROV-Health, o instanciamento do modelo de Proveniência de Dados W3C PROV, de forma a servir de base conceitual para uma estrutura que possa ser utilizada no contexto de SIS de forma genérica. Foi proposta também uma instrumentalização genérica dos serviços de saúde em SIS para uso no PROV-Health para o auxílio na captura de dados de proveniência dos SIS, com intuito de ser ajustada para diferentes SIS. Logo após foi proposta a arquitetura genérica do PROV-Health apresentando suas camadas e seus respectivos módulos, demonstrando as

interações dos componentes (*Agent, Entity, Activity*) e suas relações entre esses componentes (*wasAssociatedWith, wasAttributeTo, wasGeneratedBy, wasDerivedFrom, Used*) conforme as bases do modelo W3C PROV. Por fim, nesta seção foi apresentada a proposta do PROV-Health, a qual pode contribuir para servir de guia para os processos de captura, coleta, preparação, organização, controle e armazenamento, visualização, consulta e análise dos dados de proveniência para diferentes SIS, além de contribuir para o gerenciamento dos dados de SIS.

Por conseguinte, houve a motivação de determinar a efetividade do PROV-Health adaptando-o a um estudo de caso realizando experimentações a ser apresentado na Seção 6.

6 EXPERIMENTAÇÃO DO PROV-HEALTH

Nesta seção são apresentadas as etapas para a experimentação do PROV-Health em um cenário real de Telemedicina, o STT/RCTM, no que compete ao Telediagnóstico por imagens médicas no âmbito da Telerradiologia situado nas dependências da UFSC. Uma contextualização é apresentada sobre o STT/RCTM, sendo posteriormente apresentada a arquitetura do cenário de aplicação e logo em seguida, os consumidores e clientes do PROV-Health são definidos. Em seguida são descritas as tecnologias computacionais utilizadas, seguindo na definição dos Profiles FHIR criados para o mapeamento dos dados coletados. Logo após, a instrumentalização específica dos serviços de saúde no STT/RCTM é descrita no uso do contexto das tecnologias utilizadas para compor o PROV-Health. Nesse contexto, é descrita a metodologia de experimentação do PROV-Health no STT/RCTM com base na experimentação realizada, partindo dos pressupostos da arquitetura específica de experimentação detalhada do PROV-Health em seu contexto tecnológico aplicada ao STT/RCTM com as respectivas experimentações definidas posteriormente, onde na sequência, são apresentados os desafios e ameaças encontrados em sua experimentação, finalizando com as considerações finais da seção.

6.1 CONTEXTUALIZAÇÃO DO AMBIENTE DE EXPERIMENTAÇÃO

A Telemedicina tem um papel fundamental referente aos cuidados de saúde em níveis nacionais e internacionais, e, especificamente em Santa Catarina, o STT/RCTM tem realizado um importante papel para a saúde catarinense nas dependências da UFSC. O STT/RCTM opera por meio de aplicações *Web* para interação entre os sistemas integrados, oferecendo suporte às diversas aplicações de Telemedicina no contexto de endoscopia, cardiologia, dermatologia e radiologia. O STT/RCTM tem mais de 18 anos de operação em Santa Catarina (SC), constituído por uma gama de tecnologias que realizam processos de trabalho médico em larga escala para acompanhamento de pacientes no contexto do Sistema Único de Saúde (SUS) em parceria com a Secretaria de Estado da Saúde de Santa Catarina. O STT/RCTM é reconhecido internacionalmente como um projeto de referência no âmbito da UFSC, fornecendo suporte a mais de 80.000 exames médicos e teleconsultorias médicas por mês, sendo considerada uma das maiores infraestruturas de telediagnóstico do Hemisfério Sul e marco para a tecnologia em saúde no Brasil.

Nesse sentido, ressalta-se que o estudo de caso desta tese se volta para um ambiente de

Telemedicina de larga escala, com um nível elevado de exames absorvidos de diferentes unidades de saúde e diferentes sistemas integrados por meio de padrões de interoperabilidade. Nesse sentido, até 2021 foram realizados mais de 14 milhões de exames de Telemedicina pelo STT/RCTM em Santa Catarina, destacando o potencial trabalho desenvolvido no âmbito da UFSC em grandes quantidades, envolvendo diversos participantes no STT/RCTM, fazendo refletir sobre os processos de armazenamento desses exames. Atualmente, o STT/RCTM já ultrapassou a marca dos 14 milhões de exames realizados em 2022, sendo considerado um dos maiores ambientes de Telemedicina do Hemisfério Sul.

Dessa forma, na perspectiva do objeto desta tese, apresenta-se um estudo de caso para experimentação do PROV-Health, o qual tem o objetivo focado no gerenciamento de dados de proveniência em SIS. Nesse sentido, a experimentação aqui realizada, teve seu foco no ambiente de Telemedicina do STT/RCTM no âmbito da UFSC, relacionado ao teliagnóstico por imagens médicas, especificamente a Telerradiologia como forma de validação da eficácia do PROV-Health. Nessa perspectiva, foi desenvolvida uma estratégia computacional indispensável (PROV-Health) para o gerenciamento dos dados de proveniência dos diferentes SIS presentes no STT/RCTM, onde a execução desta estratégia foi realizada na própria rede interna do STT/RCTM com autorização dos gestores responsáveis (APÊNDICE D), sendo exclusivamente uma validação de estudo de caso por meio do PROV-Health.

Ressalta-se que os bancos de dados originais do STT/RCTM, os quais foram estudados, não foram alterados, pois o intuito é exclusivamente acadêmico e se refere à comprovação da eficácia do PROV-Health no gerenciamento dos dados descentralizados de proveniência do STT/RCTM, escopo de estudo desta tese. Importante ressaltar que o protótipo deste estudo tornou viável sua aplicação, sendo de uso exclusivo do STT/RCTM. Portanto, o estudo está delimitado a partir dos dados que chegam nos diversos PACS, através de sistemas integrados para a rede no STT/RCTM, delimitando um estudo acadêmico de aplicação interna.

Diante da falta de métodos para o problema de gerenciamento de dados de proveniência em SIS no STT/RCTM, considerou-se esse cenário real, ideal para a experimentação do PROV-Health, por se tratar de um fluxo informacional que possui vários processos descentralizados que necessitam de estratégias computacionais que sejam capazes de gerenciar os dados de proveniência advindo de diferentes SIS. Desse modo, o PROV-Health contribui com uma estratégia computacional para o rastreamento e o gerenciamento dos dados de proveniência dos diversos dados de saúde que chegam nos vários PACS distribuídos nesse cenário de larga escala.

Importante ressaltar que, no STT/RCTM, sistemas nativos realizam o gerenciamento dos

dados existentes, porém, não gerenciam especificamente a proveniência desses. Dessa forma, o PROV-Health torna-se uma ferramenta tecnológica que vem a somar com os processos de gestão dos dados no STT/RCTM, oportunizando aos gestores mais uma ferramenta ao seu dispor para tomadas de decisões mais assertivas associadas aos dados de proveniência dos diferentes SIS integrados ao STT/RCTM.

6.2 ARQUITETURA DO STT/RCTM PARA O AMBIENTE DE EXPERIMENTAÇÃO

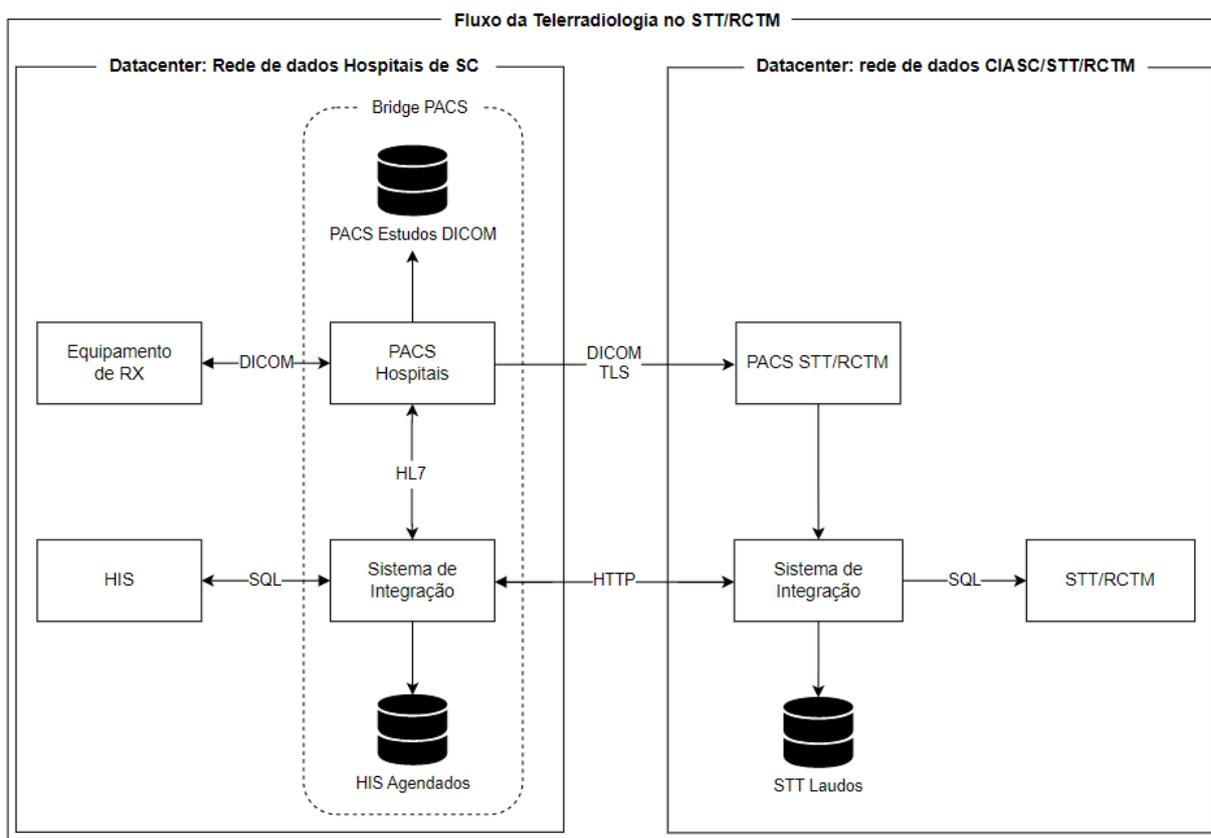
A arquitetura do STT/RCTM, a qual, para esse estudo de caso na aplicação do PROV-Health, delimita-se à Telerradiologia. Esse cenário é provido pelo Centro de Informática e Automação do Estado de Santa Catarina (CIASC), onde são executados diversos sistemas integrados, entre eles o PACS (Hospitais e STT/RCTM), que atende, simultaneamente, mais de 30 estabelecimentos de saúde (municipal e estadual), geograficamente distribuídos, por meio do uso dos padrões DICOM, HL7 e ferramentas ETL, para o fluxo de integração entre os sistemas HIS, PACS e STT/RCTM.

Nesse sentido, a estrutura de armazenamento dos dados e imagens no STT/RCTM, é disposta de forma descentralizada, contendo diversos serviços, evidenciando assim, a necessidade de gerenciar a proveniência desses dados de forma a centralizá-los em um único repositório de dados de proveniência contido na estrutura do PROV-Health, para fins de visualização, consulta e possíveis análises desses dados na rede interna do STT/RCTM. De fato, o PROV-Health possibilitou um novo repositório para centralizar os dados de proveniência dos diferentes SIS integrados ao STT/RCTM. Apesar dessa centralização de dados possibilitar um armazenamento exponencial, o repositório de proveniência local pode liberar espaço de armazenamento à medida que a Proveniência dos Dados venha a ser rastreada e analisada nos dados armazenados para a compreensão dos diferentes fluxos informacionais advindos de fontes descentralizadas. Dessa forma, o PROV-Health apresenta-se como uma possível solução que ainda não existia nesse cenário de saúde para o gerenciamento e análise dos dados de proveniência. Importante ressaltar também, que a experimentação aqui realizada, observou que os recursos do HL7 FHIR já utilizados em outros processos com dados de saúde no STT/RCTM, poderiam complementar a proveniência dos dados no PROV-Health, o qual possui bases em premissas do W3C PROV, beneficiando assim, o STT/RCTM com uma estrutura de proveniência de dados essencial aos seus processos de gerenciamento de dados.

Avaliando o contexto da experimentação do PROV-Health no STT/RCTM, a arquitetura apresentada na Figura 19 demonstra a abrangência de aplicação desta estudo de caso

no âmbito da Telerradiologia, ou seja, os exames realizados nos hospitais de Santa Catarina e integrados no STT/RCTM, utilizam equipamentos médicos e tecnologias que permitem interoperabilidade por meio do padrão DICOM, HL7 e ferramentas ETL, predominantemente em exames para a modalidade de radiologia, embora também fazem parte deste fluxo, exames das modalidades de cardiologia, endoscopia e oftalmologia. A Figura 19 apresenta com mais detalhes a arquitetura para a abrangência da experimentação do PROV-Health.

Figura 19 – Arquitetura de abrangência para a experimentação do PROV-Health no STT/RCTM



Fonte: Elaborada pelo autor (2023).

A Figura 19 apresenta a arquitetura de abrangência da experimentação do PROV-Health no STT/RCTM no âmbito da Telerradiologia, iniciando nos hospitais de Santa Catarina, utilizando-se de um datacenter para execução dos sistemas internos do HIS, sendo esse o sistema utilizado para gestão hospitalar, o qual tem como módulo principal o cadastro do paciente e realiza o gerenciamento do prontuário com todos os atendimentos e serviços realizados. Uma ponte PACS (*bridge PACS*) é delimitada, relacionando diversos processos de integrações através do protocolo de transmissão de imagens médicas DICOM com o PACS Hospitais que armazena dados na base de controle (PACS Estudos DICOM) que tem

comunicação via mensagens HL7 com o sistema de integração que armazena dados complementares na base de controle (HIS Agendados) a serem visualizados pelos médicos especialistas.

Sendo assim, na perspectiva da integração do STT/RCTM com o HIS, o fluxo tem início pelo compartilhamento dos exames agendados no HIS (HIS Agendados), que são vistos também como solicitações ou pedidos de exames realizadas por outros médicos, seja de pacientes internos ou externos, quando encaminhados de outras unidades para realização de exames. Dessa maneira, conforme a Figura 19 destaca, um sistema de integração, tem a função de permitir a coleta dos registros de agendamentos usando consultas SQL, processando-os e armazenando-os em uma base de controle para esses dados (HIS Agendados) e assim garantir a integração do PACS com o STT/RCTM.

O processamento da integração com o PACS Hospitais se dá pela transformação dos registros de agendamentos obtidos no HIS para o formato HL7, especificamente (HL7 ORM^001⁴¹), sendo então enviado para um serviço do PACS Hospitais, que por sua vez irá mapear automaticamente essa mensagem HL7 para um registro de Worklist DICOM⁴² que ficará disponível para consultas aos equipamentos.

Como prática da realização de exames de imagens, é comum a utilização do serviço de Worklist DICOM, sendo esse nos hospitais disponíveis por um serviço do PACS Hospitais. Esse serviço tem como objetivo facilitar e assegurar o preenchimento dos dados do paciente que realizou procedimentos médicos nos equipamentos, evitando erros de cadastro manual, além de garantir maior rastreabilidade dos dados entre os sistemas integrados.

Também é previsto a comunicação do equipamento de RX com o PACS Hospitais para o armazenamento das imagens realizadas pelo protocolo DICOM. Uma vez armazenado as imagens na base de controle (PACS Estudos DICOM), essas são enviadas do PACS Hospitais para o PACS STT/RCTM, e devido a fatores de infraestrutura geográfica, esse envio é enfileirado e controlado de maneira a garantir a correta integração das imagens com o PACS STT/RCTM, sendo necessário para esse caso a criptografia dos dados (DICOM TLS⁴³) para evitar exposição dos mesmos.

Os agendamentos quando recebidos no sistema de integração dos hospitais, são enviados simultaneamente para o PACS STT/RCTM e sistema de integração do STT/RCTM

⁴¹ https://hl7-definition.caristix.com/v2/HL7v2.3/TriggerEvents/ORM_001

⁴² Fornece um método para gerenciar os dados do paciente dentro do sistema de informações de radiologia de um hospital.

⁴³ <https://www.leadtools.com/help/sdk/v21/dh/to/adding-tls-security-to-a-DICOM-connection.html>

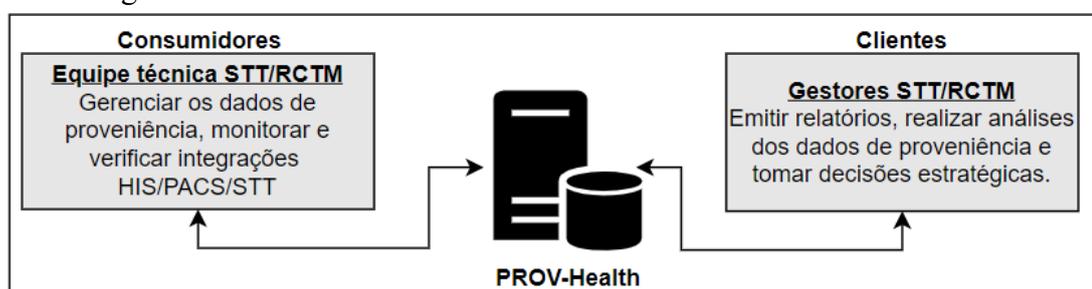
por meio do protocolo HTTP⁴⁴, sendo esse último, responsável por disponibilizar o exame no STT/RCTM ainda sem imagem, mas já visível e disponível para acessos. Quando as imagens médicas chegam no PACS STT/RCTM, um processo de integração irá verificar a existência de algum agendamento associado a essas imagens, de maneira que permita atualizar as imagens dos exames agendados no STT/RCTM e possibilite assim, a análise das imagens pelo médico especialista.

Uma vez disponíveis, os exames e respectivas imagens para acesso no STT/RCTM, é possível o diagnóstico por um médico especialista que irá emitir o laudo, o qual é previsto a integração no HIS, sendo o sistema de integração do STT/RCTM, responsável em coletar e armazenar os laudos de exames integrados na base de controle (STT Laudos) e enviar os serviços para os sistemas de integração dos Hospitais, que por sua vez, irá integrar com o HIS, finalizando assim todas as etapas de integração.

6.3 DEFINIÇÃO DOS CONSUMIDORES E CLIENTES DO PROV-HEALTH

A experimentação do PROV-Health tem o objetivo de validar sua eficiência e apresentar suas contribuições para o gerenciamento de dados de proveniência advindos dos diferentes SIS no ambiente de Telemedicina do STT/RCTM nas dependências da UFSC. Ressalta-se que o PROV-Health foi projetado para ambientes de redes de saúde internas para garantir a segurança e a Proveniência dos Dados de saúde que serão armazenados. Conforme a Figura 20, o PROV-Health destina-se a consumidores e clientes específicos dentro do STT/RCTM.

Figura 20 – Consumidores e clientes do PROV-Health no STT/RCTM



Fonte: Elaborado pelo autor (2023).

Nesse sentido, é importante destacar que, os consumidores do PROV-Health são os profissionais de TI que atuam na equipe técnica do STT/RCTM mediante o estudo de caso realizado. Esses profissionais podem atuar no gerenciamento dos dados de proveniência

⁴⁴ Protocolo de comunicação utilizado para sistemas de informação de hipermídia, distribuídos e colaborativos.

armazenados no PROV-Health, além de monitorar e verificar as integrações existentes no HIS/PACS/STT. Por se tratar de um protótipo, o PROV-Health pode sofrer uma contínua melhoria de seus processos e ampliação de seus componentes (Agentes, Entidades e Atividades) por parte da equipe técnica do STT/RCTM.

Os clientes do PROV-Health são os gestores do STT/RCTM, os quais podem emitir relatórios, realizar análises dos dados de proveniência e tomar decisões estratégicas no tocante aos dados armazenados dos diferentes SIS. A preservação digital dos dados de saúde tem um custo e o PROV-Health contribui para a redução dos custos de identificação, rastreamento, armazenamento, contribuindo também para os processos de interoperabilidade, sendo premissas observadas pelos gestores de quaisquer instituições de saúde.

6.4 TECNOLOGIAS COMPUTACIONAIS UTILIZADAS PARA COMPOR A ESTRATÉGIA COMPUTACIONAL DO PROV-HEALTH

Nessa subseção são descritas as tecnologias computacionais utilizadas na experimentação e integração do PROV-Health no STT/RCTM na perspectiva de Proveniência de Dados, adiciona novos recursos de maneira a colaborar na integração dos dados entre sistemas HIS e PACS do STT/RCTM. No Quadro 22 apresenta-se de forma detalhada cada tecnologia utilizada.

Quadro 22 – Tecnologias computacionais selecionadas para o PROV-Health no STT/RCTM

| Tecnologias computacionais | Descrição das tecnologias computacionais |
|--|---|
|  SIMPLIFIER.NET | <p>SIMPLIFIER - Plataforma de colaboração e publicação FHIR, possibilitando carregar, baixar e encontrar visualizações de recursos FHIR. Criado e gerenciado pela Firely (https://fire.ly/). O conteúdo do FHIR disponível está organizado em quatro níveis: usuários, recursos, projetos e organizações. É possível importar recursos FHIR e fazer uso da integração entre Forge by Firely e Simplifier (SIMPLIFIER.NET, 2022a). Para a experimentação do PROV-Health foi utilizado o Simplifier STT/RCTM on FHIR.</p> |
|  SIMPLIFIER.NET  <small>PROJECT</small> STT on FHIR® (R4) | <p>SIMPLIFIER - STT on FHIR® (R4) – Projeto do Sistema Integrado de Telemedicina e Telessaúde – Universidade Federal de Santa Catarina (SIMPLIFIER.NET, 2022b). Alguns perfis já existentes foram complementados com novas tratativas, de maneira a manter compatibilidades com projetos anteriores e com as novas necessidades para experimentação do PROV-Health. Todas as informações dos perfis já existentes no STT/SC\UFSC encontram-se em https://simplifier.net/STT/SC.</p> |
|  SIMPLIFIER.NET  | <p>FORGE - Com uma conta configurada no Simplifier, foi possível ter acesso a ferramenta de perfil FHIR Forge by Firely. O Forge by Firely é um aplicativo de desktop do Windows para criar perfis FHIR. Uma interface gráfica de usuário que permite criar, editar</p> |

| | |
|---|---|
|  <p>FORGE</p> | <p>e validar facilmente perfis, extensões e guias de implementação de FHIR. Oferece integração com servidores FHIR online e registros FHIR, possibilitando baixar e publicar perfis. (SIMPLIFIER.NET, 2022c). O Forge by Firely foi utilizado para criar os principais perfis necessários para experimentação no PROV-Health: SolicitacaoHospitalar, EstudosDICOM e LaudoExameHospitalar.</p> |
|  | <p>VISUAL STUDIO CODE (VS CODE) - É um editor de código-fonte com suporte integrado para JavaScript, TypeScript e Node.js com um rico ecossistema de extensões para outras linguagens e ambientes de execução como: C++, C#, Java, Python, PHP, Go, NET (VISUAL STUDIO CODE, 2022). O VS CODE foi utilizado como editor para rodar os algoritmos em Python utilizados na experimentação do PROV-Health.</p> |
|  | <p>INSOMNIA - É uma ferramenta cliente de API REST, a qual é uma interface de programação de aplicações em conformidade com as restrições da arquitetura REST (INSOMNIA, 2022), utilizado para testes de requisições no HAPI FHIR com o objetivo de validação e experimentação do PROV-Health.</p> |
|  <p>PostgreSQL</p> | <p>POSTGRES SQL - é um sistema de banco de dados relacional de objeto de código aberto que usa e estende a linguagem SQL combinada com muitos recursos que armazenam e dimensionam com segurança as cargas de trabalho de dados mais complicadas (POSTGRES.ORG, 2022). Utilizado nos hospitais para integração dos agendamentos e no STT/RCTM como base de dados principal, sendo necessário a realização de consultas para integração dos dados de proveniência com o PROV-Health BI (HAPI FHIR).</p> |
|  | <p>SQL significa <i>Structured Query Language</i>, literalmente a linguagem padrão para realizar queries. SQL é uma linguagem padrão para trabalhar com bancos de dados relacionais utilizada em diversos sistemas de bancos de dados e outras aplicações de código aberto (AWS.AMAZON.COM). O SQL foi utilizado no PROV-Health para consulta às bases de dados disponíveis nos sistemas de integração e PACS.</p> |
|  | <p>GITLAB - É uma plataforma que possibilita planejar, criar, proteger e implantar software de forma colaborativa para impulsionar os resultados mais rapidamente. O GITLAB permite que os desenvolvedores armazenem o código em seus próprios servidores, ao invés de servidores de terceiros (GITLAB, 2022). Foi utilizado o serviço códigos UFSC, sendo esse uma aplicação GITLAB para armazenar os códigos desenvolvidos para o PROV-Health.</p> |
|  | <p>PYTHON - Python é uma linguagem de programação que permite trabalhar mais rapidamente e integrar sistemas com mais eficiência. É uma linguagem de programação de alto nível, interpretada por script, imperativa, orientada a objetos, funcional, de tipagem dinâmica e forte (PYTHON.ORG, 2022). No PROV-Health foram utilizados códigos em Python para integrações com ETL.</p> |
|  | <p>JSON - é um formato leve de intercâmbio de dados. É fácil para os humanos ler e escrever. É fácil para as máquinas analisar e gerar. Ele é baseado em um subconjunto do JavaScript Programming Language Standard (JSON.ORG, 2022). Foi utilizado para mapeamento dos dados de proveniência, sendo a</p> |

| | |
|---|--|
| | <p>origem caracterizada por dados em bancos de dados SQL e o destino caracterizado por requisições em API com os dados no formato JSON, conforme previsto nas definições dos perfis desenvolvido para o PROV-Health.</p> |
|  | <p>HL7 FHIR – Software no domínio de saúde com foco na interoperabilidade de dados de saúde. Padrão que define recursos, incluindo definições de conteúdo, arquitetura, modelos e paradigmas de intercâmbio de informações em saúde. O HL7 FHIR suporta a <i>Service-oriented Architecture</i> (SOA) e <i>Representational State Transfer</i> (REST) também chamado abordagem RESTful para troca de informações. A arquitetura RESTful permite uma comunicação usando apenas o protocolo HTTP onipresente sem camadas adicionais (HL7 INTERNATIONAL INC, 2022c). O HL7 FHIR foi o padrão de interoperabilidade que instituiu o PROV-Health. O HL7 FHIR possui recursos de Proveniência de Dados baseado no modelo W3C PROV.</p> |
|  | <p>HAPI FHIR - É uma implementação completa do padrão HL7 FHIR para interoperabilidade de assistência médica em Java. É rico em recursos, incluindo: o servidor, excelente documentação. Possibilita hospedar conjuntos de dados de saúde personalizados e retê-los por um longo período, sendo possível de realizar esse processo em um servidor público (HAPI FHIR, 2023). O HAPI FHIR é onde estão alocados os dados de proveniência, ou seja, o próprio PROV-Health, sendo caracterizado como um repositório de dados de proveniência. Caracteriza por um repositório de dados pelas funcionalidades de coleta, armazenamento e gerenciamento grandes conjuntos de dados com suporte a proveniência, por utilizar estruturas FHIR. HAPI FHIR atende as necessidades do PROV-Health como repositório de dados são plataformas responsáveis por coletar, armazenar e gerenciar grandes conjuntos de dados de pesquisa.</p> |
|  | <p>DCM4CHE – coleção de aplicativos e utilitários de código aberto para empresas de saúde. No centro do projeto dcm4che está uma implementação robusta do padrão DICOM. O kit de ferramentas e biblioteca dcm4che DICOM é usado em muitos aplicativos de produção em todo o mundo. O aplicativo contém os serviços e interfaces DICOM, HL7 necessários para fornecer armazenamento, recuperação e fluxo de trabalho para um ambiente de saúde. Serviços que são disponibilizados pelo dcm4che: Web-based UI, DICOM Storage, DICOM Query/Retrieve, WADO and RID, DICOM Services, HL7 Server e IHE Services (DCM4CHE.ORG, 2023), e DCM4CHEE - contém uma interface de usuário robusta para administradores que roda inteiramente em um navegador da Web. O dcm4chee é capaz de armazenar, consultar e recuperar qualquer tipo de objeto DICOM (DCM4CHE.ORG, 2023). Ambos podem ser considerados PACS. Os exames com procedimentos de imagens são armazenados no PACS do STT/RCTM e integrados por DICOM IAN (ou SQL), e relacionando no PROV-Health a base de dados utilizada pelo dcm4chee.</p> |

| | |
|---|--|
|  | <p>ETL (Extract, Transform, Load), que significa extrair, transformar e carregar, é um processo de integração de dados que combina dados de várias fontes de dados em um único armazenamento de dados consistente que é carregado em um data warehouse ou outro sistema de destino (IBM.COM). O ETL foi utilizado no PROV-Health como abordagem para coleta de dados e para implementação do PROV-Health-BI, usando rotinas definidas em python.</p> |
|  <p>Linux</p> | <p>O Linux é um sistema operacional <i>open source</i> e gratuito, disponibilizado sob a <i>General Public License (GPL)</i>. Qualquer usuário interessado em informática pode executar, estudar, modificar e redistribuir o código-fonte, ou até mesmo vender cópias do código modificado, desde que façam isso sob a mesma licença (LINUX.ORG). O sistema operacional Linux foi utilizado no PROV-Health como servidor para execução dos serviços, aplicações e banco de dados.</p> |
|  <p>Bash</p> | <p>Bash é um acrônimo para "<i>Bourne Again Shell</i>" e foi desenvolvido em 1989 para suceder ao Bourne Shell. É um programa shell executado em uma interface de linha de comando e tem como principal função controlar sistemas operacionais, sendo comumente utilizado no Linux (OPENSOURCE.COM). O Bash foi utilizado no PROV-Health para auxiliar na automação e acesso a infraestrutura por meio de rotinas agendadas e integradas com os processos de coleta e implementação do PROV-Health-BI.</p> |
|  <p>Windows</p> | <p>Windows é um sistema operacional de multitarefas para computadores e dispositivos móveis desenvolvido pela Microsoft (MICROSOFT.COM) O Windows utilizado no PROV-Health por meio de uma licença de estudante, foi destinado para o acesso e geração dos Profiles no Forge by Firely.</p> |
|  <p>docker</p> | <p>Docker é um conjunto de produtos de plataforma como serviço que usam virtualização de nível de sistema operacional para entregar software em pacotes chamados contêineres. Os contêineres são isolados uns dos outros e agrupam seus próprios softwares, bibliotecas e arquivos de configuração (DOCKER.COM). O Docker foi utilizado no PROV-Health para execução de serviços e aplicações como PostgreSQL, Grafana, HAPI-FHIR.</p> |
|  <p>Grafana</p> | <p>Grafana é uma aplicação web de análise de código aberto multiplataforma e visualização interativa da web. Ele fornece tabelas, gráficos e alertas para a Web quando conectado a fontes de dados suportadas. É expansível através de um sistema de plug-in (GRAFANA.COM). O Grafana foi utilizado no PROV-Health para visualização dos dados de proveniência no PROV-Health-BI.</p> |

Fonte: Elaborado pelo autor (2023).

As ferramentas computacionais apresentadas no Quadro 22, serviram de base para que toda a estratégia computacional do PROV-Health fosse implementada, contribuindo para que os processos de Proveniência de Dados para experimentação no STT/RCTM pudessem ser realizados.

Importante ressaltar também que, houve a preocupação do uso da LGPD (BRASIL, 2018) em toda estratégia computacional do PROV-Health para que os processos de Proveniência de Dados realizados estivessem em consonância com a lei de proteção de dados brasileira.

6.5 DEFINIÇÃO DOS PROFILES FHIR PARA EXPERIMENTAÇÃO DO PROV-HEALTH NO STT/RCTM

O FHIR é projetado para padronizar cerca de 80% das informações de saúde e registros médicos. O padrão define a maioria dos processos e elementos de dados médicos existentes. Mas cada caso de implementação de FHIR é único, assim como os casos de uso de recursos. A especificação de recursos para a definir regras e requisitos com base no contexto utilizado é conhecido como definição de perfis (Profiles) FHIR⁴⁵. Os profiles FHIR permitem que sejam personalizadas as definições de recursos de proveniência⁴⁶ em consonância com o modelo de proveniência de dados W3C PROV, restringindo quais elementos de um recurso são necessários e quais não são. As restrições e extensões de um recurso FHIR também ajudam a melhorar a qualidade dos dados, incentivando a consistência e a capacidade de gerenciamento dos dados. Embora existam profiles FHIR com padrões prontos publicados pelo HL7, esses profiles ainda exigem que desenvolvedores de assistência médica e analistas de negócios desenvolvam seus próprios profiles FHIR que atendam aos requisitos de um projeto específico (HL7 INTERNATIONAL INC, 2022d).

A especificação base do FHIR descreve um conjunto de recursos básicos, estruturas e APIs que são utilizados em muitos contextos diferentes na área da saúde. Por esta razão, a especificação FHIR é uma “especificação de plataforma” que cria uma base comum, na qual soluções diferentes são implementadas. Como consequência, esta especificação geralmente requer maior adaptação a contextos particulares de uso (HL7 INTERNATIONAL INC, 2022d).

Logo, para auxiliar na coleta e armazenamento de dados de proveniência no STT/RCTM foi utilizada a abordagem de definição dos Profiles FHIR para as especificações do PROV-Health no instanciamento do modelo W3C PROV. Os Profiles FHIR possuem especificações de recursos de proveniência embutidos em sua caracterização⁴⁷, tornando a arquitetura do PROV-Health com relações causais já definidas intrinsecamente. Para isso, foi utilizada a

⁴⁵ <https://fhir-ru.github.io/profilelist.html>

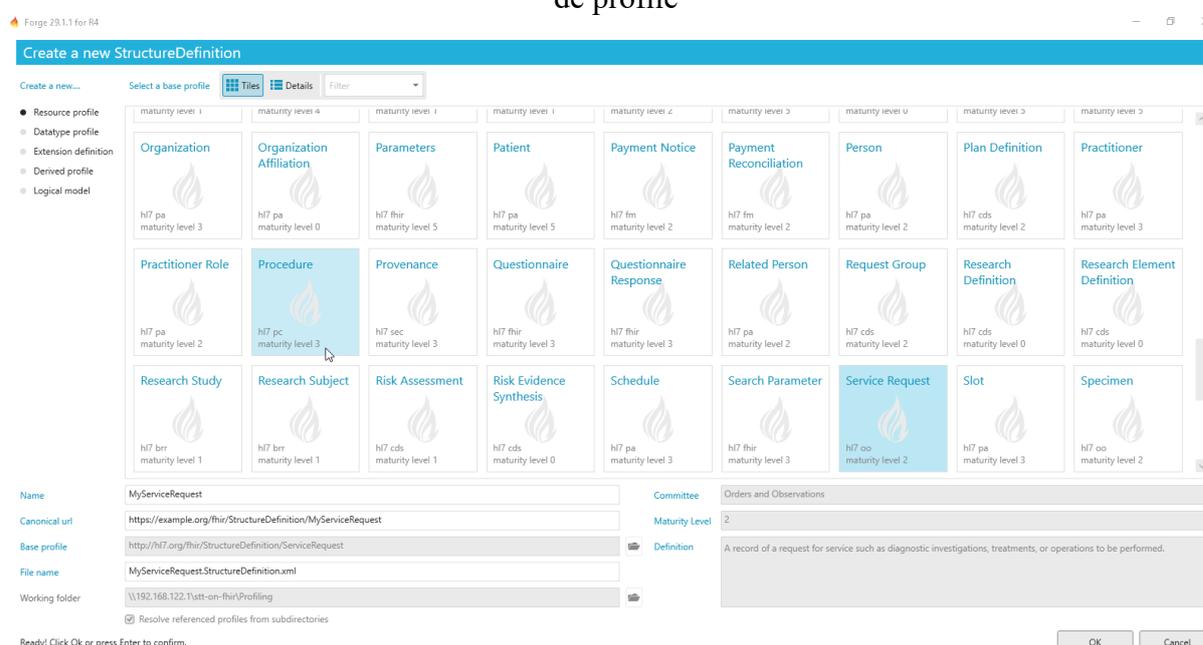
⁴⁶ <https://build.fhir.org/ig/HL7/US-Core/StructureDefinition-us-core-provenance.html>

⁴⁷ <http://hl7.org/fhir/R5/provenance-profiles.html>

ferramenta computacional Forge by Firely (um editor FHIR manual para criar perfis FHIR, modelos de dados e extensões) para a definição e validação dos perfis utilizados no PROV-Health, sendo eles: *SolicitacaoHospitalar*, *EstudoDicon* e *LaudoExameHopitalar*, respectivamente direcionados a coleta dos agendamentos de exames, imagens de exames realizados e laudos emitidos para os exames.

A seguir, a Figura 21 apresenta a tela inicial do Forge by Firely FHIR onde são criados os perfis essenciais para execução do PROV-Health no STT/RCTM.

Figura 21 – Tela principal do Forge FHIR para a criação de uma nova definição de estrutura de profile



Fonte: Forge by Firely (2023).

A Figura 21 apresenta um editor para a criação de perfis FHIR, extensões e modelos de dados FHIR, totalmente compatível com HL7 FHIR e rica em recursos. Com o editor Forge FHIR é possível: criar e editar perfis FHIR, criar e editar definições de extensão FHIR, criar e editar pacotes de conformidade FHIR, validar perfis FHIR, buscar e publicar perfis de/para um servidor FHIR, buscar e publicar perfis de/para um registro FHIR, definir restrições formais, definir vinculações de conjunto de valores e definir mapeamentos.

Nas próximas subseções são definidos os perfis: *SolicitacaoHospitalar*, *EstudosDICOM* e *LaudoExameHopitalar* que compõem o mapeamento dos dados de proveniência da estratégia computacional do PROV-Health, respectivamente direcionados à coleta dos dados de agendamentos de exames, imagens de exames realizadas e laudos emitidos para os exames.

6.5.1 Definição do profile *SolicitacaoHospitalar*

O profile *SolicitacaoHospitalar* foi definido para permitir a coleta dos dados (HIS Agendados) de agendamentos de exames provenientes dos hospitais (HIS). Para essa definição foi realizado a análise dos dados de origem e identificado no contexto do STT/RCTM, uma base de dados integrada diretamente com o HIS, sendo assim analisada e mapeada para um recurso FHIR *ServiceRequest*⁴⁸, visto como um contexto diretamente relacionado à prática de pedidos de exames e solicitações de laudos.

A análise para definição do profile *SolicitacaoHospitalar* foi realizada a partir do levantamento dos dados de agendamentos identificados na base de dados de nome “agenda_local”, a qual, todo hospital integrado ao STT/RCTM, possui tal base, e assim, foram identificados como relevante para fins de proveniência os campos conforme identificados na no Quadro 23.

Quadro 23 – Identificação dos campos de agendamentos de exames do HIS

| Campos | Descrição |
|-----------------------|---|
| cnes | Número do cadastro nacional de estabelecimento de saúde |
| accessionnumber | Número utilizado para associar solicitações de exames no HIS com estudos DICOM realizados e enviados para o PACS |
| codigora | Código utilizado pelo HIS para registro do atendimento do paciente |
| sequencialatendimento | Número incremental relacionado à quantidade de atendimentos realizadas por um paciente |
| codigoexame | Código do exame |
| examedescricao | Descrição do exame |
| descricaooservico | Descrição do serviço solicitado |
| descricaoorigem | Possui duas ocorrências, sendo “interno” quando o pedido é feito por atendimento iniciado no próprio hospital, e “externo” quando o pedido é feito por outro estabelecimento ou profissional de saúde |
| numeroptocolo | Número do protocolo |
| idequipamento | Identificação do equipamento |
| nomepaciente | Nome do paciente |
| sexo | Sexo do paciente |

⁴⁸ <https://hl7.org/fhir/r4/servicerequest.html>

| | |
|---------------------------|--|
| codigoprontuario | Código do prontuário do paciente |
| dataexame | Data do exame que o paciente realizou |
| horaexame | Hora do exame que o paciente realizou |
| data_inclusao | Data de inclusão do exame do paciente |
| nomesolicitante | Nome do profissional solicitante do exame (médico) |
| numeroconselhosolicitante | Número do conselho do profissional de saúde solicitante (CRM médico) |
| codigoexame | Código do exame do paciente |
| descricaoexame | Descrição do exame do paciente |
| dadosclinicos | Dados clínicos do paciente |

Fonte: Elaborado pelo autor (2023).

Para mapeamento dos dados identificados na base de dados “agenda_local”, foram criados e vinculados ao profile SolicitacaoHospitalar, alguns “NamingSystem⁴⁹” para contemplar a coleta dos dados de origem:

- a) NamingSystemIdentificadorSolicitacaoHospitalar: Definido pela composição do número CNES (Cadastro Nacional de Estabelecimentos de Saúde) da instituição, separador “_”, seguido do número de “accessionnumber”.
- b) NamingSystemCnes: para o CNES mantido pelo DATASUS (Departamento de Informática do SUS).
- c) NamingSystemPatientId: Pacientes atendidos pelos serviços de Telemedicina e Telessaúde.
- d) NamingSystemNumeroConselhoTrabalho: para o número do profissional de saúde no conselho de trabalho.
- e) NamingSystemIdentificadorPedidoExameSolicitacaoHospitalar: os mesmos são adotados no contexto de cada sistema envolvido. Podem ser identificadores únicos ou comuns a outras solicitações que façam parte do mesmo pedido.
- f) NamingSystemIdentificadorExameSolicitacaoHospitalar: Os mesmos são adotados no contexto de cada sistema envolvido com a persistência e/ou interoperabilidade de conteúdo.
- g) NamingSystemAccessionNumber: utilizado na integração entre sistemas RIS (Radiology Information System) e PACS (Picture Archiving and Communication System).

⁴⁹ <https://build.fhir.org/namingsystem.html>

- h) *NamingSystemOrigemPedidoExameSolicitacaoHospitalar*: utilizado para definir a origem dos pedidos de exame de uma solicitação hospitalar. Os mesmos são adotados no contexto de cada sistema envolvido com o objetivo de classificar ou categorizar solicitantes, sejam médicos ou estabelecimentos.
- i) *NamingSystemProtocoloSTT*: para protocolos de acesso a exames gerados pelo STT, utilizado por pacientes, médicos e para integrações entre sistemas de saúde.
- j) *NamingSystemDescricaoServicoSolicitacaoHospitalar*: para a descrição de serviços em uma solicitação hospitalar, adotados no contexto de cada sistema envolvido com a persistência e/ou interoperabilidade de conteúdo. O mesmo permite a atribuição de uma descrição de serviço a uma solicitação hospitalar.
- k) *NamingSystemIdentificadorEquipamentoSolicitacaoHospitalar*: para identificadores de equipamentos relacionados a uma solicitação hospitalar. Os mesmos são adotados no contexto de cada sistema envolvido com a persistência e/ou interoperabilidade de conteúdo, e servem para relacionar equipamentos a exames e solicitações e para auxiliar na rastreabilidade e integrações entre sistemas.

Outros recursos são relacionados no profile *SolicitacaoHospitalar*, além do *ServiceRequest*⁵⁰, sendo esses o *Practitioner*, *Patient* e *Organization*, os quais também possuem profiles associados, sendo, respectivamente, o *MedicoSolicitante*, *Paciente* e *Instituição*, embora já definidos anteriormente, foram realizadas adequações quanto cardinalidade para permitir seu uso também direcionado a coleta dos dados para fins de proveniência.

Assim, o profile *SolicitacaoHospitalar* foi definido com objetivo de representar e coletar os dados de exames agendados para o PROV-Health, sendo descrito sua estrutura de dados para utilização conforme o Quadro 24.

Quadro 24 – Estrutura de dados para utilização do Profile *SolicitacaoHospitalar* no PROV-Health

| Atributo | FHIR System | Dados de proveniência |
|-------------|--|---------------------------------|
| identifier | IdentificadorSolicitacaoHospitalar | cnes, accessionnumber |
| requisition | - | codigora, sequencialatendimento |
| code | - | codigoexame, examedescricao |
| orderDetail | NamingSystemAccessionNumber | Accessionnumber |
| | DescricaoServicoSolicitacaoHospitalar | descricaoservico |
| | OrigemPedidoExameSolicitacaoHospitalar | descricaorigem |

⁵⁰ <https://build.fhir.org/servicerequest.html>

| | | |
|------------------------|--|---|
| | ProtocoloSTT | numeroptocolo |
| | EquipamentoSolicitacaoHospitalar | idequipamento |
| | http://DICOM.nema.org/resources/ontology/DCM | modalidade |
| subject | Paciente | codigoprontuario, nomepaciente, sexo |
| occurrenceDateTim e | - | dataexame, horaexamedataexame, horaexame |
| authoredOn | - | dataexame, horaexame |
| requester | MedicoSolicitante | nomesolicitante, numeroconselhosolicitante |
| performer | Instituicao | cnes |
| note | - | dadosclinicos |

Fonte: Elaborado pelo autor (2023).

Dessa forma, especificações mais detalhadas sobre o profile *SolicitacaoHospitalar* dentro da rede do STT/RCTM, podem ser encontradas em PROJECT STT on FHIR® (R4): <https://simplifier.net/STT/SolicitacaoHospitalar/~overview>.

6.5.2 Definição do profile EstudoDICOM

O profile EstudoDICOM foi definido para possibilitar a coleta de estudos DICOM do PACS STT/RCTM (PACS Estudos DICOM) no PROV-Health, que abrange a realização dos exames de imagens realizados nos hospitais relacionados aos pedidos de exames na base de dados (pacs_db), sendo utilizado no profile, o recurso “*StudyImaging*⁵¹”, visto como diretamente associado à prática de realização e aquisição de exames de imagens.

Para essa definição foi realizado a análise do banco de dados utilizado pelo PACS, especificamente, a plataforma DCM4CHEE 2.18.1, a partir da qual foram relacionadas às tabelas Estudos e aos Pacientes, e selecionado os dados relevantes para proveniência, conforme descrito no Quadro 25.

⁵¹ <https://hl7.org/fhir/imagingstudy.html>

Quadro 25 – Dados de proveniência relevantes relacionadas a Pacientes

| Dados de proveniência | Descrição |
|-----------------------|---|
| pk | Chave primária do registro |
| study_iuid | Study Instance UID |
| study_id | Identificador do estudo fornecido pelo equipamento |
| accession_no | Accession Number utilizado para integração entre sistemas HIS/PACS |
| created_time | O horário em que o objeto foi criado; descrição: Uma descrição da rede; ID: O ID exclusivo do objeto no sistema |
| study_desc | Um recurso ImagingStudy refinado para representar um estudo em imagens corporativas |
| pat_id | A ID do paciente |
| pat_name | O nome do paciente |
| pat_sex | Sexo do paciente |
| pat_birthdate | Data de aniversário do paciente |
| mods_in_study | As modalidades envolvidas no estudo |

Fonte: Elaborado pelo autor (2023).

Para mapeamento dos dados identificados na base de dados “pacodb”, foram criados e vinculados ao profile EstudoDICOM, alguns “NamingSystem” e “Extensions⁵²” para contemplar a coleta dos dados de proveniência:

- a) NamingSystemStudyId: Definido para identificação do StudyID de um estudo DICOM;
- b) NamingSystemIdentificadorPacs: Definido para identificação do PACS do Hospital para o estudo DICOM integrado no STT/RCTM;
- c) ExtensionEstudoDICOMOrigemAetitle: Extensão que permite adicionar o PACS vinculado a origem do estudo DICOM;

Outros recursos são relacionados no profile *EstudoDICOM*, além do ImagingStudy⁵³, sendo esses o *Patient* e *Organization*, os quais também possuem profiles associados, sendo, respectivamente, o Paciente e Instituição, embora já definidos anteriormente, foram feitas adequações quanto a cardinalidade para permitir seu uso também direcionado a coleta dos dados para fins de proveniência.

Assim, o Profile EstudoDICOM foi definido com objetivo de representar e coletar os dados de exames de imagens para o PROV-Health, sendo descrito sua estrutura de dados para utilização conforme o Quadro 26.

⁵² <http://hl7.org/fhir/R4/extensibility.html>

⁵³ <https://build.fhir.org/imagingstudy.html>

Quadro 26 – Estrutura de dados para utilização do Profile EstudoDICOM

| Atributo | FHIR System | Dados de proveniência |
|-------------|--|--|
| identifier | urn:DICOM:uid | study_iuid |
| | NamingSystemStudyId | cnes, study_id |
| | NamingSystemAccessionNumber | cnes, accession_no |
| modality | http://DICOM.nema.org/resources/ontology/DCM | mods in study |
| subject | Reference (Patient) | cnes, pat_id, pat_sex, pat birthdate. pat_name |
| started | - | created time |
| description | - | study desc |
| extension | ExtensionEstudoDICOMOrigemAetitle | - |
| | NamingSystemIdentificadorPacs | ae_title |

Fonte: Elaborado pelo autor (2023).

As especificações mais detalhadas sobre o profile EstudoDICOM dentro da rede do STT/RCTM podem ser encontradas em (PROJECT STT on FHIR® (R4)):

<https://simplifier.net/stt/estudoDICOM>

6.5.3 Definição do profile LaudoExameHospitalar

O profile *LaudoExameHospitalar* foi definido para permitir a coleta dos dados de laudos de exames (STT Laudos) na base de dados (agenda_global) provenientes do STT/RCTM. Para essa definição foi realizada a análise dos dados de proveniência e identificado uma base de dados integrada diretamente com o STT/RCTM, sendo assim analisada e mapeada para um recurso FHIR “DiagnosticReport⁵⁴” visto como um contexto diretamente relacionado à prática de prescrição médica e entrega de resultados de exames.

A análise para definição do profile LaudoExameHospitalar foi realizada a partir do levantamento dos dados de agendamentos identificados na base de dados “agenda_global”, a qual é integrada ao STT/RCTM, e foram identificados como relevante para fins de proveniência, os dados conforme identificados no Quadro 27.

⁵⁴ <https://www.hl7.org/fhir/diagnosticreport.html>

Quadro 27 – Identificação dos campos *LaudoExameHospitalar* no STT/RCTM

| Campos | Descrição |
|----------------------------|---|
| cnes | Número do cadastro nacional de estabelecimento de saúde |
| idlaudo | Identificador do registro do laudo no STT |
| studyinstanceuid | Estudo (identificado por seu UID de Instância de Estudo) está sempre associado a um Paciente. |
| accessionnumber | Número utilizado para associar solicitações de exames no HIS com estudos DICOM realizados e enviados para o PACS. |
| numeroprotocolo | Número do protocolo do Laudo no STT/RCTM |
| descricaoexame | Descrição do exame no Laudo no STT/RCTM |
| codigoprontuario | Identificador do prontuário do paciente |
| nomepaciente | Nome do paciente |
| sexo | Sexo do paciente |
| datanascimento | Data de nascimento do paciente |
| dataexame, horaexame | Data do exame e hora do exame do paciente |
| ufmedicolaucao | Unidade Federativa do médico que realizou o Laudo |
| codigoconselhomedicolaucao | Número do Conselho de medicina do médico que realizou o laudo |
| nomemedicolaucao | Nome do médico que realizou o laudo |

Fonte: Elaborado pelo autor (2023).

Para mapeamento dos dados identificados na base de dados “agenda_global”, foram criados e vinculados ao profile *LaudoExameHospitalar*, alguns “*NamingSystem*” para contemplar a coleta dos dados de origem:

- a) *NamingSystemIdentificadorMedico*: definido para identificação do médico que realiza o laudo;
- b) *NamingSystemIdentificadorLaudoTelemedicina*: definido para identificação do Laudo na base de dados do sistema de Telemedicina do STT/RCTM.
- c) *NamingSystemProtocoloStt*: protocolos de acesso a exames gerados pelo STT/RCTM, utilizado por pacientes, médicos e para integrações entre sistemas de saúde.

Outros recursos são relacionados no profile *LaudoExameHospitalar*, além do *DiagnosticReport*, sendo esses o *Patient* e *Organization*, os quais também possuem profiles associados, sendo, respectivamente, o Paciente e Instituição, embora já definidos anteriormente, foram feitas adequações quanto cardinalidade para permitir seu uso também direcionado a coleta dos dados para fins de proveniência.

Assim, o profile *LaudoExameHospitalar* foi definido com objetivo de representar e coletar os dados de exames de imagens para o PROV-Health, sendo descrito sua estrutura de dados para utilização conforme o Quadro 28.

Quadro 28 – Estrutura de dados para utilização do *Profile LaudoExameHospitalar* no STT/RCTM

| Atributo | FHIR System | Dados de proveniência |
|--------------------|--|---|
| Identifier | NamingSystemIdentificadorLaudoTelemedicina | idlaudo |
| | urn:DICOM:uid | studyinstanceuid |
| | NamingSystemAccessionNumber | accessionnumber |
| code | NamingSystemProtocoloSTT | numeroprotocolo, descricaoexame |
| issued | NamingSystemProtocoloSTT | dataexame, horaexame |
| ResultsInterpreter | Instituicao | cnes |
| | Practitioner | ufmedicolaudado, codigoconselhomedicolaudado, nomemedicolaudado |

Fonte: Elaborado pelo autor (2023).

As especificações detalhadas sobre o profile EstudoDICOM dentro da rede do STT/RCTM podem ser encontradas em (PROJECT STT on FHIR® (R4)):

<https://simplifier.net/stt/LaudoExameHospitalar>

6.6 INSTRUMENTALIZAÇÃO ESPECÍFICA DOS SERVIÇOS DE SAÚDE PARA USO NA EXPERIMENTAÇÃO DO PROV-HEALTH

Com base na instrumentalização genérica dos serviços de saúde já definida na concepção do PROV-Health e na necessidade encontrada em sua experimentação, é apresentada a abrangência da instrumentalização específica dos serviços de saúde (ampliada) e a instrumentalização dos serviços de saúde (reduzida), inserida posteriormente na arquitetura específica do PROV-Health aplicado ao STT/RCTM.

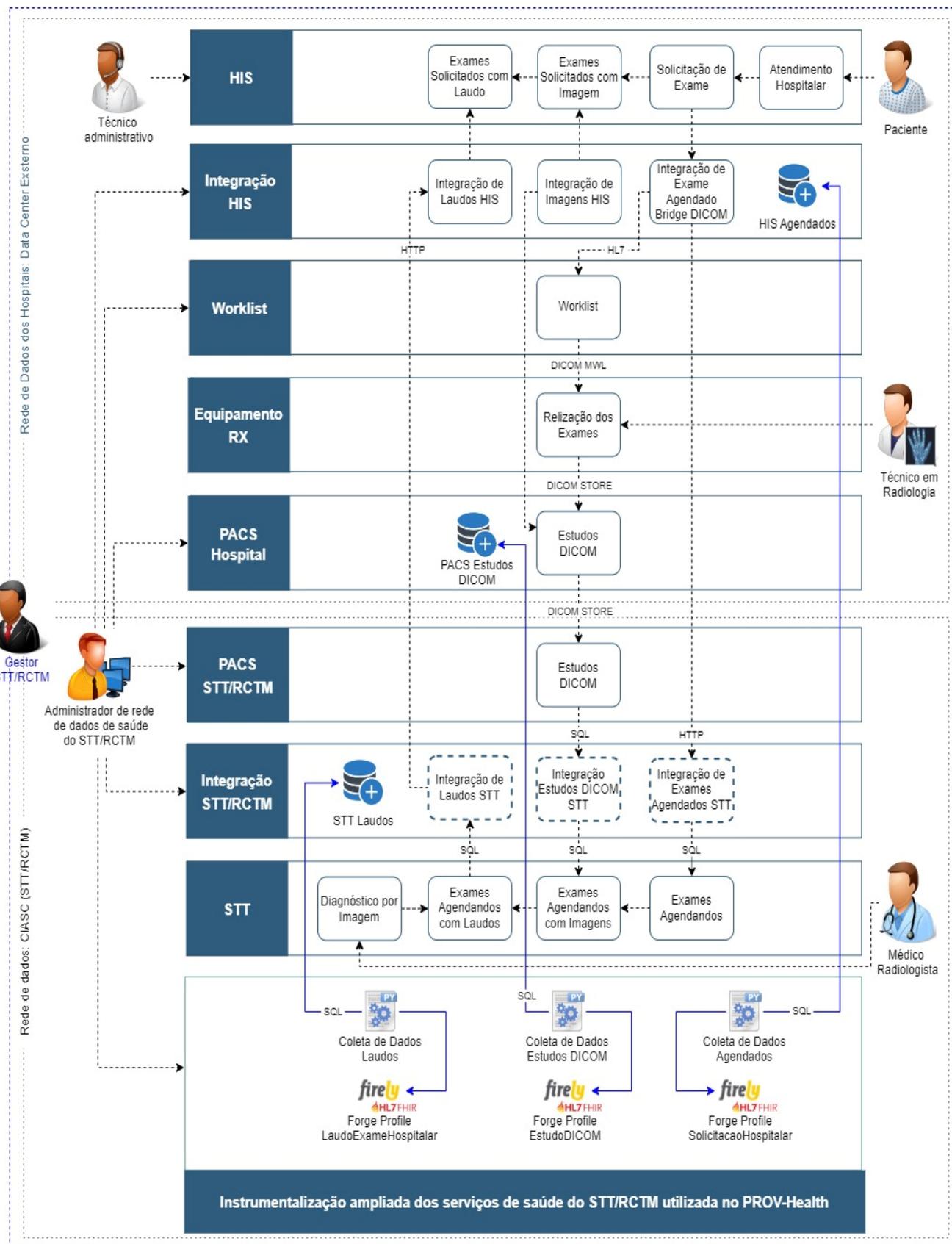
6.6.1 Abrangência da instrumentalização específica dos serviços de saúde (ampliada)

A instrumentalização específica dos serviços de saúde (ampliada) do STT/RCTM tem abrangência a diferentes sistemas e equipamentos médicos, sendo no contexto da Telerradiologia, diferentes sistemas e processos de integração, com objetivo de integrar dados de agendamentos de exames do HIS (HIS Agendados associada a genda_local), dados dos Estudos DICOM do PACS (PACS Estudos DICOM associada a pacs_db) e os dados de laudos do STT (STT Laudos associada a agenda_global). Foram identificados os agentes que fazem parte de toda a instrumentalização específica dos serviços de saúde para o contexto deste estudo

de caso para a experimentação do PROV-Health, os quais são: Gestor STT/RCTM, Administrador de redes de dados de saúde do STT/RCTM, Agente Administrativo, Técnico em Radiologia, Médico radiologista e Paciente.

Dessa forma, na Figura 22 apresenta-se a abrangência da instrumentalização específica dos serviços de saúde (ampliada) para o contexto da Telerradiologia utilizada nos processos de coleta de dados de proveniência para a experimentação do PROV-Health.

Figura 22 – Instrumentalização específica dos serviços de saúde no STT/RCTM (ampliada) para uso no PROV-Health



Fonte: Elaborada pelo autor (2023).

Conforme a Figura 22, é observado que um sistema de integração é previsto em cada hospital, contendo processos para verificação e integração dos dados de agendamentos do HIS (HIS Agendados), o quais são persistidos na base dados “agenda_local”. Para cada registro de agendamento ainda é previsto a integração com o serviço de Worklist, através do envio de uma mensagem HL7 ORM^001, a qual possui os dados do agendamento associado, necessários para criação do procedimento previsto para realização pelos equipamentos. Além da integração com o serviço de Worklist, cada registro de agendamento é integrado com outro sistema de integração disponível na infraestrutura do STT/RCTM, que por sua vez, irá integrar esse registro na base de dados “agenda_global” e controlar a integração deste, diretamente no STT/RCTM.

Os dados dos estudos DICOM (PACS Estudos DICOM), são gerados a partir da consulta dos equipamentos médicos para o serviço de Worklist, ou cadastrados manualmente pelo técnico do equipamento, e são integrados a partir do armazenamento das imagens médicas (DICOM C-STORE) no PACS do Hospital, que por sua vez é configurado para envio automático das imagens recebidas para o PACS do STT/RCTM. Quando recebido um estudo DICOM no PACS do STT/RCTM, um processo no sistema de integração do STT/RCTM, tem por objetivo, integrar registros de estudos da base de dados “pacddb”, com o STT/RCTM, sendo integrado a um exame agendado ou criado um novo exame.

Os dados dos laudos do STT/RCTM (STT Laudos) são integrados a partir do sistema de integração do STT/RCTM, que relaciona registros de laudos emitidos para exames integrados na base de dados “agenda_global” e controla por meio de outras rotinas, a integração desses registros, com o sistema de integração do Hospital, que por sua vez, irá integrar os mesmos aos respectivos exames no HIS.

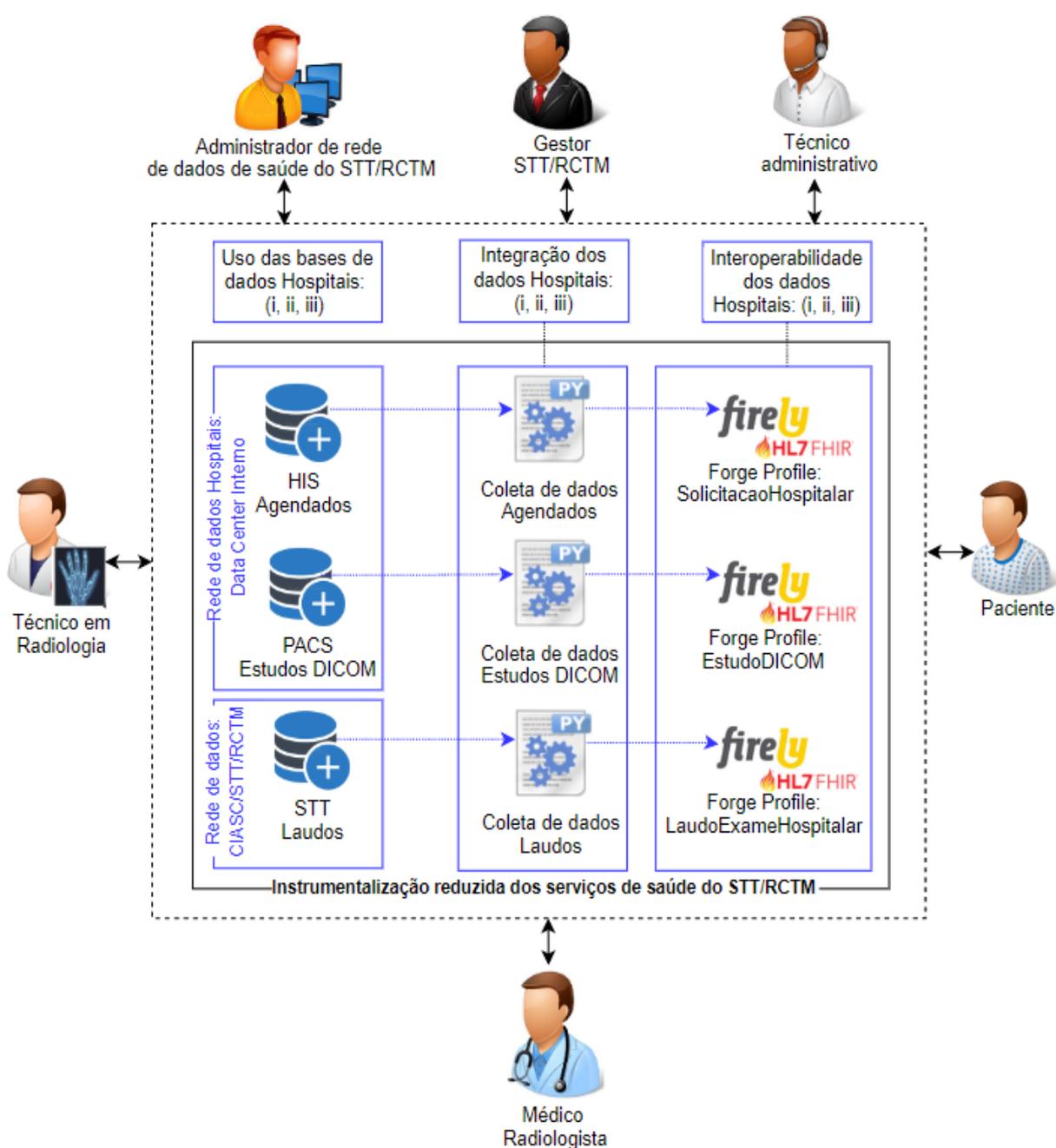
As bases de dados (HIS Agendados (agenda_local), PACS Estudos DICOM (pacs_db) e STT Laudos (agenda_global)) convergem em consultas por meio de SQL para os respectivos profiles: *SoliciracaoHospitalar*, *EstudoDICOM* e *LaudoExameHospitalar* já definidos na “subseção 6.5”.

Em relação aos agentes definidos na instrumentalização específica dos serviços de saúde no STT/RCTM no âmbito da Telerradiologia, esses serão descritos detalhadamente quando inseridos a arquitetura específica de experimentação do PROV-Health no STT/RCTM. Na próxima subseção, foi necessário apresentar a instrumentalização específica dos serviços de saúde (reduzida) para posteriormente ser inserida à arquitetura específica para experimentação do PROV-Health no STT/RCTM.

6.6.2 Instrumentalização específica dos serviços de saúde (reduzida)

Com o intuito de apresentar a instrumentalização específica dos serviços de saúde (reduzida), a Figura 23 apresenta os principais componentes e agentes necessários para posteriormente representar parte da arquitetura específica de experimentação do PROV-Health no STT/RCTM.

Figura 23 – Instrumentalização específica dos serviços de saúde (reduzida) para uso no PROV-Health



Fonte: Elaborado pelo autor (2023).

A instrumentalização específica dos serviços de saúde (reduzida) apresentada na Figura 23, juntamente com os agentes definidos e todos os processos existentes, será melhor detalhada dentro da arquitetura específica de experimentação do PROV-Health no STT/RCTM.

6.7 METODOLOGIA DE EXPERIMENTAÇÃO DO PROV-HEALTH NO STT/RCTM

Nesta subseção é realizado um detalhamento sobre a metodologia de experimentação do PROV-Health no STT/RCTM, onde experimentos automatizados foram realizados para a comprovação da eficácia e eficiência do método.

Após a contextualização do ambiente de experimentação do PROV-Health, no que diz respeito ao STT/RCTM no âmbito da Telerradiologia como delimitador deste estudo de caso, foi possível definir os consumidores e clientes do PROV-Health, assim como as ferramentas tecnológicas empregadas em sua estratégia computacional. Foi necessária a definição de Profiles FHIR para mapeamento dos dados de proveniência a serem coletados no STT/RCTM, bem como instrumentalizar os serviços de saúde do STT/RCTM apresentando-as de forma ampliada e reduzida para inserção dessa estrutura na arquitetura específica de experimentação do PROV-Health, a qual foi baseada no instanciamento do modelo de proveniência de dados W3C PROV, contendo agentes que possuem tipos diferentes de autorização de acesso ou manipulação dos dados, assim como entidades que fornecem os repositórios de dados para a realização das atividades de captura, coleta, preparação, organização, controle, armazenamento, visualização, consulta e análise dos dados de proveniência do STT/RCTM.

Para a execução da coleta dos dados de proveniência por meio do PROV-Health no cenário do STT/RCTM, o agente responsável por essa operação é o administrador de redes de dados de saúde com as credenciais e autorizações definidas pelo gestor do STT/RCTM. De modo geral, foi definido um servidor FHIR (HAPI FHIR) como repositório de dados de proveniência local, o qual contribui para a proveniência de dados em sua estrutura⁵⁵, visto que, é assegurado o mapeamento da proveniência de dados no HL7 FHIR de forma nativa em sua estrutura em compatibilidade com modelo W3C PROV⁵⁶, base para concepção do PROV-Health. Dito isto, traz à tona um importante ponto de destaque na experimentação do PROV-Health no STT/RCTM, em que todas as relações causais advindas do modelo W3C PROV estão intrinsicamente inseridas em todos os processos, pois o HL7 FHIR tem sua base de proveniência no W3C PROV de forma nativa. Assim, a plataforma HAPI FHIR foi utilizada para implantação

⁵⁵ <https://hapifhir.io/hapi-fhir/apidocs/hapi-fhir-structures-dstu3/org/hl7/fhir/dstu3/model/Provenance.html>

⁵⁶ <https://www.hl7.org/fhir/provenance-mappings.html>

do servidor FHIR e os serviços relacionados a sua execução foram instalados e configurados no Datacenter do CIASC⁵⁷, utilizando containers Docker, permitindo comunicação com todos os hospitais de Santa Catarina e outros serviços também disponíveis nesse Datacenter. Para este estudo de caso foram selecionados os seguintes hospitais: Hospital Infantil Joana de Gusmão (HIJG), Hospital Governador Celso Ramos (HGCR) e Hospital Geral de São José (HGSJ). Esses hospitais foram selecionados por se tratar de instituições de saúde que possuem os maiores volumes de dados de saúde de Santa Catarina (SC) e pela infraestrutura de rede de acesso viável para obtenção dos dados por meio de um administrador de redes de dados de saúde credenciado ao STT/RCTM.

Portanto, com objetivo da experimentação do PROV-Health, foram implementados scripts Python para contínua coleta de dados de proveniência, com rotinas de integração para extração de dados de bases SQL e mapeamento desses dados para os respectivos recursos FHIR associados já definidos. Importante destacar que, o PROV-Health tem a capacidade de ficar continuamente coletando dados de proveniência dos experimentos, porém, o mesmo foi pausado para fins de experimentação no estudo de caso desta tese no período de 01 de novembro de 2022 a 01 de janeiro de 2023.

Nesse sentido, foram realizados três experimentos para coletas de dados de proveniência na Camada de Atividades (A) da arquitetura específica para experimentação do PROV-Health, realizando assim, os testes e validações conforme descrito no Quadro 29.

Quadro 29 – Metodologia da experimentação para coleta de dados de proveniência no STT/RCTM

| Experimento 1: Coleta de dados Agendados |
|--|
| Sistema: HIS Base de dados: agenda_local Recurso FHIR: ServiceRequest Profile: SolicitacaoHospitalar Descrição: Exames agendados e integrados com o HIS (HIS Agendados) Resumo: Coleta de agendamentos (origem descentralizada por hospital, base de dados agenda_local); Armazenamento de exames agendados (pedidos de exames) no PROV-Health usando Profile SolicitacaoHospitalar que é um Resource ServiceRequest. |
| Experimento 2: Coleta de dados Estudos DICOM |
| Sistema: PACS Base de Dados: pacddb Recurso FHIR: ImagingStudy Profile: EstudoDICOM Descrição: Estudos DICOM recebidos e armazenados no PACS (PACS Estudos DICOM) Resumo: Coleta de Estudos DICOM do PACS dos Hospitais que abrange realização dos |

⁵⁷ <https://www.ciasc.sc.gov.br/nuvem-e-datacenter/>

exames de imagens realizados nos hospitais relacionados aos pedidos de exames; Armazenamento dos dados de estudos DICOM no PROV-Health usando Profile EstudoDICOM que é um Resource ImagingStudy.

Experimento 3: Coleta de dados Laudos

Sistema: STT

Base de Dados: agenda_global

Recurso FHIR: DiagnosticReport

Profile: LaudoExameHospitalar

Descrição: Laudos emitidos no STT/RCTM e integrados com os hospitais (STT Laudos)

Resumo: Coleta de laudos do STT/RCTM relacionado aos exames dos hospitais; Armazenamento no PROV-Health usando o Profile LaudoExameHospitalar que é um Resource DiagnosticReport.

Fonte: Elaborado pelo autor (2023).

Já em relação ao controle (integrações) dos dados de proveniência por meio do PROV-Health na experimentação no STT/RCTM, o mapeamento dos dados de proveniência coletados no PROV-Health, foram realizadas análises das fontes de dados e criados os respectivos perfis no padrão HL7 FHIR, seguindo abordagens e compatibilidade com o STT on FHIR: SolicitacaoHospitalar (Base de Dados "agenda_local"), EstudosDICOM (Base de Dados "pacodb") e LaudoExameHospitalar (Base de Dados "agenda_global"). Esses perfis permitem a definição de uma estrutura de dados prevista para a entrada e validação desses dados no PROV-Health experimentado no STT/RCTM. Assim, as relações causais do PROV-Health no instanciamento do modelo W3C PROV estão intrínsecas para os processos de integrações. Nesse sentido, são relacionados três experimentos direcionados às atividades do administrador de redes de dados de saúde do STT/RCTM. Esses experimentos possuem aplicações e rotinas próprias para criar e/ou atualizar os recursos no PROV-Health, sendo ajustados os Perfis para atender as necessidades de cada experimento.

Assim, foram realizados três experimentos para integrações de dados de proveniência na Camada de Atividades (B) da arquitetura específica na experimentação do PROV-Health para os testes e validações conforme descrito no Quadro 30.

Quadro 30 – Metodologia realizada para a integração dos dados de proveniência no STT/RCTM

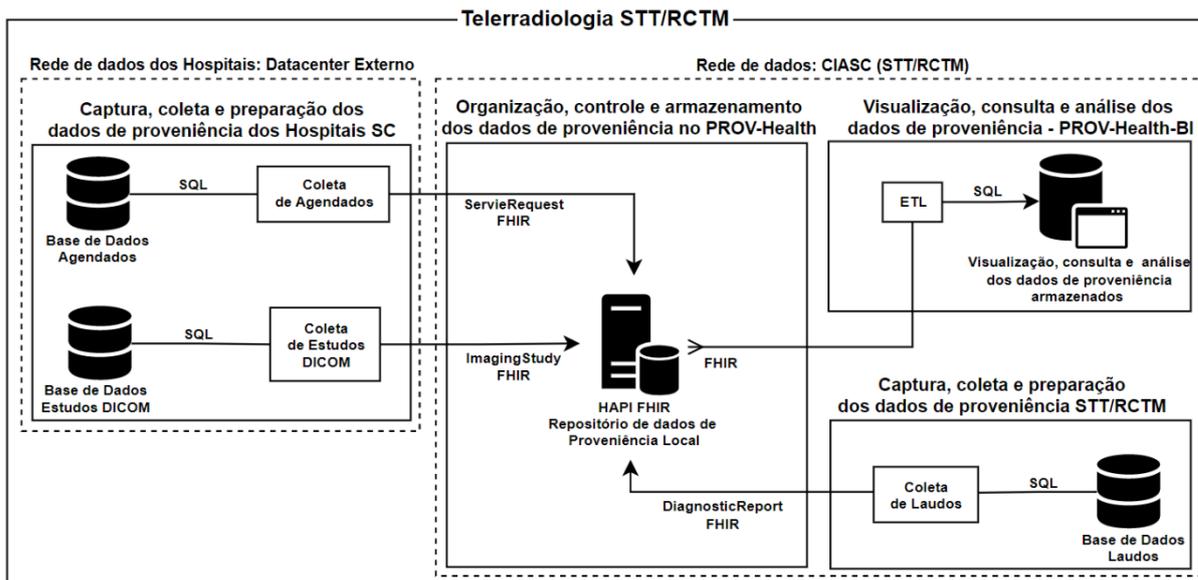
| |
|--|
| Experimento 1: Integração de solicitações hospitalares com estudos DICOM |
| Sistema: HIS Base de Dados: PROV-Health Recurso FHIR: ServiceRequest e ImagingStudy Profile: SolicitacaoHospitalar e EstudoDICOM Descrição: Exames agendados que foram vinculados no estudo DICOM Resumo: Validar se um agendamento possui estudo relacionado; |
| Experimento 2: Integração de solicitações hospitalares com laudos publicados |
| Sistema: HIS Base de Dados: PROV-Health Recurso FHIR: ServiceRequest e DiagnosticReport Profile: SolicitacaoHospitalar e LaudoExameHospitalar Descrição: Exames agendados que foram vinculados a um laudo. Resumo: Validar se um agendamento do HIS possui laudo relacionado. |
| Experimento 3: Integração de laudos com estudos DICOM |
| Sistema: STT Base de Dados: PROV-Health Recurso FHIR: DiagnosticReport e ImagingStudy Profile: LaudoExameHospitalar e EstudoDICOM Descrição: Laudos que foram vinculados a um estudo DICOM. Resumo: Validar se o laudo possui estudo DICOM relacionado. |

Fonte: Elaborado pelo autor (2023).

Após consolidação dos experimentos relacionados às coletas e integrações dos dados de proveniência no PROV-Health, um processo de ETL foi implementado em Python, com rotinas para consultas FHIR para extração dos dados de novos recursos criados. Logo em seguida são mapeados para inserção em base de dados por meio de códigos em SQL definida como “PROV-Health-BI”, pertencente a Camada de Atividades (C) da arquitetura específica para experimentação do PROV-Health, a qual é integrada com a aplicação Grafana, ferramenta para visualização, consulta e possíveis análises dos dados de proveniência do STT/RCTM. O PROV-Health-BI tem o intuito de promover possibilidades de análises dos dados de proveniência no STT/RCTM, tanto de forma técnica por parte dos administradores de redes de dados autorizados, quanto de forma gerencial por parte dos gestores, ambos relacionados ao STT/RCTM.

Sendo assim, para melhor compreensão no que se refere à execução do PROV-Health no ambiente de estudo de caso desta tese, na Figura 24 apresenta-se a arquitetura simplificada para as experimentações a serem realizadas no âmbito da Telerradiologia do STT/RCTM.

Figura 24 – Execução simplificada do PROV-Health no STT/RCTM



Fonte: Elaborado pelo autor (2023).

Observa-se na Figura 24, que a captura, coleta e preparação dos dados de proveniência do PROV-Health no STT/RCTM tem início a partir da execução dos processos Python desenvolvido e executados no mesmo servidor que o HAPI FHIR está hospedado. Dessa forma a coleta de agendados ocorre por meio de consultas SQL na base de dados de agendados (*agenda_local*) disponível no Hospital selecionado, assim como ocorre para coleta dos estudos DICOM, a qual faz as consultas SQL para base do PACS (*pacddb*) e em ambos os casos são respectivamente mapeadas para os recursos *ServiceRequest* e *ImagingStudy* e armazenadas no HAPI FHIR. Quanto à coleta de laudos, ocorre a partir da consulta SQL para base de dados na infraestrutura do CIASC (*agenda_global*) e a mesma é mapeada para recurso *DiagnosticReport* e armazenada também no HAPI FHIR.

Já em relação à organização, controle e armazenamento dos dados de proveniência na utilização do PROV-Health experimentado ao STT/RCTM, esses processos constituem a verificação de integração entre os sistemas relacionados às solicitações hospitalares, os estudos DICOM e laudos publicados no STT/RCTM.

Por fim, após os processos de coletas e integrações por meio do FHIR, ocorre o processo de visualização, consulta e análises, que por sua vez é integrado a plataforma de visualização de dados (Grafana) que conduz os gráficos dos dados de proveniência coletados pelo PROV-Health no STT/RCTM, sendo apresentados no denominado PROV-Health-BI, parte da estratégia computacional do PROV-Health.

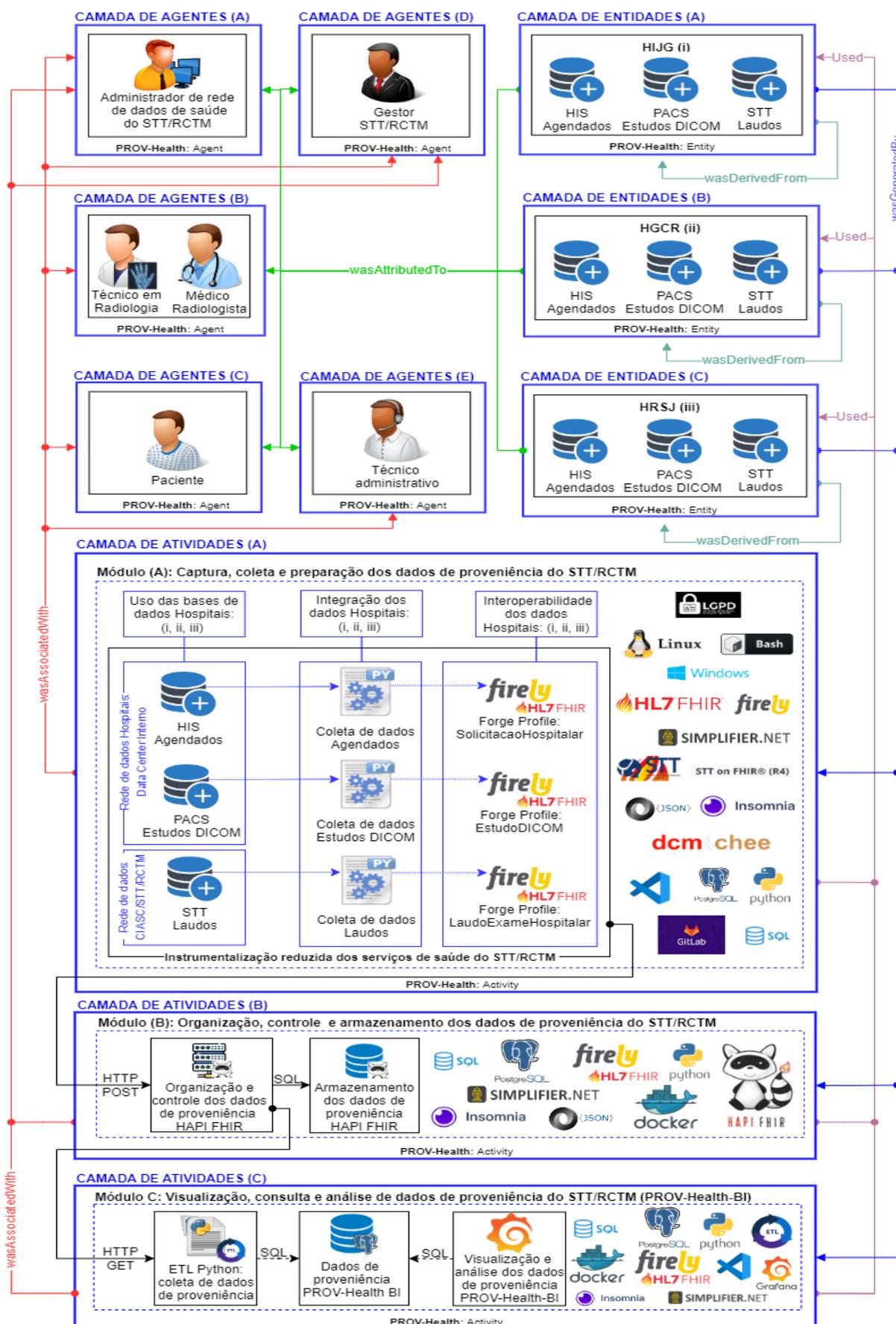
Portanto, após descrever a metodologia de experimentação do PROV-Health no STT/RCTM, é definida a arquitetura específica de experimentação do método, utilizando-se

das ferramentas computacionais já definidas na subseção 6.4, sendo possível observar detalhadamente na próxima subseção todos os processos que a constitui.

6.8 ARQUITETURA ESPECÍFICA PARA EXPERIMENTAÇÃO DO PROV-HEALTH

Adaptando a arquitetura genérica do PROV-Health para uso em diferentes SIS já descrita na Seção 5, obtém-se uma arquitetura específica para experimentação do método em estudo de caso no STT/RCTM, a qual será pautada por camadas e seus respectivos módulos, possibilitando a realização de experimentos, sendo apresentada na Figura 25.

Figura 25 – Arquitetura específica do PROV-Health para experimentação no STT/RCTM



Fonte: Elaborado pelo autor (2023).

Na arquitetura de experimentação do PROV-Health no STT/RCTM da Figura 25, preocupou-se em manter para cada camada com base na concepção do modelo de Proveniência de Dados W3C PROV apresentada na Seção 5, estipulando agentes, entidades e atividades necessárias dentro do contexto do STT/RCTM para os processos de acesso e manipulação dos dados de saúde.

Assim, conforme a Figura 25, a arquitetura específica para experimentação do PROV-Health no STT/RCTM contém as seguintes especificações:

- Cinco camadas de agentes: Camada de Agentes (A) – Administrador de redes de dados de saúde do STT/RCTM, Camada de Agentes (B) – Técnico em Radiologia e Médico Radiologista, Camada de Agentes (C) - Paciente, Camada de Agentes (D) – Gestor STT/RCTM e Camada de Agentes (E) – Técnico Administrativo;
- Três camadas de Entidades: Camada de Entidades (A) – Hospital Infantil Joana de Gusmão (HIJG), Camada de Entidades (B) – Hospital Governador Celso Ramos (HGCR) e camada de Entidades (C) – Hospital Regional de São José (HRSJ);
- Três camadas de Atividades: Camada de Atividades (A) – contendo o módulo de captura, coleta e preparação dos dados de proveniência do STT/RCTM; Camada de Atividades (B) – contendo o módulo de organização, controle e armazenamento dos dados de proveniência do STT/RCTM; e, Camada de Atividades (C) – contendo o módulo de visualização, consulta e análise dos dados de proveniência do STT/RCTM.

Nesse contexto, para melhor compreender a aplicação do PROV-Health no STT/RCTM em suas respectivas camadas apresentado na Figura 25, o mesmo será descrito de forma detalhada no uso de suas ferramentas computacionais nas próximas subseções.

6.8.1 Aplicação das Camadas de Agentes

A primeira camada de agentes a ser apresentada, é a Camada de Agentes (A), sendo definida pelo Administrador de rede de dados de saúde do STT/RCTM que pode ser visualizado na Figura 26.

Figura 26 – Camada de Agentes (A) para aplicação do PROV-Health no STT/RCTM



Fonte: Elaborado pelo autor (2023).

O Agente (A) é responsável por realizar todos os procedimentos técnicos operacionais necessários nos processos computacionais do PROV-Health, por meio do desenvolvimento de sistemas de integração para consulta aos dados de proveniência em suas respectivas bases de dados e envio desses dados para o PROV-Health, bem como, atualizações desses dados conforme atividades relacionadas a esse agente. Utiliza-se um computador com acesso liberado e autorizado (credenciais) por meio de um protocolo *Secure Socket Shell (SSH)*, que permite ao mesmo acessar e gerenciar servidores pela internet para entrar nos terminais do servidor, bem como para acessar os repositórios e bases de dados dos hospitais selecionados para esse estudo de caso (HIJG, HGCR, HRSJ) para gerenciar e realizar as manutenções devidas. O acesso às fontes de dados de diferentes instituições de saúde de Santa Catarina é realizado por servidores do STT/RCTM, os quais utilizam sistema operacional Linux e interfaces de acesso remoto via terminal Bash, suportado pelo serviço SSH, que garante comunicação criptografada mediante autorização e liberação de acesso. Além do SSH, também são utilizados os acessos às bases de dados SQL que permitem mapear e relacionar diferentes sistemas e dados médicos. Outra funcionalidade relevante, é a gestão de configuração do PACS, que permite definir fluxos de informações, coleta e exclusão dos dados, bem como o acesso a esses por meio de protocolos de interoperabilidade DICOM. Esse agente tem acesso a todas as Camadas de Entidades e Atividades da arquitetura específica do PROV-Health utilizada como base para a experimentação no STT/RCTM por meio de autorizações, sendo observado no APÊNDICE D a declaração de conhecimento do gestor do STT/RCTM para a coleta de dados e apresentada NO APÊNDICE E, o termo de compromisso para utilização desses dados.

Já na camada de Agentes (B), ela é composta pelo Técnico em Radiologia e o Médico Radiologista. O Técnico em Radiologia é o agente responsável pela operação de equipamentos de RX para execução dos exames de radiologia. Tem como atividades, o cadastro ou validação

dos dados do paciente e adicionar informações complementares aos exames realizados. Tem acesso aos dados do prontuário do paciente com os procedimentos executados. Já o Médico Radiologista é o agente responsável por realizar os laudos referentes aos exames do paciente, o qual faz o diagnóstico por análise das imagens médicas realizadas pelos equipamentos de RX. Esse agente tem acesso aos dados dos SIS dos hospitais e em especial, ao prontuário do paciente para tomadas de decisão. Vale ressaltar que, esses agentes não realizam nenhum procedimento técnico computacional no uso de ferramentas computacionais, são agentes que realizam funções clínicas e operacionais de saúde para compor o ciclo de instrumentalização dos serviços de saúde do STT/RCTM.

A Figura 27 apresenta a Camada de Agentes (B) da arquitetura específica de experimentação do PROV-Health no STT/RCTM de forma separada.

Figura 27 – Camada de Agentes (B) do PROV-Health no STT/RCTM



Fonte: Elaborado pelo autor (2023).

Foram definidos para a Camada de Agentes (B), os principais agentes envolvidos no ciclo de instrumentalização específica dos serviços de saúde do STT/RCTM já descrita anteriormente, no que se refere à manipulação e manutenção dos dados de saúde. Esses agentes são responsáveis por grande parte do ecossistema dos dados de saúde dos serviços médicos voltados à especialidade de radiologia. Esses agentes têm acesso a Camada de Atividades (A) por meio da Camada de Entidades (A), (B) e (C).

Para a Camada de Agentes (C), estão contidos os pacientes, os quais realizam exames médicos em equipamentos de RX. Esse agente é relacionado em todas as etapas do agendamento, no que compete a realização do exame ao laudo realizado no STT/RCTM. Seus dados de saúde são inseridos em seu PEP, tendo acesso posteriormente ao resultado dos procedimentos realizados, como relatórios de exames, diagnósticos, receituários, atestados e o laudo descrito pelo médico radiologista do STT/RCTM. Esse agente não tem acesso aos dados internos dos SIS.

A Figura 28 apresenta a Camada de Agentes (C) da arquitetura específica de experimentação do PROV-Health no STT/RCTM de forma separada.

Figura 28 - Camada de Agentes (C) do PROV-Health no STT/RCTM



Fonte: Elaborado pelo autor (2023).

Os pacientes fazem parte da instrumentalização específica dos serviços de saúde do STT/RCTM já descrito anteriormente, por conterem dados de saúde de seu prontuário eletrônico, ou seja, o PEP. Esses agentes possuem acessos aos dados voltados às prescrições, receituários e laudos de seus tratamentos apresentados pelo Médico Radiologista. Destaca-se que o paciente só acessa a Camada de Atividades (A) por meio da camada de Entidades (A), (B) e (C), onde estão inseridos seus dados de saúde.

Em relação à Camada de Agentes (D), essa contém o Gestor do STT/RCTM, o qual é responsável pelas tomadas de decisão do STT/RCTM no que se refere aos dados de saúde do PROV-Health.

A Figura 29 apresenta a Camada de Agentes (D) da arquitetura específica de experimentação do PROV-Health no STT/RCTM de forma separada.

Figura 29 - Camada de Agentes (D) do PROV-Health no STT/RCTM



Fonte: Elaborado pelo autor (2023).

A Camada de Agentes (D) delega funções e monitora todos os processos gerenciais do

STT/RCTM. Esse agente tem acesso a todas as camadas da arquitetura específica de experimentação do PROV-Health no STT/RCTM por meio da Camada de Agentes (A).

Já em relação à Camada de Agentes (E) está contido o Técnico Administrativo responsável pelo agendamento dos procedimentos de RX a serem realizados pelos pacientes, bem como atualizações nesses procedimentos.

A Figura 30 apresenta a Camada de Agentes (E) da arquitetura específica de experimentação do PROV-Health no STT/RCTM de forma separada.

Figura 30 – Camada de Agentes (E) do PROV-Health no STT/RCTM



Fonte: Elaborado pelo autor (2023).

Esses agentes fazem a coleta dos dados para atendimento do paciente para os procedimentos de RX, e tem um papel fundamental no auxílio ao Gestor do STT/RCTM no que diz respeito à coordenação e supervisão dos serviços de saúde, contribuindo para auxílio na compreensão das rotinas clínicas existentes no STT/RCTM. Nesse contexto, as Camadas de Agentes (D) e (E) da arquitetura específica de experimentação do PROV-Health trabalham de forma conjunta no STT/RCTM. Por fim, a Camada de Agentes (E) só tem acesso a Camada de Atividades (A) por meio da Camada de Entidades (A), (B) e (C).

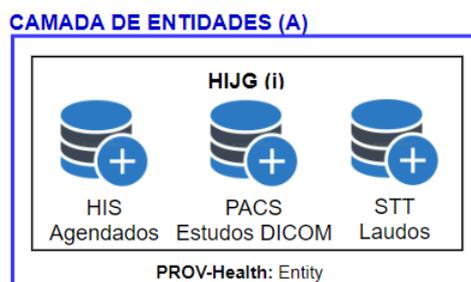
6.8.2 Aplicação das Camadas de Entidades

As camadas de entidades da arquitetura específica de experimentação do PROV-Health no STT/RCTM são compostas por três hospitais de Santa Catarina escolhidos para compor a amostra dos dados coletados para o estudo de caso aqui executado. Todas as Camadas de Entidades (A) – HIJG (i), (B) – HGCR (ii) e (C) – HRSJ (iii) contém respectivamente as bases de dados HIS agendados (agenda_local), PACS Estudos DICOM (pacs_db) e STT Laudos (agenda_global) para a execução das coletas de dados de proveniência. Estes hospitais foram selecionados por estarem entre os maiores hospitais de Santa Catarina, com maior quantidade

de dados e facilidade de acesso por parte da Camada de Agentes (A) e por estarem em servidores e redes de alta capacidade que possibilitaram uma amostragem compatível com os objetivos do PROV-Health.

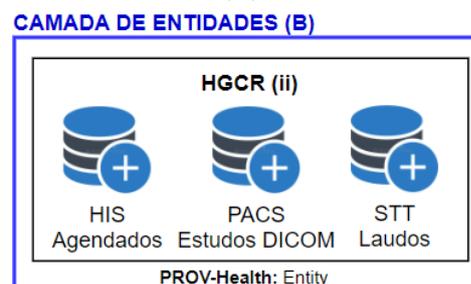
As três Camadas de Entidades (A), (B) e (C) são apresentadas de forma separadas nas Figuras 31, 32 e 33, respectivamente.

Figura 31 – Camada de Entidades (A) do PROV-Health no STT/RCTM



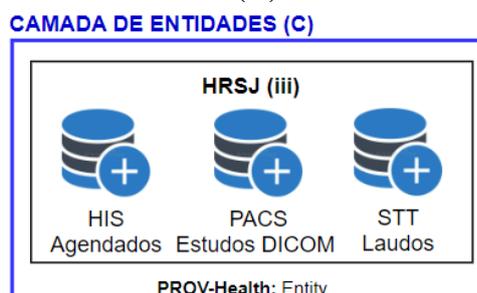
Fonte: Elaborado pelo autor (2023).

Figura 32 – Camada de Entidades (B) do PROV-Health no STT/RCTM



Fonte: Elaborado pelo autor (2023).

Figura 33 – Camada de Entidades (C) do PROV-Health no STT/RCTM



Fonte: Elaborado pelo autor (2023).

As Camadas de Entidades (A), (B) e (C) são responsáveis por fornecer os dados de proveniência da amostragem de dados do PROV-Health experimentado no STT/RCTM por meio da coleta de dados da Camada de Atividades (A) a ser apresentada nas próximas seções, realizada pela Camada de Agentes (A) já apresentada anteriormente.

6.8.3 Aplicação das Camadas de Atividades

As camadas de atividades para experimentação do PROV-Health no STT/RCTM, são divididas em três: Camada de Atividades (A) contendo o Módulo (A) de captura, coleta e preparação dos dados de proveniência do STT/RCTM, Camada de Atividades (B) composta pelo Módulo (B) de organização, controle e armazenamento dos dados de proveniência do STT/RCTM e Camada de Atividades (C) composta pelo Módulo (C) de visualização, consulta e análise dos dados de proveniência do STT/RCTM. Ferramentas computacionais já definidas na seção 6.4, constituem a estratégia computacional de todas as três camadas de atividades do PROV-Health direcionadas ao estudo de caso no STT/RCTM.

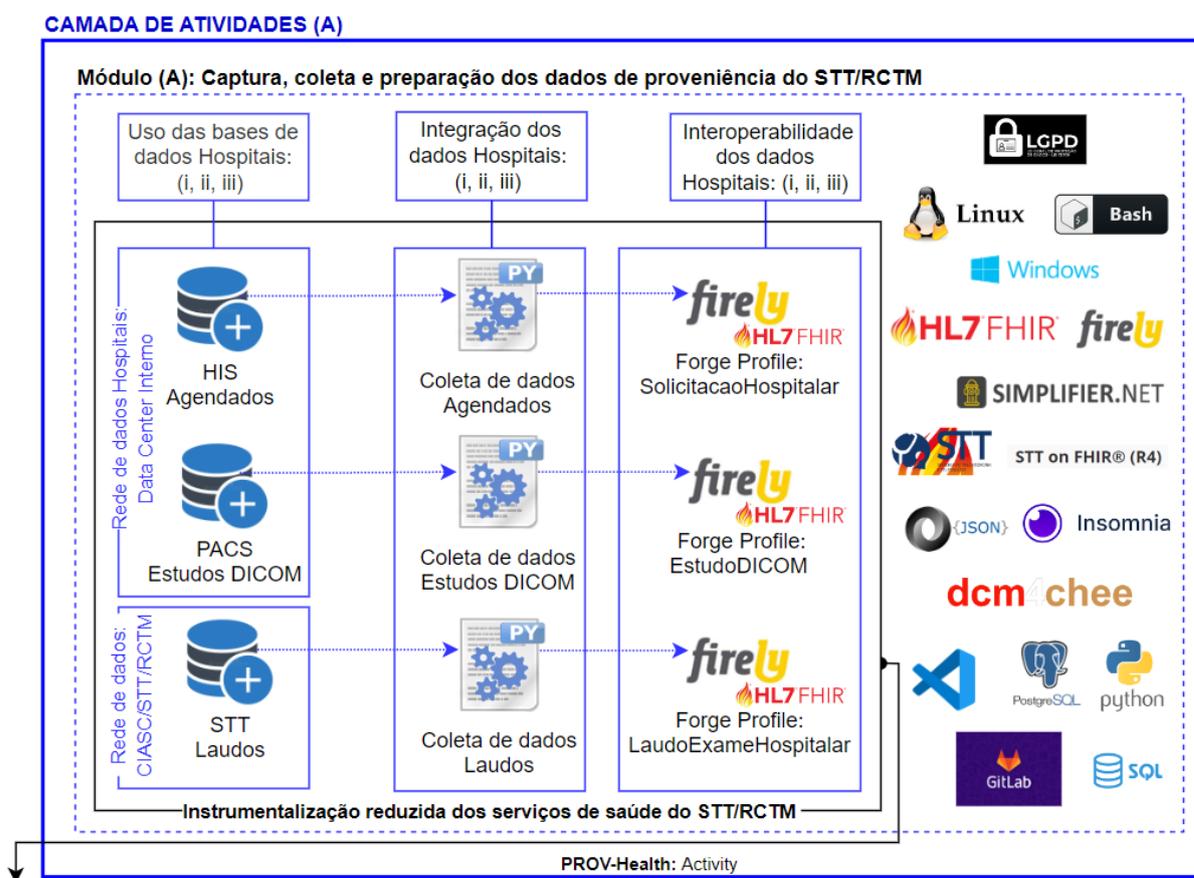
6.8.3.1 Camada de Atividades (A) – Módulo (A): captura, coleta e preparação dos dados de proveniência do STT/RCTM

A Camada de Atividades (A) contém o Módulo (A) de captura, coleta e preparação dos dados de proveniência do STT/RCTM que, por meio dos processos da instrumentalização reduzida dos serviços de saúde do STT/RCTM (já descrita na subseção 6.6.2, é capaz de traçar todo o caminho da coleta de dados necessária para que o PROV-Health tenha êxito. Importante ressaltar que, todas as Camadas de Agentes (A), (B), (C), (D) e (E) são associados a Camada de Atividades (A), porém, somente a Camada de Agentes (A) tem prioridades operacionais para manipulação dos dados de forma computacional, sendo que as Camada de Agentes (B), (C) e (E) só podem inserir ou consultar dados, exceto a Camada de Agentes (D) que pode tomar decisões operacionais em conjunto com a Camada de Agentes (A).

Assim, baseado na abrangência da instrumentalização específica dos serviços de saúde no STT/RCTM já descrita anteriormente na seção 6.6.1, a mesma foi reduzida para ser inserida na arquitetura específica de experimentação do PROV-Health, sendo possível uma melhor visualização. A instrumentalização reduzida contempla os agendamentos digitais e os processos manuais que também existem nesse cenário, contendo em sua estruturação de apoio, ferramentas computacionais que auxiliam a coleta dos dados dos hospitais referente as Camadas de Entidades (A) – HIJG (i), (B) – HGCR (ii) e (C) – HRSJ (iii). Nesse contexto, a instrumentalização reduzida dos serviços de saúde do STT/RCTM será detalhada em todos os seus processos computacionais para a coleta de dados desses hospitais selecionados para este estudo de caso.

De forma, a detalhar melhor a da Camada de Atividades (A) a Figura 34 apresenta sua estrutura contendo o Módulo (A): Captura, coleta e preparação dos dados de proveniência juntamente com a instrumentalização reduzida dos serviços de saúde do STT/RCTM para fins de experimentação deste estudo de caso. As ferramentas computacionais empregadas nessa parte da estratégia computacional do PROV-Health também são explicitadas e detalhadas posteriormente, as quais são representadas por ícones na Figura 34.

Figura 34 – Camada de Atividades (A) do PROV-Health para experimentação no STT/RCTM



Fonte: Elaborado pelo autor (2023).

De modo geral, a execução das coletas na Figura 34, foi realizada pela Camada de Agentes (A) a partir de um servidor no CIASC com acesso às bases de dados dos Hospitais (i, ii, iii). Ao executar as coletas na rede de dados do CIASC, a mesma contém a base de dados do STT Laudos, respectivamente, gerando a coleta de dados (Laudos) pelo profile (LaudoExameHospitalar). Já a coleta realizada no Datacenter interno dos Hospitais, contendo a base de dados HIS Agendados, gerando a coleta de (Agendados) por meio do Profile (SolicitacaoHospitalar) e a base de dados PACS Estudos DICOM, gerando a coleta de (Estudos DICOM) por meio do Profile (EstudoDICOM). Agendados, Estudos DICOM e Laudos fazem

parte das integrações dos Hospitais (i, ii, iii) por meio da interoperabilidade HL7 FHIR oportunizada pelos respectivos Profiles (SolicitacaoHospitalar, EstudoDICOM e LaudoExameHospitalar).

Assim, as coletas foram baseadas por três experimentos:

- Experimento 1 – Coleta de dados Agendados (*Profile: SolicitacaoHospitalar*);
- Experimento 2 – Coleta de dados Estudos DICOM (*Profile: EstudoDICOM*);
- Experimento 3 – Coleta de dados Laudos (*Profile: LaudoExameHospitalar*).

Nesse contexto, foi criado um diretório com aplicações Python, devidamente parametrizada. Os parâmetros necessários à execução das coletas por meio do PROV-Health no STT/RCTM, podem ser observados na Quadro 31.

Quadro 31 – Parâmetros necessários a execução das coletas

| Parâmetros | Descrição |
|---|--|
| organizacao_id | Possui o número do CNES da instituição em que os dados foram coletados. |
| organizacao_nome | Possui o nome da instituição em que os dados foram coletados. |
| fhir_url | Especifica a URL do serviço FHIR que os dados coletados serão armazenados após mapeados para os respectivos recursos e perfis FHIR. |
| id_instituicao_stt | Especifica o identificador da instituição no STT e utilizado na coleta de laudos para delimitar a consulta SQL por instituição. |
| db_name db_user db_pass db_host db_port | Parâmetros para acesso a base de dados utilizados por coleta, de agendamentos, estudos DICOM e laudos, e também por instituição. |
| ae_title | Especifica o nome do PACS em que os estudos DICOM foram coletados, sendo esse nome único e controla no PACS do STT, conforme cadastro de Application Entity Title (AE Title) e visa delimitar o contexto da coleta de estudos DICOM por instituição. |

Fonte: Elaborado pelo autor (2023).

Assim, foram criadas as estruturas para os três experimentos de coletas já mencionados a serem realizados no HIJG, HGCR e HRSJ, sendo as respectivas Camadas de Entidades (A), (B) e (C) do PROV-Health. Como exemplo, a estrutura do HGCR apresentada na Figura 35 em Bash Linux, é a mesma para HIJG e HRSJ.

Figura 35 – Estrutura de diretórios com aplicações de coletas parametrizadas para execução no HGCR

```

root@rctm-node005:/opt/provhealth/hgcr#
agendados_hgcr:
agendados_hgcr.sh
agendados_main.py
agendados_parses.py
agendados_services.py
agendados_template.json.j2
nohup.out
__pycache__
README.md

estudos_hgcr:
estudos_hgcr.sh
estudos_main.py
estudos_parses.py
estudos_services.py
estudos_template.json.j2
nohup.out
__pycache__
README.md

laudos_hgcr:
laudos_hgcr.sh
laudos_main.py
laudos_parses.py
laudos_services.py
laudos.sh
laudos_template.json.j2
nohup.out
__pycache__
README.md

```

Fonte: Elaborado pelo autor (2023).

A execução ocorreu por meio de scripts Bash Linux (interpretador de comando Linux) executados em linhas de comandos, por meio dos quais são executados continuamente os respectivos processos Python para coleta de um registro por vez com rede de dados instável. Como forma de exemplo, a Figura 36 apresenta essa execução para o HGCR, sendo a mesma para o HIJG e HRSJ.

Figura 36 – Execução contínua de processos Python para coleta de exames agendados do HGCR

```

#!/bin/bash

while true; do
  cd /opt/provhealth/agendados_hgcr/
  python3 agendados_main.py
done

```

Fonte: Elaborado pelo autor (2023).

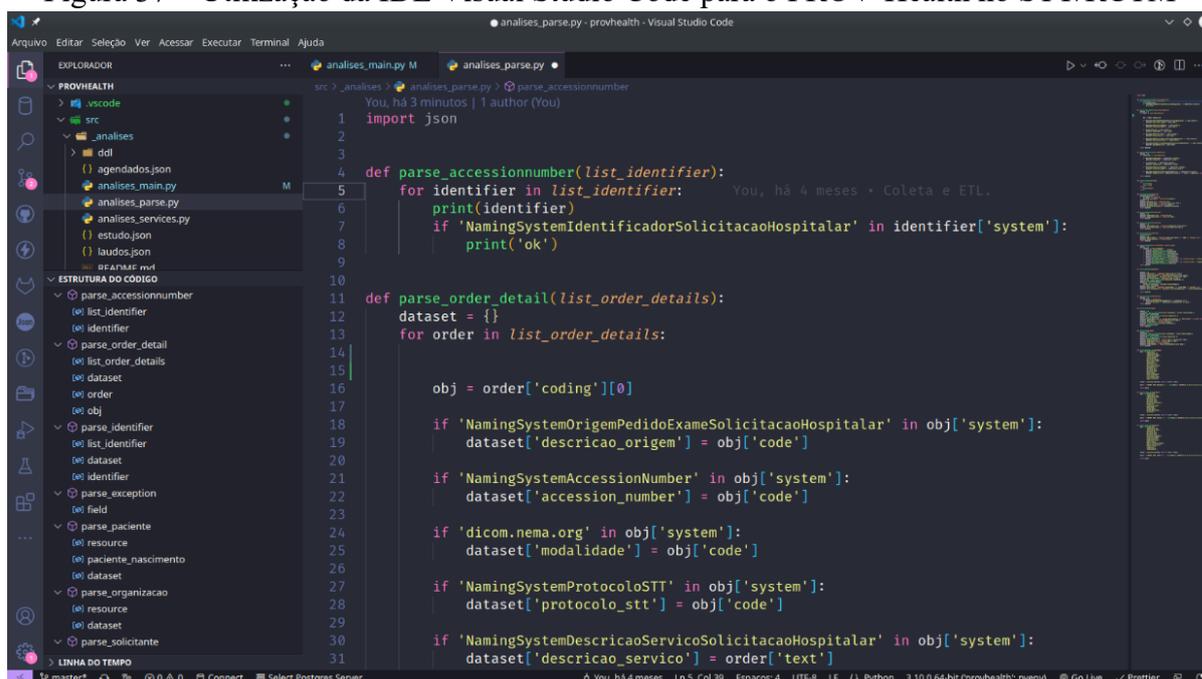
As Figuras 35 e 36 apresentam as integrações dos hospitais (i), (ii) e (iii), respectivamente (HIJG), (HGCR) e (HRSJ), relacionada às bases de dados HIS Agendados (agenda_local) e PACS Estudos DICOM (pacs_db), as quais fazem parte da infraestrutura interna dos Hospitais, enquanto a base do STT Laudos (agenda_global), relaciona a infraestrutura central disponível no CIASC. Nesse contexto, os sistemas de integração

(aplicações Python) utilizados para a coleta de dados Agendados, Estudos DICOM e Laudos, tem em comum, o acesso às bases de dados (agenda_local, pacs_db e agenda_global) e processos para consulta do mapeamento dos dados de proveniência para os respectivos perfis, conforme:

- a) O sistema de integração para Agendados (HIS Agendados), tem acesso ao banco de dados “agenda_local” disponível na infraestrutura dos Hospitais (i, ii, iii), e tem início sua rotina de integração, por meio da realização de uma consulta SQL para selecionar os dados, a qual possui os dados conforme tabela 4 já descrita anteriormente. Em seguida, os dados selecionados na consulta, são mapeados para o profile SolicitacaoHospitalar.
- b) O sistema de integração para coleta de Estudos DICOM (PACS Estudos DICOM) têm acesso ao banco de dados “pacsdb” disponível na infraestrutura dos Hospitais (i, ii, iii), e tem início sua rotina de integração, por meio da realização de uma consulta SQL para selecionar os dados das tabelas “patient” e “study”, a qual relaciona os dados conforme tabela 6 já descrita anteriormente. Em seguida, os dados selecionados na consulta, são mapeados para o profile EstudoDICOM.
- c) O sistema de integração para coleta de Laudos, (STT Laudos) tem acesso ao banco de dados “agenda_global” disponível na infraestrutura do STT/RCTM, e tem início de sua rotina de integração, por meio da realização de uma consulta SQL para selecionar os dados da tabela agenda_global, a qual possui os dados conforme o Quadro 8, já descrito anteriormente. Em seguida, os dados selecionados na consulta, são mapeados para o profile LaudoExameHospitalar.

Assim, os sistemas de integração dos Hospitais (i, ii, iii) já mencionados, são desenvolvidos em Python e utilizam bibliotecas para conexões com banco de dados PostgreSQL, além de geração de modelos pré-definidos de documentos JSON e biblioteca para requisições HTTP. Assim, para o desenvolvimento das aplicações Python, foi utilizado a IDE Visual Studio Code, definindo uma estrutura de projeto de desenvolvimento e utilizado o serviço códigos UFSC como repositório do código fonte. As Figuras 37 e 38 apresentam as interfaces de utilização do IDE Visual Studio Code e Serviço Códigos UFSC dos códigos fonte do PROV-Health.

Figura 37 – Utilização da IDE Visual Studio Code para o PROV-Health no STT/RCTM



Fonte: IDE Visual Studio Code (2023).

Nesse sentido, foi necessário estruturar o desenvolvimento do PROV-Health em diretórios e arquivos dentro do STT/RCTM, sendo que o Quadro 32 apresenta esse processo de forma detalhada.

Quadro 32 – Estrutura de desenvolvimento utilizado para o PROV-Health no STT/RCTM.

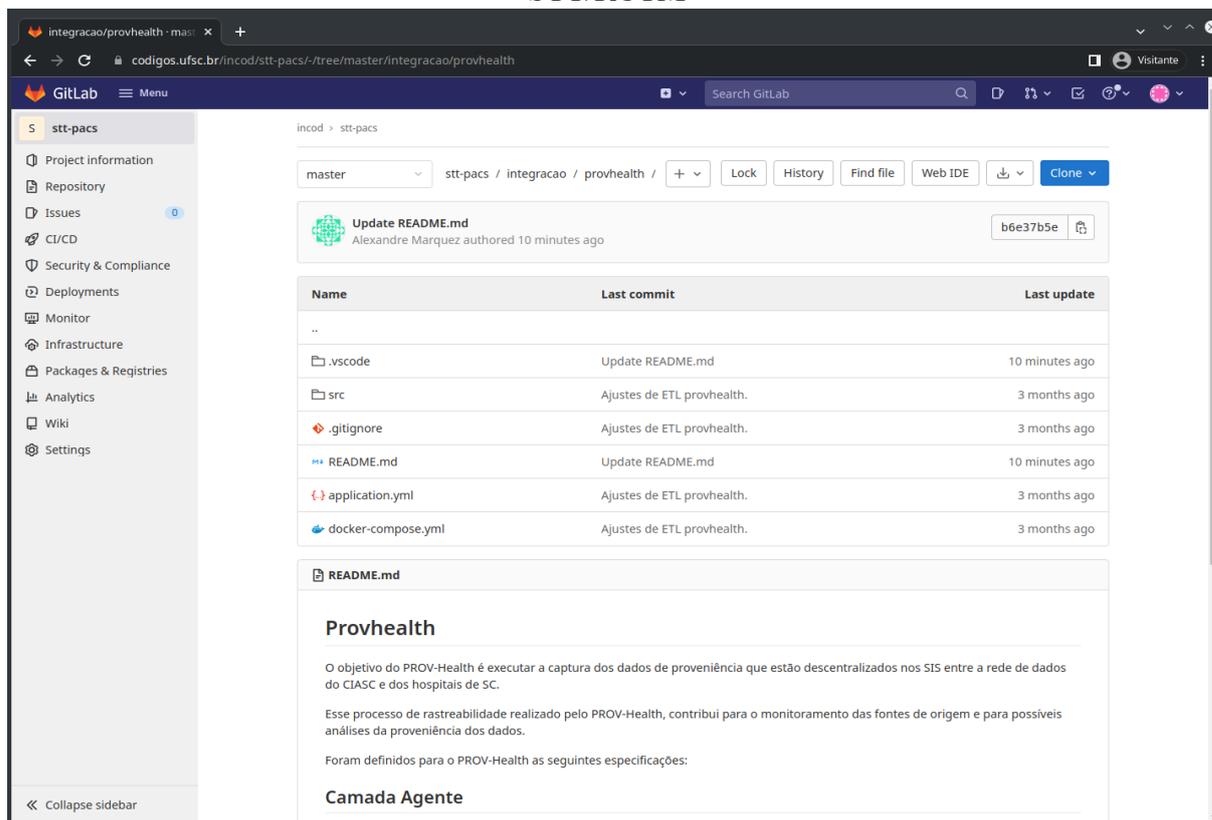
| Diretório | Arquivos | Descrição |
|--------------------------|--|--|
| provhealth | .gitignore docker-compose.yml README.md | Contém o README, arquivos de configurações e armazenamento de dados temporário para execução em ambiente de desenvolvimento. |
| provhealth/src/ | - | Delimita os arquivos que contém código diretamente relacionado ao projeto em específico. |
| provhealth/src/analises | analises_main.py analises_parse.py analises_services.py | Contém os códigos desenvolvidos para coleta dos dados de proveniência e mapeamento para base de dados do PROV-Health-BI. |
| provhealth/src/agendados | agendados_main.py agendados_parse.py agendados_services.py | Contém os códigos desenvolvidos para coleta dos dados de agendamentos do HIS e mapeamento para o recurso ServiceRequest (profile SolicitacaoHospitalar). |
| provhealth/src/estudos | estudos_main.py estudos_parse.py estudos_services.py | Contém os códigos desenvolvidos para coleta dos dados de estudos DICOM do PACS do Hospital e mapeamento para o recurso ImagingStudy (profile |

| | | |
|--------------------------|---|--|
| | | EstudoDICOM) |
| provhealth/src/estudos | laudos_main.py laudos_parse.py laudos_services.py | Contém os códigos desenvolvidos para coleta dos dados de laudos do STT e mapeamento para o recurso DiagnosticReport (profile LaudoExameHospitalar) |
| provhealth/src/controles | controles_main.py controles_services.py | Contém os códigos desenvolvidos para verificação de integrações e atualização de "status" dos recursos. |

Fonte: Elaborado pelo autor (2023).

Assim, para gerenciar os códigos fontes do PROV-Health foi utilizado o Serviço Códigos UFSC no GitLab para realizar esse processo, o qual pode ser visto na Figura 38.

Figura 38 – Serviço Códigos UFSC para gerenciamento de código do PROV-Health no STT/RCTM



Fonte: GitLab (2023).

No decorrer do desenvolvimento do PROV-Health, os perfis *SolicitacaoHospitalar*, *EstudoDICOM* e *LaudoExameHospitalar* foram desenvolvidos utilizando ferramenta *Forge by Firely* já descrita na subseção 6.5, e após validadas, foram publicadas no *Simplifier.net*, em específico para o projeto *STT on FHIR*.

A Figura 39 apresenta o STT on FHIR onde estão os perfis criados para uso no PROV-

Health (<https://simplifier.net/stt/~resources?category=Profile>).

Figura 39 – STT on FHIR onde estão alocados os profiles utilizados no PROV-Health

The screenshot shows the Simplifier.NET website interface. At the top, there is a yellow header with the Simplifier.NET logo and navigation links for SNIPPETS, FEEDBACK, LOG IN, and SIGN UP. Below the header is a search bar and icons for Query, API, and Download. The main content area displays the project information for 'STT on FHIR® (R4)', including the project description and a list of resources. The 'Resources' tab is selected, showing a list of profiles with their names, descriptions, and 'Active' status. The left sidebar shows the 'Resource Categories' menu with 'Profiles' selected.

| Resource Name | Description | Status | Last Updated |
|--------------------------------|--|--------|--------------|
| Agendamento de teleatendimento | Profile on Appointment Agendamentos realizados para eventos de teleatendimento registrados em sistemas de Telemedicina e Telessaúde. | Active | 08/04/2021 |
| Instituição | Profile on Organization Instituições (incluindo estabelecimentos de saúde) referenciadas ou que utilizam os serviços dos sistemas de Telemedicina e Telessaúde. | Active | 17/01/2021 |
| Médico | Profile on Practitioner Perfil base para médicos relacionados a eventos registrados em sistemas de Telemedicina e Telessaúde | Active | 05/05/2022 |

Fonte: Simplifier.net – STT on FHIR (2023).

Os Profiles do PROV-Health estão todos alocados ao STT on FHIR, e possivelmente farão parte da continuação e aprimoramento do método futuramente.

Em resumo, em relação às ferramentas computacionais empregadas na Camada de Atividades (A), pode-se destacar que: a aplicação do DCM4CHEE foi utilizada para o PACS nos hospitais selecionados da Camada de Entidades (A) HIJG (i), (B) HGCR (ii) e (C) HRSJ (iii); Foi realizada a execução por meio de scripts em terminal Bash por crontab Linux; Scripts Python foram desenvolvidos na IDE VS Code; Forge by Firely foi utilizado para o desenvolvimento e definição de Profiles FHIR: SolicitacaoHospitalar, EstudoDICOM e LaudoExameHospitalar com apoio do sistema operacional Windows utilizando HL7 FHIR, já descritos na “seção 6.5”; Publicação dos Profiles no Simplifier.net respectivamente no STT on FHIR® (R4) da UFSC; Códigos Python foram desenvolvidos para o mapeamento dos dados e transformações desses em estruturas conforme Profiles já definidos. Ainda, Códigos Python foram realizados para envio de recursos FHIR (POST); O Insomnia foi utilizado com um framework *Open Source* para desenvolvimento/teste de API Clients do STT/RCTM utilizado para envio de recursos FHIR (POST); Banco de dados PostgreSQL e consultas SQL foram utilizadas pelos sistemas de integração desenvolvidos em Python na conexão de bibliotecas

específicas para uso da ferramenta; JSON para armazenar informações estruturadas para transferir dados entre um servidor e um cliente do contexto STT/RCTM; GitLab foi utilizado no serviço códigos UFSC para gerenciamento de código do PROV-Health no STT/RCTM; e, por fim, a LGPD foi respeitada em todo processo de captura, coleta e preparação dos dados dentro do STT/RCTM, principalmente pela Camada de Agentes (A).

Nesse contexto, as coletas dos dados Agendados, Estudos DICOM e Laudos, explicitadas até o momento, foram base para os três experimentos já mencionados, os quais serão descritos detalhadamente nas próximas subseções.

6.8.3.1.1 Experimento 1: Coleta de dados Agendados (*Profile SolicitacaoHospitalar*)

Para coleta dos dados de exames agendados no experimento 1 são apresentados todos os processos computacionais utilizados em sua execução. Assim, primeiramente é apresentado o código em Python utilizado para a coleta de dados Agendados de forma estruturada e sequencial no Quadro 33.

Quadro 33 – Coleta de dados Agendados

```
registro = consulta_db("select * from exame_agendado where proveniencia = 'f' order by id desc limit 1")
solicitacao_hospitalar = parse_solicitacao(registro)
paciente = parse_paciente(registro)
solicitacao_hospitalar.update(paciente)
solicitante = parse_solicitante(registro)
solicitacao_hospitalar.update(solicitante)
solicitacao_hospitalar['data_exame'] = parse_data_exame(
    registro['dataexame'], registro['horaexame'])
solicitacao_hospitalar['modalidade'] = modalidade
payload = render_json(solicitacao_hospitalar)
with open(f'../tmp/solicitacao_hospitalar.json', 'w') as f:
    json.dump(json.loads(payload), f)
http_status = requisicao_fhir(payload)
if str(http_status) == "201":
    agendado_integrado(registro['id'])
    print(registro['id'])
```

Fonte: Elaborado pelo autor (2023).

Já no que diz respeito à transformação do registro para o *profile SolicitacaoHospitalar*, permite o mapeamento direto para os campos do FHIR conforme o código em Python do Quadro 34.

Quadro 34 – Mapeamento do Profile de SolicitacaoHospitalar

```
def parse_solicitacao(registro):
    solicitacao = {
        "organizacao_nome": organizacao_nome,
        "organizacao_id": organizacao_id,
        "requisicao": registro["codigora"],
        "accession_number": registro["accessionnumber"],
        "codigo_exame": registro["codigoexamesus"],
        "descricao_exame": registro["descricaoexame"],
        "descricao_origem": registro["descricaoorigem"],
        "protocolo_stt": registro["numeroprotocolo"],
        "descricao_servico": registro["descricaooservico"],
        "equipamento_id": registro["idequipamento"],
        "dados_clinicos": registro['dadosclinicos']
    }
return solicitacao
```

Fonte: Elaborado pelo autor (2023).

No que compete ao paciente nesse experimento, o mesmo tem os dados mapeados e para mapeamento do sexo do paciente foi necessário a transformação dos valores “F” para “Female” e “M” para “Male” conforme código em Python visualizado no Quadro 35.

Quadro 35 – Mapeamento do sexo do paciente

```
def parse_paciente(registro):
    sexo = parse_sexo(registro['sexo'])
    paciente = {'cpf': registro['cpf'],
               'paciente_nome': registro['nomepaciente'],
               'paciente_id': registro['codigoprontuario'],
               'paciente_sexo': sexo,
    }
    return paciente
```

Fonte: Elaborado pelo autor (2023).

Em relação aos dados do solicitante do exame neste experimento, são mapeados os dados “solicitante_nome” e “solicitante_id”, respectivamente para os dados de origem “nomesolicitante” e “numeroconselhosolicitante” conforme observado no código em Python do Quadro 36.

Quadro 36 – Mapeamento do médico solicitante

```
def parse_solicitante(registro):
    return {
        'solicitante_nome': registro['nomesolicitante'],
        'solicitante_id': registro['numeroconselhosolicitante']}
    }
```

Fonte: Elaborado pelo autor (2023).

Já em relação à data do exame utilizada neste experimento, a mesma é obtida pela composição e transformação dos dados de origem “dataexame” e “horaexame” usando o pacote “datetime” do python e função “striptime” para definir datas a partir de textos conforme o código em Python do Quadro 37.

Quadro 37 – Mapeamento da data do exame

```
def parse_data_exame(dataexame, horaexame):
    return datetime.strptime(
        f'{dataexame} {horaexame}', '%d/%m/%Y %H:%M:%S')
        .astimezone().isoformat()
```

Fonte: Elaborado pelo autor (2023).

A modalidade do exame neste experimento é obtida através do campo “idequipamento”, sendo utilizado como filtro para a consulta na tabela “modalidade” também disponível na base de dados “agenda_local” conforme código Python do Quadro 38.

Quadro 38 – Mapeamento da modalidade DICOM

```
def parse_modalidade(idequipamento):
    row = consulta_db(
        f'select modalidade from equipamento_modalidade
        where idequipamento = '{idequipamento}''')
    return row['modalidade']
```

Fonte: Elaborado pelo autor (2023).

Assim, foi possível gerar um documento JSON conforme estrutura prevista no profile SolicitacaoHospitalar, utilizado o pacote “Jinja2⁵⁸” do python, que permite gerar documentos a partir de modelos pré-definidos e parametrizados, sendo assim utilizado o modelo “agendados template.json.j2”, com os dados previstos conforme o código Python do Quadro 39.

Quadro 39 – Geração do documento JSON conforme o profile SolicitacaoHospitalar

```
def render_json(solicitacao):
    environment=Environment(loader=FileSystemLoader("./"))
    template= environment.get_template("agendados_template.json.j2")
    r = template.render(
        requisicao=solicitacao["requisicao"],
        organizacao_id=organizacao_id,
        organizacao_nome=organizacao_nome,
        accession_number=solicitacao["accession_number"],
        codigo_exame=solicitacao["codigo_exame"],
        descricao_exame=solicitacao["descricao_exame"]
            .replace("\\', ' '),
        descricao_origem=solicitacao["descricao_origem"],
        modalidade=solicitacao['modalidade'],
        protocolo_stt=solicitacao["protocolo_stt"],
        descricao_servico=solicitacao["descricao_servico"],
        equipamento_id=solicitacao["equipamento_id"],
        paciente_id=solicitacao["paciente_id"],
        paciente_nome=solicitacao["paciente_nome"],
        paciente_sexo=solicitacao["paciente_sexo"],
        data_exame=solicitacao["data_exame"],
        data_recebimento=solicitacao["data_recebimento"],
        solicitante_id=solicitacao["solicitante_id"],
```

⁵⁸ O Jinja é um *template engine* escrito em Python que facilita a criação de páginas HTML em aplicações Python.

```
solicitante_nome=solicitacao["solicitante_nome"],
dados_clinicos=solicitacao['dados_clinicos'])
return r
```

Fonte: Elaborado pelo autor (2023).

Dessa maneira, foi possível realizar a coleta de dados do experimento 1 para o armazenamento dos dados de proveniência coletados para a Camada de Atividades (B) da arquitetura específica de experimentação do PROV-Health no STT/RCTM.

6.8.3.1.2 Experimento 2: Coleta de dados Estudos DICOM (Profile EstudoDICOM)

Para coleta de Estudos DICOM no experimento 2, são detalhados todos os processos computacionais por meio de códigos em Python. Assim, primeiramente detalha-se no Quadro 40, o código em Python de forma estruturada e sequencial para a coleta de estudos DICOM.

Quadro 40 – Coleta de dados Estudos DICOM

```
registro = consulta_estudo()
estudo_DICOM = {}
estudo_DICOM['paciente_id'] = registro['pat_id']
estudo_DICOM['estudo_uid'] = registro['study_iuid'] estudo_DICOM['estudo_id'] = registro['study_id']
estudo_DICOM['accession_number'] = registro['accession_no']
estudo_DICOM['modalidade'] = registro['mods_in_study']
estudo_DICOM['estudo_descricao'] = registro['study_desc']
estudo_DICOM['paciente_nome'] = parse_paciente_nome(registro['pat_name'])
estudo_DICOM['paciente_sexo'] = parse_paciente_sexo(registro['pat_sex'])
estudo_DICOM['paciente_nascimento'] = parse_paciente_nascimento(registro['pat_birthdate'])
estudo_DICOM['estudo_data'] = parse_estudo_data(registro['study_datetime'])
with open(f'../tmp/estudo_DICOM.json', 'w') as f:
    json.dump(json.loads(payload), f)
http_status = envia_estudo(payload)
if str(http_status) == "201":
    atualiza_estudo(registro['pk'])
```

Fonte: Elaborado pelo autor (2023).

Após esse processo, foi necessário realizar consultas em SQL na função consulta_db () como mostra o Quadro 41 para a coleta de dados no Estudo DICOM.

Quadro 41 – Consulta SQL realizada para coleta de Estudo DICOM

```
select s.pk, s.study_iuid, s.study_id, s.accession_no, s.study_datetime, s.study_datetime, s.created_time,
s.study_desc, s.num_series, s.num_instances, s.ref_physician, p.pat_id, p.pat_name, p.pat_sex,
p.pat_birthdate, s.mods_in_study from study s left join patient p on p.pk = s.patient_fk where s.pk = (select
pk from study where integrado_proveniencia = 'f' order by pk desc limit 1);
```

Fonte: Elaborado pelo autor (2023).

Com relação ao Quadro 41, os dados de proveniência “pat_id”, “study_iuid”, “study_id”, “accession_no”, “mods_in_study”, “study_desc” foram mapeados de forma direta

respectivamente aos dados “paciente_id”, “estudo_iuid”, “estudo_id”, “accession_number”, “modalidade” e “estudo_descricao” para que o processo de coleta de Estudos DICOM pudesse ter êxito de execução computacional.

Porém, outros dados foram necessários nos processos de transformações, como o nome do paciente, que é estruturado conforme normatização do DICOM para nomes de pessoas (Person Name (PN) Value Representation), sendo assim realizado a adequação desse, a partir da geração de uma lista para os elementos separados pelo caracter “^” (person_name), que por sua vez são adicionados a outra lista (paciente_nome) de maneira ordenada conforme estrutura: nome, nome do meio e sobrenome, sendo observados na Quadro 42.

Quadro 42 – Transformações do nome do paciente

```
def parse_paciente_nome(pat_name):
    person_name = pat_name.split('^')
    paciente_nome = []
    paciente_nome.append(person_name[1]) # given name
    paciente_nome.append(person_name[2]) # middle name
    paciente_nome.append(person_name[0])
    return ','.join(paciente_nome).strip()
```

Fonte: Elaborado pelo autor (2023).

Outro processo computacional necessário para este experimento 2, foi a transformação do sexo do paciente, o qual reutiliza a mesma abordagem utilizada para a coleta de agendamentos, onde na origem os valores “F” e “M” são respectivamente mapeados para “female” e “male” ou então utilizado “other” para casos não mapeados.

Já em relação à data de nascimento utilizada no experimento 2, a mesma é transformada utilizando condição de exceção caso a data de nascimento não tenha sido informada e utiliza do pacote datetime do Python, que permite identificar e alterar o formato de uma data, como pode ser observado no Quadro 43.

Quadro 43 – Transformação da data de nascimento

```
def parse_paciente_nascimento(pat_birthdate):
    try:
        parse_data = datetime.strptime(f'{pat_birthdate}', "%Y%m%d")
        return datetime.strptime(parse_data, "%Y-%m-%d")
    except ValueError:
        return "
```

Fonte: Elaborado pelo autor (2023).

Já a data do Estudo DICOM no experimento 2, é obtida a partir formatação do campo

‘study_datetime’ no formato ISOFORMAT⁵⁹ como pode ser visualizado no Quadro 44.

Quadro 44 – Data do estudo DICOM

```
def parse_estudo_data(study_datetime):
    return study_datetime.isoformat()
```

Fonte: Elaborado pelo autor (2023).

Assim, foi possível gerar um documento JSON conforme estrutura prevista no profile EstudoDICOM, utilizando o pacote “jinja” do python e utilizando também o modelo “estudos_template.json.j2”, com os dados previstos conforme pode ser observado no Quadro 45.

Quadro 45 – Geração do documento JSON conforme profile EstudoDICOM

```
def render_json(estudo):
    environment = Environment(loader=FileSystemLoader("./"))
    template = environment.get_template("estudos_template.json.j2")
    r = template.render(
        organizacao_id=organizacao_id,
        organizacao_nome=organizacao_nome,
        paciente_id=estudo['paciente_id'],
        paciente_nome=estudo['paciente_nome'],
        paciente_sexo=estudo['paciente_sexo'],
        paciente_nascimento=estudo['paciente_nascimento'],
        estudo_uid=estudo['estudo_uid'],
        estudo_id=estudo['estudo_id'],
        accession_number=estudo['accession_number'],
        modalidade=estudo['modalidade'],
        estudo_data=estudo['estudo_data'],
        wado_url=wado_url,
        wado_type=wado_type,
        estudo_descricao=estudo['estudo_descricao'],
        ae_title=ae_title
    )
    return r
```

Fonte: Elaborado pelo autor (2023).

Assim, foi possível realizar a coleta de dados Estudos DICOM do experimento 2 para o armazenamento dos dados de proveniência coletados para a Camada de Atividades (B) da arquitetura específica para experimentação do PROV-Health no STT/RCTM.

6.8.3.1.3 Experimento 3: Coleta de dados Laudos (Profile: LaudoExameHospitalar)

Para coleta dos dados Laudos de exames hospitalares que compõe o experimento 3, foram detalhados todos os processos computacionais utilizados. Assim, primeiramente é

⁵⁹ Usada para retornar uma string de data, hora e deslocamento UTC para o fuso horário correspondente no formato ISO 8601

descrito o código em Python de forma estruturada e sequencial para a coleta de laudos de exames hospitalares conforme pode ser visualizado no Quadro 46.

Quadro 46 – Código em Python para coleta dos dados de laudos de exames hospitalares

```
registro = consulta_laudo("select * from exame_agendado where id_instituicao_stt = 1001 and
status_agendamento = 15 order by id desc limit 1;")
laudo = {}
"paciente_id": registro['codigoprontuario']
"paciente_nome": registro['nomepaciente']
"estudo_uid": registro['studyinstanceuid']
"accession_number": registro['accessionnumber']
"exame_descricao": registro['descricaoexame']
"exame_protocolo": registro['numeroprotocolo']
"medico_nome": registro['nomemedicolaudos']
"medico_conselho_numero":
    registro['codigoconselhomedicolaudos']
"medico_conselho_uf": registro['ufmedicolaudos']
    "laudo_stt_id": registro['idlaudo']
"paciente_sexo": parse_paciente_sexo(registro['sexo'])
"paciente_nascimento":
    parse_paciente_nascimento(registro['datanascimento'])
"laudo_data": parse_data_laudo(
    registro['dataexame'], registro['horaexame'])
payload = render_json(laudo)
with open(f./laudos.json', 'w') as f:
    json.dump(json.loads(payload), f)
http_status = envia_laudo(payload)
if str(http_status) == "201":
    atualiza_laudo(registro['id'])
    print(registro['id'])
```

Fonte: Elaborado pelo autor (2023).

Após o processo computacional executado no Quadro 46, os dados de proveniência foram mapeados de forma direta, utilizando *parsers*⁶⁰ semelhantes aos já descritos nas coletas anteriores. Assim, foi possível gerar um documento JSON que utiliza-se do modelo parametrizado “laudos_template.json.j2”, com os dados previstos conforme observado no Quadro 47.

Quadro 47 – Modelo parametrizado para “laudos_template.json.j2”

```
def render_json(laudo):
    environment = Environment(loader=FileSystemLoader("./"))
    template = environment.get_template("laudos_template.json.j2")
    r = template.render(
        organizacao_id=organizacao_id,
        organizacao_nome=organizacao_nome,
        paciente_id=laudo['paciente_id'],
        paciente_nome=laudo['paciente_nome'],
        paciente_sexo=laudo['paciente_sexo'],
        paciente_nascimento=laudo['paciente_nascimento'],
```

⁶⁰ O construtor de tipo Parser é uma mônada e representa uma computação que constrói um valor a partir de uma dada string.

```
estudo_uid=laudo['estudo_uid'],
accession_number=laudo['accession_number'],
exame_protocolo=laudo['exame_protocolo'],
exame_descricao=laudo['exame_descricao'],
medico_nome=laudo['medico_nome'],
medico_conselho_numero=laudo['medico_conselho_numero'],
medico_conselho_uf=laudo['medico_conselho_uf'],
laudo_stt_id=laudo['laudo_stt_id'],
laudo_data=laudo['laudo_data'],
laudo_base64=laudo['laudo_base64'])
return r
```

Fonte: Elaborado pelo autor (2023).

Portanto, foi possível realizar a coleta de dados dos Laudos explicitada no experimento 3 para o armazenamento dos dados de proveniência coletados para a Camada de Atividades (B) da arquitetura específica para experimentação do PROV-Health no STT/RCTM.

6.8.4.1 Camada de Atividades (B) – Módulo (B): organização, controle e armazenamento dos dados de proveniência do STT/RCTM

Essa camada é composta pelo Módulo (B) de organização, controle e armazenamento dos dados de proveniência do STT/RCTM, onde são executadas parte da estratégia computacional pela Camada de Agentes (A) por meio de ferramentas computacionais estabelecidas anteriormente na Camada de Atividades (A).

A organização, controle e armazenamento de dados nesta camada, tem por objetivo complementar os dados que são armazenados no PROV-Health, selecionados para este estudo de caso, para posterior visualização, consulta e análise desses dados na Camada de Atividades (C). Por essa razão, essa prática é considerada no PROV-Health, essencial para guiar as decisões estratégicas que podem ser tomadas com os dados de proveniência coletados pela Camada de Atividades (A).

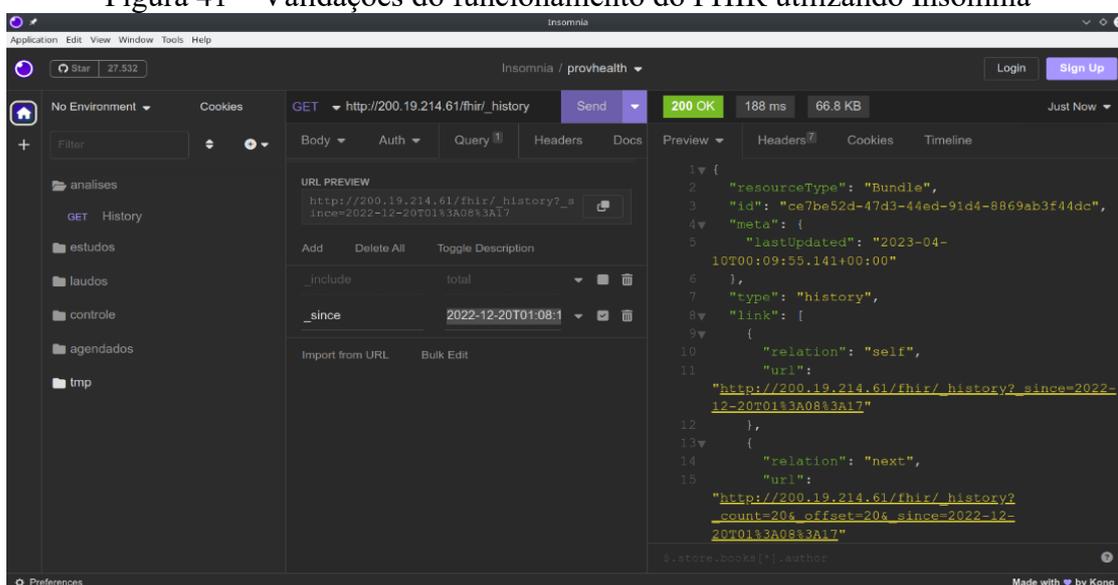
Nesse contexto, a Figura 40 apresenta a Camada de Atividades (B) de forma separada da arquitetura específica de experimentação do PROV-Health no STT/RCTM para melhor compreensão de sua estratégia computacional para desempenhar às atividades definidas pelo Módulo (B).

Figura 40 – Camada de Atividades (B) do PROV-Health para experimentação no STT/RCTM



Observa-se na Figura 40, que existem ferramentas computacionais que apoiam a estratégia computacional executada na Camada de Atividades (B), sendo que, em resumo: o Simplifier.net foi utilizado para integração de perfis já criados para o PROV-Health para o HAPI FHIR (como repositório de proveniência PROV-Health); foi utilizado Stack Docker Swarm para execução dos serviços do HAPI FHIR e PostgreSQL; Python e SQL para os códigos voltados às verificações realizadas. No Insomnia foram realizadas consultas FHIR. O Insomnia foi utilizado como uma aplicação para testes de API sendo utilizado para consultas no servidor FHIR do PROV-Health, assim, foram realizadas validações de armazenamento e consulta de documentos JSON conforme especificação FHIR. Sua utilização é demonstrada na Figura 41, que realiza a consulta (GET) para o servidor HAPI FHIR do PROV-Health, com objetivo de experimentar e validar o funcionamento das “queries” previstas para o serviço “_history”, sendo assim possível definir essa abordagem para integrar os recursos FHIR no PROV-Health-BI na próxima camada, a Camada de Atividades (C).

Figura 41 – Validações do funcionamento do FHIR utilizando Insomnia



Fonte: Insomnia (2023).

Nesse contexto do uso das ferramentas computacionais, enfatiza-se para a organização e controle, que a Camada de Atividades (B), a qual está associada a ações de configurações e customizações da plataforma HAPI FHIR é a camada destinada para o armazenamento dos dados de proveniência no STT/RCTM. Essas ações podem ser resumidas em relação à utilização de profiles advindos da Camada de Atividades (A) para validação das coletas dos sistemas integrados para os recursos relacionados, bem como, aplicações desenvolvidas em Python e definidas com objetivo de auxiliar na gestão dos dados de proveniência do STT/RCTM, sendo relacionado à verificação de integração entre os sistemas internos e externos, permitindo atualização dos recursos e também adicionando novos recursos de maneira a complementar os dados de proveniência.

Quanto ao armazenamento, é utilizado a plataforma HAPI FHIR, a qual relaciona interfaces de API REST⁶¹ para entradas e saídas de dados, utilizando-se de recursos FHIR. Por meio da implementação dos profiles FHIR, permite ter a entrada controlada e validada para determinados recursos, como ServiceRequest, ImagingStudy e DiagnosticReport respectivamente controladas pelos profiles SolicitacaoHospitalar, EstudoDICOM e LaudoExameHospitalar. O HAPI FHIR utiliza-se de uma base de dados PostgreSQL para fins de controlar os recursos armazenados e possui sua modelagem orientada ao relacionamento dos recursos com seus atributos de maneira a atender critérios para especificação do servidor FHIR do PROV-Health.

Para a configuração do HAPI FHIR foi utilizado um servidor Linux no CIASC, instalando o Docker e ativando o modo Swarm⁶², que permite o gerenciamento de containers como serviços e utilização da ferramenta Portainer⁶³ para facilitar a administração desses serviços. Inicialmente foi definida a Stack⁶⁴ com os serviços PostgreSQL e aplicação HAPI FHIR. Nessa Stack (Quadro 48), termo utilizado para composição de serviços gerenciados em uma mesma rede do Docker, foram utilizadas a imagem Docker oficial do PostgreSQL 12.3 para este serviço, mapeando volume para persistência dos dados e variáveis para configuração do mesmo. Também é utilizada a imagem oficial do projeto HAPI FHIR, bem como definido o mapeamento do arquivo com suas configurações e o diretório para armazenamento dos dados

⁶¹ Também chamada API RESTful, a API REST é uma interface de programação de aplicações em conformidade com as restrições da arquitetura REST.

⁶² É um recurso do Docker que fornece funcionalidades de orquestração de contêiner, incluindo clustering nativo de hosts do Docker e agendamento de cargas de trabalho de contêineres.

⁶³ É uma interface web que interage com o socket do docker para criar novos contêineres e monitorá-los.

⁶⁴ De forma simplificada, a evolução do Compose, com foco em clusters (utilizando swarm).

desse serviço.

Quadro 48 – Configuração dos serviços para execução do HAPI FHIR no STT/RCTM

```

version: '3'
services:
  postgresql:
    image: postgres:12.3
    volumes:
      - /dados/volumes/provhealth/postgresql/data:/var/lib/postgresql/data/
    environment:
      - POSTGRES_USER=postgres
      - POSTGRES_PASSWORD=*****
      - POSTGRES_DB=hapifhir
      - POSTGRES_HOST_AUTH_METHOD=trust
    ports:
      - 5434:5432
  hapifhir:
    image: "hapiproject/hapi:latest"
    environment:
      SPRING_CONFIG_LOCATION: 'file:///application.yml'
    ports:
      - "80:8080"
    volumes:
      - /dados/volumes/provhealth/hapifhir/data:/data/hapi/
      - /dados/volumes/provhealth/hapifhir/application.yaml:/application.yml

```

Fonte: Elaborado pelo autor (2023).

Em relação aos controles de integração executados na Camada de Atividades (B) são realizados três experimentos para as verificações relacionadas às integrações agendadas na execução de processos pelo PROV-Health no STT/RCTM, sendo elas:

- a) Experimento 1: Integração de solicitações hospitalares com estudos DICOM;
- b) Experimento 2: Integração de solicitações hospitalares com laudos publicados;
- c) Experimento 3: Integração de laudos publicados com estudos DICOM.

Para as validações das integrações relacionadas aos três experimentos, foram criados os “ExtensionIntegracaoPacs” e “ExtensionIntegracaoLaudo”, sendo essas extensões adicionadas aos recursos “ServiceRequest” e “DiagnosticReport” para verificação do status de integração de cada experimento. Importante destacar que, Agendados assume-se como SolicitacoesHospitalares para esses experimentos. Assim, os processos computacionais utilizados nos três experimentos são detalhados nas próximas subseções.

6.8.4.1.1 Experimento 1: Integração de solicitações hospitalares com estudos DICOM

Para verificação das solicitações hospitalares integradas com estudos DICOM é realizada uma consulta FHIR para os recursos ServiceRequest com valor “não verificado” para

o atributo “ExtensionIntegracaoPacs”. A partir dos registros obtidos, são realizadas novas consultas para rastrear os recursos que tenham dados relacionados, de maneira que permita atualizar o atributo “ExtensionIntegracaoPacs” para “integrado” ou “pendente”, conforme descrito de forma estruturada e sequencial. O processo Python utilizado pode ser observado no Quadro 49.

Quadro 49 – Processo Python para os processos de verificação das solicitações hospitalares integradas com estudos DICOM

```

accession_number = verifica_solicitacoes()
if not accession_number:
    print("não existem solicitações para verificar")
    exit(0)
    accession_number_estudo= consulta_estudo_DICOM(accession_number)
if not accession_number_estudo:
    update_fhir_solicitacao("pendente")
if accession_number_estudo:
    update_fhir_solicitacao("integrado")

```

Fonte: Elaborado pelo autor (2023).

Para o experimento 1 das integrações de SolicitacoesHospitalares com estudos DICOM, são realizadas comparações de dados de proveniência já coletados na Camada de Atividades (A) para apresentação e comprovação das integrações na Camada de Atividades (B).

6.8.4.1.2 Experimento 2: Integração de solicitações hospitalares com laudos publicados

Para verificação das solicitações hospitalares com laudos publicados no experimento 2 é realizado uma consulta FHIR para os recursos ServiceRequest com valor “não verificado” para o atributo “ExtensionIntegracaoLaudos”. A partir dos registros obtidos, são realizadas novas consultas para rastrear os recursos DiagnosticReport que tenham dados relacionados, de maneira que permita atualizar o atributo “ExtensionIntegracaoLaudos” para “integrado” ou “pendente”, conforme descrito de forma estruturada e sequencial. O processo Python utilizado pode ser observado no Quadro 50.

Quadro 50 – Processo Python para os processos de verificação das solicitações hospitalares integradas com laudos publicados

```

accession_number = verifica_solicitacoes()
if not accession_number:
    print("não existem solicitações para verificar")
    exit(0)
    accession_number_estudo= consulta_laudos_publicados(accession_number)
if not accession_number_laudos_publicados:
    update_fhir_solicitacao("pendente")
if accession_number_laudos_publicados:
    update_fhir_solicitacao("integrado")

```

Fonte: Elaborado pelo autor (2023).

Nesse experimento 2 de integração de solicitações hospitalares com laudos publicados são realizadas comparações de dados de proveniência já coletados na Camada de Atividades (A) para apresentação e comprovação das integrações na Camada de Atividades (B).

6.8.4.1.3 Experimento 3: Integração de laudos publicados com estudos DICOM

Para verificação dos laudos publicados com estudos DICOM integrados no experimento 3, é realizado uma consulta FHIR para os recursos DiagnosticReport com valor “não-verificado” para o atributo “ExtensionIntegracaoPacs”. A partir dos registros obtidos, são realizadas novas consultas para rastrear os recursos ImagingStudy que tenham dados relacionados, de maneira que permita atualizar o atributo “ExtensionIntegracaoPacs” para “integrado” ou “pendente”, conforme descrito de forma estruturada e sequencial. O processo Python utilizado pode ser observado no Quadro 51.

Quadro 51 – Processo Python para os processos de verificação dos laudos publicados com estudos DICOM

```

accession_number = verifica_laudos()
if not accession_number:
    print("não existem laudos para verificar")
    exit(0)
    accession_number_laudos= consulta_estudo_DICOM(accession_number)
if not accession_number_estudo:
    update_fhir_solicitacao("pendente")
if accession_number_estudo:
    update_fhir_solicitacao("integrado")

```

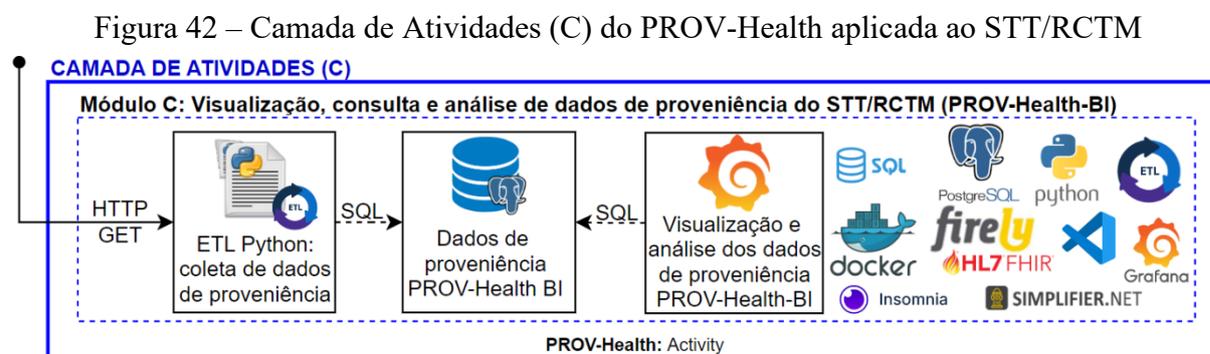
Fonte: Elaborado pelo autor (2023).

Nesse experimento 3 de integração de laudos publicados com estudos DICOM são realizadas comparações de dados de proveniência já coletados na Camada de Atividades (A) para apresentação e comprovação das integrações na Camada de Atividades (B).

6.8.5.1 Camada de Atividades (C) – Módulo (C): visualização, consulta e análise dos dados de proveniência do STT/RCTM

Essa camada é composta pelo Módulo (C) de visualização, consulta e análise dos dados de proveniência do STT/RCTM, onde são executadas parte da estratégia computacional pela Camada de Agentes (A) por meio de ferramentas computacionais estabelecidas anteriormente na Camada de Atividades (B), sendo incrementada as ferramentas ETL e Grafana. Essa camada é composta por um processo ETL desenvolvido em Python para coleta dos recursos FHIR armazenados no PROV-Health e mapeamento desses em outra base de dados, PROV-Health-BI, parte da estratégia computacional do PROV-Health, a qual é integrada ao Grafana, para visualização, consulta e análise de dados de proveniência do STT/RCTM.

A Figura 42 apresenta a Camada de Atividades (C) de forma separada da arquitetura de aplicação do PROV-Health no STT/RCTM para melhor compreensão de sua estratégia computacional para desempenhar as atividades definidas pelo Módulo (C).



Fonte: Elaborado pelo autor (2023).

Observa-se na Figura 42, que existem ferramentas computacionais que apoiam a estratégia computacional executada na Camada de Atividades (C), sendo que, em resumo: foi utilizado ETL Python para coleta de dados do PROV-Health, realizando consulta via HTTP fazendo uma requisição GET no servidor HAPI FHIR. Usa uma base de dados específica para análise no PostgreSQL conectada no Grafana, alimentada por meio de um processo ETL python; foi utilizado uma base dados PostgreSQL para tabelas utilizadas no PROV-Health-BI, agendados, estudos e laudos; Stack Docker Swarm foi utilizado para execução dos serviços do Grafana e PostgreSQL; Scripts Python para coleta de dados de proveniência do HAPI FHIR, desenvolvidos na IDE VS Code; e por fim, o Grafana permite conexão com base dados, consultas, análises e definição de dashboards para o PROV-Health-BI.

Para coleta dos recursos FHIR armazenados no PROV-Health, é descrito o código em Python de forma estruturada e sequencial conforme observado no Quadro 52.

Quadro 52 – Coleta dos recursos FHIR armazenados no PROV-Health

```

since = "
parametros = consulta_parametros_bi()
if 'valor' in parametros:
    since = parametros['valor']
if not since:
    exit()
url = f"{fhir_url}?_since={since}&_count=50"
res = consulta_fhir(url)
while True:
    if 'entry' in res:
        processamento_history(res['entry'])
    next_link = "
    for link in res["link"]:
        if link["relation"] == "next":
            next_link = link['url']
            res = consulta_fhir(next_link)
    if not next_link:
        atualiza_parametro_bi(res['meta']['lastUpdated'])
        break

```

Fonte: Elaborado pelo autor (2023).

Após esse processo, um item obtido a partir da consulta realizada no FHIR para o serviço “_history⁶⁵”, o qual tem por objetivo disponibilizar todos os recursos inseridos e alterados em ordem cronológica, é feito a partir da função “processamento_history”, e usa de três casos condicionais conforme atributo “resourceType⁶⁶”, sendo para cada um definido a transformação e mapeamento dos dados para inserção na base de dados do PROV-Health-BI, conforme observado no Quadro 53.

Quadro 53 – processamento de um item obtido a partir da consulta realizada no FHIR

```

def processamento_history(registro_entry):
    for entry in registro_entry:
        resource_type = entry['resource']['resourceType']
        if resource_type == 'ServiceRequest':
            agendado = parse_agendado(entry['resource'])
            query = parse_agendado_sql(agendado)
            insere_bi(query)
        if (resource_type == 'ImagingStudy'):
            estudo = parse_estudo(entry['resource'])
            query = parse_estudo_sql(estudo)
            insere_bi(query)
        if (resource_type == 'DiagnosticReport'):
            laudo = parse_laudo(entry['resource'])
            query = parse_laudo_sql(laudo)
            insere_bi(query)

```

Fonte: Elaborado pelo autor (2023).

⁶⁵ <https://hl7.org/fhir/history.html>

⁶⁶ <https://hl7.org/fhir/valueset-resource-types.html>

Para cada recurso obtido no processamento do “_history” são inseridos ou atualizados os dados na base de dados PROV-Health-BI, a qual possui três tabelas com os dados considerados relevantes para visualização, consulta e análise dos dados de proveniência do STT/RCTM, conforme apontados no Quadro 54.

Quadro 54 – Campos das tabelas utilizadas pelo PROV-Health-BI

| Agendados | Estudos | Laudos |
|---|--|--|
| requisicao organizacao_id organizacao_nome accession_number codigo_exame descricao_exame descricao_origem modalidade protocolo_stt descricao_servico equipamento_id paciente_id paciente_nome paciente_sexo data_exame data_recebimento solicitante_id solicitante_nome dados_clinicos integracao_pacs integracao_laudo | estudo_uid organizacao_id organizacao_nome paciente_id paciente_nome paciente_sexo paciente_nascimento estudo_id accession_number modalidade estudo_data estudo_descricao ae_title | estudo_uid organizacao_id organizacao_nome paciente_id paciente_nome paciente_sexo paciente_nascimento accession_number exame_protocolo exame_descricao medico_nome medico_conselho_numero medico_conselho_uf laudo_stt_id integracao_pacs |

Fonte: Elaborado pelo autor (2023).

Importante ressaltar que os dados do Quadro 53, foram anonimizados para fins de visualização no PROV-Health-BI. Porém, considerando que o PROV-Health executa processos na rede interna do STT/RCTM, essa anonimização foi realizada por meio do uso da LGPD para fins éticos no uso dos dados. Assim, a execução da coleta se fez a partir de um script desenvolvido em Bash, o qual executa o processo para execução da aplicação Python “analises_main.py” conforme observado no Quadro 55.

Quadro 55 – Execução da coleta se fez a partir de um script desenvolvido em Bash

```
#!/bin/bash
Cd/opt/provhealth/_analises/
python3 analises main.py
```

Fonte: Elaborado pelo autor (2023).

O agendamento é realizado diariamente por meio do serviço de “crontab⁶⁷” do Linux, sendo agendado para 2h horas todos os dias, e executado o script “analises.sh” com usuário “provhealth”, conforme indicado no Quadro 56.

Quadro 56 – Agendamento é realizado diariamente por meio do serviço de “crontab” do Linux

```
0 2 * * * provhealth analises.sh
```

Fonte: Elaborado pelo autor (2023).

Assim, o Grafana foi adicionado a Stack do Docker Swarm do PROV-Health, sendo mais um serviço provido e gerenciado no servidor Linux por meio do Docker Swarm. O serviço contém a imagem oficial do Grafana, variáveis `GF_SERVER_ROOT_URL` e `GF_SERVER_SERVE_FROM_SUB_PATH` para permitir o uso de sufixos como “/grafana” ao invés de apenas “/”, sendo tal abordagem necessária para o mapeamento da URL “<https://pacs.telemedicina.saude.sc.gov.br/grafana>”, por meio de um proxy HTTP conforme observado no Quadro 57.

Quadro 57 – Serviço Grafana adicionado na Stack do PROV-Health

```
version: '3'
services:
  postgresql:
    image: postgres:12.3
    volumes:
      - /dados/volumes/provhealth/postgresql/data:/var/lib/postgresql/data/
    environment:
      - POSTGRES_USER=postgres
      - POSTGRES_PASSWORD=Hp101Na21
      - POSTGRES_DB=hapifhir
      - POSTGRES_HOST_AUTH_METHOD=trust
    ports:
      - 5434:5432
  grafana:
    image: grafana/grafana
    container_name: grafana
    environment:
      GF_SERVER_ROOT_URL: '%(protocol)s://%(domain)s:%(http_port)s/grafana/'
      GF_SERVER_SERVE_FROM_SUB_PATH: 'true'
    volumes:
      - /dados/volumes/provhealth/grafana:/var/lib/grafana/
    ports:
      - 3568:3000
  hapifhir:
    image: "hapiproject/hapi:latest"
    environment:
      SPRING_CONFIG_LOCATION: 'file:///application.yml'
    ports:
      - "80:8080"
```

⁶⁷ Permite agendar as tarefas para que elas sejam executadas automaticamente.

volumes:

- /dados/volumes/provhealth/hapifhir/data/:/data/hapi/
- /dados/volumes/provhealth/hapifhir/application.yaml:/application.yml

Fonte: Elaborado pelo autor (2023).

Dessa forma, o Grafana foi configurado e integrado à base de dados do PROV-Health, de maneira a permitir consultas SQL executadas pela Camada de Agentes (A) definindo os gráficos para visualização dos dados de proveniência coletados e verificados nas Camadas de Atividades (A) e (B).

Sendo assim, tanto os dados de proveniência coletados pela Camada de Atividades (A) contendo três experimentos, quanto às integrações realizadas pela Camada de Atividades (B) também contendo três experimentos, ambos podem ser observados no PROV-Health-BI, o qual utiliza-se da ferramenta de visualização de dados (Grafana) no seguinte link:

<https://pacs.telemedicina.saude.sc.gov.br/grafana>

6.8.5.1.1 Visualização dos dados de proveniência dos experimentos da Camada de Atividades (A)

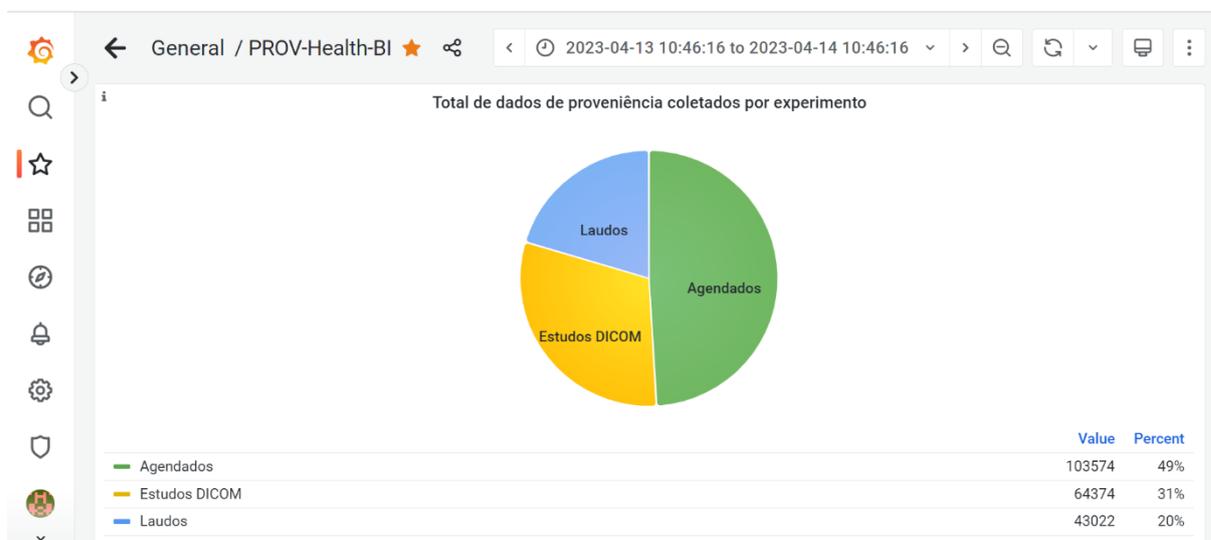
Para a Camada de Atividades (A) foram realizados três experimentos: Experimento 1: Coleta de dados Agendados (Profile SolicitacaoHospitalar); Experimento 2: Coleta de dados Estudos DICOM (Profile EstudoDICOM) e Experimento 3: Coleta de dados Laudos (Profile: LaudoExameHospitalar), ambos para realizar a coleta de dados de proveniência nos repositórios de dados dos hospitais: HIJG, HGCR e HRSJ, correspondendo as Camadas de Entidades (A), (B) e (C) respectivamente. Nesse sentido, o total de dados coletados por meio dos profiles definidos para o PROV-Health (SolicitacaoHospitalar, EstudoDICOM e LaudoExameHospitalar), foram padronizados com os seguintes nomes para fins de apresentação dos dados, sendo respectivamente: Agendados, Estudos DICOM e Laudos para melhor compreensão dos termos dos respectivos experimentos. Assim, utilizando a ferramenta de visualização de dados Grafana, a consulta em SQL apresentada no Quadro 58, possibilita gerar o gráfico visualizado na Figura 43 no PROV-Health-BI com o total de dados de proveniência coletados por experimento.

Quadro 58 – Código realizado no PostgreSQL para a totalidade dos dados por experimento

```
select
(SELECT COUNT(*) from agendados) as "Agendados",
(SELECT COUNT(*) from estudos) as "Estudos DICOM",
(SELECT COUNT(*) from laudos) as "Laudos"
```

Fonte: Elaborado pelo autor (2023).

Figura 43 – Total de dados coletados por experimento da Camada de Atividades (A)



Fonte: Grafana – PROV-Health-BI (2023).

Através do código do Quadro 58 que gerou a Figura 43 apresenta-se o total de dados de proveniência coletados por experimento, sendo: Agendados (103.574 – 49%), Estudos DICOM (64.374 – 31%) e Laudos (43.022 – 20%), correspondentes ao total de 210.970 dados coletados pelo PROV-Health no STT/RCTM na Camada de Atividades (A) nas bases de dados e sistemas integrados dos hospitais HIJG, HGCR e HRSJ, correspondentes às Camadas de Entidades (A), (B) e (C). Dos diversos dados existentes, essa amostra foi possível de apresentar a eficiência do PROV-Health para a coleta de dados de proveniência dos hospitais selecionados. Ressalta-se, que esses processos de coleta realizados na Camada de Atividades (A), foi executado pelo único agente com autorização no PROV-Health para executar processos de TI, sendo o administrador da rede de dados de saúde do STT/RCTM da Camada de Agentes (A) do PROV-Health. Nesse sentido, destaca-se que os dados coletados de Agendados são obtidos a partir de sistemas integrados com o HIS de cada hospital. Já os dados coletados em Estudos DICOM, são obtidos a partir da base de dados utilizada em cada hospital. E por fim, os dados coletados dos Laudos, são obtidos a partir do sistema integrado com o STT/RCTM.

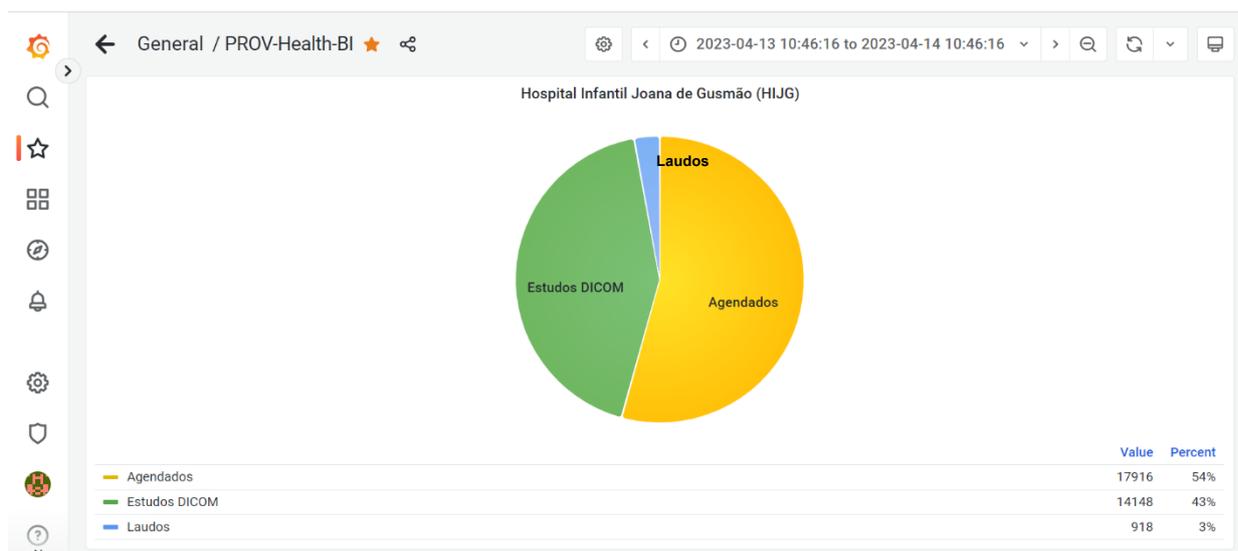
Para os experimentos Agendados, Estudos DICOM e Laudos realizados na Camada de Atividades (A), no que compete à coleta dos dados de proveniência dos hospitais HIJG, HGCR e HRSJ correspondente às Camadas de Entidades (A), (B) e (C) do PROV-Health experimentado no STT/RCTM, o PROV-Health-BI apresenta o total de dados de proveniência coletados por experimento e por hospital, como pode ser observado nos códigos em SQL dos Quadros 59, 60 e 61, gerando os respectivos gráficos das Figuras 44, 45 e 46.

Quadro 59 – Código realizado no PostgreSQL para a totalidade dos experimentos coletados do HIJG

```
select (select count(*) from estudos where organizacao_id='2691868') as "Estudos DICOM",
(select count(*) from agendados where organizacao_id='2691868') as "Agendados",
(select count(*) from laudos where organizacao_id='2691868') as "Laudos"
```

Fonte: Elaborado pelo autor (2023).

Figura 44 – Total da coleta de dados de proveniência por experimento no HIJG



Fonte: Grafana – PROV-Health-BI (2023).

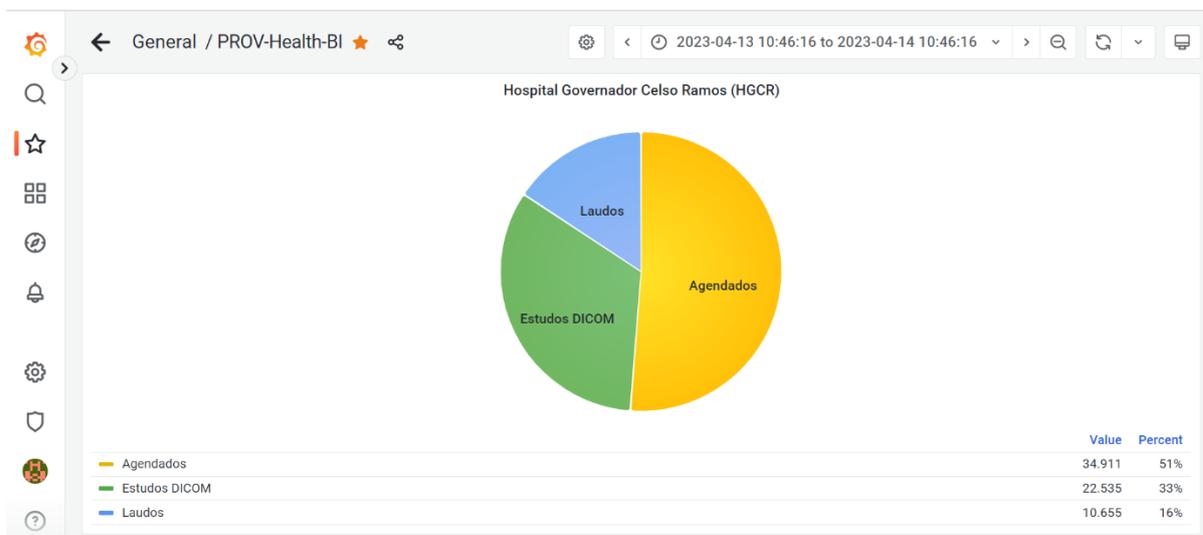
Por meio do código do Quadro 59, que gerou o gráfico da Figura 44, é apresentado o total de dados de proveniência coletados do HIJG referente aos experimentos: Agendados (17.916 – 54%), Estudos DICOM (14.148 – 43%) e Laudos (918 – 3%), correspondendo aos valores e percentuais da amostra de 32.982 dados analisados para os três experimentos desse hospital. Com relação ao objetivo de apresentar a coleta dos dados de proveniência do HIJG, o PROV-Health-BI do PROV-Health, apresentou-se eficaz para essa experimentação realizada no STT/RCTM.

Quadro 60 – Código realizado no PostgreSQL para a totalidade dos experimentos coletados do HGCR

```
select (select count(*) from estudos where organizacao_id='2691841') as "Estudos DICOM",
(select count(*) from agendados where organizacao_id='2691841') as "Agendados",
(select count(*) from laudos where organizacao_id='2691841') as "Laudos"
```

Fonte: Elaborado pelo autor (2023).

Figura 45 – Total da coleta de dados de proveniência por experimento no HGCR



Fonte: Elaborado pelo autor (2023).

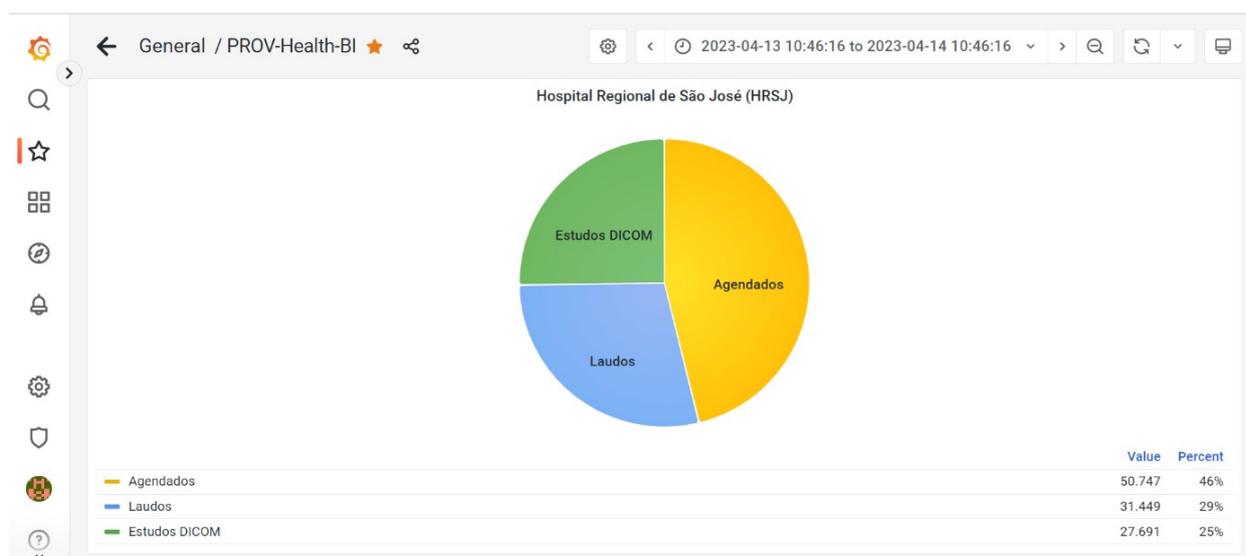
Por meio do código do Quadro 60, que gerou o gráfico da Figura 45, é apresentado o total de dados de proveniência coletados do HGCR referente aos experimentos: Agendados (34.911 – 51%), Estudos DICOM (22.535 – 33%) e Laudos (10.655 – 16%), correspondendo aos valores e percentuais da amostra de 68.101 dados analisados para os três experimentos neste hospital. Para o objetivo de apresentar a coleta dos dados de proveniência do HGCR, o PROV-Health por meio do PROV-Health-BI, apresentou-se eficaz para essa experimentação realizada no STT/RCTM.

Quadro 61 – Código realizado no PostgreSQL para a totalidade dos experimentos coletados do HRSJ

```
select (select count(*) from estudos where organizacao_id='2555646') as "Estudos DICOM",
(select count(*) from agendados where organizacao_id='2555646') as "Agendados",
(select count(*) from laudos where organizacao_id='2555646') as "Laudos"
```

Fonte: Elaborado pelo autor (2023).

Figura 46 – Total da coleta de dados de proveniência por experimento no HRSJ



Fonte: Elaborado pelo autor (2023).

Por meio do código do Quadro 61, que gerou o gráfico da Figura 46, é apresentado o total de dados de proveniência coletados do HRSJ referente aos experimentos: Agendados (50.747 – 46%), Estudos DICOM (27.691 – 25%) e Laudos (31.449 – 29%), correspondendo aos valores e percentuais da amostra de 109.887 dados analisados para os três experimentos neste hospital. Para o objetivo de apresentar a coleta dos dados de proveniência do HRSJ, o PROV-Health por meio do PROV-Health-BI, apresentou-se eficaz para essa experimentação realizada no STT/RCTM.

Portanto, no contexto da descrição da metodologia de experimentação do PROV-Health, no que compete os experimentos 1, 2 e 3 (respectivamente Agendados, Estudos DICOM e Laudos para fins de visualização no PROV-Health-BI) realizados para a Camada de Atividades (A) no Módulo (A) de captura, coleta e preparação dos dados de proveniência do STT/RCTM, foi possível atingir o objetivo de visualizar os conjuntos de dados obtidos com a execução de cada experimento em cada hospital e apresentá-los na Camada de Atividades (C) no que compete ao Módulo (C) de visualização, consulta e análise dos dados de proveniência do STT/RCTM. Assim, destaca-se que, os processos de coleta de dados de proveniência realizados nos três experimentos, constituem importantes características de proveniência de dados para o STT/RCTM, beneficiando diretamente as Camadas de Agentes (A) no auxílio do gerenciamento dos grandes volumes de dados de saúde gerados no contexto da Telerradiologia, assim como, beneficiando a Camada de Agentes (D) no auxílio das tomadas de decisões gerenciais com relação ao fluxo de dados observados. Como resultados preliminares desses três experimentos realizados quanto a coleta dos dados de proveniência do STT/RCTM, evidencia-

se a redução dos esforços necessários para a rastreabilidade de dados de proveniência, aprimorando a qualidade e veracidade de dados, características importantes dada a crescente complexidade na geração de conjuntos de dados de saúde do STT/RCTM. Porém, além dos benefícios, o PROV-Health apresentou conflitos de configuração de ferramentas computacionais no STT/RCTM, pois, como os dados estavam descentralizados em diferentes redes, ferramentas computacionais utilizadas nesses experimentos apresentaram certa complexidade de configuração na coleta dos dados de proveniência. Na seção 7, uma análise mais detalhada será traçada, observando diferentes abordagens em relação aos três experimentos realizados em no tocante à coleta de dados de proveniência no STT/RCTM.

6.8.5.1.2 Visualização dos dados de proveniência dos experimentos da Camada de Atividades (B)

Para a Camada de Atividades (B) foram realizados três experimentos: Experimento 1: Integração de solicitações hospitalares com estudos DICOM; Experimento 2: Integração de solicitações hospitalares com laudos publicados e Experimento 3: Integração de laudos com estudos DICOM. Esses experimentos realizaram as verificações das integrações existentes nos repositórios de dados dos hospitais: HIJG, HGCR e HRSJ, correspondendo as Camadas de Entidades (A), (B) e (C) respectivamente. Nesse sentido, o total de integrações realizadas, assim como as integrações realizadas em cada experimento para cada hospital, são apresentados nos códigos referentes aos Quadros 62, 63 e 64, que possibilitaram a geração dos gráficos das Figuras 47, 48 e 49, respectivamente, onde “t” igual a “true” são consideradas as integrações realizadas “Integrado” e “f” igual a “false” as integrações não realizadas “Não integrado”. O PROV-Health-BI, utilizando-se da ferramenta de visualização de dados Grafana, demonstra essas integrações por meio dos códigos e gráficos mencionados.

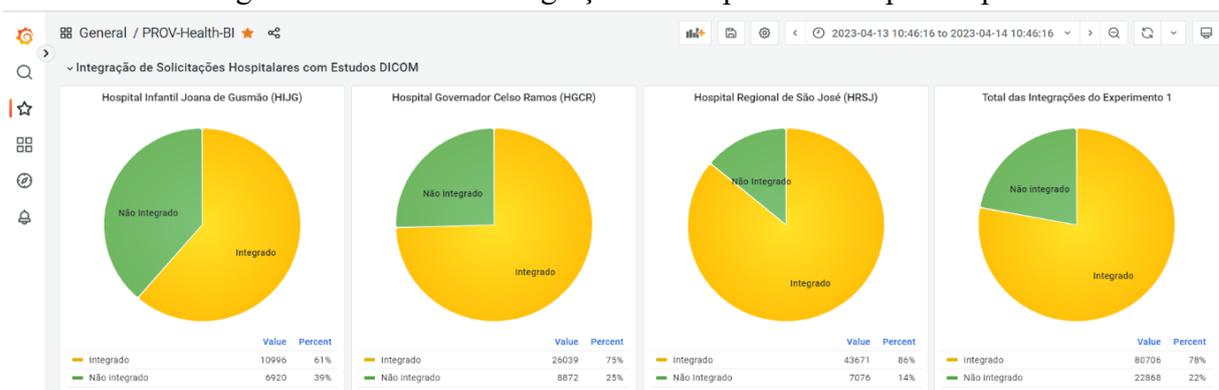
Quadro 62– Códigos realizados no PostgreSQL para as integrações do experimento 1 por hospital

| Código | Hospital |
|--|----------|
| select (select count(*) from agendados where integrado_pacs ='f' and organizacao_id='2691868') as "Não integrado", (select count(*) from agendados where integrado_pacs ='t' and organizacao_id='2691868') as "Integrado" | HIJG |
| select (select count(*) from agendados where integrado_pacs ='f' and organizacao_id='2691841') as "Não integrado", (select count(*) from agendados where integrado_pacs ='t' and organizacao_id='2691841') as "Integrado" | HGCR |
| select (select count(*) from agendados where integrado_pacs ='f' and | HRSJ |

| | |
|--|-----------------------|
| organizacao_id='2555646') as "Não integrado", (select count(*) from agendados where integrado_pacs ='t' and organizacao_id='2555646') as "Integrado" | |
| select (select count(*) from agendados where integrado_pacs ='f') as "Não integrado", (select count(*) from agendados where integrado_pacs ='t') as "Integrado" | Total das integrações |

Fonte: Elaborado pelo autor (2023).

Figura 47 – Total das integrações no Experimento 1 por hospital



Fonte: Elaborado pelo autor (2023).

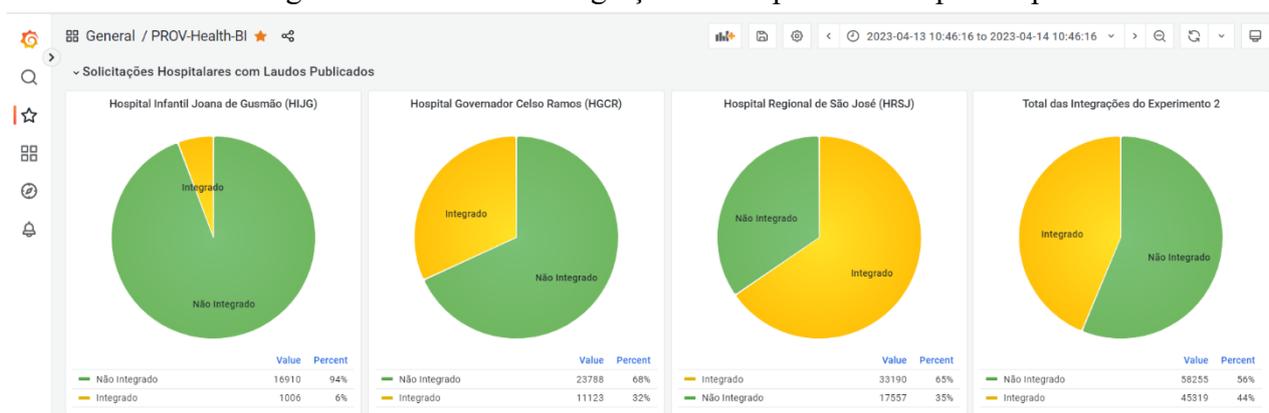
Conforme demonstrado na Figura 47, as integrações realizadas em cada hospital, por meio dos códigos do Quadro 62, estão discriminadas da seguinte maneira: HIJG (10.996 – Integrado (61%) / 6.920 – Não integrado (39%) correspondendo aos 17.916 em Agendados para esse hospital); HGCR (26.039 – Integrado (75%) / 8.872 – Não integrado (25%) correspondendo aos 34.911 Agendados para esse hospital); HRSJ (43.671 – Integrado (86%) / 7.076 – Não integrado (14%) correspondendo aos 50.747 dos Agendados para esse hospital), totalizando 80.706 integrados e 22.868 não integrados.

Quadro 63 – Códigos realizados no PostgreSQL para as integrações do experimento 2 por hospital

| Código | Hospital |
|--|-----------------------|
| select (select count(*) from agendados where laudo_publicado ='f' and organizacao_id='2691868') as "Não integrado", (select count(*) from agendados where laudo_publicado ='t' and organizacao_id='2691868') as "Integrado" | HIJG |
| select (select count(*) from agendados where laudo_publicado ='f' and organizacao_id='2691841') as "Não integrado", (select count(*) from agendados where laudo_publicado ='t' and organizacao_id='2691841') as "Integrado" | HGCR |
| select (select count(*) from agendados where laudo_publicado ='f' and organizacao_id='2555646') as "Não integrado", (select count(*) from agendados where laudo_publicado ='t' and organizacao_id='2555646') as "Integrado" | HRSJ |
| select (select count(*) from agendados where laudo_publicado ='f') as "Não integrado", (select count(*) from agendados where laudo_publicado ='t') as "Integrado" | Total das integrações |

Fonte: Elaborado pelo autor (2023).

Figura 48 – Total das integrações do Experimento 2 por hospital



Fonte: Elaborado pelo autor (2023).

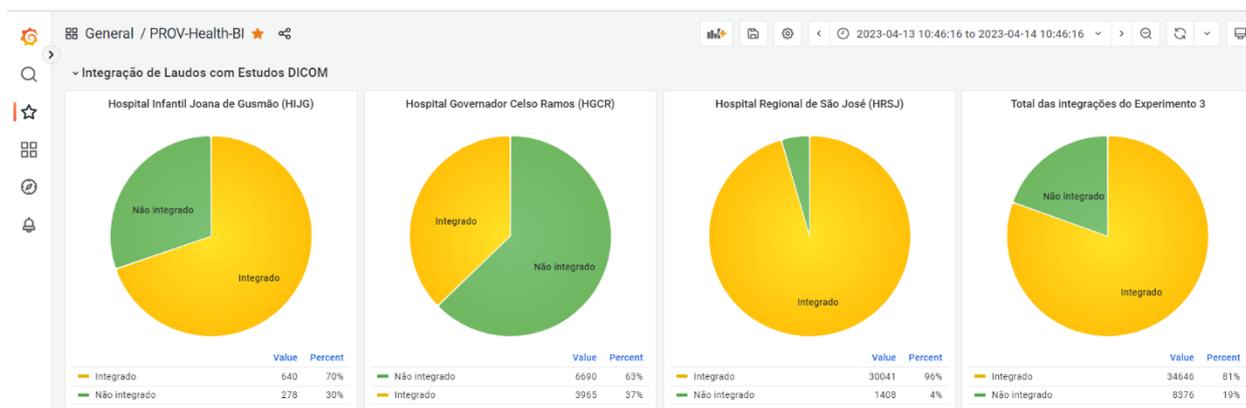
As integrações realizadas em cada hospital, conforme demonstrado na Figura 48 por meio dos códigos do Quadro 63, estão discriminadas da seguinte maneira: HIJG (16.910 – Não integrado (94%)/1.006 – Integrado (6%) correspondendo aos 17.916 Agendados para esse hospital), HGCR (23.778 – Não integrado (68%)/11.123 – Integrado (32%) correspondendo aos 34.911 Agendados para esse hospital), HRSJ (33.190 – Integrado/17.557 – Não integrado (35%) correspondendo aos 50.747 Agendados para esse hospital), totalizando 45.319 integrados e 58.255 não integrados.

Quadro 64 – Códigos realizados no PostgreSQL para as integrações do experimento 3 por hospital

| Código | Hospital |
|--|-----------------|
| select (select count(*) from laudos where integrado_pacs = 'f' and organizacao_id='2691868') as "Não integrado", (select count(*) from laudos where integrado_pacs = 't' and organizacao_id='2691868') as "Integrado" | HIJG |
| select (select count(*) from laudos where integrado_pacs = 'f' and organizacao_id = '2691841') as "Não integrado", (select count(*) from laudos where integrado_pacs = 't' and organizacao_id = '2691841') as "Integrado" | HGCR |
| select (select count(*) from laudos where integrado_pacs = 'f' and organizacao_id='2555646') as "Não integrado", (select count(*) from laudos where integrado_pacs = 't' and organizacao_id='2555646') as "Integrado" | HRSJ |
| select (select count(*) from laudos where integrado_pacs = 'f') as "Não integrado", (select count(*) from laudos where integrado_pacs = 't') as "Integrado" | Total Hospitais |

Fonte: Elaborado pelo autor (2023).

Figura 49 – Total das integrações do Experimento 3 por hospital



Fonte: Elaborado pelo autor (2023).

As integrações realizadas em cada hospital, conforme a Figura 49 por meio dos códigos do Quadro 64, estão discriminadas da seguinte maneira: HIJG (640 – Integrados (70%)/278 – Não integrados (30%) correspondendo aos 918 Laudos para esse hospital), HGSR (3.965 – Integrados (37%)/6.690 – Não integrados (63%) correspondendo aos 10.655 Laudos para esse hospital), HRSJ (30.041 – Integrado (96%)/1.408 – Não integrado (4%) correspondendo aos 31.449 Laudos para esse hospital), totalizando 34.446 integrados e 8.376 não integrados.

Portanto, no contexto da descrição da metodologia de experimentação do PROV-Health no STT/RCTM, no que compete aos experimentos 1, 2 e 3 realizados pela Camada de Atividades (B) no Módulo (B) organização, controle e armazenamento dos dados de proveniência do STT/RCTM, foi possível atingir o objetivo de visualizar os conjuntos de dados obtidos com a execução da cada experimento em cada hospital e apresentá-los na Camada de Atividades (C) no que compete ao Módulo (C) de visualização, consulta e análise dos dados de proveniência do STT/RCTM. Os processos de controle (integrações) realizados nos três experimentos, constituem importantes características de proveniência de dados no STT/RCTM, beneficiando diretamente a Camada de Agentes (A) no monitoramento e análises das integrações existentes, possibilitando escalonar tarefas, detectar falhas e realizar adaptações que promovem o controle do fluxo informacional existente na Telerradiologia do STT/RCTM. Dessa forma, como resultados preliminares dos três experimentos realizados em relação às integrações dos dados de proveniência do STT/RCTM, evidencia-se desempenhos distintos, porém, integrações satisfatórias de um ponto de vista técnico, dada a crescente complexidade na geração de conjuntos de dados de saúde do STT/RCTM. Por outro lado, em alguns casos, a não utilização do Worklist e o não cadastramento de exames nos SIS integrados ao STT/RCTM. Ainda é possível constatar, que o PROV-Health pode não trazer todas as integrações existentes dos resultados das coletas anteriormente realizadas, ou seja, a granularidade dos dados

coletados é alta, podendo apontar diferentes tipos de falhas. Na Seção 7 uma análise mais detalhada é traçada, observando diferentes abordagens em relação aos três experimentos realizados no quesito de controle (integrações) dos dados de proveniência no STT/RCTM.

6.9 DESAFIOS E AMEAÇAS ENCONTRADOS NA IMPLEMENTAÇÃO E EXPERIMENTAÇÃO DO PROV-HEALTH NO STT/RCTM

Mediante as experimentações realizadas no PROV-Health até o momento no contexto do STT/RCTM, foram encontrados alguns desafios e ameaças, que influenciam para a possível continuação e aprimoramento do método por outros pesquisadores, conforme listado a seguir:

- a) Apesar das tecnologias computacionais utilizadas na estratégia computacional do PROV-Health serem *open source*, houve dificuldades de implementação e conexões entre as diferentes tecnologias já existentes na infraestrutura de TI do STT/RCTM, sendo essas, em alguns casos, complexas de estruturação para compor a arquitetura computacional do PROV-Health;
- b) A estratégia computacional apresentada pelo PROV-Health em suas camadas, pode sofrer mudanças para aplicações em diferentes cenários de saúde, necessitando a inserção de tecnologias mais abrangentes do que as propostas e alguns casos a necessidade de sistemas proprietários que dependam de licença, inviabilizando os custos do PROV-Health para instituições públicas de saúde;
- c) Considerando que as bases de dados dos SIS estão em um contexto descentralizado, são necessários ajustes nas estratégias de coleta dos dados pelo PROV-Health para englobar os vários subsistemas que compõem os dados de saúde de todas as práticas médicas desenvolvidas nos diferentes SIS;
- d) Os processos de interoperabilidade no uso do HL7 FHIR utilizado pelo PROV-Health contribuiu em partes para que seja possível relacionar a proveniência dos dados, sendo que desafios nesse quesito ainda continuam persistindo com os diferentes formatos de dados contidos em toda a estrutura computacional dos diferentes SIS;
- e) A disposição dos dados nem sempre está acessível em formato digital, sendo que em alguns casos, procedimentos médicos são inseridos manualmente nos SIS dos hospitais onde foram coletados os dados, não sendo identificados tão facilmente nas coletas realizadas, havendo a necessidade de novas coletas por parte do PROV-Health;
- f) As redes de dados de saúde dos hospitais, podem exigir acessos restritos que impossibilitem a coleta completa de dados de proveniência por parte do PROV-Health,

- inviabilizando o conhecimento do fluxo informacional ao qual se pretende gerenciar;
- g) A complexidade e a velocidade em que as ações em saúde devem ser processadas e analisadas, além da alta taxa de mudança nos processos econômicos e tecnológicos, tornam o PROV-Health ainda mais crítico, fazendo com que o foco embutido em suas camadas com princípios de gestão e interoperabilidade de dados de proveniência seja visto como um grande desafio e ameaça para sua estrutura tecnológica por utilizar-se de tecnologias computacionais *open source*.

Por fim, esses pontos foram a realidade encontrada até o momento dos experimentos realizados no PROV-Health, enfatizando que houve complexidades significativas na experimentação, pois nem todas as ferramentas computacionais se conectam facilmente ao ambiente do estudo de caso de grande escala (STT/RCTM), sendo necessária a compreensão aprofundada de todo o complexo fluxo informacional existente para que os experimentos fossem passíveis de execução. Assim, é necessário dar continuidade na ampliação do PROV-Health com novas ferramentas, buscando encontrar desafios e ameaças que podem surgir em novos experimentos, garantindo às necessidades de incorporação de outras tecnologias à medida que a descentralização da gestão dos dados de proveniência se torna cada vez mais concreta.

6.10 CONSIDERAÇÕES DA SEÇÃO

Nesta seção foi apresentado o estudo de caso para experimentação do PROV-Health no STT/RCTM como forma de validação do método. Para tanto, foram descritas na arquitetura específica de experimentação do PROV-Health as Camadas de Agentes (A), (B), (C) e (D), Camadas de Entidades (A), (B) e (C) e Camadas de Atividades (A), (B) e (C), sendo nessas últimas camadas especificado cada módulo. As Camadas de Atividades (A) e (B), apresentaram experimentos relacionados com base na estrutura da metodologia de experimentação definida para o STT/RCTM, ambiente de estudo de caso desta tese no âmbito da Telerradiologia. Nesse contexto, na Camada de Atividades (A), a estratégia computacional foi detalhada sendo realizado três experimentos referente a coleta de dados de proveniência, apresentando uma amostra total de 210.970 dados coletados pelo PROV-Health no STT/RCTM nas bases de dados e sistemas integrados dos hospitais estudados.

Desse modo, foi possível capturar, coletar e preparar os dados de proveniência e centralizá-los no servidor HAPI FHIR do PROV-Health, possibilitando por meio dos três experimentos executados, apresentar a geração dos grandes volumes de dados oriundos de

diferentes SIS integrados de forma descentralizada no STT/RCTM. Já os três experimentos executados na Camada de Atividades (B) relacionados às integrações por meio da estratégia computacional detalhada, foram significativos para a validação das integrações mais relevantes para o estudo de caso no STT/RCTM no uso do PROV-Health, observando um contexto geral de integrações realizadas e não realizadas com base em comparações executadas nos dados de proveniência do PROV-Health. Ainda, desafios e ameaças foram observados na experimentação do PROV-Health, evidenciando a continuação desse método no âmbito de desenvolvimento tecnológico do STT/RCTM, constituindo novas oportunidades de conhecimento com base nos dados de proveniência que forem gerados.

Por fim, vale destacar que todos os resultados encontrados na experimentação do PROV-Health em suas camadas e as relações causais do modelo W3C PROV foram levadas em consideração, pois partiram do uso do HL7 FHIR com base nas premissas do W3C PROV e suas características.

7 ANÁLISES E AVALIAÇÕES DA EXPERIMENTAÇÃO DO PROV-HEALTH

Nesta seção são efetuadas as análises dos experimentos realizados na Seção 6, assim como são apresentados os resultados da pesquisa realizada com os especialistas para avaliar o PROV-Health. As análises dos experimentos executados nas Camadas de Atividades (A) e (B) e visualizados na Camada de Atividades (C) da arquitetura de experimentação do PROV-Health no STT/RCTM foram estudados no contexto de coleta, integração e visualização dos dados de proveniência, os quais possuem suas relações causais intrinsicamente embutidos no PROV-Health por meio do instanciamento do modelo W3C PROV. Foram realizadas análises dos dados de proveniência para todos os experimentos das Camadas de Atividades (A) e (B), assim como analisados os resultados encontrados e enviados para a Camada de Atividades (C).

A análise dos dados de proveniência armazenados pelo PROV-Health, voltou-se ao quesito operacional de tarefas executadas pela Camada de Agentes (A). Por sua vez, a análise voltada para as principais características da Proveniência de Dados, foi realizada com base no estudo de Sembay, Marquez Filho e Macedo (2022a), no qual os autores utilizam de parte da taxonomia de Simmhan, Plale e Gannon (2005) para identificar as principais características da Proveniência de Dados, sendo essas características já definidas no referencial teórico desta tese, quais sejam: qualidade dos dados, trilhas de auditoria, receitas de replicação, atribuição e informacional.

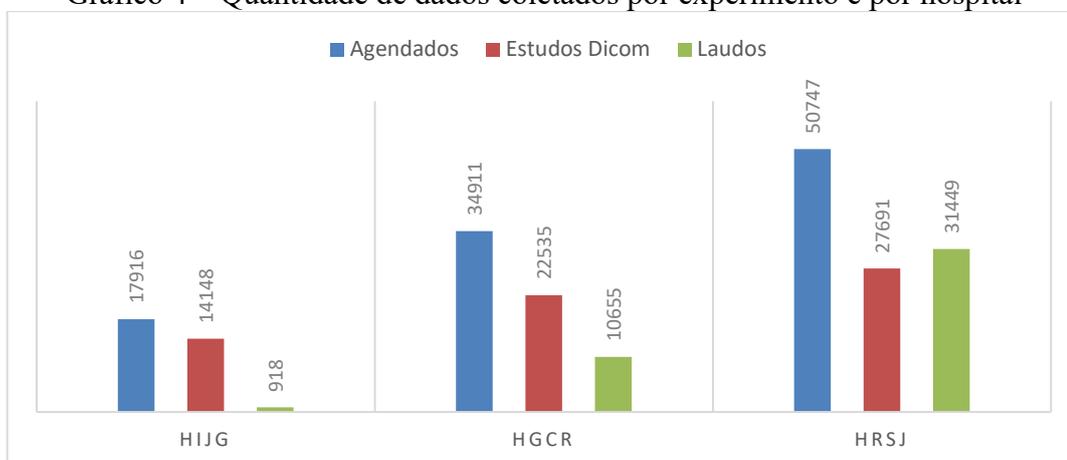
Já no que diz respeito à pesquisa realizada com os especialistas para avaliação do PROV-Health, são descritas as respostas guiadas pelo formulário eletrônico do *Google Forms* enviado por e-mail, apresentando, assim, as percepções fornecidas pela avaliação dos especialistas convidados a avaliar o PROV-Health no STT/RCTM.

7.1 ANÁLISES DOS EXPERIMENTOS DE COLETAS NO PROV-HEALTH

Foram realizadas as análises pertinentes dos três experimentos realizados em relação à Camada de Atividades (A) no Módulo (A) de captura, coleta e preparação dos dados de proveniência no STT/RCTM — Coleta de dados Agendados (*Profile SolicitacaoHospitalar*), Coleta de dados Estudos DICOM (*Profile: EstudoDICOM*) e Coleta de dados Laudos (*Profile: LaudoExameHospitalar*) —, executadas com base na instrumentalização específica dos serviços de saúde reduzida para uso no STT/RCTM já descrita na subseção 6.6.2 da Seção 6. Assim, observa-se no Gráfico 4, a quantidade de dados coletados pelos três experimentos e por hospital, no que refere às Camadas de Entidades (A), (B) e (C), respectivamente para os

hospitais: HIJG, HGCR e HRSJ.

Gráfico 4 – Quantidade de dados coletados por experimento e por hospital



Fonte: Elaborado pelo autor (2023).

Partindo do pressuposto de que todos os exames agendados (Agendados) devem conter imagens (Estudos DICOM) e laudos (Laudos), observa-se as seguintes análises do Gráfico 4:

- k) Observa-se que o HIJG é o hospital que apresenta maior quantidade de exames agendados (17.916) em relação aos Estudos DICOM (14.148), ou seja, esse hospital tem 78,9% dos exames agendados associados com imagens, o que é satisfatório para as práticas médicas hospitalares. Porém, a baixa quantidade de Laudos publicados (918) corresponde apenas a 6,5% dos exames agendados com imagens, evidenciando, assim, que os profissionais especialistas desse hospital não têm o STT/RCTM como sistema preferencial para os laudos. Esse resultado também pode estar relacionado aos altos volumes de exames que dispensam a necessidade de laudo, como por exemplo, exames em que o médico especialista se utiliza de equipamentos médicos portáteis para diagnosticar uma certa enfermidade radiológica sem a necessidade de um laudo. Dessa forma, evidencia-se que hospitais com esse perfil de utilização têm o interesse na utilização do STT/RCTM direcionado ao armazenamento dos exames e imagens médicas para melhor precisão de seus dados estatísticos.
- l) Quanto ao HGCR, observa-se que existe uma diferença de 64,5% entre exames agendados (34.911) e Estudos DICOM (22.535), sendo um possível fator de evidência da parametrização do processo de integração da Worklist para agrupar os exames agendados por atendimento, de maneira que um estudo DICOM seja vinculado a múltiplos exames agendados, considerando esses, relacionados a um atendimento com múltiplos pedidos de exames. A prática de agrupamento de exames agendados para um

estudo DICOM visa atender uma demanda dos próprios hospitais para integração do HIS com o PACS. Para os laudos (10.655), assume-se 47,3% dos exames agendados com imagens, evidenciando que o HGCR utiliza da publicação de laudos com especificidades para determinados exames de radiologia computadorizada, sendo assim, uma possível causa para o caso de laudos não integrados.

- m) Para o HRSJ, observa-se que dos exames agendados (50.747), 54,5% possuem Estudos DICOM (27.691), ou seja, estão associados às imagens. Já os laudos (31.449) ultrapassam os exames com imagens, sendo 61,9% dos exames agendados. Nesse contexto, é evidenciado o agrupamento de exames agendados para um estudo DICOM, como mostra-se a quantidade de laudos nesse hospital superior a quantidade de estudos DICOM, de maneira que um estudo DICOM também pode estar associado a múltiplos exames e laudos no STT/RCTM, visto que esses são diretamente associados aos exames agendados. O HRSJ destaca-se também pelo maior volume de exames agendados e melhor aproveitamento da integração.

Assim, o processo de proveniência encontrado nesses experimentos quanto à coleta está voltado para a centralização dos dados dos hospitais: HIJG, HGCR e HGSJ no servidor de dados de proveniência HAPI FHIR do PROV-Health, localizado na rede interna do STT/RCTM para fins de visualização, consultas e análises desses dados para tomadas de decisão por parte da Camada de Agentes (D). Desse modo, os três experimentos relacionados à coleta dos dados de proveniência por meio do PROV-Health contribuem para o setor de saúde pública, em que o gestor do STT/RCTM da Camada de Agentes (D), de posse desses dados, pode compreender a complexidade dos setores de saúde pública das regiões de Santa Catarina no âmbito da Telerradiologia, servindo para outras áreas, visualizando dados populacionais de saúde em uma única tela, além de identificar rapidamente onde e como atuar para evitar ou melhorar problemas de saúde pública. Isso de fato não acontece em bases de dados descentralizadas, nas quais essas informações teriam limitações para tomadas de decisão em um contexto mais amplo de saúde.

Ainda, em relação aos três experimentos voltados à coleta de dados de proveniência, ressalta-se que esses proporcionaram possibilidades analíticas e de monitoramento que beneficiaram a compreensão dos fluxos informacionais dos diferentes SIS integrados ao STT/RCTM, que para este estudo de caso foram os hospitais: HIJG, HGCR e HGSJ. O PROV-Health utiliza a base do padrão de interoperabilidade HL7 FHIR, o qual tem compatibilidade com o modelo de proveniência de dados W3C PROV, o que possibilitou questões de interoperabilidade por meio das ferramentas computacionais utilizadas na Camada de

Atividades (A), oportunizando os conceitos práticos de proveniência de dados. Apesar dos benefícios do PROV-Health ao STT/RCTM, as ferramentas utilizadas na Camada de Atividades (A) apresentaram conflitos de configuração nos ambientes de estudo de caso, pois as tecnologias nativas das redes onde foram coletados os dados necessitam de ajustes para que fosse possível realizar os processos de coletas de dados pela Camada de Agentes (A). As ferramentas computacionais utilizadas na estratégia computacional do PROV-Health na Camada de Atividades (A), foram: Simplifier, Simplifier - STT/RCTM on FHIR® (R4), Forge by Firely, Visual Studio Code (VS CODE), Insomnia, Postgresql, SQL, GitLab, Python, Json, HL7 FHIR, DCM4CHE, Linux (Ubuntu), Windows e Bash, as quais tiveram complexidades de configurações no ambiente tecnológico do STT/RCTM, porém, atenderam o objetivo computacional dos experimentos da coleta de dados de proveniência do PROV-Health. Destaca-se que desafios e ameaças, conforme descrito na Seção 6, crescem na medida em que as tecnologias utilizadas e ajustadas no PROV-Health buscam otimizar seu desempenho, o que gera a demanda de ambientes cada vez mais customizáveis.

Nesse contexto, como forma de complementar as análises realizadas nesta subseção, e aprofundar a importância da proveniência de dados aos dados coletados pelo PROV-Health no STT/RCTM, foi realizada uma análise para identificar a existência das principais características da proveniência de dados com base no estudo de Sembay, Marquez Filho e Macedo (2022a) de parte da taxonomia de Simmhan, Plale e Gannon (2005). Portanto, no Quadro 65, é apresentada a existência das principais características da proveniência de dados nos experimentos 1 – Coleta de dados Agendados (profile: SolicitacaoHospitalar), experimento 2 – Coleta de dados Estudos DICOM (profile: EstudoDICOM) e experimento 3 – Coleta de dados Laudos (profile: LaudoExameHospitalar) da Camada de Atividades (A) apresentados nesta subseção.

Quadro 65 – Identificação da existência das principais características da proveniência de dados no PROV-Health experimentado no STT/RCTM (Experimentos de Coletas)

| | | Experimentos de coletas do PROV-Health no STT/RCTM | | |
|---|------------------------|--|--------------------------------------|-------------------------------|
| | | Experimento 1 - Agendados (HIJG) | Experimento 2 - Estudos DICOM (HGCR) | Experimento 3 - Laudos (HRSJ) |
| Principais características da proveniência de dados | Qualidade dos dados | ⊙ | ⊙ | ⊙ |
| | Trilhas de auditoria | ⊙ | ⊙ | ⊙ |
| | Receitas de replicação | ⊙ | ⊙ | ⊙ |
| | Atribuição | ⊙ | ⊙ | ⊙ |
| | Informacional | ⊙ | ⊙ | ⊙ |
| ⊙ Altamente relevante: considerando as características da proveniência de dados que tem efeito direto nos dados; ⊙ Relevante: considerando as características da proveniência de dados que tem efeito indireto sobre os dados; ○ Não identificado: nenhuma característica definida. | | | | |

Fonte: Com base no estudo de Sembay, Marquez Filho e Macedo (2022a).

Em relação à qualidade dos dados para os experimentos 1, 2 e 3, essa característica é altamente relevante, pois os dados coletados dos hospitais, são qualificados pelo PROV-Health sendo possível a comprovação das transformações e derivações existentes quando comparados com a base de dados original.

Já em relação às trilhas de auditoria para os experimentos 1, 2 e 3, essa característica é altamente relevante, considerando que a coleta dos dados no PROV-Health possibilita coletar os dados dos hospitais de forma a centralizá-los, ocasionando o rastreamento para realização de possíveis auditorias conforme a necessidade operacional.

No que diz respeito às receitas de replicação para os experimentos 1, 2 e 3, essa característica se mostra relevante, pois o PROV-Health tem o objetivo de centralizar dados de bases descentralizadas sem a necessidade de repetir a derivação desses dados, apesar de os mesmos se manterem atualizados na base de origem.

Com relação à atribuição para os experimentos 1, 2 e 3, essa característica é altamente relevante, pois a proveniência dos dados no PROV-Health é preservada dentro da rede interna do STT/RCTM possibilitando estabelecer os direitos autorais e suas propriedades, assim como possibilitar a identificação dos responsáveis pelos dados incorretos.

No que se refere à informacional para os experimentos 1, 2 e 3, essa característica é altamente relevante, pois os dados coletados pelo PROV-Health vêm acompanhado de seus metadados para a descoberta de dados, possibilitando compreender e interpretar os dados de proveniência.

Por fim, não houve características não identificadas nas análises aqui realizadas nos

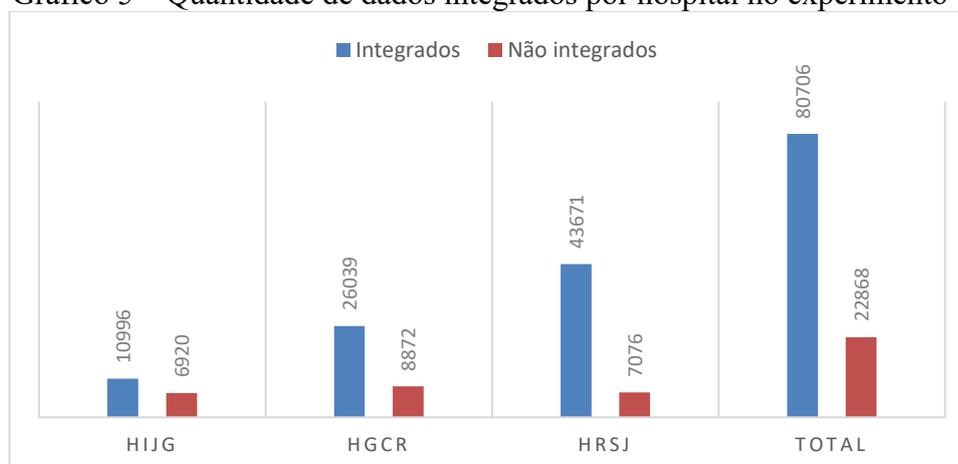
experimentos 1, 2 e 3 em relação aos dados coletados, confirmando a proveniência de dados em todos os processos de coleta do PROV-Health aplicados ao STT/RCTM. Porém, vale ressaltar que os dados dos três experimentos analisados em relação à coleta, foram devidamente selecionados como relevantes, e que a descoberta de quais dados precisam ser analisados nos processos de coleta do PROV-Health e a decisão de quais parâmetros de análise utilizar, podem ser tarefas complexas, uma vez que os dados coletados são de fontes descentralizadas de diferentes SIS e instituições de saúde, podendo haver falhas na interpretação do contexto de proveniência de dados em sua centralização no PROV-Health.

7.2 ANÁLISES DOS EXPERIMENTOS DE INTEGRAÇÕES DO PROV-HEALTH

No que compete aos três experimentos realizados, com relação às integrações dos dados de proveniência na Camada de Atividades (B) no Módulo (B) organização, controle e armazenamento dos dados de proveniência do STT/RCTM da arquitetura de experimentação do PROV-Health — Integração de solicitações hospitalares com estudos DICOM, integração de solicitações hospitalares com laudos publicados e integração de laudos com estudos DICOM —, observa-se no Gráfico 5 os parâmetros encontrados nos hospitais referente às Camadas de Entidades (A), (B) e (C), respectivamente HIJG, HGCR e HGSJ.

Assim, em relação ao experimento 1: Integração de solicitações hospitalares com estudos DICOM, o Gráfico 5 apresenta os totais encontrados por hospitais.

Gráfico 5 – Quantidade de dados integrados por hospital no experimento 1



Fonte: Elaborado pelo autor (2023).

Para esse experimento foram atualizadas as solicitações hospitalares (Agendados) para controlar o estado de integração com estudos DICOM, com objetivo de avaliar a integração dos

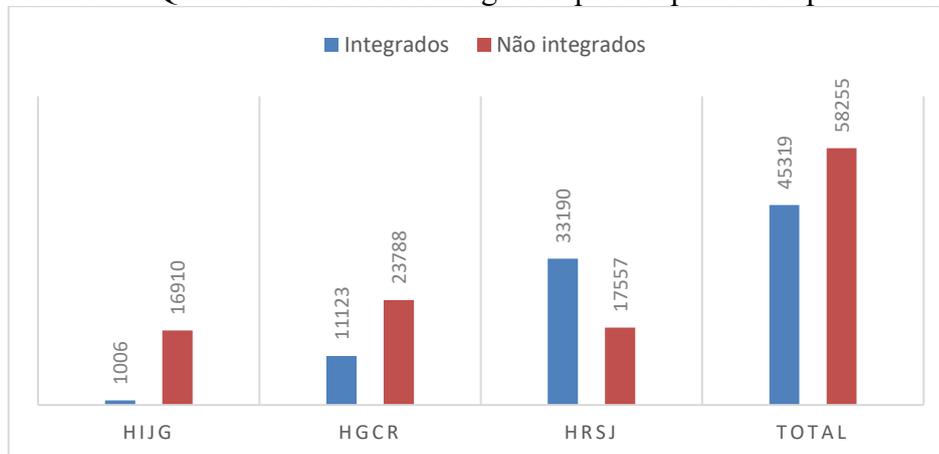
agendamentos nos equipamentos a partir do serviço de Worklist. Dessa maneira, quando um exame agendado possui vínculo com um estudo DICOM, o serviço de Worklist foi utilizado para realização do exame, enquanto, para casos de exames agendados sem estudos DICOM relacionados, ou não foram realizados, ou foram cadastrados manualmente e não serão integrados. Assim, o Gráfico 5 apresenta os exames integrados e não integrados em relação ao experimento 1 executado na Camada de Atividades (B) do PROV-Health, contendo as seguintes especificações:

- a) HIJG: do total de 17.996 registros de exames agendados, 10.996 estudos DICOM foram integrados (61%) e 6.920 não integraram (39%);
- b) HGCR: do total de 34.911 registros de exames agendados, 26.039 estudos DICOM foram integrados (75%) e 8.872 não integraram (25%);
- c) HRSJ: do total de 50.747 registros de exames agendados, 43.671 estudos DICOM foram integrados (86%) e 7.076 não integraram (14%).

Assim, dos 103.574 exames agendados, 80.706 foram integrados e 22.868 não foram integrados, sendo, no geral, um resultado satisfatório, se for considerada a complexidade dos processos de saúde existentes no âmbito da Telerradiologia do STT/RCTM. Destaca-se o HIJG que apresentou o melhor índice de integração no experimento 1. Assim, o experimento 1 permitiu avaliar e verificar ações de melhorias necessárias por parte da Camada de Agentes (D) para melhor aproveitamento do fluxo de integração, buscando compreender as falhas ocorridas durante o processo de integração dos exames.

Por sua vez, o experimento 2 – integração de solicitações hospitalares com laudos publicados, são apresentados no Gráfico 6 os totais encontrados por hospitais.

Gráfico 6 – Quantidade de dados integrados por hospital no experimento 2



Fonte: Elaborado pelo autor (2023).

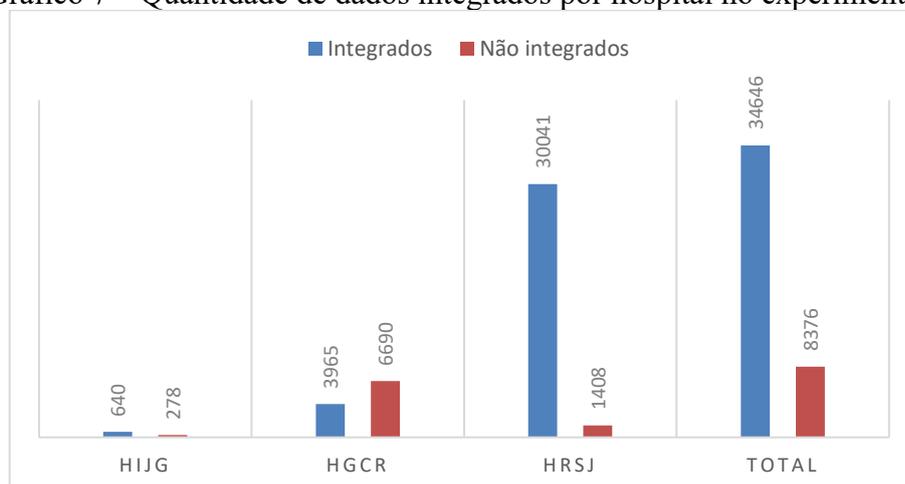
Para esse experimento, foram atualizadas as solicitações hospitalares para controlar o estado de integração com laudos publicados. Dessa maneira, quando um exame agendado possui vínculo com um laudo publicado no STT/RCTM, conclui-se todas as etapas de integrações previstas, enquanto, para casos de exames agendados sem laudos relacionados, ou não foram integrados ou não foram publicados. Assim, o Gráfico 6 apresenta os exames integrados e não integrados em relação ao experimento 2 executado na Camada de Atividades (B) do PROV-Health, contendo as seguintes especificações:

- a) HIJG: do total de 17.996 registros de exames agendados, 1.006 laudos foram integrados (6%) e 16.910 não integraram (94%);
- b) HGCR: do total de 34.911 registros de exames agendados, 11.123 laudos foram integrados (32%) e 23.788 não integraram (68%);
- c) HRSJ: do total de 50.747 registros de exames agendados, 33.190 laudos foram integrados (65%) e 17.557 não integraram (35%).

À vista disso, dos 103.574 exames agendados, 45.319 foram integrados e 58.255 não foram integrados, sendo esse resultado considerado insatisfatório perante os processos de saúde existentes no âmbito da Telerradiologia do STT/RCTM. Apesar das falhas ocorridas, o único hospital que seguiu com as integrações de forma cronológica, foi o HRSJ. Assim, o experimento 2 permitiu avaliar desempenho da integração de laudos por instituição de saúde, de maneira a colaborar para a Camada de Agentes (D) nas verificações e avaliações de melhorias para melhor desempenho dessa integração, buscando compreender as falhas ocorridas durante o processo de integração dos exames no âmbito da Telerradiologia no STT/RCTM.

Em relação ao experimento 3 – integração de laudos com estudos DICOM, no Gráfico 7 são apresentados os totais encontrados por hospitais.

Gráfico 7 – Quantidade de dados integrados por hospital no experimento 3



Fonte: Elaborado pelo autor (2023).

Para esse experimento, foram atualizados os recursos *Diagnostic Report* de laudos, para controlar o estado de integração com estudos DICOM. Dessa maneira, quando um laudo do STT/RCTM possui vínculo com um estudo DICOM, conclui-se que a integração do exame agendado com estudo DICOM também ocorreu com sucesso, enquanto, para casos de laudos no STT/RCTM sem estudos DICOM vinculados, evidencia-se que mesmo não ocorrendo a integração, o exame foi realizado visto que o médico especialista publicou um laudo relacionado e que, geralmente, também está relacionado a essa falha de integração a falta de suporte do equipamento para realizar consultas no serviço de Worklist. Assim, o Gráfico 7 apresenta os exames integrados e não integrados em relação ao experimento 3 executado na Camada de Atividades (B) do PROV-Health, contendo as seguintes especificações:

- a) HIJG: do total de 918 registros de laudos integrados, 640 laudos foram integrados (70%) e 278 não integraram (30%);
- b) HGCR: do total de 10.655 registros de exames agendados, 3.965 laudos foram integrados (37%) e 6.690 não integraram (63%);
- c) HRSJ: do total de 31.449 registros de exames agendados, 30.041 laudos foram integrados (96%) e 1.408 não integraram (4%).

Dessa forma, dos 43.022 exames de laudos, 34.646 foram integrados e 8.376 não foram integrados, sendo considerado um resultado satisfatório perante os processos de saúde existentes no âmbito da Telerradiologia do STT/RCTM. Nesse sentido, esse experimento permitiu verificar a quantidade de laudos publicados que tiveram as imagens corretamente vinculadas. Embora seja previsto o laudo para exames com imagens, os exames agendados mesmo que sem imagens vinculadas, são liberados para laudos visto que há a necessidade de integrar o laudo do STT/RCTM no HIS para fins de faturamento. Algumas funcionalidades foram adicionadas, para permitir ao médico especialista, vincular as imagens de outros exames oriundos do PACS (cadastro manual) com exames agendados disponíveis no STT/RCTM e sem imagens. Torna-se ainda mais relevante essa verificação e contínua avaliação por melhorias desse processo de integração, devido aos casos recorrentes em que usuários não encontram as imagens relacionadas a um exame agendado no STT/RCTM e abrem chamado para o suporte do sistema.

Assim, o experimento 3 permitiu avaliar desempenho da integração de laudos com as respectivas imagens, de maneira a colaborar para a Camada de Agentes (D) nas verificações e avaliações de melhorias para melhorar o desempenho dessa integração, buscando compreender as falhas ocorridas durante o processo de integração dos exames no âmbito da Telerradiologia

no STT/RCTM.

Nesse contexto, as integrações realizadas no estudo de caso desta tese, por meio do PROV-Health aplicado ao STT/RCTM, comprovam sua eficácia ao permitir que a Camada de Agentes (D) encontre agrupamentos relacionados a processos de integrações vinculados às características dos experimentos realizados, podendo propor ações preventivas, de forma a evitar não só problemas operacionais, como alocar os recursos necessários de saúde e logística para melhorar a qualidade de vida da população em diversas regiões de Santa Catarina. Assim, os processos de integração dos três experimentos realizados na Camada de Atividades (B) considerando o quesito de armazenar dados de sistemas heterogêneos e distribuídos no servidor HAPI FHIR do PROV-Health com intuito de gerenciar o grande volume desses dados em um ambiente de execução de forma centralizada sem modificar as bases de dados originais, é uma característica de proveniência de dados que beneficia os processos de gerenciamentos de dados do STT/RCTM.

Embora as ferramentas computacionais utilizadas na camada de Atividades (B), — Simplifier, Forge by Firely, Insomnia, Postgresql, SQL, Python, Json, Docker e HAPI FHIR — tenham apresentado complexas instruções e configurações para realizar as três experimentações relacionadas às integrações desses experimentos no PROV-Health, elas atenderam os objetivos computacionais previstos para os experimentos realizados no estudo de caso. Salienta-se que para novos experimentos, ajustes dessas ferramentas computacionais serão necessários. Entretanto, o PROV-Health mostra-se um grande aliado nos processos computacionais de saúde do STT/RCTM, contribuindo para qualificar e oportunizar melhorias dos processos de atendimento da Telerradiologia e servindo de base para extensão de outras modalidades médicas, no quesito de contribuir na redução de tempo de atendimento e custo, que são pontos oportunizados no uso da proveniência de dados em saúde.

Nesse contexto, como forma de complementar as análises realizadas nesta subseção, e aprofundar a importância da proveniência de dados nas integrações realizadas pelo PROV-Health no STT/RCTM, foi realizada a mesma análise dos experimentos anteriores para identificar as principais características da proveniência de dados com base no estudo de Sembay, Marquez Filho e Macedo (2022a) de parte da taxonomia de Simmhan, Plale e Gannon (2005). Portanto, no Quadro 66, é apresentado a existência das principais características da proveniência de dados nos experimentos 1 – Integração de Solicitações Hospitalares com Estudos DICOM, experimento 2 – Integração de Solicitações Hospitalares com Laudos publicados e experimento 3 – Integração de Laudos publicados com Estudos DICOM, ambos analisados nesta subseção.

Quadro 66 – Identificação da existência das principais características da proveniência de dados no PROV-Health experimentado no STT/RCTM (Experimentos de integrações)

| | | Experimentos de integrações do PROV-Health no STT/RCTM | | |
|---|------------------------|--|--|--|
| | | Experimento 1 – Solicitações Hospitalares com Estudos DICOM (HIJG) | Experimento 2 – Solicitações Hospitalares com Laudos publicados (HGCR) | Experimento 3 – Laudos publicados com Estudos DICOM (HRSJ) |
| Principais características da proveniência de dados | Qualidade dos dados | ⊙ | ⊙ | ⊙ |
| | Trilhas de auditoria | ⊙ | ⊙ | ⊙ |
| | Receitas de replicação | ⊙ | ⊙ | ⊙ |
| | Atribuição | ⊙ | ⊙ | ⊙ |
| | Informacional | ⊙ | ⊙ | ⊙ |
| ⊙ Altamente relevante: considerando as características da proveniência de dados que tem efeito direto nos dados; ⊙ Relevante: considerando as características da proveniência de dados que tem efeito indireto sobre os dados; ○ Não identificado: nenhuma característica definida. | | | | |

Fonte: Com base no estudo de Sembay, Marquez Filho e Macedo (2022a).

Em relação à qualidade dos dados para os experimentos 1, 2 e 3, essa característica é altamente relevante, pois as integrações entre Solicitações Hospitalares com Estudos DICOM dos hospitais, são qualificados pelo PROV-Health quando comparados os dados para a realização das integrações, sendo observadas as transformações e derivações existentes quando comparados com a base de dados original.

Já em relação às trilhas de auditoria para os experimentos 1, 2 e 3, essa característica é altamente relevante, considerando que as integrações realizadas no PROV-Health possibilitam aos hospitais o rastreamento dos exames a serem integrados para realização de possíveis auditorias conforme a necessidade operacional.

No que diz respeito às receitas de replicação para os experimentos 1, 2 e 3, essa característica é relevante, pois as integrações realizadas no PROV-Health têm o objetivo de comparar dados de bases descentralizadas armazenados de forma centralizada no PROV-Health sem a necessidade de replicar a derivação desses dados.

No que se refere à atribuição para os experimentos 1, 2 e 3, essa característica é altamente relevante, pois a proveniência dos dados no PROV-Health é preservada às integrações realizadas dentro da rede interna do STT/RCTM, possibilitando estabelecer os direitos autorais dos dados, assim como a identificação dos responsáveis pelos dados incorretos como consequência das integrações não realizadas.

A característica informacional para os experimentos 1, 2 e 3, é altamente relevante, pois as integrações realizadas pelo PROV-Health vêm acompanhadas de seus metadados que possibilitam identificar as integrações possíveis de realização, oportunizando também, a compreensão e interpretação dos dados de proveniência.

Por fim, não houve características não identificadas nas análises aqui realizadas nos experimentos 1, 2 e 3 em relação às integrações, confirmando a proveniência de dados em todos os processos de integrações do PROV-Health aplicado ao STT/RCTM. Vale ressaltar, que dos três experimentos analisados em relação às integrações, foram devidamente selecionados como relevantes, sendo necessário frisar que a descoberta de quais dados precisam ser analisados nos processos de integração do PROV-Health e a decisão de quais parâmetros de análise utilizar, pode ser uma tarefa complexa, uma vez que os dados coletados são de fontes descentralizadas de SIS e instituições de saúde heterogêneas, podendo haver falhas na interpretação do contexto de proveniência de dados em sua centralização no PROV-Health.

7.3 ANÁLISES DAS VISUALIZAÇÕES DOS EXPERIMENTOS: COLETAS E INTEGRAÇÕES

No tocante à visualização dos experimentos realizados na Camada de Atividades (A) e (B) referente às Camadas de Entidades (A), (B) e (C) por meio da execução operacional da Camada de Agentes (A), a Camada de Atividades (C) no Módulo (C) de visualização, consulta e análise de dados de proveniência do STT/RCTM - PROV-Health-BI, foi possível monitorar e realizar análises dos dados de proveniência coletados e integrados, os quais foram visualizados no PROV-Health-BI utilizando-se a ferramenta de visualização de dados Grafana. A Camada de Atividades (C), no contexto denominado PROV-Health-BI, tem o objetivo de comprovar a existência dos dados coletados e integrados por meio da estratégia computacional do PROV-Health no STT/RCTM, possibilitando possíveis análises que forem suportadas pela ferramenta Grafana.

Assim, foi possível na camada de Atividades (C), por meio das ferramentas computacionais utilizadas, oportunizar na Camada de Agentes (A) e (D) a visualização dos dados de proveniência coletados e integrados pelo PROV-Health para compreensão do fluxo informacional da Telerradiologia do STT/RCTM com fins de auxílio para descobertas de possíveis falhas existentes nos dados provenientes dos diferentes SIS e instituições de saúde integradas ao STT/RCTM, tornando-se, assim, mais uma ferramenta que contribui para possíveis tomadas de decisão por parte da Camada de Agentes (D). A Camada de Atividades

(C) em relação à proveniência dos dados do STT/RCTM, contribui para: a garantia da qualidade dos dados; verificação dos dados; possíveis auditorias dos dados; a manutenção das bases de dados dos hospitais; informação sobre os dados coletados e armazenados; análises de tendências entre outros benefícios em relação à compressão dos dados de saúde advindas de fontes heterogêneas integradas ao STT/RCTM.

As ferramentas computacionais utilizadas na estratégia computacional do PROV-Health na Camada de Atividades (C), — Simplifier, Forge by Firely, Visual Studio Code (VS CODE), Insomnia, Postgresql, SQL, Python, HL7 FHIR, Docker, ETL e Grafana — tiveram complexidades de configurações no ambiente tecnológico do STT/RCTM, porém, atenderam o objetivo computacional dos experimentos realizados nas Camadas de Atividades (A) e (B) para a visualização de dados de proveniência no PROV-Health, em específico no ambiente PROV-Health-BI.

Como forma de complementar as análises realizadas nesta subseção sobre os experimentos efetivados nas Camadas de Atividades (A) e (B), os quais foram visualizados na Camada de Atividades (C), e aprofundar a importância da proveniência de dados no quesito de visualização desses dados concretizada pelo PROV-Health (PROV-Health-BI) no STT/RCTM, foi realizada a mesma análise dos experimentos anteriores para encontrar as principais características da proveniência de dados com base no estudo de Sembay, Marquez Filho e Macedo (2022a) de parte da taxonomia de Simmhan, Plale e Gannon (2005).

Nos Quadros 67 e 68, é evidenciada a existência das principais características da proveniência de dados nos experimentos da Camada de Atividades (A) e (B), ambos já analisados nesta subseção.

Quadro 67 – Identificação da existência das principais características da proveniência de dados no PROV-Health experimentado no STT/RCTM (Visualização das coletas)

| | | Visualização dos experimentos de coletas do PROV-Health no STT/RCTM | | |
|---|------------------------|---|--------------------------------------|-------------------------------|
| | | Experimento 1 - Agendados (HIJG) | Experimento 2 - Estudos DICOM (HGCR) | Experimento 3 - Laudos (HRSJ) |
| Principais características da proveniência de dados | Qualidade dos dados | ⊙ | ⊙ | ⊙ |
| | Trilhas de auditoria | ⊙ | ⊙ | ⊙ |
| | Receitas de replicação | ⊙ | ⊙ | ⊙ |
| | Atribuição | ⊙ | ⊙ | ⊙ |
| | Informacional | ⊙ | ⊙ | ⊙ |
| ⊙ Altamente relevante: considerando as características da proveniência de dados que tem efeito direto nos dados; ⊙ Relevante: considerando as características da proveniência de dados que tem efeito indireto sobre os dados; ○ Não identificado: nenhuma característica definida. | | | | |

Fonte: Elaborado com base no estudo de Sembay, Marquez Filho e Macedo (2022a).

Em relação à qualidade dos dados para os experimentos 1, 2 e 3, essa característica é altamente relevante, pois os dados coletados dos hospitais e visualizados pelo PROV-Health, são qualificados, oportunizando realizar análises de transformações ou derivações dos dados que ocorreram quando comparados com as bases de dados originais.

No que se refere às trilhas de auditoria para os experimentos 1, 2 e 3, salienta-se que essa característica é altamente relevante, considerando que a visualização dos dados coletados no PROV-Health possibilita observar os dados de forma centralizada, oportunizando possíveis auditorias.

A característica receitas de replicação para os experimentos 1, 2 e 3, também se apresenta altamente relevante, pois o PROV-Health pode oportunizar novas análises através da visualização dos dados coletados e armazenados e possivelmente repetir essas análises replicando os dados.

Em relação à atribuição para os experimentos 1, 2 e 3, essa característica é altamente relevante, pois a visualização dos dados coletados de proveniência no PROV-Health contém os direitos autorais e suas propriedades no servidor do PROV-Health, sendo assim, possível de identificar os responsáveis pelos dados incorretos na rede interna do STT/RCTM.

E no que diz respeito à informacional para os experimentos 1, 2 e 3, essa característica é altamente relevante, pois os dados coletados e visualizados pelo PROV-Health possibilitam compreender e interpretar os dados de proveniência do fluxo informacional do STT/RCTM.

No Quadro 68 apresenta-se a análise da existência das principais características da

proveniência de dados na visualização dos dados de controle do PROV-Health no STT/RCTM.

Quadro 68 – Identificação da existência das principais características da proveniência de dados no PROV-Health experimentado no STT/RCTM (Visualização das integrações)

| | | Visualização dos experimentos de integrações do PROV-Health no STT/RCTM | | |
|---|------------------------|---|--|--|
| | | Experimento 1 – Solicitações Hospitalares com Estudos DICOM (HIJG) | Experimento 2 – Solicitações Hospitalares com Laudos publicados (HGCR) | Experimento 3 – Laudos publicados com Estudos DICOM (HRSJ) |
| Principais características da proveniência de dados | Qualidade dos dados | ⊙ | ⊙ | ⊙ |
| | Trilhas de auditoria | ⊙ | ⊙ | ⊙ |
| | Receitas de replicação | ⊙ | ⊙ | ⊙ |
| | Atribuição | ⊙ | ⊙ | ⊙ |
| | Informacional | ⊙ | ⊙ | ⊙ |
| ⊙ Altamente relevante: considerando as características da proveniência de dados que tem efeito direto nos dados; ⊙ Relevante: considerando as características da proveniência de dados que tem efeito indireto sobre os dados; ○ Não identificado: nenhuma característica definida. | | | | |

Fonte: Com base no estudo de Sembay, Marquez Filho e Macedo (2022a).

Em relação à qualidade dos dados para os experimentos 1, 2 e 3, essa característica é altamente relevante, pois a visualização das integrações entre Solicitações Hospitalares com Estudos DICOM, são qualificados pelo PROV-Health oportunizando análises de comparações entre esses dados para as possíveis integrações, sendo observadas as transformações e derivações existentes quando comparados com a base de dados original.

Concernente às trilhas de auditoria para os experimentos 1, 2 e 3, essa característica é altamente relevante, pois ao visualizar as integrações realizadas pelo PROV-Health, isso possibilita a compreensão dos fluxos de integração que podem apresentar falhas que possam ser auditadas conforme a necessidade operacional.

No que diz respeito às receitas de replicação para os experimentos 1, 2 e 3, essa característica é altamente relevante, pois a visualização das integrações realizadas no PROV-Health oportuniza compreender os agrupamentos existentes nas integrações das bases de dados descentralizadas armazenados de forma centralizada no PROV-Health possibilitando interpretar a derivação desses dados.

Em relação à atribuição para os experimentos 1, 2 e 3, essa característica é altamente relevante, pois a visualização das integrações no PROV-Health, possibilitando estabelecer os direitos autorais dos especialistas solicitantes e a identificação dos responsáveis pelos dados

incorretos por consequência das integrações não realizadas.

E no que diz respeito à informacional para os experimentos 1, 2 e 3, essa característica é altamente relevante, pois a visualização das integrações realizadas pelo PROV-Health traz os metadados que possibilitam compreender as integrações realizadas por meio dos dados de proveniência.

Portanto, a visualização dos dados das Camadas de Atividades (A) e (B) auxiliam na compreensão e no uso da proveniência de forma intuitiva, considerando os aspectos de complexidade dos processos do STT/RCTM. No entanto, apenas visualizar a proveniência dos dados do PROV-Health não garante uma adequada compreensão dos dados de proveniência, sendo necessária a utilização de diferentes técnicas para exploração desses dados, sendo que a Camada de Atividades (C) oportuniza esse processo identificando informações importantes para tomadas de decisão para a Camada de Agentes (D). Desse modo, ainda na Camada de Atividades (C), a mesma oportuniza a realização de diferentes análises e consultas por parte da camada de Agentes (A) por meio de solicitações gerenciais da Camada de Agentes (D), sendo que alguns exemplos podem ser observados no link: <https://pacs.telemedicina.saude.sc.gov.br/grafana>. Esses exemplos estão na aba “Exemplos de gráficos de possíveis análises no PROV-Health-BI”, onde esses gráficos possibilitam consultas para análises específicas da seguinte forma: Solicitações Hospitalares por Hospital, Quantidade de Laudos por Hospital, Solicitações Hospitalares por Modalidade Médica, Estudos DICOM por Modalidade Médica, Quantidade de Solicitações Hospitalares por Descrição do Exame, Quantidade de Laudos por Descrição do Exame, entre outros tipos de gráficos por meio de consultas executadas no PostgreSQL que podem ser realizadas pela Camada de Agentes (A) para análise da Camada de Agentes (D) que forem necessárias. Esses gráficos são apenas exemplos como forma de demonstrar a infinidade de análises que o PROV-Health pode realizar referente aos dados de proveniência coletados e integrados no STT/RCTM.

Por fim, destaca-se o PROV-Health no STT/RCTM, como uma ferramenta de gerenciamento de dados de proveniência para tomadas de decisões gerenciais que contribuem para os processos de análises de dados de proveniência no âmbito da Telerradiologia.

7.4 DISCUSSÕES FINAIS DAS ANÁLISES DOS EXPERIMENTOS DE COLETAS E INTEGRAÇÕES

Nesta subseção são apresentadas as discussões finais dos experimentos de coleta e

integrações do PROV-Health realizados nas Camadas de Atividades (A) e (B) visualizados na Camada de Atividades (C) do PROV-Health experimentado no STT/RCTM no ambiente de Telerradiologia. Na Tabela 2 pode ser visualizado o estrato final dos experimentos de coletas e integrações em relação às quantidades dos dados de proveniência encontradas, contemplando-se as principais características da proveniência de dados encontradas para cada conjunto de experimentos. O objetivo da Tabela 1 é mostrar, de forma visual e ampla, os resultados analisados nos experimentos realizados nas Camadas de Atividades (A) e (B) da arquitetura específica de experimentação do PROV-Health, posteriormente visualizados na Camada de Atividades (C), considerando as principais características da proveniência de dados identificadas para os mesmos.

Tabela 2 – Extrato geral das análises dos experimentos por meio do PROV-Health no STT/RCTM

| Total de dados de proveniência coletados por meio do PROV-Health no STT/RCTM Camada de Atividades (A) | | | | |
|--|--|--|-------------------------------------|---------|
| | Experimento 1 (HIJG) | Experimento 2 (HGCR) | Experimento 3 (HRSJ) | Total |
| Agendados | 17.916 | 34.911 | 50.747 | 103.574 |
| Estudos DICOM | 14.148 | 22.535 | 27.691 | 64.374 |
| Laudos | 918 | 10.655 | 31.449 | 43.022 |
| Principais características da proveniência de dados: coleta de dados de proveniência – Camada de Atividades (A) | | | | |
| | Experimento 1 - Agendados (HIJG) | Experimento 2 - Estudos DICOM (HGCR) | Experimento 3 - Laudos (HRSJ) | |
| Qualidade dos dados | ⊙ | ⊙ | ⊙ | |
| Trilhas de auditoria | ⊙ | ⊙ | ⊙ | |
| Receitas de replicação | ⊙ | ⊙ | ⊙ | |
| Atribuição | ⊙ | ⊙ | ⊙ | |
| Informacional | ⊙ | ⊙ | ⊙ | |
| Total de dados de proveniência integrados por meio do PROV-Health no STT/RCTM Camada de Atividades (B) | | | | |
| Experimento 1 - Solicitações Hospitalares com Estudos DICOM | | | | |
| | HIJG | HGCR | HRSJ | Total |
| Integrado | 10.996 | 26.039 | 43.671 | 80.706 |
| Não integrado | 6.920 | 8.872 | 7.076 | 22.868 |
| Experimento 2 - Solicitações Hospitalares com Laudos publicados | | | | |
| | HIJG | HGCR | HRSJ | Total |
| Integrado | 1.006 | 11.123 | 33.190 | 45.319 |
| Não integrado | 16.910 | 23.788 | 17.557 | 58.255 |
| Experimento 3 - Laudos publicados com Estudos DICOM | | | | |
| | HIJG | HGCR | HRSJ | Total |
| Integrado | 640 | 3.965 | 30.041 | 34.646 |
| Não integrado | 278 | 6.690 | 1.408 | 8.376 |
| Principais características da proveniência de dados: integrações dos dados de proveniência – Camada de Atividades (B) | | | | |

| | Experimento 1 | Experimento 2 | Experimento 3 |
|---|---------------|---------------|---------------|
| Qualidade dos dados | ⊙ | ⊙ | ⊙ |
| Trilhas de auditoria | ⊙ | ⊙ | ⊙ |
| Receitas de replicação | ⊙ | ⊙ | ⊙ |
| Atribuição | ⊙ | ⊙ | ⊙ |
| Informacional | ⊙ | ⊙ | ⊙ |
| Principais características da proveniência de dados: visualização dos dados de proveniência (Coleta e Integrações) – Camada de Atividades (C) | | | |
| | Experimento 1 | Experimento 2 | Experimento 3 |
| Qualidade dos dados | ⊙ | ⊙ | ⊙ |
| Trilhas de auditoria | ⊙ | ⊙ | ⊙ |
| Receitas de replicação | ⊙ | ⊙ | ⊙ |
| Atribuição | ⊙ | ⊙ | ⊙ |
| Informacional | ⊙ | ⊙ | ⊙ |

Fonte: Elaborado pelo autor (2023).

Considerando o grande volume de dados de proveniência de diferentes SIS e instituições de saúde coletados pelo PROV-Health no STT/RCTM, sendo o HRSJ o hospital com as maiores quantidades de dados nos três experimentos da Camada de Atividades (A), há uma preocupação quanto ao processamento desses dados para gerar informações necessárias e relevantes, pois todo o processamento deve ser ágil com o menor delay possível, devendo ser tratados em tempo hábil. Os três experimentos realizados no PROV-Health para a coleta de dados de proveniência, consideraram apenas três hospitais de Santa Catarina com um grande volume de dados para fins de estudo de caso desta tese, sendo possível processar os dados em um tempo hábil. Porém, novos experimentos para uma gama maior de bases de dados de hospitais precisam ser testados para verificar a questão do tempo de processamento que o PROV-Health levaria utilizando as ferramentas descritas e utilizadas em suas camadas e respectivos módulos. Em relação às principais características da proveniência analisadas para as coletas de dados de proveniência, o PROV-Health atende a todas de forma relevante e altamente relevante conforme apresentado na Tabela 2.

Já para as análises referentes aos três experimentos de integrações na Camada de Atividades (B) realizadas pelo PROV-Health no STT/RCTM, foram observadas algumas falhas, principalmente por motivos evidenciados de uso de equipamentos médicos portáteis e por não utilizarem com frequência os serviços de Worklist. Evidencia-se que agrupamentos de exames são realizados, porém, o PROV-Health não tem esse objetivo de prever esses processos. Nesse sentido, as falhas evidenciadas podem ser fatores de risco para o andamento promissor de atendimento e de disponibilização de recursos, sendo um dos pontos destacados pelo uso do PROV-Health nas análises realizadas para esses experimentos. Os três hospitais selecionados no estudo de caso desta tese possuem disparidade na integração dos dados por motivos já citados nas análises, mas evidencia-se de modo geral, que o STT/RCTM tem sobrecarga de

dados pela complexa estrutura descentralizada, sendo evidente o uso de uma ferramenta que gerencie os dados de proveniência e os centralize como é o caso do PROV-Health. Assim, observa-se que o PROV-Health atende as principais características da proveniência de dados de forma relevante e altamente relevante conforme visualizado na Tabela 2.

Em relação às análises da visualização dos dados de proveniência na Camada de Atividades (C) oriundos das Camadas de Atividades (A) e (B) da arquitetura do PROV-Health, esses foram analisados perante as principais características da proveniência de dados, sendo de forma unânime a relevância e alta relevância das mesmas contidas nessa camada conforme visualizado na Tabela 2. O PROV-Health no quesito de visualização dos dados, proporciona consultas e análises realizadas pela Camada de Agentes (A) por solicitações da Camada de Agentes (D), servindo como uma ferramenta que soma ao STT/RCTM para beneficiar o gerenciamento de dados de proveniência de diferentes SIS e instituições de saúde integradas.

Em relação às ferramentas utilizadas na estratégia computacional do PROV-Health nas Camada de Atividades (A), (B) e (C) — Simplifier, Simplifier - STT/RCTM on FHIR® (R4), Forge by Firely, Visual Studio Code (VS CODE), Insomnia, Postgresql, SQL, GitLab, Python, Json, HL7 FHIR, HAPI FHIR, DCM4CHE, ETL, Linux (Ubuntu), Windows, Bash, Docker e Grafana —, ressalta-se que estas apresentaram complexidades de configurações no ambiente tecnológico do STT/RCTM. Foram necessários vários testes de configuração para os ajustes para alcançar os objetivos propostos pelas Camadas de Atividades (A), (B) e (C), porém, atenderam o objetivo computacional dos experimentos para o estudo de caso por meio do PROV-Health no STT/RCTM no âmbito da Telerradiologia.

Outro ponto importante a ser destacado em relação às análises realizadas nos experimentos executados no PROV-Health para o STT/RCTM nas Camadas de Atividades (A), (B) e (C), é que as mesmas utilizam-se dos padrões de interoperabilidade HL7 FHIR, o qual é baseado no modelo de proveniência W3C PROV. Considerando que o HL7 FHIR é um padrão de interoperabilidade para troca de informações de saúde entre sistemas e fornece suporte para expressão de informações de proveniência de recursos, compactua com o W3C PROV, o qual é um padrão de uso geral para informações de proveniência em termos de agentes, entidades e atividades, as quais serviram de base para a concepção do PROV-Health. Nesse sentido, destaca-se que a parte de proveniência do HL7 FHIR estende o W3C PROV, tornando o processo de proveniência de dados do PROV-Health no cenário de saúde de STT/RCTM uma ferramenta viável para fins de continuação e ampliação para outras modalidades médicas, além da Telerradiologia.

Porém, é importante enfatizar que os desafios e ameaças descritos na seção 6.9, são

pontos considerados limitantes, mesmo que de forma direta ou indireta, afetaram a realização dos experimentos realizados no PROV-Health no STT/RCTM, assim como a interpretação das análises aqui apresentadas. Com base nesse pressuposto, destaca-se a importância do aprimoramento do PROV-Health em diferentes frentes de trabalho e diferentes modalidades médicas associadas ao STT/RCTM. Por fim, apesar dos desafios encontrados e ameaças previstas na experimentação do PROV-Health no STT/RCTM, as discussões finais das análises dos experimentos realizados nas Camadas de Atividades (A) e (B) apresentadas na Camada de Atividades (C) da arquitetura específica de experimentação do PROV-Health no STT/RCTM, apresentaram os seguintes benefícios na experimentação do método, como:

a) melhoria na qualidade dos dados de proveniência gerenciados nos diferentes SIS avaliados no STT/RCTM através da otimização do processo de rastreabilidade para controle dos dados de proveniência e tomada de decisão por parte dos gestores do STT/RCTM;

b) redução do tempo de rastreabilidade dos dados de proveniência nos diversos sistemas integrados e descentralizados do STT/RCTM, reduzindo custos para viabilização de tomadas de decisão dos gestores do STT/RCTM;

c) possibilidade de melhoria dos processos e fluxos informacionais que envolvem a gestão dos dados de proveniência no STT/RCTM por parte do administrador de redes de dados de saúde do STT/RCTM, otimizando, também, a tomada de decisões mais rápidas em relação a diversas operações técnicas necessárias nos diversos repositórios de dados existentes no contexto de radiologia.

d) contribuição para garantir a segurança, confidencialidade, integridade e a gestão focada em dados de saúde que sejam apresentados como insumos para geração de conhecimento dentro dos SIS do STT/RCTM;

e) contribuir para a qualidade dos dados no STT/RCTM, além de melhorar a interpretação dos dados sem função do reconhecimento da fonte;

f) contribuir para a justificativa do uso de um determinado dado, reduzindo a possibilidade de erros no seu uso;

g) permitir identificar o processo utilizado na criação dos dados de proveniência do STT/RCTM, atualizando-os constantemente a partir de visões relacionais de consulta e controle;

h) possibilitar o uso de fontes de dados que possuem valores agregados para tomadas de decisão no STT/RCTM.

7.5 AVALIAÇÃO DO PROV-HEALTH

Esta seção apresenta os resultados da avaliação da experimentação do PROV-Health realizada com os especialistas da equipe técnica do STT/RCTM. Para esta avaliação foram convidados especialistas com o critério de experiência no gerenciamento de dados de saúde com formação na área de TI e conhecimento das ferramentas empregadas no PROV-Health, além de fazerem parte da equipe técnica do STT/RCTM conhecendo a estrutura tecnológica do cenário de experimentação. Essa avaliação ocorreu por meio de formulário eletrônico do *Google Forms* enviado aos e-mails dos especialistas com a estrutura de questões fechadas e abertas (semiestruturadas).

Inicialmente são apresentados os objetivos da avaliação (Seção 7.5.1) e em seguida o planejamento e preparação da avaliação (Seção 7.5.2). Na seção 7.5.3, expõe-se como a avaliação foi aplicada e a seção 7.5.4 em diante, aborda a análise e os resultados da avaliação.

7.5.1 Objetivos da avaliação

A avaliação do PROV-Health foi realizada utilizando o formulário eletrônico do *Google Forms* e contemplou os seguintes objetivos:

- a) verificar se o PROV-Health atende as necessidades de gerenciamento de dados de proveniência do STT/RCTM;
- b) verificar os possíveis benefícios da aplicação do PROV-Health no gerenciamento de dados de proveniência no STT/RCTM;
- c) verificar os possíveis riscos da aplicação do PROV-Health no gerenciamento de dados de proveniência STT/RCTM;
- d) verificar se o PROV-Health apresenta de forma clara e suficientes os dados de proveniência a serem gerenciados no STT/RCTM;
- e) verificar se os dados de proveniência coletados nas diferentes bases de dados pelo PROV-Health, são suficientes para realizar o gerenciamento desses dados no STT/RCTM;
- f) verificar se o método pode ser adaptado para a aplicação em outros contextos de saúde;
- g) verificar a contribuição dos especialistas avaliadores, sobre debilidades ou melhorias para o método (Contribuição/Recomendação).

7.5.2 Planejamento e preparação da avaliação

Para que o processo de avaliação estivesse em consonância com as normas de pesquisa na participação com seres humanos, fez-se necessário submeter a pesquisa, bem como o processo de avaliação para a apreciação do Comitê de Ética em Pesquisas com Seres Humanos (CEPSH) da UFSC, sendo aprovado mediante o parecer para a realização da avaliação emitido sob o processo de nº 5.063.11, conforme disponível do APÊNDICE C.

No que compete ao perfil dos especialistas que participaram da avaliação, não foram estabelecidas restrições em relação à sua formação, pois na área de TI, atuam profissionais com formações diversas. Porém, um critério adotado para o convite aos especialistas, foi que os mesmos atuassem em projetos que envolvem o gerenciamento de dados de saúde no âmbito da equipe técnica do STT/RCTM no contexto do Sistema Único de Saúde (SUS) em sistemas de grande escala.

Em relação aos objetivos definidos para a avaliação do PROV-Health, foi elaborado um questionário no *Google Forms* a ser preenchido pelos especialistas, o qual foi enviado por e-mail, sendo organizado em sete etapas. A primeira etapa apresentou o Termo de Consentimento Livre e Esclarecido (TCLE)⁶⁸ — disponível no APÊNDICE F; na segunda etapa foi apresentado um breve vídeo de explicação do formulário de avaliação do PROV-Health⁶⁹. A terceira etapa, teve como intuito identificar os participantes da pesquisa, ou seja, seu perfil, contendo sete questões com instrução (fechadas e abertas). Na quarta etapa, foi descrita brevemente a conceituação do método para compreensão de sua arquitetura em camadas e módulos. Para a quinta etapa, foi descrita brevemente a metodologia de execução do método para o cenário de aplicação. Já na sexta etapa foi disponibilizado o link de acesso, contendo login e senha para visualização dos dados de proveniência do método. E, por fim, na sétima etapa, foram disponibilizadas sete questões para avaliação do método conforme disponível no APÊNDICE G.

Os objetivos apresentados na seção anterior originaram as questões para avaliação do método, conforme apresentado no APÊNDICE G, contendo questões fechadas e abertas para a identificação do perfil do participante e questões abertas para avaliação do PROV-Health.

Para a avaliação do PROV-Health foram convidados sete especialistas e desses, cinco concordaram em participar da pesquisa. Para o convite foram realizados contatos por e-mail

⁶⁸<https://docs.google.com/forms/d/e/1FAIpQLSecKjtR6nre9i4Ju-E3p4roRh9BbpQvEhnUBSkIwXV-h2X1wQ/viewform?vc=0&c=0&w=1&flr=0>

⁶⁹ <https://youtu.be/XxI8ZN3IZL4>

enfatizando os passos e a importância da pesquisa no cenário de aplicação do estudo de caso, sendo o mesmo cenário de trabalho dos participantes.

O tempo necessário para a realização do processo de avaliação foi de 40 minutos, incluindo o vídeo de apresentação dos passos de execução do PROV-Health, os conteúdos descritivos sobre a estrutura do mesmo e o preenchimento da avaliação.

7.5.3 Aplicação da avaliação

A avaliação foi realizada individualmente através de um formulário eletrônico de avaliação elaborado no *Google Forms*⁷⁰ enviado ao e-mail dos participantes. Ao acessar o formulário eletrônico de avaliação, o especialista realiza a leitura do resumo do termo de consentimento livre esclarecido, solicitando logo após se deseja participar ou não da pesquisa. Em seguida, um breve vídeo de explicação do formulário eletrônico de avaliação do PROV-Health é apresentado com aproximadamente 9 minutos. Em seguida, sete questões são expostas com intuito de identificar os participantes da pesquisa (questões abertas e fechadas). Logo após, o participante realiza a leitura de uma breve conceituação do método em relação à sua estrutura em camadas e módulos, seguindo de uma breve descrição metodológica de execução simplificada do método para o cenário de aplicação. Após esses procedimentos, o link⁷¹ de acesso para visualização dos dados de proveniência coletados e armazenados pelo método é disponibilizado para que os especialistas tenham amplo conhecimento da execução computacional do método no quesito dos dados de proveniência. Logo após, sete questões abertas são disponibilizadas para avaliação dos especialistas para que descrevam suas respostas em relação à aplicação do método. Por fim, uma solicitação de aprovação ou reprovação da aplicação do método é solicitada ao especialista avaliador.

7.5.4 Apresentação e análise dos resultados

Após a primeira etapa, na qual o participante fez a leitura e aceitação do TCLE, em seguida na segunda etapa, o participante assiste ao breve vídeo de explicação do formulário de avaliação do método para compreender quais os passos serão necessários. Logo após, na terceira etapa, foram coletadas informações referentes à identificação (perfil) dos cinco

⁷⁰https://docs.google.com/forms/d/e/1FAIpQLSfl1tS2H_6bVkJQ1zOxRGdSWDikpYQPga9PqUGl0E5kTiGX27w/viewform?vc=0&c=0&w=1&flr=0

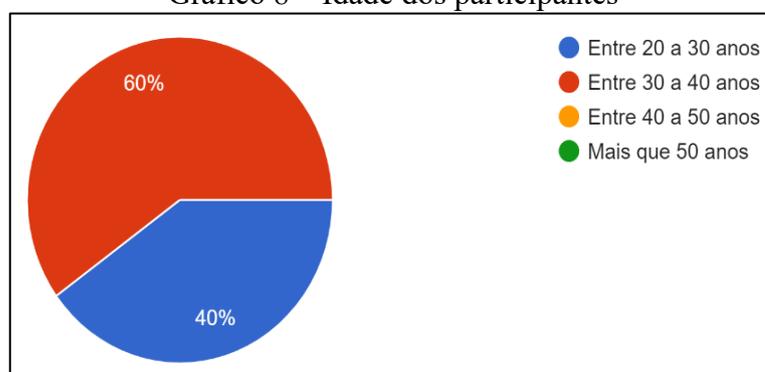
⁷¹ Disponível em: <https://pacs.telemedicina.saude.sc.gov.br/grafana>

participantes da equipe técnica do STT/RCTM, as quais foram relevantes para confirmar a experiência dos especialistas que participaram da pesquisa.

Em relação à identificação/codinome, mesmo sendo opcional, todos os cinco especialistas inseriram seus nomes, porém, por uma questão de confidencialidade descrita no TCLE, foram anonimizados.

No que diz respeito à idade dos participantes, dois deles possuem entre 20 e 30 anos e três deles possuem entre 30 e 40 anos, sendo todos do sexo masculino como visto no Gráfico 8.

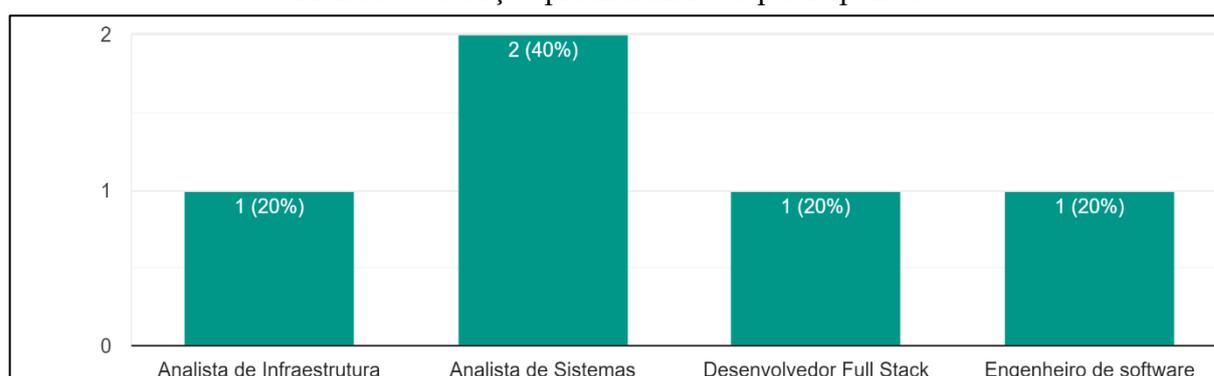
Gráfico 8 – Idade dos participantes



Fonte: Elaborado pelo autor (2023).

A função profissional dos participantes é apresentada no Gráfico 9.

Gráfico 9 – Função profissional dos participantes



Fonte: Elaborado pelo autor (2023).

Sobre as funções profissionais apresentadas no Gráfico 9, destaca-se que os participantes são profissionais voltados à tecnologia da informação, visto que apresentaram as seguintes funções: Participante 1: Analista de Infraestrutura (20%); Participante 2 e 3: Analista de Sistemas (40%); Participante 4: Desenvolvedor Full Stack (20%); e, Participante 5: Engenheiro de Software (20%).

Quanto ao tempo de atuação na equipe dos profissionais/especialistas do STT/RCTM, com unanimidade, todos apresentam mais de três anos de atuação. Em relação à formação acadêmica, a maioria dos participantes possui nível superior, sendo uma característica comum desses profissionais que trabalham diretamente com o mercado de tecnologia, visto que apresentaram as seguintes formações: Participante 1: Graduado em Engenharia Elétrica; Participante 2: Graduação em Tecnologia em Sistemas de Telecomunicações e Mestre em Ciência da Informação; Participante 3 e 4: Graduação e Mestrado em Ciências da Computação; e, Participante 5: Graduação em Sistemas de Informação.

Todos os cinco participantes possuem mais de cinco anos de experiência profissional na área de atuação.

Na quarta etapa, os especialistas realizaram a leitura da breve conceituação do método para compreensão de sua arquitetura em camadas e módulos. Em seguida na quinta etapa, os participantes realizaram a leitura da breve metodologia de execução do PROV-Health no STT/RCTM. Já na sexta etapa foi disponibilizado o link de acesso (<https://pacs.telemedicina.saude.sc.gov.br/grafana>) para visualização dos dados de proveniência do PROV-Health.

Já para a sétima etapa da avaliação, o propósito foi verificar se a percepção dos especialistas está alinhada aos objetivos de avaliação do PROV-Health mencionado na seção 7.5.1, seja em relação ao seu funcionamento ou à sua contribuição para a área. Nesse contexto, o processo de avaliação do PROV-Health é norteado por sete questões abertas para que os especialistas possam inserir suas contribuições.

A primeira questão, “De modo geral, em sua visão, o PROV-Health atende as necessidades de gerenciamento de dados de proveniência no STT/RCTM?” é descrita no Quadro 69.

Quadro 69 – Respostas para a questão 1 da avaliação do PROV-Health

| Participante | Respostas |
|----------------|--|
| Especialista 1 | Sim. |
| Especialista 2 | Sim. |
| Especialista 3 | Acredito que o PROV-Health possa melhorar a rastreabilidade dos dados. |
| Especialista 4 | Se pensarmos apenas em dados de exames radiológicos (hospitais, como costumamos chamar), sim, mas existem outras modalidades de exames no âmbito do STT/RCTM que não passam pela rede de hospitais, como exames dermatológicos, eletrocardiograma etc. Para esses, todos os dados estão armazenados no contexto da rede do próprio STT/RCTM e seria necessário adaptar a coleta dos dados para obter dados relevantes desses exames. |
| Especialista 5 | Sim. |

Fonte: Elaborado pelo autor (2023).

Em análise às respostas inseridas, verificou-se que os especialistas 1, 2 e 5 concordam que o PROV-Health atende as necessidades de gerenciamento de dados de proveniência no STT/RCTM, mesmo porque gerenciadores de dados nativos não realizam esse processo. Já os especialistas 3 e 4, apresentaram uma percepção diferente, mais voltada às limitações do PROV-Health, pois o método está limitado aos dados de exames de diagnóstico por imagem da Telerradiologia do STT/RCTM. A percepção do especialista 3 está voltada ao melhoramento da rastreabilidade do PROV-Health, sendo considerado um argumento importante, pois o estudo de caso baseou-se apenas em três hospitais de Santa Catarina. Assim, a percepção do especialista 4 deixa clara a necessidade de realizar mais testes no PROV-Health para atingir outras modalidades de exames e obter um conjunto de dados de proveniência mais completo para melhorar a rastreabilidade e alcance dos dados para possíveis análises e tomadas de decisão.

Ao avaliar os benefícios do PROV-Health, procurou-se identificar na percepção dos participantes as respostas direcionadas à segunda questão: “Em sua visão, descreva quais os possíveis benefícios que podem ser observados com a aplicação do PROV-Health para o gerenciamento dos dados de proveniência no STT/RCTM”. As respostas são descritas no Quadro 70.

Quadro 70 – Respostas para a questão 2 da avaliação do PROV-Health

| Participante | Respostas |
|---------------------|--|
| Especialista 1 | No contexto back-end/infra: Padrão de projeto e interoperabilidade. No contexto front-end: fácil visualização dos dados para tomadas de decisões. |
| Especialista 2 | Análise, auditoria e monitoramento dos dados integrados entre os sistemas médicos integrados. O PROV-Health permite maior controle e validação da integração entre os sistemas observados por meio da coleta dos dados de origem, assegurando um repositório para consultas e recuperação dos dados quando necessário, além de permitir o contínuo monitoramento do fluxo de integração por meio de análises e relatórios. |
| Especialista 3 | Com a aplicação do PROV-Health, pode-se melhorar a rastreabilidade dos dados e melhorar o planejamento de infraestrutura com base nos gráficos do visualizador Grafana. Além disso, é possível vermos uma visão mais “macro” de como os hospitais estão enviando exames para o sistema. |
| Especialista 4 | Vejo duas grandes vantagens: 1- Visão ampla e centralizada dos dados de todas as unidades hospitalares do estado, facilitando o estudo das áreas com maior incidência de determinado tipo de solicitação, achados nos laudos etc. 2- Melhor gerenciamento de recursos, principalmente de origem humana, dado que se sabe com maior precisão qual tipo de profissional é necessário em uma unidade hospitalar a partir dos achados. |
| Especialista 5 | Consolidação dos dados. |

Fonte: Elaborado pelo autor (2023).

O especialista 1 volta sua percepção de benefícios do PROV-Health aos quesitos de back-end e front-end, ou seja, consolida os objetivos do método em sua estrutura tecnológica e menciona a interoperabilidade existente como positiva, considerando a complexidade dos sistemas integrados aos STT/RCTM. Menciona também, a facilidade de se obter a visão do conjunto de dados para o auxílio a tomadas de decisão, ponto crucial no quesito de visualização dos dados de proveniência do PROV-Health. Com relação à observação do especialista 2, ao compreender seu apontamento, observa-se que a preocupação do mesmo está relacionada com a centralização dos dados de proveniência no PROV-Health, possibilitando análise, auditoria e monitoramento dos dados integrados no STT/RCTM, indicando uma das principais características positivas do método. A possibilidade de monitorar os dados integrados de forma centralizada, oportuniza recuperar dados para qualificá-los por meio de análises, possibilitando interpretar certas falhas existentes, sendo uma das principais características da proveniência de dados. Já na visão do especialista 3, o mesmo menciona que o PROV-Health contribui para melhorar a rastreabilidade dos dados e o planejamento de infraestrutura com a percepção do conjunto de dados dos hospitais. Isso mostra que o método contribui para as possíveis melhorias na gestão dos recursos na área da saúde pública, priorizando um melhor atendimento dos setores. Em relação às percepções do especialista 4, o mesmo destaca a centralização dos dados de proveniência como um benefício contribuidor para a compreensão dos dados com maior incidência em determinados setores de saúde, sendo que esse benefício tem grande importância gerencial. Menciona também uma contribuição no fator humano, em compreender os tipos de profissionais de saúde necessários em determinados setores de atuação, considerando um ponto positivo no que diz respeito à identificação das responsabilidades existentes voltadas aos profissionais exigidos em cada setor de saúde, principalmente na inserção dos dados de saúde. E finalizando as percepções dessa questão, o especialista 5 ressalta a consolidação dos dados no PROV-Health como sendo o principal objetivo do método, contribuindo para que dados descentralizados possam ser gerenciados de forma centralizada dentro da rede interna do STT/RCTM.

Em relação aos riscos do PROV-Health, o questionamento teve como objetivo verificar se na opinião dos participantes o método apresenta complexidades que podem ser prejudiciais, sendo relacionado na terceira questão: “Em sua visão, descreva quais os riscos que podem ser possíveis de observação na aplicação do PROV-Health para o gerenciamento dos dados de proveniência no STT/RCTM.” As respostas dos especialistas foram descritas no Quadro 71.

Quadro 71 – Respostas para a questão 3 da avaliação do PROV-Health

| Participante | Respostas |
|----------------|---|
| Especialista 1 | Nem todas as ferramentas de visualização utilizam FHIR, precisaria sempre criar um adaptador para visualização dos dados em alguns casos. |
| Especialista 2 | A coleta de dados é sensível a erros e pode prejudicar toda aplicação uma vez que estiver com dados duplicados, faltantes ou errados. A coleta dos dados de origem deve ser preferencialmente nos sistemas geradores desses, pois quando ocorre em bases integradas como no HIS, perde-se uma etapa dessa integração. |
| Especialista 3 | A inclusão de mais camadas de tratamento de dados pode trazer problemas como “bugs” no tratamento dos dados (que podem ocasionar a análise equivocada das informações coletadas). Temos ainda outros serviços rodando em paralelo com o STT/RCTM, isto acarreta a necessidade de ampliação da infraestrutura utilizada. |
| Especialista 4 | A necessidade da verificação dos dados disponibilizados, para que o sistema não leve o administrador a ter interpretações errôneas ao administrar os recursos existentes. |
| Especialista 5 | Segurança dos dados, uma vez que são adicionados pontos de armazenamento de dados sensíveis. Se a coleta dos dados parar por qualquer motivo sem que seja percebido, levando a interpretações equivocadas. Considerando que são redes distintas, nem sempre sua gestão é centralizada e eventos como a troca de um servidor, por exemplo, passa despercebida. Neste caso o agente de coleta precisaria ser restabelecido e muitas vezes quem faz esse trabalho não é o administrador da rede. |

Fonte: Elaborado pelo autor (2023).

As percepções dos especialistas quanto aos riscos do método foram significativas do ponto de vista das limitações do PROV-Health. Nesse sentido, o especialista 1 destaca que nem todas as aplicações de saúde utilizam FHIR, sendo considerado um risco importante a ser observado na estratégia computacional do PROV-Health que se baseou nos recursos FHIR. De fato, FHIR não é um padrão de interoperabilidade que está pronto, estando em processo contínuo de desenvolvimento e de atualização e pode acarretar riscos dependendo do avanço das tecnologias existentes na complexidade dos sistemas de Telemedicina. Já o especialista 2, volta sua atenção aos erros existentes na coleta de dados de proveniência no contexto de dados incorretos, além da duplicação e falta dos mesmos em bases integradas no HIS. Isso mostra que o risco está voltado a perda de algumas etapas de integração, o que pode prejudicar toda a análise ao final da execução do PROV-Health. O especialista 3 menciona o risco voltado ao tratamento dos dados de proveniência coletados, sendo necessário, em sua visão, ampliar a infraestrutura de experimentação do PROV-Health para melhor compreender essa questão, o que de fato é essencial para a descoberta de possíveis ajustes necessários para sanar o risco de análises errôneas com base nos dados coletados, considerando os diversos serviços associados no STT/RCTM. Na visão do especialista 4, a preocupação está voltada para a verificação dos dados disponibilizados, ponto importante levantado, pois os dados existentes devem estar em

conformidade com as bases de dados originais para não correr o risco de interpretar a proveniência dos dados de forma incorreta levando a análises falhas dos dados de proveniência em saúde. Em relação à percepção do especialista 5, o mesmo menciona riscos vinculados à segurança, apesar de que o armazenamento dos dados de proveniência do PROV-Health é alocado na rede interna do STT/RCTM, sempre existe um ponto de preocupação com a segurança desses dados. Destaca também, o risco de infraestrutura tecnológica que pode falhar no momento da coleta, uma vez que falhas de conexão da rede ou infraestrutura de computadores possam comprometer a coleta de dados do PROV-Health ocasionando em falta de dados necessários. A percepção destacada é que em alguns casos o responsável por retomar esse processo não seja um administrador de redes, pode consequentemente levar a falta de experiência em como tratar essa falha, ocasionando no processo de coleta riscos de falhas que também comprometam as análises dos conjuntos de dados de proveniência coletados.

Referente à questão 4, a qual verifica a percepção dos participantes em relação aos dados de proveniência para visualização, no que relaciona: “O PROV-Health apresenta os dados de proveniência do STT/RCTM de forma clara por meio da ferramenta de visualização de dados escolhida? Em sua visão, é suficiente?” Para esta questão foram descritas as respostas apresentadas no Quadro 72.

Quadro 72 – Respostas para a questão 4 da avaliação do PROV-Health

| Participante | Respostas |
|----------------|--|
| Especialista 1 | Dependendo do usuário, a visualização dos dados por meio de dashboard resolveria muitos dos casos. Existem usuários que solicitam além do dashboard, a seleção de variáveis com filtro de tempo, no qual o usuário pode exportar para serem trabalhadas no Python ou Excel. |
| Especialista 2 | A ferramenta de visualização apresenta uma visão básica e inicial, mas demonstra potencial. Para visualizações mais avançadas, seria necessário complementar o PROV-Health com verificações e validações do relacionamento entre as entidades coletadas, de maneira a permitir identificar quais agendamentos foram de fato realizados, e quais estudos DICOM de fato possuem laudos, entre outras relações possíveis. |
| Especialista 3 | Sim, a visualização dos dados está clara. Poderiam ter mais informações exibidas, como por exemplo: (a) % de exames laudados, (b) número de imagens por exame ... |
| Especialista 4 | Para a finalidade que se designa, por enquanto sim. Mas há muita capacidade de expansão, como controle de laudos emitidos por período, agrupadores com instituições que mais emitem laudo por descrição, filtros, etc. |
| Especialista 5 | De modo geral, os gráficos são de fácil visualização, mas é necessário permitir cortes no tempo, por exemplo, visualização de dados relativos etc. |

Fonte: Elaborado pelo autor (2023)

Avaliando as respostas dos especialistas, com relação à visualização dos dados de proveniência de forma clara, observou-se de modo geral, que todos os especialistas tiveram

uma percepção de amplitude sobre a visualização dos dados de proveniência. Como são especialistas com formação em tecnologia, é comum que ao coletar dados esses sejam visualizados para tomadas de decisão. Nesse contexto, os especialistas 1 e 5 mencionam que a visualização poderia gerar variáveis de filtro de tempo para exportar para outros visualizadores integrados no STT/RCTM. Para os especialistas 2, 3 e 4, mencionam que a consultas poderiam explorar melhor os exames agendados, estudos DICOM e Laudos, cruzando mais informações para geração de gráficos mais específicos. Diante das percepções dos especialistas, observa-se a necessidade de ampliar os experimentos do PROV-Health para alcançar a proveniência de dados em diferentes perspectivas de consultas e estratégias que possibilitem novas informações pertinentes ao STT/RCTM. O PROV-Health é um protótipo que confirma a importância de seu uso baseado nas respostas dos especialistas que atuam na equipe técnica do STT/RCTM, sendo necessário projetos futuros para a sua continuação.

Partindo do pressuposto que o PROV-Health obteve uma amostra de considerável de dados de proveniência em sua aplicação, no quesito de ser uma amostra suficiente, sendo assim, a questão 5: “De modo geral, os dados de proveniência do STT/RCTM que foram coletados pelo PROV-Health são suficientes?”. Para esta questão foram descritas as respostas apresentadas no Quadro 73.

Quadro 73 – Respostas para a questão 5 da avaliação do PROV-Health

| Participante | Respostas |
|----------------|--|
| Especialista 1 | Acredito que sim. Em alguns casos, a ETL pode requerer cruzamento com fontes externas. |
| Especialista 2 | Os dados coletados do PROV-Health representam os dados em suas origens, sendo suficientes para análises de início ao fim desta pesquisa, porém, também existem sistemas integrados intermediários, que fazem parte do fluxo e podem colaborar se coletados ou pelo menos verificados e validados a correta integração dos dados nesses sistemas. |
| Especialista 3 | No geral, sim. Mas creio que outras informações como as citadas na questão 4 são passíveis de serem coletados/exibidos no dashboard do Grafana. |
| Especialista 4 | Para a finalidade que se designa, sim. |
| Especialista 5 | A visualização de números absolutos sem a variável “tempo” é um dado importante a ser observado. No uso da ferramenta por parte do gestor, essa variável pode contribuir para tomadas de decisão. |

Fonte: Elaborado pelo autor (2023).

Analisando as respostas dos especialistas sobre se a quantidade de dados coletados e apresentados pelo PROV-Health é suficiente, observa-se que os especialistas 1, 2, 3 e 4 concordam que os dados são suficientes, apesar de que outras fontes externas poderiam ser incorporadas a amostra, potencializando ainda mais as consultas e geração de gráficos no Grafana, ou seja, os dados são suficientes, mas a visualização desses dados deveria ser mais

bem explorada. Em específico, o especialista 3 frisa a importância das recomendações descritas anteriormente na questão 4. Já no que compete à percepção do especialista 5, ele menciona a variável tempo já indicada na questão 4 e destaca um fator importante para tomadas de decisão do gestor do STT/RCTM.

Comcerrente à adaptação do método em outros cenários de saúde, para a questão 6: “Em sua visão, o PROV-Health pode ser adaptado para a aplicação em outros contextos de saúde?” foram descritas as respostas apresentadas no Quadro 74.

Quadro 74 – Respostas para a questão 6 da avaliação do PROV-Health

| Participante | Respostas |
|----------------|---|
| Especialista 1 | Sim. |
| Especialista 2 | Sim. O PROV-Health demonstra que a proveniência pode colaborar com validações de fluxos de integração entre sistemas comum no contexto da saúde. |
| Especialista 3 | Sim. Com a utilização do padrão HL7-FHIR, o uso do PROV-Health pode ser adaptado para outras aplicações da área da saúde. |
| Especialista 4 | Sim, não somente no âmbito radiológico, mas também no telediagnóstico (Eletrocardiograma, Dermatologia, entre outros) e quaisquer outras frentes de trabalho que demandem análise de dados. |
| Especialista 5 | Sim. |

Fonte: Elaborado pelo autor (2023).

Com unanimidade, os especialistas afirmam que o PROV-Health pode ser adaptado a outros cenários de saúde confirmando o objetivo da questão 6, apontando assim, a potencialidade do método aplicado ao STT/RCTM. Para os especialistas 1 e 5, a adaptação do PROV-Health é algo certo. Já o especialista 2, menciona que a proveniência de dados pode contribuir para os fluxos de integração, considerando a complexidade do STT/RCTM, isto comprova a eficácia do PROV-Health. No que diz respeito à percepção do especialista 3, o uso do HL7 FHIR, padrão internacionalmente utilizado, outras aplicações de saúde podem ser beneficiadas com o PROV-Health. Lembrando que o HL7 compactua o modelo W3C PROV, sendo o mesmo base para a concepção do PROV-Health, o que confirma sua importância no quesito de proveniência avaliado por esses especialistas. E para o especialista 4, a adaptação do PROV-Health pode ir além da Telerradiologia, enfatizando que outras modalidades médicas associadas ao STT/RCTM podem fazer uso das adaptações do PROV-Health.

A última questão da avaliação, complementa os propósitos do PROV-Health buscando a verificação da percepção dos especialistas em relação às debilidades e melhorias para o aperfeiçoamento do PROV-Health para a sua continuidade, qual seja a questão 7: “Gostaria de descrever algo mais para contribuir com a pesquisa em relação às debilidades ou sugerir melhorias para o PROV-Health?”. Essa questão tem suas respostas descritas no Quadro 75.

Quadro 75 – Respostas para a questão 7 da avaliação do PROV-Health

| Participante | Respostas |
|----------------|--|
| Especialista 1 | Uma forma de exportar em .csvs ou excel os dados brutos da coleta de dados utilizado no Grafana. Verificar a possibilidade de cruzar os dados coletados com fontes externas para melhorar alguma informação, como por exemplo, geolocalização (lonlat ⁷²) dos pacientes. |
| Especialista 2 | Melhorias podem ser feitas no uso de padrões de interoperabilidade e arquitetura de sistemas para coleta dos dados de proveniência, como HL7, FHIR e DICOM, bem como mecanismos de notificação de novos registros ao invés de consultas periódicas às bases de dados. |
| Especialista 3 | Acho interessante o desempenho do PROV-Health, porém precisa ser indicado suas limitações para prever otimizações, visto o impacto direto no uso do STT/RCTM por seus usuários. |
| Especialista 4 | Se for ampliado, creio que possa ser utilizado um conjunto de filtros para maximizar a análise, como filtros por período, região da instituição etc. |
| Especialista 5 | Necessidade de coleta de mais variáveis, assim como a necessidade de monitoramento por parte dos agentes que coletam os dados de proveniência, de modo a identificar pausas, falhas e dificuldades na obtenção do dado. |

Fonte: Elaborado pelo autor (2023).

Ao avaliar a percepção dos especialistas nos quesitos de debilidades e melhorias para o aperfeiçoamento do PROV-Health, observou-se que todos os especialistas mencionaram melhorias significativas para continuidade do PROV-Health no âmbito geral do STT/RCTM que compactuam com o objetivo da questão 7. Assim, o especialista 1, menciona a possibilidade de exportação dos dados em diferentes formatos para uso em diferentes plataformas de análises e o cruzamento desses dados com fontes externas para realização de novas consultas, o que de fato, pode ser ajustado no PROV-Health em sua estratégia computacional. O especialista 2, indica a possibilidade de ampliação dos recursos de interoperabilidade utilizados pelo PROV-Health e aprofundar além das consultas periódicas às bases de dados. Esses pontos são possíveis de execução no PROV-Health ajustando sua estratégia computacional sem muitas complexidades. O especialista 3 aponta um fator de importância, as limitações do PROV-Health, sendo que algumas limitações já foram descritas nesta tese, porém, outras podem surgir conforme novas experimentações no cenário do STT/RCTM. Já o especialista 4, aponta aprimoramentos nas consultas e análises dos dados de proveniência do método, sendo pontos a serem acatados para estudos futuros e continuação do PROV-Health no STT/RCTM. O especialista 5 menciona que, além das ampliações de consultas e análises, existe a preocupação com as possíveis falhas de infraestrutura tecnológica, um ponto significativo para estabelecer novos padrões de segurança e interoperabilidade na continuação do PROV-Health no cenário

⁷² Em linguagens de programação, lonlat significa uma biblioteca de localização (longitude e latitude).

do STT/RCTM.

Ao final da avaliação, foi solicitado aos especialistas se os mesmos aprovavam ou desaprovavam o PROV-Health para aplicação no STT/RCTM, sendo que de forma unânime, todos aprovaram.

Assim, de modo geral, as análises realizadas com base na avaliação dos especialistas e com a confirmação das respostas perante os objetivos da avaliação do PROV-Health, concluiu-se que o método atende seu propósito. Vale ressaltar que os 5 especialistas respondentes fazem parte da equipe técnica do STT/RCTM, os quais motivados e interessados em contribuir com a pesquisa, apresentaram excelentes e significativas contribuições que foram acatadas para continuação e aperfeiçoamento do PROV-Health no STT/RCTM. Observou-se que a maioria das sugestões apresentadas pelos especialistas está voltada a ampliação do PROV-Health, o qual estará disponível para o STT/RCTM realizar o contínuo aperfeiçoamento e adaptações necessárias, sendo consideradas para esta tese como trabalhos futuros.

Diante do exposto, conclui-se que a avaliação realizada pelos especialistas foi considerada satisfatória e atingiu os objetivos definidos. Considerando a avaliação realizada com o perfil dos especialistas, pode-se observar que os resultados obtidos compactuam com os propósitos do PROV-Health e estão alinhados com as percepções encontradas, confirmando assim, que o método pode contribuir com o gerenciamento dos dados de proveniência. Dessa forma, a Proveniência de Dados no contexto da saúde pode contribuir para os diversos processos de gestão de dados. Sendo assim, a concretização da experimentação do PROV-Health no STT/RCTM de forma positiva, gera oportunidades de continuação e aprimoramento do método no intuito de contribuir para o gerenciamento dos dados de proveniência desse cenário.

Por fim, a contribuição dos resultados promissores encontrados, vai além da percepção prática do uso de tecnologias que relacionam a Proveniência de Dados no cenário de saúde aqui escolhida, pode instigar outras instituições de saúde a refletirem sobre a importância em gerenciar seus dados de proveniência, apontando o PROV-Health como uma ferramenta promissora para suas tomadas de decisões por parte dos gestores dessas instituições.

7.6 CONSIDERAÇÕES DA SEÇÃO

Nesta seção foi abordado um conjunto de análises dos experimentos realizados. A seção foi dividida em duas partes, sendo a primeira delas a análise dos experimentos de coletas referentes à Camada de Atividades (A), análises dos experimentos de integrações referente à

Camada de Atividades (B) e análise da visualização dos dados de proveniência na Camada de Atividades (C) advindo das duas camadas de atividades anteriormente mencionadas, sendo essas camadas pertencentes a arquitetura de experimentação do PROV-Health no STT/RCTM. As análises realizadas foram positivas no que refere à experimentação do método no cenário de Telemedicina, especificamente na Telerradiologia no contexto de diagnóstico por imagem. O PROV-Health apresentou-se útil em relação aos objetivos de suas camadas, possibilitando na prática concluir êxito em sua experimentação que, apesar dos desafios e ameaças encontrados, apresentou um desempenho experimental satisfatório. Salienta-se que as análises realizadas na experimentação dos PROV-Health no STT/RCTM, que o mesmo pode contribuir com resultados para além do bom funcionamento dos fluxos informacionais do cenário de saúde em experimentação no quesito de gerenciamento de dados de proveniência, com vistas a beneficiar baixos custos de rastreamento, interoperabilidade e manutenção, o que de fato, foi um dos pontos mais significativos do método, considerando que o cenário de experimentação é associado ao SUS.

Na segunda parte, foi realizada a avaliação do PROV-Health com especialistas da equipe técnica do STT/RCTM, os quais, com experiência no ambiente de experimentação do método, puderam apontar debilidades e melhorias significativas para o PROV-Health. A avaliação foi considerada satisfatória e as sugestões foram acatadas para a continuidade e aprimoramento do PROV-Health no âmbito do STT/RCTM, sendo considerado para esta tese como trabalhos futuros.

Por fim, do ponto de vista dos desafios e ameaças encontrados nas experimentações e na avaliação positiva dos especialistas, ambos voltados ao PROV-Health no STT/RCTM, as contribuições do método, tanto no quesito teórico ou prático, oportunizam novas frentes de pesquisa que contribuem para os processos de saúde pública do STT/RCTM.

8 CONCLUSÕES E TRABALHOS FUTUROS

Esta tese foi motivada pela necessidade de desenvolver um método adequado para contribuir na solução dos problemas relacionados ao gerenciamento de dados de proveniência e falta de interoperabilidade em diferentes SIS com fontes de dados descentralizados, frente ao aumento expressivo de dados de saúde nesses sistemas nos últimos anos. Tal cenário demanda com atenção a proposição de novas ações de gerenciamento de dados de proveniência que podem ser úteis em diversas situações relacionadas a tomadas de decisões importantes voltadas aos serviços de saúde humana, assim como, a interoperabilidade desses dados. E isso só é possível mediante a propostas de novos métodos que sejam capazes de reunir dados descentralizados em repositórios de dados de proveniência, centralizados em redes internas seguras para compreender a proveniência desses dados e suas transformações. De modo que, o acesso a estes dados de proveniência só será possível mediante autorização por parte dos gestores de instituições de saúde, comprometidos em compreender os fluxos informacionais dos diferentes SIS para tomadas de decisão estratégicas e gerenciais em prol do melhor atendimento dos serviços de saúde à população.

Dessa forma, esta tese apresentou o PROV-Health, um método baseado no instanciamento do modelo de Proveniência de Dados (W3C PROV) em conformidade com o HL7 FHIR para auxiliar no gerenciamento de dados de proveniência em diferentes SIS descentralizados e nas questões de interoperabilidade. Assim, o PROV-Health simplificou a forma de gerenciar dados de proveniência de SIS descentralizados em uma estratégia computacional, a qual utiliza a rede interna da instituição de saúde, centralizando os dados em um repositório de proveniência local de forma segura e interoperável para tomadas de decisões necessárias cabíveis com o estudo de caso realizado nesta tese e com as autorizações administradas para a realização do mesmo. A centralização dos dados de proveniência no PROV-Health também teve o intuito de amostragem para as análises realizadas, sendo que o cenário ideal para a ampliação dos experimentos seria o uso de tecnologias de Cloud Computing. Entretanto, ao final das experimentações e análises dos resultados gerados na experimentação do PROV-Health, o método mostrou-se contribuir para o gerenciamento e rastreabilidade dos dados de proveniência em SIS e para o quesito de interoperabilidade, apresentando uma estrutura compatível e otimizada, apesar dos desafios encontrados pelas ferramentas computacionais utilizadas e pelas ameaças que podem limitar o método.

Nesse contexto, para esta tese foram definidos os seguintes objetivos específicos: a) apresentar os principais pressupostos necessários para o gerenciamento de dados de

proveniência em SIS; b) identificar métodos, técnicas, modelos ou metodologias computacionais que apoiam o gerenciamento dos dados de proveniência em SIS; c) elaborar um método para gerenciamento de dados de proveniência em SIS com base no instanciamento do modelo W3C PROV; d) aplicar o método em um cenário de saúde real para o estudo de caso; e) validar o método com especialistas em gerenciamento de dados em SIS.

Quanto aos objetivos específicos propostos nesta tese, primeiramente buscou-se alcançar o objetivo “a” — apresentar os principais pressupostos necessários para o gerenciamento de dados de proveniência em SIS — foram analisados os pressupostos apresentados no referencial teórico com base nas necessidades reais de gerenciamento de dados de proveniência em SIS que não eram realizados pelos sistemas nativos incorporados nesses cenários. Nesse sentido, a literatura apontou como principal pressuposto de maior prioridade a preocupação em relação ao volume de dados coletados nos SIS que aumentam exponencialmente a cada novo recurso tecnológico adicionado aos seus processos, sendo que quanto mais dados de saúde de diversas fontes descentralizadas são coletados maior é a necessidade de gerenciamento de dados de proveniência nesses sistemas, considerando também os complexos aspectos relacionados à interoperabilidade desses dados. Com isso, foi possível apontar o principal pressuposto do PROV-Health, fundamentando o propósito do método em relação ao gerenciamento de dados de proveniência em SIS e na atenção ao que se refere à interoperabilidade dos dados de saúde.

Para atingir o objetivo “b” — identificar métodos, técnicas, modelos ou metodologias computacionais que apoiam o gerenciamento dos dados de proveniência em SIS —, indo além da literatura, foram analisados alguns estudos relevantes por meio da RSL onde foram identificados métodos, técnicas, modelos ou metodologias computacionais utilizadas no gerenciamento de dados de proveniência em diferentes SIS. Os estudos mostraram que existe um entrelaçamento de estratégias computacionais para que se possa realizar os processos de gerenciamento dos dados de proveniência, assim como a interoperabilidade desses dados nos mais variados cenários de saúde, porém, não padronizando nenhuma estrutura específica. Assim, foi realizada em complemento a RSL, a técnica de *Snowballing (backward and forward)* buscando ampliar os resultados, encontrando algumas respostas satisfatórias em aplicações reais que nortearam as estruturas computacionais do PROV-Health. Com essas análises foi possível estabelecer a estrutura utilizada na estratégia computacional essencial para compor o PROV-Health, assim como fundamentar o método com base no modelo de Proveniência de Dados W3C PROV, sendo o modelo mais utilizado em SIS pelo padrão internacional HL7 FHIR.

Para o alcance do objetivo “c” — elaborar um método para gerenciamento de dados de proveniência em SIS com base no instanciamento do modelo W3C PROV —, considerando a estrutura estabelecida na etapa anterior, o método foi elaborado por meio de arquiteturas, sendo apresentada a arquitetura de instrumentalização dos serviços de saúde em SIS, a qual contém o fluxo de integração, interoperabilidade e diferentes redes e bases de dados de saúde, a arquitetura segmentada em camadas, a qual é baseada nos componentes definidos pelo modelo de Proveniência de Dados W3C PROV (Agentes, Entidades e Atividades). Na camada de agentes, são definidos os principais agentes que representam o ciclo de instrumentalização dos serviços de saúde em SIS, os quais tem responsabilidades diferenciadas no tocante ao acesso, uso e manutenção dos dados de proveniência nesses sistemas. Já na camada de entidades, estão contidas as instituições de saúde com diferentes SIS, contendo repositórios de dados descentralizados, as quais compactuam com a camada de agentes e a camada de atividades. Em relação à camada de atividades, as quais compactuam com as camadas de agentes e entidades, possui detalhadamente toda as estruturações dos dados de entrada e saída do PROV-Health, sendo possível a realização de adaptações que venham ser necessárias dependendo da aplicação necessária. Assim, todas as camadas relacionadas no PROV-Health, segmentam o fluxo informacional essencial para o gerenciamento dos dados de proveniência em SIS elaborado para o método.

Buscando alcançar o objetivo “d” — aplicar o método em um cenário de saúde real para o estudo de caso; —, o PROV-Health foi experimentado em um cenário de saúde real com dados reais, sendo o cenário o STT/RCTM situado nas dependências da UFSC, especificamente no âmbito da Telerradiologia no contexto de diagnóstico por imagens médicas. Durante a realização dos experimentos, observou-se significativa contribuição para o gerenciamento de dados de proveniência nos SIS integrados ao STT/RCTM, contribuindo para a melhoria e segurança da rastreabilidade e interoperabilidade dos dados de proveniência gerados nos SIS integrados, possibilitando posteriormente a consulta e análise desses dados de proveniência, algo inexistente até então no STT/RCTM, constituindo sucesso na experimentação. A composição geral do PROV-Health para experimentação no STT/RCTM, consiste na utilização do padrão de interoperabilidade de dados HL7 FHIR para facilitar a troca de dados de saúde entre os diferentes SIS. O HL7 FHIR é baseado no modelo de Proveniência de Dados W3C PROV. Seu uso no PROV-Health oportunizou os processos de proveniência em consonância com o padrão HL7 FHIR 5Ws, responsável por atribuir os recursos de proveniência entre o HL7 FHIR e o W3C PROV. Os dados de proveniência gerenciados pelo PROV-Health foram alocados em um servidor local de dados de saúde HAPI FHIR, sendo esse servidor, o local para

armazenamento dos dados de proveniência na rede interna do STT/RCTM para posterior análises necessárias. Dessa maneira, na composição geral da estratégia computacional empregada no PROV-Health, foram empregadas ferramentas computacionais *Open Source*, contribuindo para sua estruturação, — Simplifier, Simplifier - STT/RCTM on FHIR® (R4), Forge by Firely, Visual Studio Code (VS CODE), Insomnia, Postgresql, SQL, GitLab, Python, Json, HL7 FHIR, HAPI FHIR, DCM4CHE, ETL, Linux (Ubuntu), Bash, Docker e Grafana —, sendo respeitada a LGPD. Foi utilizado o sistema operacional Windows (única ferramenta com versão de estudante paga). Ao final dos experimentos, no uso das ferramentas computacionais descritas e das etapas de análises dos resultados, o PROV-Health mostrou-se como mais uma ferramenta de contribuição significativa para o STT/RCTM.

Para o alcance do objetivo “e” — validar o método com especialistas em gerenciamento de dados em SIS —, foram traçados os objetivos da avaliação para que esse processo, por meio de formulário eletrônico do *Google Forms*, fosse realizado e aplicado. Participaram da avaliação, 5 especialistas, os quais forneceram sua opinião sobre a experimentação do PROV-Health no STT/RCTM, avaliando além da arquitetura e estratégia computacional, os resultados do processo de gerenciamento de dados de proveniência nos SIS integrados ao STT/RCTM. Em atenção as opiniões descritas pelos especialistas, observa-se que o PROV-Health contribuiu para o gerenciamento e a interoperabilidade dos dados de proveniência nos SIS integrados do STT/RCTM. Foram apresentadas sugestões de natureza técnica e operacional pelos especialistas, as quais foram consideradas importantes para o processo de gerenciamento e interoperabilidade dos dados de proveniência em SIS no estágio atual do PROV-Health. Dessa forma, o PROV-Health, após sua validação no STT/RCTM, apresentou-se significativo ao contribuir para sanar os problemas definidos nesta tese, considerando também, que por se tratar de um processo inexistente nesse cenário, tornou-se uma das principais contribuições desta tese.

Em relação à questão de pesquisa definida para esta tese — 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? —, considerando o grande volume de dados gerados pelos SIS (210.970), dados no estudo de caso e a complexa interoperabilidade existente, esta questão tem sua resposta pautada na estruturação do PROV-Health, o qual tem a finalidade de gerenciar grandes fluxos de dados provenientes de SIS descentralizados, contribuindo para identificar e rastrear mudanças que ocorreram no ciclo de vida dos dados, analisando-os e reaproveitando-os para a compreensão dos fluxos informacionais dos cenários de saúde, processos pelos quais, oportuniza tomadas de decisão

estratégicas.

Com relação à hipótese definida para esta tese: **H1** – É possível afirmar que um método para gerenciamento de dados de proveniência em SIS, constituído de processos de interoperabilidade em consonância com as principais características da proveniência de dados, pode ser ajustado para diferentes contextos informacionais de saúde, contribuindo para a descrição da proveniência, das ações e dos envolvidos na construção e alteração desses dados, conclui-se que com o desenvolvimento do PROV-Health a hipótese foi confirmada, pois considerando que o método utilizou-se de ferramentas computacionais *open source*, diminuindo custos operacionais e contribuindo para os processos de interoperabilidade entre os diferentes SIS descentralizados estudados em sua experimentação, permitiu que os dados coletados e armazenados pudessem ser visualizados e possivelmente analisados e avaliados de forma estatística para compreensão do fluxo informacional do cenário de estudo de caso. Ainda no contexto da hipótese desta tese, importante destacar que o PROV-Health permite apresentar a derivação dos dados, identificando erros introduzidos de forma incorreta que aumentam à medida que se propagam em outras derivações, processo que tem implicações significativas na avaliação da qualidade de dados, constituindo um dos pontos importantes da Proveniência de Dados nesta tese.

Entretanto, esta tese revela que existem desafios e ameaças a serem ultrapassados em relação ao gerenciamento e interoperabilidade de dados de proveniência em SIS conforme a experimentação do PROV-Health no cenário de estudo de caso, sendo esses: complexas integrações entre as tecnologias computacionais existentes no STT/RCTM com as diferentes tecnologias computacionais previstas no PROV-Health; mudanças drásticas podem ocorrer na estratégia computacional do PROV-Health dependendo da estrutura de TI a qual necessita-se a coleta dos dados, sendo necessário em alguns casos, o uso de ferramentas proprietárias, inviabilizando os custos do PROV-Health para instituições públicas de saúde; nem todos os subsistemas estão integrados nos SIS dos hospitais; o HL7 FHIR contribui em partes no quesito de interoperabilidade, pois diferentes formatos de dados são muitas vezes complexos de integração, inviabilizando alguns processos da Proveniência de Dados do PROV-Health; a disponibilidade dos dados nem sempre estão em formato digital, sendo necessários realizar novas coletas por meio do PROV-Health; as redes de dados de saúde dos hospitais podem dispor de acesso restrito, inviabilizando a totalidade da coleta dos dados por parte do PROV-Health; e, processos econômicos e tecnológicos devido a velocidade e complexidade das ações de saúde, podem ser vistos como um grande desafio e ameaça para a estrutura tecnológica do PROV-Health. Com isso, observou-se que a transformação digital na saúde também é um dos

grandes desafios ainda encontrados, pois a coleta de dados nos diferentes SIS não deve ter impeditivos manuais que recue o processo de digitalização do ecossistema de saúde. O que se espera é que a transformação digital nos setores de saúde se torne uma realidade, proporcionando uma medicina mais acurada e orientada a dados, utilizando-se de tecnologias integradas, como é o caso da Telemedicina, a qual, tem contribuído para a melhora da saúde da população, contribuindo para integração dos serviços de saúde e reduzindo custos. Nesse sentido, dados de saúde digital, tornam-se um eficiente recurso de saúde, que de fato, deveria ser um processo natural utilizado por todas as instituições de saúde como um fator contribuinte no que compete ao gerenciamento e interoperabilidade de dados nos diferentes SIS. Assim, esses foram alguns desafios e ameaças encontradas no atual estado do PROV-Health, sendo necessária a continuação dos experimentos em outros cenários de saúde para identificar novos obstáculos para seu aprimoramento.

Como limitação desta tese, identifica-se a aplicação apenas no âmbito que compete ao telediagnóstico por imagens médicas no contexto da Telerradiologia do STT/RCTM. Outras fontes externas que não passam nem pelo PACS e nem pela rede do STT/RCTM não foram experimentadas, pois o ambiente de estudo de caso é extremamente complexo e de grande escala. Isso se deu devido ao tempo que se teve para sua experimentação e o escopo definido nesta tese. Outro ponto limitante, está em relação à avaliação do PROV-Health realizada por cinco especialistas, sendo considerada uma limitação da pesquisa, pois apesar do alto nível de experiência que justifica a seleção dos especialistas, uma maior quantidade de avaliadores poderia contribuir com novos apontamentos, oportunizando futuros estudos de ampliação do PROV-Health. O PROV-Health não realiza análises aprofundadas nos dados de proveniência armazenados na rede interna do STT/RCTM, por se tratar de um método proposto ainda como protótipo, sendo uma limitação temporária à medida que seus processos possam ser aprimorados por outros profissionais da equipe técnica do STT/RCTM. Ainda, a não formalização dos dados de saúde, o uso do padrão HL7 FHIR nos processos de interoperabilidade possuindo alguns desafios e a falta de inserção imediata dos dados em formato digital nos SIS, podem ser fatores limitantes para o bom funcionamento do PROV-Health.

Apesar dos desafios e ameaças encontrados no PROV-Health e suas respectivas limitações, espera-se que esta pesquisa de tese possa contribuir para: pesquisadores da comunidade acadêmico-científica, além da equipe técnica do STT/RCTM que visam dar continuidade nos processos do PROV-Health, visto que o método apresentou abordagens propostas na literatura com base em um dos principais modelos de proveniência utilizados no

domínio de saúde, o W3C PROV baseado em HL7 FHIR; instituições públicas de saúde, pois o PROV-Health utiliza-se de ferramentas computacionais *open source* sem custos a estas instituições, podendo apoiá-las em tomadas de decisões gerenciais baseadas nos dados de proveniência armazenados em sua rede interna, de forma segura para posteriores avaliações para esse fim; profissionais da área da Ciência da Informação e áreas multidisciplinares, como uma oportunidade de atuação profissional na área de gestão de dados de proveniência em SIS, pois inúmeras possibilidades de pesquisas no campo da saúde estão absorvendo esses profissionais; além de evidenciar uma das contribuições mais importante desta tese relacionada à atuação conjunta da Ciência da Informação com a Ciência da Computação, no que se refere ao gerenciamento de dados de proveniência em SIS, pois favorece a construção de uma estrutura tecnológica e informacional, que contempla as metodologias de tratamento informacionais das duas grandes áreas do conhecimento. Assim, a principal contribuição do PROV-Health para a Ciência da Informação, está voltada aos princípios de uma ciência que investiga os processos de geração, disseminação, coleta, organização, armazenamento, recuperação, interpretação e uso da informação compactuadas com as tecnologias computacionais empregadas nos princípios da Proveniência de Dados da Ciência da Computação, ambas as ciências, congruentes desses princípios, principalmente associadas a área de saúde.

Portanto, mediante a experimentação do PROV-Health em um cenário de saúde real e com dados reais, pode-se corroborar que a Ciência da Computação fornece as ferramentas tecnológicas responsáveis pelo uso de métodos, técnicas, modelos e metodologias de Proveniência de Dados, e a Ciência da Informação fornece os meios necessários de conhecimento no tratamento informacional, fazendo uso das aplicações tecnológicas.

Finalmente, apesar do fato da escassez da abordagem no que tange ao tema desta tese, considera-se que o PROV-Health teve significativo reconhecimento pela comunidade científica, devido a artigos científicos sobre o tema desta tese, sendo publicados em revistas e eventos nacionais e internacionais, tanto da área de Ciência da Informação, Ciência da Computação e áreas interdisciplinares.

8.1 TRABALHOS FUTUROS

Como sugestão de trabalhos futuros, sugere-se que o PROV-Health seja estendido, com integrações a outras frentes de trabalho do STT/RCTM além do âmbito da Telerradiologia, como Telecardiologia, Teledermatologia, entre outros contextos para aprimorar o alcance do

método em relação aos demais dados de proveniência existentes em todo cenário de saúde de Telemedicina.

Além do cenário de Telemedicina, trabalhos futuros também são sugeridos para a continuidade do PROV-Health a ser experimentado em diferentes contextos de saúde, com o intuito de comprovar o quesito de proveniência dos dados no apoio a futuras aplicações que necessitem de meios interoperáveis entre diferentes SIS com ferramentas e estruturas computacionais mais atualizadas e abrangentes do que as propostas nesta tese.

Outro ponto a ser destacado como projeto futuro, é a implementação do blockchain nos processos do PROV-Health. A tecnologia Blockchain tem sido muito discutida pela sociedade científica no que se refere aos benefícios de seu uso no contexto de saúde, e pode ser ajustada aos processos do PROV-Health para desenvolver uma capacidade ainda maior na Proveniência de Dados, garantindo um fluxo informacional em diferentes SIS. Assim, a inserção do Blockchain nas estruturas do PROV-Health associada às bases técnicas de Proveniência de Dados pode contribuir para registrar os fluxos de informação em diferentes redes de dados de saúde, possibilitando a imutabilidade e o gerenciamento de dados de proveniência em SIS descentralizados em diversos cenários de aplicação, contribuindo também para a segurança dos dados de proveniência.

Além do Blockchain, o uso de Cloud Computing, Edge Computing e Fog Computing destacados na RSL desta tese, também são encorajados como projeto futuro para o PROV-Health no quesito de Proveniência dos Dados. Estas três tecnologias estão se destacando nos setores de saúde e podem coexistir dependendo do tipo de aplicação, principalmente quando o fator é desempenho computacional, o que de fato, seria útil ao PROV-Health. Sendo assim, em relação a essas três tecnologias para uso futuro no PROV-Health, destaca-se: a) em relação ao grande volume de dados processados e armazenados, o custo para operacionalizar a Cloud Computing é aceitável, sendo este tipo de aplicação ideal; b) já o uso da Edge Computing, pode ser um fator impulsionador para inserção de dispositivos de IoHT; e, c) o uso da fog computing pode até parecer a junção das outras duas tecnologias, mas está focada em recursos distribuídos, o que poderia contribuir para questões como confiabilidade e confiança da arquitetura do método. Com o uso dessas ferramentas, a centralização dos dados em um repositório de proveniência local por parte do PROV-Health, pode ter uma abordagem diferenciada, principalmente no que diz respeito aos diferentes mecanismos de armazenamento de dados de proveniência por meio de diferentes dispositivos médicos.

Ainda, em relação à estruturação do PROV-Health, sugere-se que novas estratégias computacionais possam ser adicionadas ao método de forma a facilitar o alcance dos dados de

proveniência em contextos descentralizados em SIS de larga escala, possibilitando a centralização necessária para obter diversas melhorias quanto ao uso e análise dos dados relevantes para tomadas de decisão em contextos de saúde, além de contribuir para evitar a perda de dados e compreender suas modificações e demais situações ou problemas que podem prejudicar o funcionamento dos SIS de uma forma geral.

Por fim, seria interessante experimentar o PROV-Health em instituições de saúde que possuam perfis diferentes das analisadas nesta tese, bem como a utilização de outras técnicas associadas ao método para um maior alcance experimental entre diferentes SIS.

8.2 ARTIGOS PUBLICADOS

Durante o período de doutoramento (2018–2023) foram elaborados e publicados artigos científicos que relacionavam a temática desta tese, os quais contribuíram no desenvolvimento da pesquisa. Nesse sentido, foram estudos realizados e publicados em eventos e periódicos nacionais e internacionais, buscando aprimorar a estratégia de gerenciamento de dados de proveniência perante a comunidade científica, determinando cenários diferentes até moldar uma estratégia computacional voltada para os SIS. A seguir, são listados os artigos publicados e submetidos até o momento atual.

8.2.1 Artigos completos publicados em periódicos nacionais e internacionais

Em 2019 com o estudo de Freund, Sembay e Macedo (2019) buscando compreender a Proveniência de Dados e Segurança da Informação oriunda da Ciência da Computação inserida interdisciplinarmente no campo da Ciência da Informação:

- 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.

Já em 2020, um método genérico de captura de dados de proveniência foi publicado no estudo de Sembay, Macedo e Dutra (2020b) buscando construir os primeiros passos do método atual desta tese:

- SEMBAY, Márcio J.; MACEDO, Douglas Dyllon J de; DUTRA, Moisés Lima. A Proposed Approach for Provenance Data Gathering. **Mobile Networks and Applications**, v. 26, n. 1, p. 304-318, 2020b.

Em 2023 com o estudo de Sembay, Macedo, Pioli Júnior, Braga e Cabezuelo, buscando apresentar o ineditismo desta tese e da importância da temática em nível mundial no campo da Ciência da Computação e contribuir para o campo da Ciência da Informação:

- SEMBAY, M. J.; MACEDO, D. D. J. de; JÚNIOR, L. P., BRAGA; R. M. M.; SARASA-CABEZUELO, A. Provenance Data Management in Health Information Systems: A Systematic Literature Review. **Journal of Personalized Medicine**, v. 13, n. 6, p. 991, 2023.

8.2.2 Artigos completos publicados em anais de congressos

Em 2022, no estudo de Sembay, Marquez Filho e Macedo (2022c) foi realizada uma pesquisa em relação aos meios de interoperabilidade em SIS que contribuem para a Proveniência de dados por meio do modelo W3C PROV. Esse estudo foi base para fundamentar os estudos dos processos de interoperabilidade para o método desta tese:

- SEMBAY, Márcio José; MARQUEZ FILHO, Alexandre Augusto Gimenez; MACEDO, Douglas Dyllon Jeronimo. HL7 FHIR baseado em W3C PROV para alcançar a proveniência de dados em sistemas de informação em saúde. *In: WORKSHOP DE DADOS, INFORMAÇÃO E TECNOLOGIA (WIDAT 2022)*, 5., 2022, Vitória-ES. **Anais [...]**. Vitória: UFES, 2022. v. 1. p. 108-114.

8.2.3 Resumos publicados em anais de congressos

Em 2022, no estudo de Sembay, Marquez Filho e Macedo (2022b) sobre a Proveniência de Dados em SIS, apresentando um resumo da ideia desta tese no maior evento nacional de pesquisa em Ciência de Informação do Brasil, o artigo foi premiado em 1º lugar na categoria de resumo expandido do GT11 (Informação & Saúde):

- SEMBAY, Márcio José; MACEDO, Douglas Dyllon Jeronimo de; MARQUEZ FILHO, Alexandre Augusto Gimenez. Proveniência de dados em Sistemas de Informação em Saúde. *In: XXII ENCONTRO NACIONAL DE PESQUISA E PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO*, 22., 2022, Porto Alegre-RS. **Anais [...]**. Porto Alegre: ANCIB, 2022. v. 1. p. 1-11.

8.2.4 Artigos publicados em capítulos de livros

Em 2020, no estudo de Sembay, Macedo e Dutra (2020a), é proposto um método de

coleta de dados de proveniência com base no modelo de proveniência (PROV-DM) a ser aplicado em centros de hemoterapia brasileiros:

- 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. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering. 1. ed. **Springer International Publishing**, 2020. p. 89-102.

Em 2021, no estudo de Sembay e Macedo (2021) foi apresentada uma proposta de um método na expansão do modelo de Proveniência de Dados (PROV-DM) para SIS:

- SEMBAY, M. J.; MACEDO, D. D. J. Sistemas de informação em saúde: proposta de um método de gerenciamento de dados de proveniência no instanciamento do modelo W3C PROV-DM. *In*: Thiago Magela Rodrigues Dias. (org.). Informação, Dados e Tecnologia. **Advanced Notes in Information Science**. 1. ed. Estônia: ColNes, 2022. v. 2, p. 192-201.

Já em 2022, no estudo de Sembay, Marquez Filho e Macedo (2022a), buscou-se compreender as relações da Proveniência de Dados e Blockchain, sendo tecnologias compatíveis para o gerenciamento de dados em SIS:

- SEMBAY, Márcio José; MARQUEZ FILHO, Alexandre Augusto Gimenez; MACEDO, Douglas Dyllon Jeronimo de. Identification of the Relationships Between Data Provenance and Blockchain as a Contributing Factor for Health Information Systems. *In*: PINTO, Adilson Luiz; ARENCIBIA-JORGE, Ricardo. (org.). **Data and Information in Online Environments**. 1. ed. Cham: Switzerland, 2022. v. 1. p. 258-272.

Ainda em 2022, no estudo de Sembay e Macedo (2022), foi realizada uma pesquisa com base no instanciamento do modelo de Proveniência de Dados PROV-DM em busca de encontrar elementos que fossem relevantes para mapear os processos de proveniência em Sistemas de Informação em Saúde desta tese:

- SEMBAY, Márcio José; DE MACEDO, Douglas Dyllon Jeronimo. Health information systems: proposal of a provenance data management method in the instantiation of the W3C PROV-DM model. **Advanced Notes in Information Science**, v. 2, p. 192-201, 2022.

8.2.5 Artigos submetidos e em análise

O estudo intitulado “Data Provenance and Blockchain: an approach in the context of Health Information Systems” dos autores Márcio José Sembay (UFSC), Douglas Dyllon Jeronimo de Macedo (UFSC) e Alexandre Augusto Gimenes Marquez Filho (STT/UFSC), enviado no dia 02 de maio de 2023, ainda está em avaliação pelo Journal Mobile Networks and

Applications.

O estudo intitulado “Gerenciamento de dados de proveniência em Sistemas de Informação em Saúde” dos autores Márcio José Sembay (UFSC), Douglas Dyllon Jeronimo de Macedo (UFSC) e Alexandre Augusto Gimenes Marquez Filho (STT/UFSC), extensão do artigo premiado no ENANCIB 2022, está em processo de submissão na Revista Tendências da Pesquisa Brasileira em Ciência da Informação.

O estudo intitulado “PROV-Health: method for provenance data management in health information systems” dos autores Márcio José Sembay (UFSC), Douglas Dyllon Jeronimo de Macedo (UFSC) e Alexandre Augusto Gimenes Marquez Filho (STT/UFSC), está em processo de submissão no Journal of Biomedical Informatics.

REFERÊNCIAS

- ASSOCIAÇÃO BRASILEIRA PARA O DESENVOLVIMENTO DA INDÚSTRIA (ABDI). Centro de Gestão e Estudos Estratégicos Ciência, Tecnologia e Inovação. **Agenda Tecnológica Setorial ATS: Complexo Industrial da Saúde TELEMEDICINA PANORAMA TECNOLÓGICO**. 2016. Disponível em: <https://docplayer.com.br/46196280-Agenda-tecnologica-setorial-ats-complexo-industrial-da-saude-t-elemedicina-panorama-tecnologico.html>. Acesso em: jul. 2023.
- ABOUZHR, C.; BOERMA, T. Health information systems: the foundations of public health. **Bulletin of the World Health Organization**, v. 83, p. 578-583, 2005.
- ALBUQUERQUE, E. A. Y. *et al.* 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**, v. 7, n. 2, 2017. ISSN: 2236-1103. DOI <https://doi.org/10.18816/r-bits.v7i2.11074>. Disponível em: <https://periodicos.ufrn.br/reb/article/view/11074/8973>. Acesso em: 06 abr. 2019.
- ALKMIM, M. B. M. *et al.* **Incorporación de la telesalud en el sistema público de salud de Minas Gerais, Brasil**. Santiago de Chile Naciones Unidas; 2012.
- ALVARENGA, L. A teoria do conceito revisada em conexão com ontologias e metadados no contexto das Bibliotecas tradicionais e digitais. **Data Grama Zero: Revista de Ciência da Informação**, Rio de Janeiro, v. 2, n. 6, dez. 2001. Disponível em: <http://www.brapci.inf.br/index.php/article/download/7457>. Acesso em: 02 jan. 2019.
- ÁLVAREZ, S. *et al.* Applying Provenance in distributed organ transplant management. **IPAW**, p. 28-36, 2006. Disponível em: https://link.springer.com/content/pdf/10.1007/11890850_4.pdf. Acesso em: 10 jan. 2019.
- ANNAS, G. J. HIPAA regulations: a new era of medical-record privacy? **New England Journal of Medicine**, v. 348, p. 1486, 2003.
- ARAÚJO, C. A. Á. Ciência da informação, biblioteconomia, arquivologia e museologia: relações institucionais e teóricas. **Enc. Bibli: R. Eletr. Bibliotecon. Ci. Inf, Florianópolis**, v. 16, n. 31, p. 110-130, 2011. ISSN 1518-2924. DOI 10.5007/1518-2924.2011v16n31p110. Disponível em: <https://periodicos.ufsc.br/index.php/eb/article/viewFile/1518-924.2011v16n31p110/17765>. Acesso em: 13 fev. 2019.
- AVGEROU, C. The significance of context in information systems and organizational change. **Information Systems Journal**, v. 11, p. 43-63, 2000. DOI: 10.1046/j.1365-2575.2001.00095.x. Disponível em: https://www.researchgate.net/publication/30522670_The_Significance_of_Context_in_Information_Systems_and_Organizational_Change. Acesso em: 10 mar. 2019.
- AWS.AMAZON.COM. **O que é SQL?** Disponível em: <https://aws.amazon.com/pt/what-is/sql/>. Acesso em: 01 jun. 2022.
- BADAMPUDI, D.; WOHLIN, C.; PETERSEN, K. **Experiences from using snowballing and database searches in systematic literature studies**. *In*: INTERNATIONAL

CONFERENCE ON EVALUATION AND ASSESSMENT IN SOFTWARE ENGINEERING (EASE'15), 19., Apr. 2015. p. 1–10. DOI <https://doi.org/10.1145/2745802.2745818>

BAKKER, A. B. *et al.* Job resources boost work engagement, particularly when job demands are high. **Journal of educational Psychology**, Washington, v. 99, p. 274-284, Apr. 2007. DOI 10.1037/0022-0663.99.2.274. Disponível em: https://www.researchgate.net/publication/46688181_Job_Resources_Boost_Work_Engagement_Particularly_When_Job_Demands_Are_High. Acesso em: 15 mar. 2019.

BAKKER, A. R. HIS and RIS and PACS. *In*: HUANG, H. K. *et al.* (ed.). **Nato Ansi Series: Picture Archiving and Communications System (PACS) in Medicine**. Berlin: Springer-Verlag, 1991. v. 74. p. 157-162.

BANSAL, Srividya K. Towards a semantic extract-transform-load (ETL) framework for big data integration. *In*: **2014 IEEE International Congress on Big Data**. IEEE, 2014. p. 522-529.

BEECHAM, S. *et al.* Motivation in software engineering: A systematic literature review. **Information and software technology**, Elsevier, v. 50, n. 9-10, p. 860-878, jul. 2008. DOI 10.1016/j.infsof.2007.09.004. Disponível em: https://www.researchgate.net/publication/221956324_Motivation_in_Software_Engineering_A_systematic_literature_review. Acesso em: 15 nov. 2019.

BERNARDINI, A. *et al.* (2003). IHE: Integrating the Healthcare Enterprise, towards complete integration of healthcare information systems. **Rays**, v. 28, n. 1, p. 83-93, jan./mar. 2003. PMID: 14509182. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/14509182/>. Acesso em: 12 nov. 2019.

BIEGER, G. R.; GLOCK, M. D. The information content of picture-text instructions. **Journal of Experimental Education**, v. 1984-85, n. 53, p. 68-76. Disponível em: <https://www.tandfonline.com/doi/abs/10.1080/00220973.1985.10806364>. Acesso em: 12 nov. 2019.

BLICK, K. E. Decision-making laboratory computer system as essential tools for achievement of total quality. **Clinical Chemistry**, v. 43, n. 5, p. 908-12, 1997. DOI 10.1093/clinchem/43.5.908. Disponível em: https://www.researchgate.net/publication/14052979_Decision-making_laboratory_computer_systems_as_essential_tools_for_achievement_of_total_quality. Acesso em: 15 abr. 2019.

BOOCHEVER, S. S. HIS/RIS/PACS Integration: getting to the Gold Standard. **Radiology Management**, v. 26, n. 3, p. 16-44, maio/jun. 2004. Disponível em: https://www.researchgate.net/publication/313154397_HISRISPACS_integration_getting_to_the_gold_standard. Acesso em: 18 maio 2019.

BOOTH, A. Bridging the research-practice gap? The role of evidence based librarianship. **New Review of Information and Library Research**, v. 9, n. 1, p. 3-23, 2003. Disponível em: <https://www.tandfonline.com/doi/abs/10.1080/13614550410001687909>. Acesso em: 22 nov. 2019.

BOOTH, A. Formulating answerable questions. *In*: BOOTH, A.; BRICE, A. (ed.). **Evidence based practice for information professionals**. Facet Publishing, London, 2004. p. 61-70. Disponível em: https://www.academia.edu/2184681/Formulating_answerable_questions. Acesso em: 23 jun. 2019.

BORKO, H. Information Science: What is it? **American Documentation**, v. 19, n. 1, p. 3-5, jan. 1968. ISSN 2330-1643 Versão online. DOI <https://doi.org/10.1002/asi.5090190103>. Disponível em: <https://onlinelibrary.wiley.com/doi/abs/10.1002/asi.5090190103>. Acesso em: 09 out. 2018.

BOSE, R.; FREW, J. Lineage retrieval for scientific data processing: a survey. **ACM Comput. Surv.**, v. 37, n. 1, p. 1-28, mar. 2005. ISSN 0360-0300. Disponível em: <http://citeseerx.ist.psu.edu/viewdoc/download;jsessionid=C1746A7A546FD1B64E83540C6ED55688?doi=10.1.1.103.4855&rep=rep1&type=pdf>. Acesso em: 29 jan. 2019.

BRAGA, J. C.; BANON, G. J. F. Data Provenance: theory and application to image processing. **Latin America Transactions**, v. 6, n. 2, p. 207-214, 2008. Disponível em: <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=4609919>. Acesso em: 28 maio 2018.

BRAGHETTO, K. R.; CORDEIRO, D. Introdução à modelagem e execução de workflows científicos. *In*: **Atualizações em Informática**. 1. ed. Porto Alegre: SBC, 2014. p. 1-40. Disponível em: <http://www.each.usp.br/dc/papers/jai-cap-workflows.pdf>. Acesso em: 28 maio 2018.

BRASIL. Conselho Federal de Medicina. **Resolução de nº 1.638/2002a**. Disponível em: <https://sistemas.cfm.org.br/normas/visualizar/resolucoes/BR/2002/1638>. Acesso em: 10 nov. 2019.

BRASIL. Conselho Federal de Medicina. Resolução nº 1.639 de 10 de julho de 2002a. **Diário Oficial**, Brasília, 10 jul. 2002. Cap. 1, p. 124-5. Disponível em: http://www.portalmedico.org.br/resolucoes/cfm/2002/1639_2002.htm. Acesso em: 10 nov. 2019.

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 jan. 2019.

BRAUNSTEIN, M. L. **Health Informatics on FHIR: How HL7's New API is Transforming Healthcare**. Cham: Springer International Publishing, 2018a. DOI: 10.1007/978-3-319-93414-3.

BRITO, J. F. *et al.* Arquitetura da informação no contexto da informação em saúde: um olhar para o website da Covid-19 no Brasil. **AtoZ: Novas Práticas em Informação e Conhecimento**, v. 9, n. 2, p. 183-195, 2020. Disponível em: <https://revistas.ufpr.br/atoz/article/view/75091>. Acesso em: 20 nov. 2021.

BUCKLAND, M. K. Information as thing. **Journal of the American Society for**

Information Science, [S. l.], v. 42, n. 5, p. 351-360, June 1991.

BUNEMAN, P.; KHANNA, S. E; CHIEW, W. Why and where: a characterization of data provenance. *In*: VAN DEN BUSSCHE, J., VIANU, V. (ed.). **Database theory: ICDT 2001**. Lecture Notes in Computer Science, v. 1973. Berlin, Heidelberg: Springer, 2001. Disponível em: <https://pdfs.semanticscholar.org/b647/2ee11749ef70713bfb0e322a9ec27523ed88.pdf>. Acesso em: 02 maio 2019.

BUNEMAN, P.; TAN, W-C. **Provenance in databases**: tutorial outline. *In*: ACM SIGMOD INTERNATIONAL CONFERENCE ON MANAGEMENT OF DATA, Beijing, China, 11-14 jun. 2007. **Proceedings** [...]. Beijing, China: ACM, 2007. Disponível em: <https://pdfs.semanticscholar.org/87c1/84ab63bf6b03cff9ca2afb385d07e2a8faff.pdf>. Acesso em: 15 maio 2019.

CAPURRO, R.; HJORLAND, B. O conceito de informação. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 12, n. 1, p. 148-207, jan./abr. 2007. ISSN 1981-5344. Disponível em: <http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/54/47>. Acesso em: 07 set. 2018.

CAFEZEIRO, I.; COSTA, L. C. da; KUBRUSLY, R. da S. Ciência da Computação, Ciência da Informação, Sistemas de Informação: uma reflexão sobre o papel da informação e da interdisciplinaridade na configuração das tecnologias e das ciências. **Perspectivas em Ciência da Informação**, v. 21, p. 111-133, 2016.

CAPURRO, R. Epistemologia e Ciência da Informação. *In*: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 5., Belo Horizonte, 10 nov. 2003. **Anais** [...]. Belo Horizonte: ENANCIB, 2003. Disponível em: http://www.capurro.de/enancib_p.htm. Acesso em: 06 abr. 2018.

CAPURRO, R.; HJORLAND, B. O conceito de informação. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 12, n. 1, p. 148-207, jan./abr. 2007. ISSN 1981-5344. Disponível em: <http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/54/47>. Acesso em: 07 set. 2018.

CHIASSON, M. W.; DAVIDSON, E. Pushing the contextual envelope: developing and diffusing IS theory for health information systems research. **Information and Organization**, v. 14, p. 155-188, 2004.

CHOWDHURY, A. R; FALCHUK, B.; MISRA, A. **MediAlly**: a provenance-aware remote health monitoring middleware. *In*: 2010 IEEE INTERNATIONAL CONFERENCE ON PERSVASIVE COMPUTING AND COMMUNICATIONS (PERCOM), p. 1-10, mar. 2010. IEEE. DOI <http://dx.doi.org/10.1109/percom.2010.5466985>.

CLARKE, M.; HORTON, R. Bringing it all together: Lancet-Cochrane collaborate on systematic reviews. **Lancet**, v. 357, n. 1728, jun. 2001. Disponível em: <https://europepmc.org/abstract/med/11403806>. Acesso em: 14 maio 2018.

COOK, D.; MULROW, C. D.; HAYNES; R. B. Systematic reviews: synthesis of best evidence for clinical decisions. **Ann Intern Med.**, v. 126, p. 376-80, 1997. Disponível em: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.733.1479&rep=rep1&type=pdf>.

Acesso em: 10 jul. 2019.

COSTA, C. G. A. **Desenvolvimento e avaliação tecnológica de um sistema de prontuário eletrônico do paciente, baseado nos paradigmas de world wide web e da engenharia de software**. 2001. Dissertação (Mestrado em Engenharia Elétrica) – Universidade Estadual de Campinas, Campinas, SP, 2001. Disponível em: http://repositorio.unicamp.br/jspui/bitstream/REPOSIP/260174/1/Costa_ClaudioGiullianoAlvesda_M.pdf. Acesso em: 10 nov. 2019.

CRAIG, J; PATTERSON, V. Introduction to the practice of telemedicine. **Journal of Telemedicine and Telecare**, v. 11, n. 1, p. 3-9, 2005. DOI 10.1258/1357633053430494. Disponível em: <https://europepmc.org/abstract/med/15829036>. Acesso em: 21 nov. 2019.

CUNHA, M. B.; CAVALCANTI, C. R. O. **Dicionário de biblioteconomia e arquivologia**. Brasília: Briquet de Lemos, 2008. ISBN 978-85-85637-35-4. Disponível em: <http://repositorio.unb.br/handle/10482/34113>. Acesso em: 20 maio 2019.

CUPANI, A. **Filosofia da tecnologia: um convite**. 2. ed. Florianópolis: Ed. da UFSC, 2013. ISBN 9788532806574.

CURCIN, V. *et al.* Templates as a method for implementing data provenance in decision support systems. **Journal of Biomedical InformaTIC**, v. 65, p. 1-21, jan. 2017. DOI <http://dx.doi.org/10.1016/j.jbi.2016.10.022>. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1532046416301599>. Acesso em: 01 jun. 2018.

CURRELL, R. *et al.* Telemedicine versus face to face patient care: effects on professional practice and health care outcomes. **Cochrane Database of Systematic Reviews**, v. 2, Art. n° CD002098, 2000. Disponível em: <https://www.ncbi.nlm.nih.gov/pubmed/10796678>. Acesso em: 21 nov. 2019.

DAVIDSON, S. B; FREIRE, J. 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 jun. 2018.

DCM4CHE.ORG (2022). **Open Source Clinical Image and Object Management**. Disponível em: www.dcm4che.org/. Acesso em: 02 ago. 2022.

DELANEY, B. C. *et al.* Translational medicine and patient safety in Europe: TRANSFoRm-Architecture for the Learning Health System in Europe. **BioMed Research International**, n. 961526, 2015. DOI \bibinfo{doi}{10.1155/2015/961526}.

DEMIRIS, G. Integration of telemedicine in graduate medical informaTIC education. **J Am Med Inform Assoc**. v. 10, n. 4, p. 310-4, 2003. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC181980/>. Acesso em: 21 nov. 2019.

DIXON, E.; HAMEED, M.; SUTHERLAND, F. Evaluating meta-analyses in the general surgical literature: a critical appraisal. **Ann Surg**, v. 241, p. 450-459, 2005. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1356983/>. Acesso em: 19 jun. 2019.

DOCKER.COM. **Develop faster**. Run anywhere. Disponível em: <https://www.docker.com/>. Acesso em: 20 ago 2022.

FEATHER, J; STURGES, P. (ed.). **International encyclopedia of information and library science**. 2. ed. New York: Routledge, 2003. ISBN 0-203-40330-4 Master e-book. Disponível em: http://mlisuok.weebly.com/uploads/2/6/9/0/26907671/international_encyclopedia_of_information_ind_library_science.pdf. Acesso em: 15 mar. 2019.

FONG, H. F. Association between health literacy and parental self-efficacy among parents of newborn children. **Journal of Pediatrics**, v. 202, p. 265–271, nov. 2018.

FOSKETT, D. J. Informática. *In*: GOMES, H. E. (org.). **Ciência da Informação ou informática?** Rio de Janeiro: Calunga, 1980.

FRANÇA, F. H; GAUNT, P. N. The need for security - a clinical view. **Int J Biomed Comput**, v. 35, Suppl 1, p. 189-194, 1994. Disponível em: <https://www.ncbi.nlm.nih.gov/pubmed/8188413>. Acesso em: 22 nov. 2019.

FREIRE, G. H. A.; SILVA, J. L. C. A configuração do campo da Ciência da Informação: marcas de uma identidade. **Informação e Sociedade: estudos**, João Pessoa, v. 22, p. 161-174, número especial, 2012. ISSN 1809-4783. Disponível em: <http://www.periodicos.ufpb.br/ojs/index.php/ies/article/view/13775/8216>. Acesso em: 20 out. 2018.

FREIRE, J. *et al.* Provenance for computational tasks: a survey. **Journal Computing in Science and Engineering**, v. 10. n. 3, p. 11-21, 2008. ISSN 15219615. Disponível em: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.147.3801&rep=rep1&type=pdf>. Acesso em: 24 maio 2018.

FREUND, G. P.; SEMBAY, M. J.; MACEDO, D. 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**, Brasília, 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.

FRIEDMAN, C. *et al.* Toward a science of learning systems: a research agenda for the high-functioning Learning Health System. **Journal of the American Medical Informatics Association**, v. 22, n. 1, p. 43-50, 2015.

GALVÃO, M. C. D; RICARTE, I. L. M. O prontuário eletrônico do paciente do século XXI: contribuições necessárias da Ciência da Informação. **IncID: R. Ci. Inf. e Doc.**, Ribeirão Preto, v. 2, n. 2, p. 77-100, jul./dez. 2011. Disponível em: <http://www.revistas.usp.br/incid/article/view/42353/46024>. Acesso em: 23 nov. 2019.

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

GIL, A. C. **Como elaborar projetos de pesquisa**. 4. ed. São Paulo: Atlas, 2007.

GIL, Y.; MILES, S. **PROV Model Primer**. Disponível em: <https://www.w3.org/TR/2013/NOTE-prov-primer-20130430/>. Acesso em: 10 mar. 2020.

GITLAB. **About Us**. 2022. Disponível em: <https://about.gitlab.com/company/>. Acesso em: 06 fev. 2022.

GLAVIC, B; DITTRICH, K. **Data provenance: a categorization of existing approaches**. 2007. Disponível em: <https://pdfs.semanticscholar.org/7f05/966e238a80697f0dec94edadc1c3b99c4d4b.pdf>. Acesso em: 20 ago. 2018.

GONG, J.; LIN, S.; LI, J. Research on personal health data provenance and right confirmation with smart contract. *In: IEEE ADVANCED INFORMATION TECHNOLOGY, ELECTRONIC AND AUTOMATION CONTROL CONFERENCE (IAEAC)*, 4., p. 0-1, dez. 2019. DOI <http://dx.doi.org/10.1109/iaeac47372.2019.8997930>.

GOOGLE. **Formulários**. 2020. Disponível em: <https://workspace.google.com.br/intl/pt-BR/products/forms/>. Acesso em: 08 nov. 2020.

GRAFANA.COM. **Grafana Labs**. Disponível em: <https://grafana.com/>. Acesso em: 20 set 2022

GROTH, P; MOREAU, L. PROV-overview. An overview of the PROV family of documents. 2013. Disponível em: <https://www.w3.org/TR/prov-overview/>. Acesso em: 20 nov. 2021.

GROUP, C. **Cyclops Group**. Disponível em: <http://www.cyclops.ufsc.br>. Acesso em: 29 nov. 2019.

HAPI FHIR. **Global Good: Powering Interoperability Around the World for 18 Years**. Disponível em: <https://hapifhir.io/hapi-fhir/>. Acesso em: 01 ago. 2022.

HÅKANSSON, S.; GAVELIN, C. What do we really know about the cost-effectiveness of telemedicine? **J Telemed Telecare**, v. 6, suppl. 1, p. 133-136, 2000. Disponível em: <https://www.ncbi.nlm.nih.gov/pubmed/10793998>. Acesso em: 21 nov. 2019.

HARDIN, T.; KOTZ, D. Amanuensis: Information provenance for health-data systems. **Information Processing & Management**, v. 58, n. 2, p. 102460, 2021.

HAUX, R. Health information systems-past, present, future. **International Journal of Medical InformaTIC**, v. 75, p. 268-281, 2006.

HERSH, W. R. *et al.* Implementation and evaluation of a medical informatics distance education program. **J Am Med Inform Assoc**, v. 8, p. 570-584, 2001. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC130067/>. Acesso em: 21 nov. 2019.

HEY, T.; TANSLEY, S.; TOLLE, K. **Fourth paradigm: data-intensive scientific discovery**. EUA: Microsoft Corporation, 2009. ISBN 978-0-9825442-0-4. Disponível em: <http://research.microsoft.com/en-us/collaboration/fourthparadigm/>. Acesso em: 14 abr. 2015.

HILLS, D. J. *et al.* **The importance of data set provenance for science**. Dec. 2015.

Disponível em: <https://eos.org/opinions/the-importance-of-data-set-provenance-for-science>. Acesso em: 15 fev. 2019.

HL7 International Inc. **Provenance**. 2022a. Disponível em: <https://www.hl7.org/fhir/provenance.html>. Acesso em: 11 jan 2022.

HL7 International Inc. **5Ws Pattern**. 2022b. Disponível em: <https://build.fhir.org/fivews.html>. Acesso em: 10 jan 2022.

HL7 International Inc. **Fast Healthcare Interoperability Resources Release 3(STU)**. 2022c. Disponível em: <http://fhir.hl7.org>. Acesso em: 10 jan 2022.

HL7 International Inc. **Profiling FHIR**. 2022d. Disponível em: <https://hl7.org/fhir/profiling.html>. Acesso em: 13 jan. 2023.

HONG, J.; MORRIS, P; SEO, J. **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.*

HONEYMAN, J. C. Information Systems Integration in Radiology. **Journal of Digital Imaging**. v. 12, n. 2, supl. 1, p. 218-222, maio 1999.

HOERBST, A.; AMMENWERTH, E. Electronic health records. **Methods of information in medicine**, v. 49, n. 04, p. 320-336, 2010.

HUANG, H. K. **PACS and imaging informaTIC**: basic principles and applications. Wiley-Liss, 2004. 1073 p.

HUSSAIN, M. A.; LANGER, S. G.; KOHLI, M. Learning HL7 FHIR using the HAPI FHIR server and its use in medical imaging with the SIIM dataset. **Journal of digital imaging**, v. 31, n. 3, p. 334-340, 2018.

INSOMNIA. **Build APIs that work**, 2022. Disponível em: <https://insomnia.rest/>. Acesso em: 10 fev. 2021.

IBM.COM. ETL (Extract, Transform, Load). Disponível em: <https://www.ibm.com/topics/etl>. Acesso em: 03 ago. 2022.

JAIGIRDAR, F.T; RUDOLPH, C; BAIN, C. **Prov-IoT**: A Security-Aware IoT Provenance Model – 2020. *In: IEEE INTERNATIONAL CONFERENCE ON TRUST, 19., 2020.*

JALALI, S.; WOHLIN, C. **Systematic literature studies**: database searches vs. backward snowballing. *In: PROCEEDINGS OF THE 6TH INTERNATIONAL SYMPOSIUM ON EMPIRICAL SOFTWARE ENGINEERING AND MEASUREMENT (ESEM'12), 2012. p. 29–38. DOI <https://doi.org/10.1145/2372251.2372257>*

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

JSON.ORG. **Introducing JSON**. 2022. Disponível em: <https://www.json.org/json-en.html>.

Acesso em: 25 jan 2022.

KILINTZIS, V. *et al.* Supporting integrated care with a flexible data management framework built upon Linked Data, HL7 FHIR and ontologies. **Journal of Biomedical Informatics**, Elsevier, v. 94, Apr., jun. 2019. DOI: 10.1016/j.jbi.2019.103179.

KIM, Y. S. Telemedicine in the USA with focus on clinical applications and issues. **Yonsei Med J.**, v. 45, n. 5, p. 761-75, 2004. Disponível: <https://www.ncbi.nlm.nih.gov/pubmed/15515185> Acesso em: 21 nov. 2019.

KITCHENHAM, B. **Procedure for undertaking systematic reviews.** Computer Science Department, Keele University (TRISE-0401) and National ICT Australia Ltd. Joint Technical Report. 2004.

KITCHENHAM, B.; CHARTERS, S. **Guidelines for performing systematic literature reviews in software engineering.** Citeseer, 2007.

KLING, J. What is Thecnology? *In*: DUSEK, V. (ed.). **Philosophy the technology: the technological condition an anthology.** Malden: Blackwell Publishing, 2006.

KNECHTEL, M. R. **Metodologia da pesquisa em educação: uma abordagem teórico-prática dialogada.** Curitiba: Intersaberes, 2014.

KOLTAY, T. Data literacy for researchers and data librarians. **Journal of Librarianship and Information Science**, v. 49, n. 1, 2017. ISSN 1741-6477 versão online. DOI <https://doi.org/10.1177/0961000615616450>. Disponível em: <http://journals.sagepub.com/doi/abs/10.1177/0961000615616450>. Acesso em: 4 maio 2019.

KOCK-SCHOPPENHAUER, A. *et al.* Practical Extension of Provenance to Healthcare Data Based on the W3C PROV Standard. **Studies In Health Technology and Informatics**, 253, p. 28–32, 2018. IOS Press. DOI <http://dx.doi.org/10.3233/978-1-61499-896-9-28>.

KOOP, D.; FREIRE, J. **Reorganizing workflow evolution provenance.** *In*: USENIX WORKSHOP ON THE THEORY AND PRACTICE OF PROVENANCE (TaPP 2014), 6., 13-13 June 2014, Cologne, Germany.

KORHONEN, I.; PARKKA, J.; VAN GILS, M. Health monitoring in the home of the future. **IEEE Engineering in medicine and biology magazine**, v. 22, n. 3, p. 66-73, 2003.

KOVALCHUK, Y. *et al.* **Provenance-aware pervasive computing in clinical applications.** *In*: 2013 IEEE INTERNATIONAL CONFERENCE ON WIRELESS AND MOBILE COMPUTING, NETWORKING AND COMMUNICATIONS (WIMOB), 9., p. 1-10, out. 2013. IEEE. DOI <http://dx.doi.org/10.1109/wimob.2013.6673376>. Acesso em: 15 maio 2019.

LAW, M. Y. Y.; ZHOU, Z. (2003). New Direction in PACS Education and Training. **Computerized Medical Imaging and Graphics**, v. 27, p. 147–156, 2003.

LI, Q. *et al.* **User-centric Annotation Management for Biological Data.** *In*: INTERNATIONAL PROVENANCE AND ANNOTATION WORKSHOP (IPAW'08), 2., Salt Lake City, Utah, p. 54–61, jun. 2008.

LIANG, X. *et al.* **Integrating blockchain for data sharing and collaboration in mobile healthcare applications.** *In:* 2017 IEEE ANNUAL INTERNATIONAL SYMPOSIUM ON PERSONAL, 28., Indoor, and Mobile Radio Communications (PIMRC), Montreal, Qc, Canada, 2017. p. 1–25.

LIKERT, R. A technique for the measurement of attitudes. **Archives of Psychology**, v. 22, n. 140, p. 1–55, 1932.

LINDE, K.; WILLICH, S. N. How objective are systematic reviews? Differences between reviews on complementary medicine. **J R Soc Med.**, v. 96, p. 17–22, 2003. Disponível em: <https://www.ncbi.nlm.nih.gov/pubmed/12519797>. Acesso em: 11 jul. 2019.

LINUX.ORG. **Download Linux.** Disponível em: <https://www.linux.org/pages/download/>. Acesso em: 02 ago. 2022.

LOMOTÉY, R. K.; DETERS, R. Mobile-Based Medical Data Accessibility in mHealth. **2014 IEEE International Conference On Mobile Cloud Computing, Services, And Engineering**, 2., [S. l.], p. 1-10, abr. 2014. IEEE. DOI <http://dx.doi.org/10.1109/mobilecloud.2014.24>.

MACEDO, D. D. J. **Um modelo distribuído de armazenamento hierárquico de conhecimento médico.** 2014. 214 f. Tese (Doutorado) – Curso de Programa de Pós-graduação em Engenharia e Gestão do Conhecimento, Departamento de Engenharia do Conhecimento, Universidade Federal de Santa Catarina (UFSC), Florianópolis, 2014. Disponível em: <https://repositorio.ufsc.br/xmlui/bitstream/handle/123456789/123290/326446.pdf?sequence=1&isAllowed=y>. Acesso em: 10 nov. 2019.

MACHLUP, F, MANSFIELD, U, ed. **The study of information:** interdisciplinary messages. New York: John Wiley & Sons, 1983. ISBN 0-471-88717-X.

MAIA, R.; WANGENHEIM, A. V.; NOBRE, L. **A statewide telemedicine network for public health in Brazil.** *In:* IEEE INTERNATIONAL SYMPOSIUM ON COMPUTER-BASED MEDICAL SYSTEMS, CBMS 2006, 19., [S.l.: s.n.], 2006. p. 495-500.

MALDONADO, J. M. S. V; MARQUES, A. B; CRUZ, A. Telemedicina: desafios à sua difusão no Brasil. **Cad. Saúde Pública**, Rio de Janeiro, v. 32, Sup 2:e00155615, 2016. Disponível em: http://www.scielo.br/pdf/csp/v32s2/pt_1678-4464-csp-32-s2-e00155615.pdf. Acesso em: 22 nov. 2019.

MALLET, R. B. Teledermatology in practice. **Clin Exp Dermatol**, v. 28, n. 4, p. 356-9, jul. 2003. Disponível em: <https://www.ncbi.nlm.nih.gov/pubmed/12823290>. Acesso em: 21 nov. 2019.

MARCONI, M. A.; LAKATOS, E. M. **Fundamentos da metodologia científica.** 5. ed. São Paulo: Atlas, 2003.

MARGHERI, A. *et al.* Decentralised provenance for healthcare data. **International Journal of Medical InformaTIC**, v. 141, p. 1-21, set. 2020. Elsevier BV. DOI

<http://dx.doi.org/10.1016/j.ijmedinf.2020.104197>.

MASSI, M. *et al.* **Using PROV and blockchain to achieve health data provenance.** Technical report, 2018.

MASSAD, E.; MARIN, H. F.; AZEVEDO, R. S. Prontuário Eletrônico do Paciente: definições e conceitos. *In:* MASSAD, E.; MARIN, H. F.; AZEVEDO, R. S. (Ed.). **O prontuário eletrônico do paciente na assistência, informação e conhecimento médico.** São Paulo: USP, 2003. p. 1-20. Disponível em: http://www.sbis.org.br/biblioteca_virtual/prontuario.pdf. Acesso em: 19 nov. 2019.

MICHALA, A. L.; ATTAR, H.; VOURGANAS, I. Secure Data Transfer and Provenance for Distributed Healthcare. *In:* **Intelligent Healthcare.** Singapore: Springer, 2022. p. 241-260. DOI 10.1007/978-981-16-8150-9_11

MICROSOFT.COM. **The best Windows yet.** Disponível em: <https://www.microsoft.com/en-us/windows/>. Acesso em: 20 jul. 2022.

MILDENBERGER, P.; EICHELBERG, M.; MARTIN, E. Introduction to the DICOM standard. **European radiology,** Springer, v. 12, n. 4, p. 920-927, 2002.

MINISTÉRIO DA SAÚDE. **Diretrizes operacionais: pactos pela vida, em defesa do SUS e de gestão.** 1. ed. Brasília: Ministério da Saúde; 2006. p. 76.

MINISTÉRIO DA SAÚDE. Portaria no 2073, de 31 de agosto de 2011. Regulamenta o uso de padrões de interoperabilidade e informação em saúde para sistemas de informação em saúde no âmbito do Sistema Único de Saúde, nos níveis Municipal, Distrital, Estadual e Federal, e para os sistemas privados e do setor de saúde suplementar [Internet]. **Diário Oficial da União,** 1º set. 2011, Brasília, DF. Disponível em: https://bvsms.saude.gov.br/bvs/saudelegis/gm/2011/prt2073_31_08_2011.html. Acesso em: 6 dez. 2021.

MINISTÉRIO DA SAÚDE. Portaria no 1434, de 28 de maio de 2020. Institui o Programa Conecte SUS e altera a Portaria de Consolidação nº 1/GM/MS, de 28 de setembro de 2017, para instituir a Rede Nacional de Dados em Saúde e dispor sobre a adoção de padrões de interoperabilidade em saúde [Internet]. **Diário Oficial da União,** 29 maio 2020, Brasília, DF. Disponível em: <https://www.in.gov.br/en/web/dou/-/portaria-n-1.434-de-28-de-maio-de-2020-259143327>. Acesso em: 6 dez. 2021.

MIRANDA, N, J, O; PINTO, V. B. Prontuário eletrônico do paciente: padronização e interoperabilidade. *In:* ENCONTRO NACIONAL DE PESQUISA EM PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO, 16., 2015. João Pessoa. **Anais [...].** João Pessoa: ANCIB, 2015. Disponível em: <https://repositorio.ufc.br/handle/riufc/19187>. Acesso em: 6 dez. 2021.

MOERSCHBERGER, M. S.; CRUZ, F. R.; LANGARO, F. Reflexões acerca da ética e da qualidade dos registros psicológicos em prontuário eletrônico multiprofissional. **Revista SBPH,** Rio de Janeiro, v. 20, n. 2, p. 89–100, dez. 2017. Disponível em: http://pepsic.bvsalud.org/scielo.php?script=sci_arttext&pid=S1516-08582017000200006&lng=pt&nrm=iso. Acesso em: 24 nov. 2019.

- MOREAU, I. *et al.* The open provenance model core specification (v. 1.1). **Future Generation Computer Systems**, v. 27, n. 6, p. 743–756, jun. 2011. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0167739X10001275>. Acesso em: 30 dez. 2018.
- MOREAU, L. Usage of Provenance: a Tour of Babel. *In: INTERNATIONAL PROVENANCE AND ANNOTATION WORKSHOP (IPAW'06)*, Chicago, Illinois, USA, 3-5, May 2006. **Proceedings** [...]. Disponível em: <https://eprints.soton.ac.uk/409382/1/babel.pdf>. Acesso em: 27 jul. 2018.
- MOREAU, L.; FOSTER, I. Provenance and annotation of data. *In: FOSTER, I. (ed.), Springer Science & Business Media, 2006. v. 4145. In: INTERNATIONAL PROVENANCE AND ANNOTATION WORKSHOP, IPAW 2006, Chicago, IL, USA, May 3-5, 2006, revised selected papers.* DOI 10.1007/11890850. ISBN 978-3-540-46303-0. Disponível em: <https://www.springer.com/br/book/9783540463023>. Acesso em: 06 jul. 2018.
- MOREAU, L.; GROTH, P. Provenance: an introduction to PROV. **Synthesis lectures on the semantic web: theory and technology**, v. 3, n. 4, p. 1–129, 2013. DOI <https://10.2200/s00528ed1v01y201308wbe007>
- MOTA, E.; CARVALHO, D. M. T. Sistemas de informação em saúde. *In: Epidemiologia & Saúde*, 6. ed. Rio de Janeiro: MEDSI, 2003. p. 708.
- MOTA, F. R. L.; BABETTO, H. S. **Processamento e compartilhamento da informação em prontuários eletrônicos**. Minas Gerais: UFMG, 2004. Disponível em: https://www.academia.edu/24214266/Processamento_e_Compartilhamento_da_Informa%C3%A7%C3%A3o_em_Prontu%C3%A1rios_Eletr%C3%B4nicos. Acesso em: 11 nov. 2019.
- MOTTA, G. H. M. B. **Um modelo de autorização contextual para o controle de acesso ao prontuário eletrônico do paciente em ambientes abertos e distribuídos**. 2003. Tese (Doutorado em Sistemas Eletrônicos) – Escola Politécnica, Universidade de São Paulo, 2004. Disponível em: <https://teses.usp.br/teses/disponiveis/3/3142/tde-05042004-152226/pt-br.php> Acesso em: 26 jun. 2020.
- NADKARNI, P. M. Clinical research information systems. *In: Clinical Research Informatics*. Springer, Cham, 2019. p. 171-190.
- NATIONAL SCIENCE BOARD (NSB). **Digital research data sharing and management**, 2011. Disponível em: <https://www.nsf.gov/nsb/publications/2011/nsb1124.pdf>. Acesso em: 23 maio 2019.
- NEMA, A. C. R. **Digital imaging and communications in medicine (DICOM) Part 3: Information object definitions**. Part, 2007. Disponível em: http://DICOM.nema.org/DICOM/2004/04_03PU3.PDF. Acesso em: 11 ago. 2019.
- NEVES, B. C.; BRAZ, M. I. Interlocação entre saúde e Ciência da Informação: proposta para o diagrama multidisciplinar da CI. **Informação & Informação**, v. 23, n. 3, p. 100-121, 2018.
- NORRIS, A. C. **Essentials of Telemedicine and Telecare**. Baffins Lane, England: John Wiley & Sons, 2002.

O'BRIEN, J. A. **Sistemas de informações e as decisões gerenciais na era da Internet**. 9. ed. São Paulo: Saraiva, 2004.

OLARONKE, I; SORIYAN, A; GAMBO, I; OLALEKE, J. 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.

OOSTERWIJK, H. **DICOM Basics**. 2. ed. EUA: Otech, 2002. 158 p.

OPEN PROVENANCE MODEL (OPM). 2010. Disponível em: <https://openprovenance.org/opm/old-index.html>. Acesso em: 02 jun. 2018.

OPENSOURCE.COM. What is Bash? Disponível em: <https://opensource.com/resources/what-bash>. Acesso em: 05 ago 2022

OXFORD ENGLISH DICITIONARY (OED) Online. **datum**, n., 2015. Oxford University Press. Disponível em: <http://www.oed.com/view/Entry/47434>. Acesso em: 15 jun. 2019.

PETERSEN, K.; VAKKALANKA, S.; KUZNIARZ, L. Guidelines for conducting systematic mapping studies in software engineering: An update. **Information and Software Technology**, v. 64, p. 1-18, 2015.

PETTICREW, M.; ROBERTS, H. **Systematic reviews in the social sciences**. Malden, MA: Blackwell Publishing, 2006.

PHYTON.ORG. **Phyton** (2022). Disponível em: <https://www.python.org/>. Acesso em: 01 fev. 2022.

PINHEIRO, L. V. R. Campo interdisciplinar da Ciência da Informação: fronteiras remotas e recentes. *In*: PINHEIRO, L. V. R. **Ciência da Informação, Ciências Sociais e interdisciplinaridade**. Brasília: IBICT, 1999. p. 155-182.

PINTO, V. B. Prontuário eletrônico do paciente: documento técnico de informação e comunicação do domínio da Saúde. **Enc. Bibli: R. Eletr. Bibliotecon. Ci. Inf.**, Florianópolis, n.21, 1º sem. 2006. Disponível em: <https://www.hmtj.org.br/arquivos.hmtj/prontuario.pdf>. Acesso em: 20 jun. 2020.

POSTGRESQL.ORG. **About PostgreSQL**. What is PostgreSQL? 2022. Disponível em: <https://www.postgresql.org/about/>. Acesso em: 20 jan. 2022.

PRABHUNE, A. *et al.* Prov2ONE: an algorithm for automatically constructing ProvONE provenance graphs. *In*: **International Provenance and Annotation Workshop**. Springer, Cham, 2016. p. 204-208. DOI 10.1007/978-3-319-40593-3_22

PRASAD, A. *et al.* **Provenance framework for mHealth**. *In*: INTERNATIONAL CONFERENCE ON COMMUNICATION SYSTEMS AND NETWORKS (COMSNETS), 5., p. 1-10, jan. 2013. IEEE. DOI <http://dx.doi.org/10.1109/comsnets.2013.6465599>.

RAHMAN, M. A. *et al.* Secure and Provenance Enhanced Internet of Health Things Framework: A Blockchain Managed Federated Learning Approach. **IEEE**, v. 8, p. 205071–205087. Nov. 2020. <http://10.1109/ACCESS.2020.3037474>. Acesso em: 10 jan. 2021

RAMESH, P. *et al.* **Insight**: semantic provenance and analysis platform for multi-center neurology healthcare research. *In*: INTERNATIONAL CONFERENCE ON BIOINFORMATIC AND BIOMEDICINE (BIBM), IEEE, p. 1-10, nov. 2015. <http://dx.doi.org/10.1109/bibm.2015.7359776>. Acesso em: 02 jun. 2018.

RAO, B; LOMBARDI, A. II. Telemedicine: current status in developed and developing countries. **Journal of Drugs in Dermatology**, v. 8, n. 4, p. 371-375, 2009. Disponível em: <https://www.ncbi.nlm.nih.gov/pubmed/19363855>. Acesso em: 21 nov.2019.

REDE NACIONAL DE PESQUISA (RNP). **Quem Somos**. 2019. Disponível em: <https://www.rnp.br/>. Acesso em: 22 nov. 2019.

REDE UNIVERSITÁRIA DE TELEMEDICINA (RUTE). **O que é a Rede Universitária de Telemedicina (Rute)?** 2019. Disponível em: <https://rute.rnp.br/arute>. Acesso em: 22 nov. 2019.

RENEKER, M. 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.

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

ROINE, R.; OHINMAA, A.; HAILEY, D. Assessing telemedicine: a systematic review of the literature. **CMAJ**, v. 165, n. 6, p.765-71, 2001. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC81454/>. Acesso em: 21 nov. 2019.

SABBATINI, R. M. E. **A Telemedicina no Brasil**: evolução e perspectivas. 2012. Disponível em: <http://www.sabbatini.com/renato/papers/TelemedicinaBrasil.pdf>. Acesso em: 22 nov. 2019.

SÁNCHEZ, Y. K. R; DEMURJIAN, S. A.; BAIHAN, M. S. Achieving RBAC on RESTful APIs for mobile apps using FHIR. *In*: **2017 IEEE International Conference on Mobile Cloud Computing, Services, and Engineering (MobileCloud)**, 5., IEEE, 2017. p. 139-144.

SANTOS, A. R. dos. **Metodologia científica**: a construção do conhecimento. 6. ed. Rio de Janeiro: DP&A, 2004.

SANTOS, E. M. dos; REINHARD, N. Electronic Government Interoperability: Identifying the Barriers for Frameworks Adoption. **Social Science Computer Review**, v. 30, p. 71-82, 2011.

SARACEVIC, T. Ciência da Informação: origem, evolução e relações. **Perspectivas em Ciência da Informação**, v. 1, n. 1, p. 41-62, jan./jun. 1996. ISSN 1981-5344. Disponível em: <http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/235/22>. Acesso em: 23 out.

2018.

SARACEVIC, T. Information Science. *In*: BATES, M. J.; MAACK, M. N. (ed.). **Incyclopedia of Library and Information Science**. New York: Taylor & Francis, 2010. p. 2570-2586. Disponível em: <https://comminfo.rutgers.edu/~tefko/SaracevicInformationScienceELIS2009.pdf>. Acesso em: 14 ago. 2020.

SCHREIBER, A.; STRUMINKSI, R. Visualizing the provenance of personal data using comics. **Computers**, v. 7, n. 1, p. 12-24, 1 fev. 2018. MDPI AG. DOI <http://dx.doi.org/10.3390/computers7010012>.

SEMBAY, M. J; MACEDO, D. J; DUTRA, M. L. A 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. Springer International Publishing, 2020a. v. 1, p. 89-102. DOI 10.1007/978-3-030-50072-6_8.

SEMBAY, M. J; MACEDO, D. D. J; DUTRA, M. L. A proposed approach for provenance data gathering. **Mobile Networks & Applications**, p. 1-13, 2020b. DOI 10.1007/s11036-020-01648-7

SEMBAY, M. J., MACEDO, D. D. J. 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, Tallinn, Estonia: ColNes Publishing, 2021. DOI <https://doi.org/10.47909/anis.978-9916-9760-3-6.101>.

SEMBAY, M. J., MACEDO, D. D. J., MARQUEZ FILHO, A. A. Identification of the relationships between Data Provenance and Blockchain as a contributing factor for Health Information Systems. *In*: **Lecture Notes of the Institute for Computer Sciences, Social InformaTIC and Telecommunications Engineering**. 1. ed. [S. l.]: Springer International Publishing, 2022a.

SEMBAY, M. J.; MACEDO, D. D. J. de; MARQUEZ FILHO, A. A. G. Proveniência de dados em Sistemas de Informação em Saúde. *In*: ENCONTRO NACIONAL DE PESQUISA E PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO, 2022, Porto Alegre. 2002b. **ENCONTRO NACIONAL DE PESQUISA E PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO**, 22., Porto Alegre: ANCIB, 2022. v. 1. p. 1-11. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/200577>. Acesso em: 29 mar. 2023.

SEMBAY, M. J.; MACEDO, D. D. J.; MARQUEZ FILHO, A. A. G. HL7 FHIR baseado em W3C PROV para alcançar a proveniência de dados em sistemas de informação em saúde. Workshop de informação, dados e tecnologia (WIDAT 2022), 2022. **Anais [...]**. Vitória (ES): PPGCI/UFES, 2022c. 247p.

SEMBAY, M. J.; MACEDO, D. D. J. de, JÚNIOR, L. P.; BRAGA, R. M. M.; SARASACABEZUELO, A. Provenance Data Management in Health Information Systems: a Systematic Literature Review. **Journal of Personalized Medicine**, v. 13, n. 6, p. 991, 2023.

SENEVIRATNE, O.; KAGAL, L. **Enabling privacy through transparency**. *In*: ANNUAL INTERNATIONAL CONFERENCE ON PRIVACY, SECURITY AND TRUST, 20., p. 1-

10, jul. 2014. IEEE. DOI <http://dx.doi.org/10.1109/pst.2014.6890931>.

SHORTLIFFE, E. H. *et al.* **Medical InformaTIC**: computer applications in health care and biomedicine. 2. ed. New York: Springer Science, 2000. 854 p.

SILVA, A.; RIBEIRO, F. Information Science and Philosophy of Information: Approaches and Differences. *In*: DEMIR, H. (Ed.). **Luciano Floridi's Philosophy of Technology: Critical Reflections.**, Heidelberg: Springer Science+Business Media Dordrecht, 2012. (Philosophy of Engineering and Technology 8).

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

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

SIMMHAN, Y.; PLALE, B.; GANNON, D. A framework for collecting provenance in data-centric scientific workflows. *In*: INTERNATIONAL CONFERENCE ON WEB SERVICES, ICWS'06, Chicago, EUA, 2006. **Anais** [...]. Chicago, EUA: IEEE Computer Society, 2006. Disponível em: <https://www.cs.indiana.edu/dde/papers/SimmhanICWS06.pdf>. Acesso em: 01 jun. 2018.

SIMPLIFIER.NET. **Where FHIR profiles are published**. (2022a). Disponível em: <https://simplifier.net/>. Acesso em: 03 jan. 2022.

SIMPLIFIER.NET. **STT/RCTMon FHIR® (R4)**. Sistema Integrado de Telemedicina e Telessaúde - Universidade Federal de Santa Catarina (2022b). Acesso em: 05 set. 2022. Disponível em: <https://simplifier.net/STT/SC>

SIMPLIFIER.NET. **Forge**. (2022c). Disponível em: <https://simplifier.net/forge>. Acesso em: 3 jan. 2022.

SOUZA-JUNIOR, M.; CUNHA, M. X. C.; DORNELAS, J. S. **Examinando as dimensões burocráticas dos sistemas de informação em saúde pública a partir das lentes teóricas de Max Weber**. *In*: ENCONTRO DE ADMINISTRAÇÃO DA INFORMAÇÃO, 3., Porto Alegre: EnADI 2011.

SUN, Y.; LU, T.; GU, N. **A method of electronic health data quality assessment**: enabling data provenance. *In*: IEEE INTERNATIONAL CONFERENCE ON COMPUTER SUPPORTED COOPERATIVE WORK IN DESIGN (CSCWD), 21., p. 1-13, abr. 2017. DOI <http://dx.doi.org/10.1109/cscwd.2017.8066700>.

TAN, W. C. Provenance in databases: past, current and future. **IEEE Data Eng. Bull**, v. 30, n. 4, p. 3-12, 2007. Disponível em: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.89.3653&rep=rep1&type=pdf>. Acesso em: 24 maio 2018.

TELEMEDICINA – SANTA CATARINA. STT/SC/SC - Sistema Integrado Catarinense de Telemedicina e Telessaúde. **Teledermatologia em Florianópolis completa um ano de implantação com resultados positivos**. 2016. Disponível em: <http://site.Telemedicina.ufsc.br/teledermatologia-em-florianopolis-completa-um-ano-de-implantacao-com-resultados-positivos/>. Acesso em: 22 nov. 2019.

TELEMEDICINA – SANTA CATARINA. STT/SC/SC - Sistema Integrado Catarinense de Telemedicina e Telessaúde. **Início Telemedicina**. 2020. Disponível em: <http://site.Telemedicina.ufsc.br/>. Acesso em: 22 out. 2020.

TELESSAÚDE. Ministério da Saúde. Programa Telessaúde Brasil Redes. **Saúde Digital e Telessaúde**, 2019. Disponível em: <http://www.saude.gov.br/telessaude>. Acesso em: 22 nov. 2019.

TOBAR, F.; ROMANO YALOUR, M. **Como fazer teses em saúde pública**. Rio de Janeiro: Fiocruz, 2001. ISBN 8575410040.

UNIVERSIDADE FEDERAL DE SANTA CATARINA (UFSC). Programa de Pós-Graduação em Ciência da Informação. **Sobre o Programa**. Disponível em: <https://pgcin.ufsc.br/objetivo/>. Acesso em: 15 jun. 2022a.

UNIVERSIDADE FEDERAL DE SANTA CATARINA (UFSC). Programa de Pós-Graduação em Ciência da Informação. **Linhas de Pesquisa: Dados, Inteligência e Tecnologia**. Disponível em: <https://pgcin.ufsc.br/areas-de-pesquisa/>. Acesso em: 15 jun. 2022b.

VERGARA, S. C. **Projetos e relatórios de pesquisa em administração**. São Paulo: Atlas, 1997.

VISUAL STUDIO CODE. **Getting Started**, 2022. Disponível em: <https://code.visualstudio.com/docs>. Acesso em: 07 nov. 2021.

VUCKOVIC, I. *et al.* The principles of telemedicine in practice. **Bosn J Basic Med Sci.**, v. 3, n. 4, p. 54-60, 2003. Disponível em: <https://www.ncbi.nlm.nih.gov/pubmed/16232139>. Acesso em: 21 nov. 2019.

W3C. PROV-DM: **The PROV Data Model**, 2013. Disponível em: <http://www.w3.org/TR/prov-dm/>. Acesso em: 02 jun. 2018.

WALLAUER, J. *et al.* Building a national telemedicine network. **IT Professional**, v. 10, n. 2, p. 12-17, mar./apr. 2008. ISSN 1520-9202.

WANG, Y.; HU, X. Fuzzy reasoning of accident provenance in pervasive healthcare monitoring systems. **IEEE Journal of Biomedical And Health InformaTIC**, v. 17, n. 6, p. 1015-1022, nov. 2013. Institute of Electrical and Electronics Engineers (IEEE). DOI <http://dx.doi.org/10.1109/jbhi.2013.2274518>.

WANGENHEIM, A. V. *et al.* Ways to implement large scale telemedicine: the Santa Catarina experience. **Latin-American Journal of Telehealth**, v. 3, p. 364-376, 2009.

WERDER, Karl; RAMESH, Balasubramaniam; ZHANG, Rongen. Establishing Data Provenance for Responsible Artificial Intelligence Systems. **ACM Transactions on Management Information Systems (TMIS)**, v. 13, n. 2, p. 1-23, 2022.

WHITTEN, P; COLLINS, B. The diffusion of telemedicine. **Sci Commun**, v. 19, p. 21-40, 1997. Disponível em: <https://journals.sagepub.com/doi/abs/10.1177/1075547097019001002>. Acesso em: 21 nov. 2019.

WHO-GLOBAL OBSERVATORY FOR eHEALTH. (2010). Telemedicine: Opportunities and developments in Member States. **Observatory**, v. 2, p. 96. DOI 10.4258/hir.2012.18.2.153. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3402558/>. Acesso em: 21 nov. 2019.

WIGGINS, G.; MCTIGHE, J. **Understanding by Design**. 2. ed. ASCD: Association for Supervision and Curriculum Development, 2005. Disponível em: <http://www.vanharen.net/blog/van-haren-publishing/best-practice-model-framework-method-guidance-standard-towards-consistent-use-terminology/>. Acesso em: 21 nov. 2019.

WILCOX, A. B. *et al.* Use and impact of a computer-generated patient summary worksheet for primary care. In: AMIA ANNU SYMP, 824-828, 2005. **Proceedings** [...]. Disponível em: <https://www.ncbi.nlm.nih.gov/pubmed/16779155>. Acesso em: 10 nov. 2019.

WILSON, L. S; MAEDER, A. J. Recent directions in telemedicine: review of trends in research and practice. **Healthcare informATIC research**, v. 21, n. 4, p. 213-222, 2015.

WOHLIN, C. **Guidelines for snowballing in systematic literature studies and a replication in software engineering**. In: INTERNATIONAL CONFERENCE ON EVALUATION AND ASSESSMENT IN SOFTWARE ENGINEERING (EASE'14), 18., article n. 38, May 2014. DOI <https://doi.org/10.1145/2601248.2601268>.

WOOTTON, R; MENZIES, J; FERGUSON, P. Follow-up data for patients managed by store and forward telemedicine in developing countries. **Journal of Telemedicine and Telecare**, 2009, v. 15, n. 2, p. 83-88. Disponível em: <https://www.ncbi.nlm.nih.gov/pubmed/19246608>. Acesso em: 21 nov. 2019.

WOOTTON, R. Recent advances: Telemedicine. **BMJ**, v. 323, n. 7312, p. 557-60, 2001. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1121135/>. Acesso em: 21 nov. 2019.

WORLD HEALTH ORGANIZATION (WHO). **Developing Health Management Information Systems: a practical for developing countries**. Manila: Regional Office for the Western Pacific, 2004. Disponível em: <https://iris.wpro.who.int/handle/10665.1/5498>. Acesso em: 20 nov. 2019.

WORLD HEALTH ORGANIZATION (WHO). **Framework and standards for country health information systems**. 2. ed. Geneva: WHO, 2008. ISBN 978 92 4 159594 0. Disponível em: https://www.who.int/healthinfo/country_monitoring_evaluation/who-hmn-framework-standards-chi.pdf. Acesso em: 15 out. 2019.

W3SCHOOLS.COM (2022). **Introduction to XML**. Disponível em:

https://www.w3schools.com/xml/xml_what_is.asp. Acesso em: 28 jan. 2022.

YIN, R. K. **Estudo de caso: planejamento e métodos**. 2. ed. Porto Alegre: Bookman. 2001. ISBN 978-8582602317.

ZANINI, M. **Formulário eletrônicos**. 2007. 21 p. Universidade Federal de Santa Catarina, Florianópolis. Disponível em: https://projetos.inf.ufsc.br/arquivos_projetos/projeto_698/tcc-slides.pdf. Acesso em: 6 abr. 2017.

APÊNDICE A – FOLHA DE ROSTO PARA PESQUISA ENVOLVENDO SERES HUMANOS



MINISTÉRIO DA SAÚDE - Conselho Nacional de Saúde - Comissão Nacional de Ética em Pesquisa – CONEP

FOLHA DE ROSTO PARA PESQUISA ENVOLVENDO SERES HUMANOS

| | | | |
|--|-----------------------------|--|--------------------------------------|
| 1. Projeto de Pesquisa: Proposta de um método para gerenciamento dos dados de proveniência em sistemas de informação em saúde | | | |
| 2. Número de Participantes da Pesquisa: 5 | | | |
| 3. Área Temática: | | | |
| 4. Área do Conhecimento: Grande Área 6. Ciências Sociais Aplicadas | | | |
| PESQUISADOR RESPONSÁVEL | | | |
| 5. Nome: DOUGLAS DYLLON JERONIMO DE MACEDO | | | |
| 6. CPF: 633.987.372-34 | | 7. Endereço (Rua, n.º): LAURO LINHARES 561/1017 TRINIDADE BL 4 AP 308 FLORIANOPOLIS SANTA CATARINA 88036001 | |
| 8. Nacionalidade: BRASILEIRO | 9. Telefone: 48999997669 | 10. Outro Telefone: | 11. Email: douglas.macedo@ufsc.br |
| <p>Termo de Compromisso: Declaro que conheço e cumprirei os requisitos da Resolução CNS 466/12 e suas complementares. Comprometo-me a utilizar os materiais e dados coletados exclusivamente para os fins previstos no protocolo e a publicar os resultados sejam eles favoráveis ou não. Aceito as responsabilidades pela condução científica do projeto acima. Tenho ciência que essa folha será anexada ao projeto devidamente assinada por todos os responsáveis e fará parte integrante da documentação do mesmo.</p> | | | |
| Data: <u>29</u> / <u>09</u> / <u>2021</u> | |  <p>Documento assinado digitalmente Douglas Dyllon Jeronimo de Macedo Data: 29/09/2021 09:15:16-0300 CPF: 633.987.372-34 Verifique as assinaturas em https://v.ufsc.br</p> | |
| Assinatura | | | |
| INSTITUIÇÃO PROPONENTE | | | |
| 12. Nome: Universidade Federal de Santa Catarina | | 13. CNPJ: 83.899.526/0001-82 | 14. Unidade/Órgão: |
| 15. Telefone: (48) 3721-9206 | 16. Outro Telefone: | | |
| <p>Termo de Compromisso (do responsável pela instituição): Declaro que conheço e cumprirei os requisitos da Resolução CNS 466/12 e suas Complementares e como esta instituição tem condições para o desenvolvimento deste projeto, autorizo sua execução.</p> | | | |
| Responsável: <u>Edgar Bisset Alvarez</u> | | CPF: <u>236.516.228.27</u> | |
| Cargo/Função: <u>Coordenador (PGCIN/UFSC)</u> | |  <p>Documento assinado digitalmente Edgar Bisset Alvarez Data: 29/09/2021 19:49:49-0300 CPF: 236.516.228-27 Verifique as assinaturas em https://v.ufsc.br</p> | |
| Data: <u>29</u> / <u>09</u> / <u>2021</u> | | Assinatura | |
| PATROCINADOR PRINCIPAL | | | |
| Não se aplica. | | | |

APÊNDICE B – PROJETO DE PESQUISA ENVOLVENDO SERES HUMANOS



MINISTÉRIO DA SAÚDE - Conselho Nacional de Saúde - Comissão Nacional de Ética em Pesquisa – CONEP
PROJETO DE PESQUISA ENVOLVENDO SERES HUMANOS

Projeto de Pesquisa:
 Proposta de um método para gerenciamento dos dados de proveniência em sistemas de informação em saúde

Informações Preliminares

Responsável Principal

| | |
|-------------------------------|---|
| CPF/Documento: 633.987.372-34 | Nome: DOUGLAS DYLLON JERONIMO DE MACEDO |
| Telefone: 4899997669 | E-mail: douglas.macedo@ufsc.br |

Instituição Proponente

| | |
|--------------------------|---|
| CNPJ: 83.899.526/0001-82 | Nome da Instituição: Universidade Federal de Santa Catarina |
|--------------------------|---|

É um estudo internacional? Não

Assistentes

| CPF/Documento | Nome |
|----------------|--------------------|
| 026.086.439-06 | Márcio José Sembay |

Área de Estudo

Grandes Áreas do Conhecimento (CNPq)

- Grande Área 6. Ciências Sociais Aplicadas

Título Público da Pesquisa: Proposta de um método para gerenciamento dos dados de proveniência em sistemas de informação em saúde

Contato Público

| CPF/Documento | Nome | Telefone | E-mail |
|----------------|--------------------|-------------|-------------------------------|
| 026.086.439-06 | Márcio José Sembay | 49999449101 | marcio.sembay@posgrad.ufsc.br |

Contato Científico: DOUGLAS DYLLON JERONIMO DE MACEDO

Desenho de Estudo / Apoio Financeiro

Desenho:

O universo de pesquisa se volta ao Sistema Catarinense de Telemedicina e Telessaúde (STT) da Rede Catarinense de Telemedicina (RCTM) em atividade nas dependências do Centro de Ciências da Saúde da Universidade Federal de Santa Catarina (UFSC). Desta forma, a população selecionada, tem duas vertentes: computacional e humano. No quesito computacional, serão os dados digitais e manuais inseridos nos Sistemas de Informação em Saúde (SIS) integrados no STT/RCTM. Já no quesito humano, serão os profissionais/especialistas que atuam no gerenciamento dos dados na mesma rede. Em relação a amostra, também seguem os mesmos quesitos da população. Nesse sentido, a amostra no quesito computacional será com base no repositório de proveniência local (banco de dados local que receberá dados do STT/RCTM por meio de estratégias computacionais a serem criadas dentro da mesma rede sem acesso externo ou por pessoas não autorizadas). Já no que compete a amostra no quesito humano, serão selecionados 5 profissionais/especialistas que atuam no STT/RCTM, os quais responderão um formulário eletrônico (entrevista semiestruturada) gerada a partir do Google Forms para validação do método proposto no projeto de pesquisa. As tabelas, gráficos entre outras análises se aterão ao repositório de proveniência local, sendo essas análises apresentadas para os gestores e profissionais/especialistas autorizados.

Apoio Financeiro

| CNPJ | Nome | E-mail | Telefone | Tipo |
|------|------|--------|----------|-----------------------|
| | | | | Financiamento Próprio |

Palavra Chave

| Palavra-chave |
|---------------------------------|
| Proveniência de Dados |
| Dados de Proveniência |
| Sistemas de Informação em Saúde |
| Telemedicina |

Detalhamento do Estudo

Resumo:

Os Sistemas de Informação em Saúde (SIS) são utilizados em diferentes processos de gerenciamento de dados por diversas tecnologias em diferentes especialidades médicas. A telemedicina utiliza diferentes tipos de SIS para o processo de gerenciamento de dados de saúde. Como qualquer outro cenário de saúde, a telemedicina gera grandes volumes de dados que precisam ser gerenciados. Com o objetivo de propor um método de gerenciamento de dados de proveniência em SIS, optou-se por realizar o estudo de caso especificamente no Sistema Catarinense de Telemedicina e Telessaúde (STT) da Rede Catarinense de Telemedicina (RCTM) em atividade nas dependências do Centro de Ciências da Saúde (CCS) da Universidade Federal de Santa Catarina (UFSC). Serão utilizadas estratégias computacionais para propor um método com técnicas de proveniência de dados associadas a normas e padrões internacionais para a representação, formatação, transferência e interoperabilidade de dados na realização do processo de captura dos dados na rede em estudo. Os dados de proveniência serão capturados por meio da integração do método aqui proposto dentro de uma arquitetura restrita de acesso e com autorização da instituição responsável, sem expor dados fora do âmbito da rede. Logo após, será enviado um formulário eletrônico por e-mail para os especialistas/profissionais que desempenham seu trabalho de gestão dos dados no STT/RCTM para validar a eficiência do método aqui proposto e aprimorá-lo em uma versão concreta e ideal para aplicação em diferentes SIS.

Introdução:

Os Sistemas de Informação (SI) e as Tecnologias de Informação e Comunicação (TIC) nos proporcionaram uma nova forma de comunicação, como viver no espaço cibemético, mensagens instantâneas, comunicar-se com alguém em qualquer lugar, entre outras formas de comunicação que estão mudando não apenas o estilo de vida, mas também o modo como acontecem os negócios em todos os setores (WHO-GLOBAL OBSERVATORY FOR eHEALTH, 2010).

Nesse sentido, realçando a importância das TIC, entende-se que essas tecnologias proporcionam a construção de um dos principais cenários tecnológicos voltados à saúde humana, como é o caso dos Sistemas de Informação em Saúde (SIS), 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 SIS integrados é a telemedicina, que se sustenta em uma gama de equipamentos médicos e sistemas integrados possibilitando gerar diagnósticos e tratamentos independentemente da distância geográfica, gerando benefícios para os pacientes, como, por exemplo, viagens reduzidas ou acesso mais rápido aos conhecimentos adequados (HÅKANSSON; GAVELIN, 2000; WHITTEN; COLLINS, 1997).

Nessa perspectiva, no estado de Santa Catarina, a principal iniciativa é o Sistema Catarinense de Telemedicina e Telessaúde da Rede Catarinense de Telemedicina (STT/RCTM), situado em Florianópolis, capital do estado. Essa infraestrutura foi aplicada inicialmente em 1999, como um projeto de aplicação de algumas tecnologias que estavam naquele momento sendo desenvolvidas pelo Cyclops Group da Universidade Federal de Santa Catarina (UFSC) (GROUP, 2013). Essas tecnologias estavam sendo desenvolvidas em parceria com instituições da Alemanha e ainda contavam com financiamento do CNPq e da Fundação Volkswagen (Wolfsburg/German) (MAIA; WANGENHEIM; NOBRE, 2006; WALLAUER et al., 2008; WANGENHEIM et al., 2009; MACEDO, 2014).

Dessa forma, em consonância com este projeto de pesquisa, o uso de técnicas de proveniência de dados como instrumento tecnológico pode contribuir na gestão dos dados de proveniência em SIS. Entretanto, o termo "proveniência" de dados diz respeito à origem ou procedência dos dados, e pode também estar relacionado à auditoria, triagem, linhagem e origem dos dados (DAVIDSON; FREIRE, 2008).

Nesse sentido, questiona-se: como desenvolver um método que contribua para o gerenciamento dos dados de proveniência em SIS?

Com base nesse questionamento, as justificativas desse projeto de pesquisa são: social, científica e tecnológica.

A justificativa social para o desenvolvimento deste projeto de pesquisa se dá pela importância do tema citado para as mais variadas áreas, inclusive a Medicina. O Conselho Federal de Medicina (CFM), por meio da Resolução nº 1.821/2007, autoriza as instituições de saúde a utilizarem os SIS para realizar os processos informacionais necessários de armazenamento e manipulação de dados de pacientes no âmbito da saúde, disponibilizando esses dados quando solicitados pelo paciente. Nesse sentido, as autoridades de saúde procuram por dados que as guiam para novas tomadas de decisão, no intuito de melhorar a saúde da população na tentativa de criar SIS que deem conta do que entendem como especificidades em âmbitos políticos, administrativos, técnicos ou sanitários (BRANCO,

Data de Submissão do Projeto: 10/10/2021

Nome do Arquivo: PB_INFORMAÇÕES_BÁSICAS_DO_PROJETO_1831892.pdf

Versão do Projeto: 1

1995).

Já em relação à justificativa científica, com foco voltado para os SIS, a proveniência de dados contribui para a interpretação e o entendimento de resultados, e, em alguns casos, para reproduzir o resultado (FREIRE et al., 2008). Esses processos aplicados a SIS são primordiais para as tomadas de decisões e a geração de conhecimento nesses cenários. Logo, este projeto de pesquisa pode proporcionar discussões sobre a aplicação da proveniência de dados em SIS, as quais poderão ser ampliadas em outros contextos de saúde.

No que compete à justificativa tecnológica, ressalta-se que a ciência se apoia cada vez mais nos dados. Novas tecnologias aumentaram a eficiência na coleta de dados e, conseqüentemente, a quantidade dos dados gerados. As tecnologias também permitiram que os dados pudessem ser processados, analisados e armazenados em infraestruturas computacionais. Esse cenário gerou uma mudança nos desafios científicos. O antigo problema da escassez dos dados foi substituído pela dificuldade em gerenciar o seu excesso, sua variedade e distribuição (HEY; TANSLEY; TOLLE, 2009). Em relação às tecnologias e ferramentas tecnológicas a serem utilizadas neste projeto de pesquisa, vão seguir os mesmos padrões estabelecidos no local de coleta de dados, sem desconsiderar a segurança já existente no cenário em estudo.

Desta forma, a relevância do tema aqui proposto é vital para o bom funcionamento dos processos tecnológicos que envolvem a saúde no Brasil e que utilizam SIS para o gerenciamento de seus dados. O crescimento exponencial de armazenamento dos dados de saúde no mundo é um problema crítico quando se depende do rastreamento da origem dos dados.

Portanto, a aplicação do método proposto neste projeto de pesquisa pode contribuir para garantir o gerenciamento dos dados de proveniência em SIS que necessitam rastrear diversos dados de forma descentralizada, contribuindo para a tomadas de decisão e geração de conhecimento no cenário de saúde em estudo, oportunizando novas pesquisas.

Hipótese:

Esta pesquisa busca subsídios dentro do contexto da saúde, mais especificamente em diferentes SIS integrados a uma rede de telemedicina, para observar se o método de gerenciamento de dados de proveniência aqui proposto possibilitará, de modo eficaz, o aprimoramento dos serviços prestados e fortalecimento das tomadas de decisão necessárias.

Objetivo Primário:

Propor um método para contribuir para o gerenciamento dos dados de proveniência em SIS no STT/RCTM.

Objetivo Secundário:

a) Conhecer o fluxo dos dados de proveniência gerados pelos SIS integrados do STT/RCTM; b) Constatar inexistências de gerenciamento de dados de proveniência nos SIS integrados do STT/RCTM; c) Aplicar o método de gerenciamento dos dados de proveniência, tanto em dados digitais quanto em dados inseridos manualmente nos SIS integrados do STT/RCTM; d) Validar o método proposto com os profissionais/especialistas que atuam no STT/RCTM.

Metodologia Proposta:

Em relação à sua natureza, este projeto de pesquisa classifica-se como aplicada. De acordo com Gerhardt e Silveira (2009, p. 35), "este tipo de pesquisa tem como objetivo gerar conhecimentos para aplicação prática, dirigidos à solução de problemas específicos". O teor de aplicação deste projeto de pesquisa se refere a proposta de um método para o gerenciamento de dados de proveniência em SIS no STT/RCTM. Do ponto de vista dos procedimentos técnicos, este projeto de pesquisa caracteriza-se como experimental, sendo observadas as variáveis em um estudo de caso a ser realizado no STT/RCTM. Segundo Gil (2007), a pesquisa experimental consiste em determinar um objeto de estudo, selecionar as variáveis que seriam capazes de influenciá-lo, definir as formas de controle e de observação dos efeitos que a variável produz no objeto. Do ponto de vista em relação a abordagem desse projeto de pesquisa, classifica-se este estudo como quantitativo, que segundo Knechtel (2014) atua sobre um problema humano ou social, que neste caso está em relação a quantidade de dados a serem capturados pelo método e qualitativo que segundo Reneker (1993) o pesquisador desenvolve conceitos, ideias e entendimentos a partir de padrões encontrados nos dados, ao invés de coletar dados para comprovar teorias, hipóteses e modelos preconcebidos. Para Frew (2005), o próprio uso de técnicas de proveniência de dados qualifica os dados, sendo um dos benefícios da proveniência. Optou-se pela realização de um estudo de caso neste projeto de pesquisa, pois permitirá a aplicação do método em um cenário de saúde real com dados reais, posteriormente sendo possível validar o método juntamente com os profissionais/especialistas que atuam no STT/RCTM. O estudo de caso é o método que visa compreender fenômenos sociais complexos, preservando as características holísticas e significativas dos eventos da vida real (YIN, 2001). Em relação ao universo, ou população, segundo Vergara (1997) é o conjunto de elementos que possuem as características que serão objeto do estudo, e a amostra, ou população amostral, é uma parte do universo escolhido, selecionada a partir de um critério de representatividade. Nesse sentido, o universo de pesquisa se volta ao STT/RCTM no qual se pretende aplicar o método. Desta forma, a população selecionada, segue duas vertentes: computacional e humano. No quesito computacional, serão os dados digitais e manuais inseridos nos SIS integrados no STT/RCTM. Já no quesito humano, serão os profissionais/especialistas que atuam no gerenciamento dos dados na mesma rede. Em relação a amostra, também seguem os mesmos quesitos da população. Nesse sentido, a amostra no quesito computacional será com base no repositório de proveniência local (banco de dados local que receberá dados do STT/RCTM por meio de estratégias computacionais a ser criado dentro da mesma rede sem acesso externo ou por pessoas não autorizadas). Já no que compete a amostra no quesito humano, serão selecionados 5 profissionais/especialistas que atuam no STT/RCTM que responderão um formulário eletrônico (entrevista semiestruturada) gerada a partir do Google Forms para validação do método proposto neste projeto de pesquisa.

Riscos:

Os potenciais riscos desse projeto de pesquisa estão relacionados em dois momentos. No primeiro momento, está relacionado a aplicação do método, onde podem surgir alguns problemas de acesso aos dados de proveniência do STT/RCTM por algum motivo técnico, fazendo com que o profissional responsável pelo acesso aos dados possa despendar tempo para ajustar o controle de acesso. Já no segundo momento, após aplicar o método no STT/RCTM, está relacionado aos 5 participantes da pesquisa que irão validar o método por meio do formulário eletrônico, onde os riscos em participar da pesquisa podem estar relacionados com desconforto, perda de tempo, constrangimento, receio, intimidação, angústia, fadiga, insatisfação, irritação ou algum mal estar frente aos questionamentos decorrentes da pesquisa.

Benefícios:

Em relação aos potenciais benefícios desse projeto de pesquisa, está relacionado em um primeiro momento na aplicação do método como contribuição para o gerenciamento dos dados de proveniência no STT/RCTM para possíveis auditorias no intuito de tomadas de decisões mais precisas. No segundo momento, realizar o diagnóstico das contribuições da aplicação do método aqui proposto juntamente com 5 profissionais/especialistas da equipe técnica atuantes no STT/RCT, servirá como importante fator de contribuição

para melhorarias nos processos de rastreamento e gerenciamento dos dados de proveniência dentro da rede em estudo. Benefícios indiretos sempre serão gerados pela produção de conhecimento associada ao desenvolvimento dessa pesquisa, oportunizando a novos pesquisadores aprimorarem o método aqui em questão em novas vertentes de pesquisa no âmbito do STT/RCTM e futuramente em outros cenários de saúde.

Metodologia de Análise de Dados:

Em relação a coleta de dados para realizar as análises, a mesma segue duas formas: a primeira forma de coleta de dados se dará pelo método aqui proposto de forma prática e aplicada, por meio de intervenção de softwares que, conforme explica Santos (2004, p. 28), "são os métodos práticos utilizados para juntar as informações necessárias à construção dos raciocínios em torno de um fato/fenômeno/problema". O método a ser aplicado terá premissas em sua estrutura para analisar quais dados serão analisados e posteriormente capturados. Os dados coletados por intermédio do método será nos bancos de dados dos SIS integrados no STT/RCTM, onde serão armazenados em um repositório de dados proveniência local já mencionado anteriormente neste projeto de pesquisa. Dessa forma, esses dados posteriormente serão 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 de saúde em estudo. Internamente, sem acesso externo, os dados serão utilizados pelos gestores e profissionais autorizados no STT/RCTM para controle, monitoramento e rastreamento das informações técnicas especializadas para contribuir em tomadas de decisão e possíveis auditorias. Porém, é de suma importância ressaltar, que os dados pessoais e dados sensíveis serão anonimizados conforme Lei Geral de Proteção de Dados Pessoais (BRASIL, 2018). Esse processo se dará com auxílio de ferramentas de software e posteriormente esses dados serão apagados do repositório de dados de proveniência local, pois o intuito dessa coleta e armazenamento é a aplicação e validação do método juntamente com equipe técnica. Já a segunda forma de coleta de dados, será por meio de formulário eletrônico gerado no Google Forms enviado por e-mail para 5 profissionais/especialistas que atuam no STT/RCTM que experimentarão e validarão o método. Para Zanini (2007), existem várias vantagens associadas à utilização do formulário eletrônico quando comparado ao formulário convencional (com utilização de papel), entre elas, a facilidade na busca de dados, a utilização de armazenamento físico diminuído e distribuição fácil e rápida. Nesse aspecto, uma alternativa disponível para criação de formulários eletrônicos on-line é a plataforma Google Forms, uma ferramenta que oferece suporte para a criação de formulários personalizados de forma simples (GOOGLE, 2020). Em relação ao formulário eletrônico, são apresentadas questões de cunho técnico, operacional e gerencial, referente a experimentação do método para a sua possível validação pelos 5 profissionais/especialistas atuantes no STT/RCTM. Após as aplicações dos formulários eletrônicos para descobrir a eficiência do método, as respostas serão transcritas e tabuladas para análises detalhadas pelos gestores do STT/RCTM, para então aprimorá-lo e desenhá-lo em uma versão ideal para aplicação em diversos cenários da saúde.

Desfecho Primário:

Diagnóstico da aplicação do método, o qual será encaminhado juntamente com os resultados e benefícios encontrados para os gestores do STT/RCTM para apreciação.

Tamanho da Amostra no Brasil: 5

Países de Recrutamento

| País de Origem do Estudo | País | Nº de participantes da pesquisa |
|--------------------------|--------|---------------------------------|
| Sim | BRASIL | 5 |

Outras Informações

Haverá uso de fontes secundárias de dados (prontuários, dados demográficos, etc)?

Não

Informe o número de indivíduos abordados pessoalmente, recrutados, ou que sofrerão algum tipo de intervenção neste centro de pesquisa:

5

Grupos em que serão divididos os participantes da pesquisa neste centro

| ID Grupo | Nº de Indivíduos | Intervenções a serem realizadas |
|---|------------------|--------------------------------------|
| Profissionais/especialistas que atuam no STT/RCTM | 5 | Entrevista via formulário eletrônico |

O Estudo é Multicêntrico no Brasil?

Não

Propõe dispensa do TCLE?

Não

Haverá retenção de amostras para armazenamento em banco?

Sim

Justificativa:

Serão armazenados os dados capturados pelo método em um banco de dados local (repositório de proveniência local, inserido internamente no STT/RCTM) sem acesso externo ou profissionais não autorizados. Importante ressaltar, que esse banco de dados local será apenas utilizado para aplicação e validação do método, sendo que após esse processo a equipe técnica poderá excluir os dados ou mantê-los para pesquisas internas.

Cronograma de Execução

| Identificação da Etapa | Início (DD/MM/AAAA) | Término (DD/MM/AAAA) |
|--|---------------------|----------------------|
| Acesso às fontes de dados para captura e | 20/11/2021 | 20/12/2021 |

Data de Submissão do Projeto: 10/10/2021

Nome do Arquivo: PB_INFORMAÇÕES_BÁSICAS_DO_PROJETO_1831892.pdf

Versão do Projeto: 1

armazenamento dos dados de proveniência

| | | |
|---|------------|------------|
| Transcrição das análises dos dados de proveniência e das respostas do formulário eletrônico | 01/02/2022 | 28/02/2022 |
| Confecção do documento final para ser inserido na tese | 01/03/2022 | 20/03/2022 |
| Preparação e manipulação dos dados de proveniência para geração de análises | 20/12/2021 | 20/01/2022 |
| Envio de formulário eletrônico para realizar as entrevistas | 20/01/2022 | 31/01/2022 |

Orçamento Financeiro

| Identificação de Orçamento | Tipo | Valor em Reais (R\$) |
|----------------------------|---------|----------------------|
| Internet | Custeio | R\$ 100,00 |
| Energia Elétrica | Custeio | R\$ 20,00 |
| Total em R\$ | | R\$ 120,00 |

Outras informações, justificativas ou considerações a critério do pesquisador:

O pesquisador terá o custo apenas da energia elétrica e internet. Os dados serão capturados por acesso remoto ao STT/RCTM e a entrevista com os profissionais/especialistas será enviada via e-mail em formato de formulário eletrônico, ambos os processos não geram gastos.

Bibliografia:

BOSE, R.; FREW, J. Lineage retrieval for scientific data processing: a survey. *ACM Comput. Surv.*, v. 37, n. 1, p. 1-28, mar. 2005. ISSN 0360-0300. Disponível em: <http://citeseerx.ist.psu.edu/viewdoc/download?sessionid=C1746A7A546FD1B64E83540C6ED55688?doi=10.1.1.103.4855&rep=rep1&type=pdf>. Acesso em: 29 jan. 2019. BRANCO, M. A. F. Sistema de Informação em Saúde em Âmbito Local e Organização de Interesses Sociais: um estudo de caso no Município do Rio de Janeiro. 1995. 129f. Dissertação (Mestrado) Instituto de Medicina Social, Universidade do Estado do Rio de Janeiro, 1995. 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. Acesso em: 10 jun. 2021. Disponível em: <http://bit.ly/1mTMIS3>. Acesso em: 29 jan. 2019. 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. Acesso em: 12 jun. 2021. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/L13709.htm. Acesso em: 29 jan. 2019. BUNEMAN, P.; KHANNA, S. E.; CHIEW, W. Why and where: a characterization of data provenance. In: VAN DEN BUSSCHE, J., VIANU, V. (Ed.). Database theory. ICDT 2001. Lecture Notes in Computer Science, v. 1973. Berlin, Heidelberg: Springer, 2001. Disponível em: <https://pdfs.semanticscholar.org/b647/2ee11749ef70713bfb0e322a9ec27523ed88.pdf>. Acesso em: 02 maio 2019. CRAIG, J.; PATTERSON, V. Introduction to the practice of telemedicine. *Journal of Telemedicine and Telecare*, v. 11, n. 1, p. 3-9, 2005. DOI 10.1258/1357633053430494. Disponível em: <https://europepmc.org/abstract/med/15829036>. Acesso em: 21 nov. 2019. FREIRE, J. et al. Provenance for computational tasks: a survey. *Journal Computing in Science and Engineering*, v. 10, n. 3, p. 11-21, 2008. ISSN 15219615. Disponível em: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.147.3801&rep=rep1&type=pdf>. Acesso em: 24 maio 2018. GERHARDT, T. E.; SILVEIRA, D.; T. Métodos de pesquisa. Porto Alegre: Ed. da UFRGS, 2009. GIL, A. C. Como elaborar projetos de pesquisa. 4. ed. São Paulo: Atlas, 2007. GOOGLE. Formulários. 2020. Disponível em: <https://workspace.google.com/intl/pt-BR/products/forms/>. Acesso em: 08 nov. 2020. GROTH, P.; MOREAU, L. PROV-Overview: an overview of the prov family of documents. 2013. Disponível em: <https://www.w3.org/TR/prov-overview/>. Acesso em: 20 jun. 2021. GROUP, C. Cyclops Group. Disponível em: <http://www.cyclops.ufsc.br>. Acesso em: 29 nov. 2019. HAKANSSON, S.; GAVELIN, C. What do we really know about the cost-effectiveness of telemedicine? *J Telemed Telecare*, v. 6, suppl. 1, p. 133-136, 2000. Disponível em: <https://www.ncbi.nlm.nih.gov/pubmed/10793998>. Acesso em: 21 nov. 2019. HAUX, R. Health information systems-past, present, future. *International Journal of Medical Informatics*, v. 75, p. 268-281, 2006. HEY, T.; TANSLEY, S.; TOLLE, K. Fourth paradigm: data-intensive scientific discovery. EUA: Microsoft Corporation, 2009. ISBN 978-0-9825442-0-4. Disponível em: <http://research.microsoft.com/en-us/collaboration/fourthparadigm/>. Acesso em: 14 abr. 2015. KIM, Y. S. Telemedicine in the USA with focus on clinical applications and issues. *Yonsei Med J.*, v. 45, n. 5, p. 761-75, 2004. Disponível em: <https://www.ncbi.nlm.nih.gov/pubmed/15515185>. Acesso em: 21 nov. 2019. KNECHTEL, M. R. Metodologia da pesquisa em educação: uma abordagem teórico-prática dialogada. Curitiba: Intersaberes, 2014. MACEDO, D. D. J. Um modelo distribuído de armazenamento hierárquico de conhecimento médico. 2014. 214 f. Tese (Doutorado) Curso de Programa de Pós-graduação em Engenharia e Gestão do Conhecimento, Departamento de Engenharia do Conhecimento, Universidade Federal de Santa Catarina - UFSC, Florianópolis, 2014. Disponível em: <https://repositorio.ufsc.br/xmlui/bitstream/handle/123456789/123290/326446.pdf?sequence=1&isAllowed=y>. Acesso em: 10 nov. 2019. MAIA, R.; WANGENHEIM, A. V.; NOBRE, L. A statewide telemedicine network for public health in Brazil. In: IEEE INTERNATIONAL SYMPOSIUM ON COMPUTER-BASED MEDICAL SYSTEMS, CBMS 2006, 19., [S.l.: s.n.], 2006. p. 495-500. MOREAU, L. Usage of provenance: a tour of babel. In: INTERNATIONAL PROVENANCE AND ANNOTATION WORKSHOP (IPAW'06), Chicago, Illinois, USA, 3-5, May 2006. Proceedings [...]. Disponível em: <https://eprints.soton.ac.uk/409382/1/babel.pdf>. Acesso em: 27 jul. 2018. OXFORD ENGLISH DICTIONARY (OED) Online. datum, n., 2015. Oxford University Press. Disponível em: <http://www.oed.com/view/Entry/47434>. Acesso em: 15 jun. 2019. RAHMAN, M. A. et al. Secure and Provenance Enhanced Internet of Health Things Framework: A Blockchain Managed Federated Learning Approach. *IEEE*, v. 8, p. 205071-205087, nov. 2020. <http://10.1109/ACCESS.2020.3037474>. Acesso em: 10 jan. 2021. RENEKER, M. 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. ROINE, R.; OHINMAA, A.; HAILEY, D. Assessing telemedicine: a systematic review of the literature. *CMAJ*, v. 165, n. 6, p. 765-71, 2001. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC81454/>. Acesso em: 21 nov. 2019. SANTOS, A. R. dos. Metodologia científica: a construção do conhecimento. 6. ed. Rio de Janeiro: DP&A, 2004. SIMMHAN, Y. L.; PLALE, B.; GANNON, D. A survey of data provenance techniques. Technical Report TR-618: Computer Science Department, Indiana University, 2005. Disponível em: <http://ftp.extreme.indiana.edu/pub/techreports/TR618.pdf>. Acesso em: 24 maio 2018. VERGARA, S. C. Projetos e relatórios de pesquisa em administração. São Paulo: Atlas, 1997. VICTORA, C. G. et al. Saúde de mães e crianças no Brasil: progressos e desafios. *The Lancet*, London, p. 32-46, 2011. WALLAUER, J. et al. Building a national telemedicine network. *IT Professional*, v. 10, n. 2, p. 12-17, mar./apr. 2008. ISSN 1520-9202. WANGENHEIM, A. V. et al. Ways to implement large scale telemedicine: the Santa Catarina experience. *Latin-American Journal of Telehealth*, v. 3, p. 364-376, 2009. WHITTEN, P.; COLLINS, B. The diffusion of telemedicine. *Sci Commun.*, v. 19, p. 21-40, 1997. Disponível em: <https://journals.sagepub.com/doi/abs/10.1177/1075547097019001002>. Acesso em: 21 nov. 2019. WHO-GLOBAL OBSERVATORY FOR eHEALTH. (2010). Telemedicine: Opportunities and developments in Member States. *Observatory*, v. 2, p. 96. DOI 10.4258/hir.2012.18.2.153. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3402558/>. Acesso em: 21 nov. 2019. WORLD HEALTH ORGANIZATION (WHO). Developing Health Management Information Systems: a practical for developing countries.

Data de Submissão do Projeto: 10/10/2021

Nome do Arquivo: PB_INFORMAÇÕES_BÁSICAS_DO_PROJETO_1831892.pdf

Versão do Projeto: 1

Manila: Regional Office for the Western Pacific, 2004. Disponível em: <https://iris.wpro.who.int/handle/10665.1/5498>. Acesso em: 20 nov. 2019. YIN, R. K. Estudo de caso: planejamento e métodos. 2. ed. Porto Alegre: Bookman, 2001. ISBN 978-8582602317. ZANINI, M. Formulário eletrônico. 2007. 21 p. Universidade Federal de Santa Catarina, Florianópolis. Disponível em: https://projetos.inf.ufsc.br/arquivos_projetos/projeto_698/tcc-slides.pdf. Acesso em: 6 abr. 2017.

Upload de Documentos

Arquivo Anexos:

| Tipo | Arquivo |
|---|--|
| TCLE / Termos de Assentimento / Justificativa de Ausência | TERMO_DE_CONSENTIMENTO_LIVRE_E_ESCLARECIDO_sem_assinatura_do_pesquisador.pdf |
| Outros | ROTEIRO_PARA_A_ENTREVISTA_SEMIESTRUTURADA.pdf |
| Projeto Detalhado / Brochura Investigador | PROJETO_DE_PESQUISA.pdf |
| Outros | ROTEIRO_PARA_A_ENTREVISTA_SEMIESTRUTURADA.pdf |
| Folha de Rosto | Folha_de_Rosto_Assinada.pdf |
| TCLE / Termos de Assentimento / Justificativa de Ausência | TERMO_DE_CONSENTIMENTO_LIVRE_E_ESCLARECIDO_sem_assinatura.pdf |
| Outros | ROTEIRO_PARA_A_ENTREVISTA_SEMIESTRUTURADA.pdf |
| Declaração de Instituição e Infraestrutura | DECLARACAO_INSTITUICAO_PARA_COLETA_DE_DADOS_assinado.pdf |
| TCLE / Termos de Assentimento / Justificativa de Ausência | TERMO_DE_CONSENTIMENTO_LIVRE_E_ESCLARECIDO_com_assinatura_pesquisador.pdf |
| Outros | TERMO_DE_COMPROMISSO_PARA_UTILIZACAO_DE_DADOS_assinado.pdf |

Finalizar

 Manter sigilo da integra do projeto de pesquisa: Não

Data de Submissão do Projeto: 10/10/2021

Nome do Arquivo: PB_INFORMAÇÕES_BÁSICAS_DO_PROJETO_1831892.pdf

Versão do Projeto: 1

APÊNDICE C – PARECER CONSUBSTANCIADO DO CEP

UNIVERSIDADE FEDERAL DE
SANTA CATARINA - UFSC



PARECER CONSUBSTANCIADO DO CEP

DADOS DO PROJETO DE PESQUISA

Título da Pesquisa: Proposta de um método para gerenciamento dos dados de proveniência em sistemas de informação em saúde

Pesquisador: DOUGLAS DYLLON JERONIMO DE MACEDO

Área Temática:

Versão: 1

CAAE: 52672021.4.0000.0121

Instituição Proponente: Universidade Federal de Santa Catarina

Patrocinador Principal: Financiamento Próprio

DADOS DO PARECER

Número do Parecer: 5.063.111

Apresentação do Projeto:

Doutorado de Márcio José Sembay no PPG em Ciência da Informação da UFSC, sob orientação de Douglas D. J. de Macedo.

Segundo os pesquisadores, no formulário da Plataforma Brasil:

Desenho:

O universo de pesquisa se volta ao Sistema Catarinense de Telemedicina e Telessaúde (STT) da Rede Catarinense de Telemedicina (RCTM) em atividade nas dependências do Centro de Ciências da Saúde da Universidade Federal de Santa Catarina (UFSC). Desta forma, a população selecionada, tem duas vertentes: computacional e humano. No quesito computacional, serão os dados digitais e manuais inseridos nos Sistemas de Informação em Saúde (SIS) integrados no STT/RCTM. Já no quesito humano, serão os profissionais/especialistas que atuam no gerenciamento dos dados na mesma rede. Em relação a amostra, também seguem os mesmos quesitos da população. Nesse sentido, a amostra no quesito computacional será com base no repositório de proveniência local (banco de dados local que receberá dados do STT/RCTM por meio de estratégias computacionais a serem criadas dentro da mesma rede sem acesso externo ou por pessoas não autorizadas). Já no que compete a amostra no quesito humano, serão selecionados 5 profissionais/especialistas que

Endereço: Universidade Federal de Santa Catarina, Prédio Reitoria II, R: Desembargador Vitor Lima, nº 222, sala 401

Bairro: Trindade

CEP: 88.040-400

UF: SC

Município: FLORIANOPOLIS

Telefone: (48)3721-6094

E-mail: cep.propesq@contato.ufsc.br

UNIVERSIDADE FEDERAL DE
SANTA CATARINA - UFSC



Continuação do Parecer: 5.063.111

atuam no STT/RCTM, os quais responderão um formulário eletrônico (entrevista semiestruturada) gerada a partir do Google Forms para validação do método proposto no projeto de pesquisa. As tabelas, gráficos entre outras análises se aterão ao repositório de proveniência local, sendo essas análises apresentadas para os gestores e profissionais/especialistas autorizados.

Resumo:

Os Sistemas de Informação em Saúde (SIS) são utilizados em diferentes processos de gerenciamento de dados por diversas tecnologias em diferentes especialidades médicas. A telemedicina utiliza diferentes tipos de SIS para o processo de gerenciamento de dados de saúde. Como qualquer outro cenário de saúde, a telemedicina gera grandes volumes de dados que precisam ser gerenciados. Com o objetivo de propor um método de gerenciamento de dados de proveniência em SIS, optou-se por realizar o estudo de caso especificamente no Sistema Catarinense de Telemedicina e Telessaúde (STT) da Rede Catarinense de Telemedicina (RCTM) em atividade nas dependências do Centro de Ciências da Saúde (CCS) da Universidade Federal de Santa Catarina (UFSC). Serão utilizadas estratégias computacionais para propor um método com técnicas de proveniência de dados associadas a normas e padrões internacionais para a representação, formatação, transferência e interoperabilidade de dados na realização do processo de captura dos dados na rede em estudo. Os dados de proveniência serão capturados por meio da integração do método aqui proposto dentro de uma arquitetura restrita de acesso e com autorização da instituição responsável, sem expor dados fora do âmbito da rede. Logo após, será enviado um formulário eletrônico por e-mail para os especialistas/profissionais que desempenham seu trabalho de gestão dos dados no STT/RCTM para validar a eficiência do método aqui proposto e aprimorá-lo em uma versão concreta e ideal para aplicação em diferentes SIS. Metodologia Proposta:

Metodologia Proposta:

Em relação à sua natureza, este projeto de pesquisa classifica-se como aplicada. De acordo com Gerhardt e Silveira (2009, p. 35), "este tipo de pesquisa tem como objetivo gerar conhecimentos para aplicação prática, dirigidos à solução de problemas específicos". O teor de aplicação deste projeto de pesquisa se refere a proposta de um método para o gerenciamento de dados de proveniência em SIS no STT/RCTM. Do ponto de vista dos procedimentos técnicos, este projeto de

Endereço: Universidade Federal de Santa Catarina, Prédio Reitoria II, R: Desembargador Vitor Lima, nº 222, sala 401
Bairro: Trindade **CEP:** 88.040-400
UF: SC **Município:** FLORIANOPOLIS
Telefone: (48)3721-6094 **E-mail:** cep.propesq@contato.ufsc.br

UNIVERSIDADE FEDERAL DE
SANTA CATARINA - UFSC



Continuação do Parecer: 5.063.111

pesquisa caracteriza-se como experimental, sendo observadas as variáveis em um estudo de caso a ser realizado no STT/RCTM. Segundo Gil (2007), a pesquisa experimental consiste em determinar um objeto de estudo, selecionar as variáveis que seriam capazes de influenciá-lo, definir as formas de controle e de observação dos efeitos que a variável produz no objeto. Do ponto de vista em relação a abordagem desse projeto de pesquisa, classifica-se este estudo como quantitativo, que segundo Knechtel (2014) atua sobre um problema humano ou social, que neste caso está em relação a quantidade de dados a serem capturados pelo método e qualitativo que segundo Reneker (1993) o pesquisador desenvolve conceitos, ideias e entendimentos a partir de padrões encontrados nos dados, ao invés de coletar dados para comprovar teorias, hipóteses e modelos preconcebidos. Para Frew (2005), o próprio uso de técnicas de proveniência de dados qualifica os dados, sendo um dos benefícios da proveniência. Optou-se pela realização de um estudo de caso neste projeto de pesquisa, pois permitirá a aplicação do método em um cenário de saúde real com dados reais, posteriormente sendo possível validar o método juntamente com os profissionais/especialistas que atuam no STT/RCTM. O estudo de caso é o método que visa compreender fenômenos sociais complexos, preservando as características holísticas e significativas dos eventos da vida real (YIN, 2001). Em relação ao universo, ou população, segundo Vergara (1997) é o conjunto de elementos que possuem as características que serão objeto do estudo, e a amostra, ou população amostral, é uma parte do universo escolhido, selecionada a partir de um critério de representatividade. Nesse sentido, o universo de pesquisa se volta ao STT/RCTM no qual se pretende aplicar o método. Desta forma, a população selecionada, segue duas vertentes: computacional e humano. No quesito computacional, serão os dados digitais e manuais inseridos nos SIS integrados no STT/RCTM. Já no quesito humano, serão os profissionais/especialistas que atuam no gerenciamento dos dados na mesma rede. Em relação a amostra, também seguem os mesmos quesitos da população. Nesse sentido, a amostra no quesito computacional será com base no repositório de proveniência local (banco de dados local que receberá dados do STT/RCTM por meio de estratégias computacionais a ser criado dentro da mesma rede sem acesso externo ou por pessoas não autorizadas). Já no que compete a amostra no quesito humano, serão selecionados 5 profissionais/especialistas que atuam no STT/RCTM que responderão um formulário eletrônico (entrevista semiestruturada) gerada a partir do Google Forms para validação do método proposto neste projeto de pesquisa.

Estão previstos 5 participantes, profissionais/especialistas que atuam no STT/RCTM, que responderão a um formulário eletrônico.

Endereço: Universidade Federal de Santa Catarina, Prédio Reitoria II, R: Desembargador Vitor Lima, nº 222, sala 401
Bairro: Trindade **CEP:** 88.040-400
UF: SC **Município:** FLORIANOPOLIS
Telefone: (48)3721-6094 **E-mail:** cep.propesq@contato.ufsc.br

UNIVERSIDADE FEDERAL DE
SANTA CATARINA - UFSC



Continuação do Parecer: 5.063.111

Objetivo da Pesquisa:

Segundo os pesquisadores, no formulário da Plataforma Brasil:

Objetivo Primário:

Propor um método para contribuir para o gerenciamento dos dados de proveniência em SIS no STT/RCTM.

Objetivo Secundário:

- a) Conhecer o fluxo dos dados de proveniência gerados pelos SIS integrados do STT/RCTM;
- b) Constatar inexistências de gerenciamento de dados de proveniência nos SIS integrados do STT/RCTM;
- c) Aplicar o método de gerenciamento dos dados de proveniência, tanto em dados digitais quanto em dados inseridos manualmente nos SIS integrados do STT/RCTM;
- d) Validar o método proposto com os profissionais/especialistas que atuam no STT/RCTM.

Avaliação dos Riscos e Benefícios:

Segundo os pesquisadores, no formulário da Plataforma Brasil:

Riscos:

Os potenciais riscos desse projeto de pesquisa estão relacionados em dois momentos. No primeiro momento, está relacionado a aplicação do método, onde podem surgir alguns problemas de acesso aos dados de proveniência do STT/RCTM por algum motivo técnico, fazendo com que o profissional responsável pelo acesso aos dados possa despende tempo para ajustar o controle de acesso. Já no segundo momento, após aplicar o método no STT/RCTM, está relacionado aos 5 participantes da pesquisa que irão validar o método por meio do formulário eletrônico, onde os riscos em participar da pesquisa podem estar relacionados com desconforto, perda de tempo, constrangimento, receio, intimidação, angústia, fadiga, insatisfação, irritação ou algum mal estar frente aos questionamentos decorrentes da pesquisa.

Benefícios:

Endereço: Universidade Federal de Santa Catarina, Prédio Reitoria II, R: Desembargador Vitor Lima, nº 222, sala 401
Bairro: Trindade **CEP:** 88.040-400
UF: SC **Município:** FLORIANOPOLIS
Telefone: (48)3721-6094 **E-mail:** cep.propesq@contato.ufsc.br

UNIVERSIDADE FEDERAL DE
SANTA CATARINA - UFSC



Continuação do Parecer: 5.063.111

Em relação aos potenciais benefícios desse projeto de pesquisa, está relacionado em um primeiro momento na aplicação do método como contribuição para o gerenciamento dos dados de proveniência no STT/RCTM para possíveis auditorias no intuito de tomadas de decisões mais precisas. No segundo momento, realizar o diagnóstico das contribuições da aplicação do método aqui proposto juntamente com 5 profissionais/especialistas da equipe técnica atuantes no STT/RCT, servirá como importante fator de contribuição para melhorias nos processos de rastreamento e gerenciamento dos dados de proveniência dentro da rede em estudo. Benefícios indiretos sempre serão gerados pela produção de conhecimento associada ao desenvolvimento dessa pesquisa, oportunizando a novos pesquisadores aprimorarem o método aqui em questão em novas vertentes de pesquisa no âmbito do STT/RCTM e futuramente em outros cenários de saúde.

Comentários e Considerações sobre a Pesquisa:

V. campo de considerações finais e lista de pendências.

Considerações sobre os Termos de apresentação obrigatória:

V. campo de considerações finais e lista de pendências.

Recomendações:

V. campo de considerações finais e lista de pendências.

Conclusões ou Pendências e Lista de Inadequações:

Folha de rosto assinada pelo pesquisador responsável e pela coordenação do PPG em Ciência da Informação da UFSC.

Consta do protocolo declaração de anuência de Sérgio Fernando Torres de Freitas, como representante legal do STT/RCTM.

Consta do protocolo termo de compromisso para utilização de dados, assinado pelo doutorando e pelo "profissional responsável pelo acesso aos bancos de dados do STT/RCTM" (Alexandre Augusto Gimenes Marquez Filho).

Endereço: Universidade Federal de Santa Catarina, Prédio Reitoria II, R: Desembargador Vitor Lima, nº 222, sala 401
Bairro: Trindade **CEP:** 88.040-400
UF: SC **Município:** FLORIANOPOLIS
Telefone: (48)3721-6094 **E-mail:** cep.propesq@contato.ufsc.br

UNIVERSIDADE FEDERAL DE
SANTA CATARINA - UFSC



Continuação do Parecer: 5.063.111

Consta do projeto o roteiro para a entrevista semi-estruturada a ser feita com os participantes.

O cronograma informa que o "Acesso às fontes de dados para captura e armazenamento dos dados de proveniência" se dará a partir de 20/11/2021 e que o "Envio de formulário eletrônico para realizar as entrevistas" se dará a partir de 20/01/2022.

Orçamento de R\$ 120,00 com financiamento próprio.

TCLE atende essencialmente todas as exigências da res. 466/12. Recomendamos, entretanto, que em próximas submissões não seja solicitado do participante dados não essenciais à pesquisa (tais como RG, CPF etc.) e que a finalização do TCLE não inclua textos de teor declaratório ("Declaro que concordo..."), o que pode constranger alguns tipos de participantes.

No projeto, os pesquisadores esclarecem que "os dados pessoais e dados sensíveis serão anonimizados conforme Lei Geral de Proteção de Dados Pessoais (BRASIL, 2018). Esse processo se dará com auxílio de ferramentas de software e posteriormente esses dados serão apagados do repositório de dados de proveniência local, pois o intuito dessa coleta e armazenamento é a aplicação e validação do método juntamente com os profissionais/especialistas atuantes no STT/RCTM."

Considerações Finais a critério do CEP:

Este parecer foi elaborado baseado nos documentos abaixo relacionados:

| Tipo Documento | Arquivo | Postagem | Autor | Situação |
|---|---|------------------------|--------------------|----------|
| Informações Básicas do Projeto | PB_INFORMAÇÕES_BÁSICAS_DO_PROJETO_1831892.pdf | 10/10/2021 20:54:57 | | Aceito |
| Projeto Detalhado / Brochura Investigador | PROJETO_DE_PESQUISA.pdf | 10/10/2021 20:53:33 | Márcio José Sembay | Aceito |
| Outros | ROTEIRO_PARA_A_ENTREVISTA_SEMIESTRUTURADA.pdf | 10/10/2021 20:49:42 | Márcio José Sembay | Aceito |
| Outros | TERMO_DE_COMPROMISSO_PARA_ | 10/10/2021 | Márcio José Sembay | Aceito |

Endereço: Universidade Federal de Santa Catarina, Prédio Reitoria II, R: Desembargador Vitor Lima, nº 222, sala 401
Bairro: Trindade **CEP:** 88.040-400
UF: SC **Município:** FLORIANOPOLIS
Telefone: (48)3721-6094 **E-mail:** cep.propesq@contato.ufsc.br

UNIVERSIDADE FEDERAL DE
SANTA CATARINA - UFSC



Continuação do Parecer: 5.063.111

| | | | | |
|---|---|------------------------|--------------------|--------|
| Outros | ILIZACAO_DE_DADOS_assinado.pdf | 20:44:11 | Márcio José Sembay | Aceito |
| Declaração de Instituição e Infraestrutura | DECLARACAO_INSTITUICAO_PARA_COLETA_DE_DADOS_assinado.pdf | 06/10/2021 16:08:57 | Márcio José Sembay | Aceito |
| TCLE / Termos de Assentimento / Justificativa de Ausência | TERMO_DE_CONSENTIMENTO_LIVRE_E_ESCLARECIDO_com_assinatura_pesquisador.pdf | 30/09/2021 21:33:48 | Márcio José Sembay | Aceito |
| Folha de Rosto | Folha_de_Rosto_Assinada.pdf | 30/09/2021 20:57:08 | Márcio José Sembay | Aceito |

Situação do Parecer:

Aprovado

Necessita Apreciação da CONEP:

Não

FLORIANOPOLIS, 26 de Outubro de 2021

Assinado por:
Luciana C Antunes
(Coordenador(a))

Endereço: Universidade Federal de Santa Catarina, Prédio Reitoria II, R: Desembargador Vitor Lima, nº 222, sala 401
Bairro: Trindade **CEP:** 88.040-400
UF: SC **Município:** FLORIANOPOLIS
Telefone: (48)3721-6094 **E-mail:** cep.propesq@contato.ufsc.br

APÊNDICE D – DECLARAÇÃO (INSTITUIÇÃO PARA COLETA DE DADOS)

DECLARAÇÃO (INSTITUIÇÃO PARA COLETA DE DADOS)



Universidade Federal de Santa Catarina
 Programa de Pós-Graduação em Ciência da Informação
 Campus Professor João David Ferreira Lima – Trindade – Florianópolis –
 Santa Catarina – Brasil – CEP: 88.040-900

Declaro para os devidos fins e efeitos legais que, objetivando atender as exigências para a obtenção de parecer do Comitê de Ética em Pesquisa com Seres Humanos, e como representante legal do projeto do **Sistema Catarinense de Telemedicina e Telessaúde (STT) da Rede Catarinense de Telemedicina (RCTM) em atividade no Centro de Ciências da Saúde nas dependências da Universidade Federal de Santa Catarina (UFSC), R. Delfino Conti, S/N – Trindade, Florianópolis – SC, 88040-370, telefone (48) 3721 4448**, tomei conhecimento do projeto de pesquisa: **Proposta de um método para gerenciamento dos dados de proveniência em sistemas de informação em saúde**, e cumprirei os termos da Resolução CNS 466/12 e suas complementares, sob responsabilidade de **Márcio José Sembay** e como esta instituição tem condição para o desenvolvimento deste projeto, autorizo a sua execução nos termos propostos.

Florianópolis, ___ / ___ / _____



Documento assinado digitalmente
Sergio Fernando Torres de Freitas
 Data: 06/12/2022 10:53:18-0300
 CPF: ***.828.037-**
 Verifique as assinaturas em <https://v.ufsc.br>



Documento assinado digitalmente
MARCIO JOSE SEMBAY
 Data: 06/12/2022 12:43:34-0300
 CPF: ***.086.439-**
 Verifique as assinaturas em <https://v.ufsc.br>

ASSINATURA: _____

NOME: Dr. Sérgio Fernando Torres de Freitas

CARGO: Professor Titular da Universidade Federal de Santa Catarina no Centro de Ciência da Saúde e representante legal do STT/RCTM.

CARIMBO DO/A RESPONSÁVEL

APÊNDICE E – TERMO DE COMPROMISSO PARA UTILIZAÇÃO DE DADOS (TCUD)

1/2



Universidade Federal de Santa Catarina
Programa de Pós-Graduação em Ciência da Informação
Campus Professor João David Ferreira Lima – Trindade – Florianópolis –
Santa Catarina – Brasil – CEP: 88.040-900

TERMO DE COMPROMISSO PARA UTILIZAÇÃO DE DADOS (TCUD)

Os pesquisadores da **Universidade Federal de Santa Catarina (UFSC)**, do **Programa de Pós-graduação em Ciência da Informação (PGCIN/UFSC)**, envolvidos no projeto de pesquisa intitulado “**Proposta de um método para gerenciamento dos dados de proveniência em sistemas de informação em saúde**” comprometem-se com a utilização dos dados contidos no **Sistema Catarinense de Telemedicina e Telessaúde (STT) da Rede Catarinense de Telemedicina (RCTM) em atividade no Centro de Ciências da Saúde nas dependências da UFSC** para fins de pesquisa científica, reiterando que tal acesso somente dar-se-á após aprovação pelo Comitê de Ética em Pesquisa com Seres Humanos da Universidade Federal de Santa Catarina (CEPSH-UFSC).

Todos os pesquisadores e colaboradores envolvidos na pesquisa comprometem-se com a manutenção da privacidade e a confidencialidade dos dados acessados e/ou coletados, sejam individuais ou institucionais, preservando sempre o anonimato, conforme preconizados na Resolução 466/12, bem como suas complementares, do Conselho Nacional de Saúde e a Diretriz 12 das Diretrizes Éticas Internacionais para Pesquisas Biomédicas Envolvendo Seres Humanos (CIOMS 2018).

Declaramos entender que a integridade das informações, a garantia da confidencialidade dos dados e a privacidade dos indivíduos que terão suas informações acessadas estão sob nossa responsabilidade. Também declaramos que não repassaremos os dados coletados, o banco de dados em sua íntegra (ou em parte) a pessoas não envolvidas na equipe do presente projeto de pesquisa. Ainda, os pesquisadores confirmam haver ciência de que os dados obtidos na pesquisa somente serão utilizados para este projeto. Todo e qualquer outro uso que venha a ser planejado, será, obrigatoriamente,

2/2

objeto de um novo projeto de pesquisa, o qual será submetido à apreciação do CEPESH-UFSC/CONEP.

Devido à impossibilidade de obtenção do Termo de Consentimento Livre e Esclarecido de parte ou totalidade dos indivíduos, todos os pesquisadores e colaboradores envolvidos na manipulação dos dados assinam esse Termo de Consentimento de Uso de Banco de Dados, de modo a salvaguardar os direitos dos participantes.

 Documento assinado digitalmente
Márcio José Sembay
Data: 06/10/2021 16:35:01-0300
CPF: 026.086.439-06
Verifique as assinaturas em <https://v.ufsc.br>

Me. Márcio José Sembay (Doutorando)

 Documento assinado digitalmente
Alexandre Augusto Gimenes Marquez Filho
Data: 06/10/2021 11:05:26-0300
CPF: 372.905.156-80
Verifique as assinaturas em <https://v.ufsc.br>

Me. Alexandre Augusto Gimenes Marquez Filho
(Profissional responsável pelo acesso aos Bancos de dados do STT/RCTM)

Florianópolis, 06 de outubro de 2021.

APÊNDICE F – TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO (TCLE)



Universidade Federal de Santa Catarina
Programa de Pós-Graduação em Ciência da Informação
Campus Professor João David Ferreira Lima – Trindade – Florianópolis – Santa
Catarina – Brasil – CEP: 88.040-900

TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO (TCLE)

Você está sendo convidado(a) como voluntário(a) a participar da pesquisa: **Proposta de um método para gerenciamento dos dados de proveniência em sistemas de informação em saúde**. Esta pesquisa está associada ao projeto de doutorado do Me. Márcio José Sembay, do Programa de Pós-Graduação em Ciência da Informação da Universidade Federal de Santa Catarina.

Esta pesquisa tem como objetivo capturar os dados de proveniência gerados pelos Sistemas de Informação em Saúde (SIS) integrados especificamente no Sistema Catarinense de Telemedicina e Telessaúde (STT) da Rede Catarinense de Telemedicina (RCTM) em atividade nas dependências da Universidade Federal de Santa Catarina (UFSC). Nesse sentido, será utilizado um método computacional com técnicas de proveniência de dados associadas a normas e padrões internacionais para a representação, formatação, transferência e interoperabilidade de dados para realizar esse processo. Os dados de proveniência serão capturados por meio da integração do método dentro de uma arquitetura restrita de acesso e com autorização do profissional responsável, sem expor dados fora do STT/RCTM.

A coleta de dados ocorrerá por meio de aplicação de um questionário em formato eletrônico gerado a partir do *Google Forms*. Nesse sentido, para responder a essa pesquisa será necessário um computador ou qualquer outro dispositivo móvel (celular, tablet, entre outros) com acesso à internet. Em caso de indisponibilidade, o pesquisador responsável dessa pesquisa poderá providenciar a infraestrutura necessária para responder ao questionário.

Toda pesquisa tem riscos, mesmo que mínimos como cansaço ou aborrecimento e eventuais limitações tecnológicas relacionadas a questionários de formulários eletrônicos. Para minimizar qualquer desconforto, o questionário no formulário eletrônico foi elaborado de forma a ser respondido rapidamente, levando em média 10 minutos. Em caso de desconforto, sinta-se à vontade para deixar de participar da pesquisa a qualquer momento, sem ter que apresentar

qualquer justificativa. No entanto, destacam-se alguns possíveis benefícios que a pesquisa pode proporcionar ao STT/RCTM, em especial na aplicação do método como contribuição para o gerenciamento dos dados de proveniência na rede para possíveis auditorias no intuito de tomadas de decisões mais precisas. Em um segundo momento, mapear melhorias nos processos de rastreamento dos dados de proveniência dentro do STT/RCTM a fim de elaborar uma versão do método de forma ideal para atender as demandas de rastreamento de dados de proveniência em diferentes SIS.

A você é assegurado o acompanhamento e assistência, caso tenha alguma dúvida sobre os procedimentos ou sobre a pesquisa. Você poderá entrar em contato com o pesquisador a qualquer momento pelo telefone ou e-mail.

Você é livre para recusar-se a participar, retirar seu consentimento ou interromper a participação a qualquer momento. A sua participação é voluntária e a recusa em participar não irá acarretar qualquer penalidade.

O pesquisador Me. Márcio José Sembay e seu orientador Prof. Dr. Douglas Dyllon Jeronimo de Macedo irão tratar a sua identidade com padrões profissionais de sigilo. Tomarão todas as providências necessárias para manter o sigilo, mas sempre existe a remota possibilidade da quebra do sigilo, mesmo que involuntário e não intencional, cujas consequências serão tratadas nos termos da lei. Os resultados deste trabalho poderão ser apresentados em encontros ou revistas científicas. Apresentarão apenas os resultados obtidos como um todo, sem revelar seu nome, instituição ou qualquer informação relacionada à sua privacidade.

Duas vias deste documento serão rubricadas e assinadas por você e pelo pesquisador responsável. Recomenda-se guardar cuidadosamente a sua via, pois é um documento que traz importantes informações de contato e garante os seus direitos como participante da pesquisa.

A participação na pesquisa não acarretará custos para você e não será oferecida nenhuma compensação financeira adicional. No entanto, caso você tenha qualquer despesa decorrente da participação nessa pesquisa, como transporte e alimentação, eles serão assumidos pelos pesquisadores e reembolsados para você. De igual maneira, caso sofra algum dano material ou imaterial devidamente comprovado e decorrentes diretamente dessa pesquisa, os pesquisadores garantem indenizá-lo conforme determina a lei vigente.

O pesquisador responsável, que também assina esse documento, compromete-se a conduzir a pesquisa de acordo com o que preconiza a Resolução 466/12 de 12/06/2012, que trata dos preceitos éticos e da proteção aos participantes da pesquisa.

Declaro que fui informado(a) dos objetivos da pesquisa acima de maneira clara e detalhada e esclareci minhas dúvidas. Sei que em qualquer momento poderei solicitar novas

informações e motivar minha decisão se assim o desejar.

O pesquisador Me. Márcio José Sembay e o professor orientador Prof. Dr. Douglas Dyllon Jeronimo de Macedo certificaram-me de que todos os dados desta pesquisa serão confidenciais.

Em caso de dúvidas poderei chamar o pesquisador Me. Márcio José Sembay no telefone (49) 99944-9101 ou pelo e-mail *marcio.sembay@posgrad.ufsc.br*, o professor orientador Prof. Dr. Douglas Dyllon Jeronimo de Macedo no telefone (48) 3721-4075 ou pelo e-mail *douglas.macedo@ufsc.br* ou o Comitê de Ética em Pesquisa com Seres Humanos (CEPSH) da Universidade Federal de Santa Catarina (UFSC), sito à Rua Desembargador Vitor Lima, nº 222, sala 401, Trindade – Florianópolis, SC. Telefone (48) 3721-6094 ou e-mail *cep.propesq@contato.ufsc.br*.

Declaro que concordo em participar desta pesquisa. Recebi uma via deste termo de consentimento livre e esclarecido e me foi dada a oportunidade de ler e esclarecer as minhas dúvidas.

APÊNDICE G – ROTEIRO PARA ENTREVISTA SEMIESTRUTURADA



Universidade Federal de Santa Catarina
 Programa de Pós-Graduação em Ciência da Informação
 Campus Professor João David Ferreira Lima – Trindade – Florianópolis – Santa Catarina –
 Brasil – CEP: 88.040-900

PARTE I - IDENTIFICAÇÃO DO PARTICIPANTE DA PESQUISA

1. Identificação/codínome (Opcional): ____
2. Idade:
 Entre 20 a 30 anos; Entre 30 a 40 anos; Entre 40 a 50 anos; Mais que 50 anos
3. Sexo:
 Masculino; Feminino; Outro
4. Profissão: _____
5. Quanto tempo de atuação na equipe dos profissionais/especialistas do STT/RCTM?
 Menos de 1 ano; 1 ano; 2 anos; 3 anos; Mais que 3 anos
- 6 - Especifique sua escolaridade:
 Graduação em:
 Especialização em:
 Mestrado na área de:
 Doutorado na área de:
 Pós-doutor(a) na área de:
7. Quanto tempo de experiência na área de atuação profissional?
 Menos de 1 ano (); Até um 1 ano; De 1 ano até 5 anos; Mais de 5 anos

PARTE II – ENTREVISTA SEMIESTRUTURADA

- 1) De modo geral, em sua visão o PROV-Health atende as necessidades de gerenciamento de dados de proveniência no STT/RCTM?
- 2) Em sua visão, descreva quais os possíveis benefícios que podem ser observados com a aplicação do PROV-Health para o gerenciamento dos dados de proveniência no STT/RCTM.

- 3) Em sua visão, descreva quais os riscos que podem ser possíveis de observação na aplicação do PROV-Health para o gerenciamento dos dados de proveniência no STT/RCTM.
- 4) O PROV-Health apresenta os dados de proveniência do STT/RCTM de forma clara por meio da ferramenta de visualização de dados escolhida? Em sua visão, é suficiente?
- 5) De modo geral, os dados de proveniência do STT/RCTM que foram coletados pelo PROV-Health são suficientes?
- 6) Em sua visão, o PROV-Health pode ser adaptado para a aplicação em outros contextos de saúde?
- 7) Gostaria de descrever algo mais para contribuir com a pesquisa em relação a debilidades ou sugerir melhorias para o PROV-Health?

Ao final do parecer, solicito que os participantes informem se **aprovam** ou **desaprovam** o PROV-Health para aplicação no STT/RCTM.

- Aprovo
- Desaprovo

APÊNDICE H – SOLICITAÇÃO DE REGISTRO DO PROV-HEALTH NO INPI

| | | | | | |
|--|---------------------|-----------|----------|------------|--------|
|  BRASIL | Acesso à informação | Participe | Serviços | Legislação | Canais |
|--|---------------------|-----------|----------|------------|--------|

Instituto Nacional da
Propriedade Industrial
 Ministério da Economia

Consulta à Base de Dados do INPI

[Início | Ajuda?]

» Consultar por: Pesquisa Base Programas | Finalizar Sessão 1/1

Programa de Computador

Nº do Pedido: **BR 51 2023 001481 8**
 Data do Depósito: 26/05/2023
 Linguagem: JSON / PYTHON / SQL
 Campo de Aplicação: IF-07 / IN-02 / TC-01
 Tipo Programa: AP-03 , CD-01 , GI-02 , GI-08 , TI-01
 Título: PROV-HEALTH
 Nome do Titular: UNIVERSIDADE FEDERAL DE SANTA CATARINA
 Nome do Autor: ALEXANDRE AUGUSTO GIMENES MARQUEZ FILHO / DOUGLAS DYLLON JERONIMO DE MACEDO / MÁRCIO JOSÉ SEMBAY
 Nome do Procurador: CLARISSA STEFANI TEIXEIRA

Petições ?

| Pgo | Protocolo | Data | Img | Serviço | Cliente | Delivery | Data |
|-----|-------------|------------|-----|---------|--|----------|------|
| ✓ | 87023004422 | 26/05/2023 | - | 730 | UNIVERSIDADE FEDERAL DE SANTA CATARINA | | - |

Publicações ?

| RPI | Data RPI | Despacho | Img | Complemento do Despacho |
|------|-------------------|----------|---|-------------------------|
| 2735 | 06/06/2023 | 730 |  | |

Dados atualizados até **22/08/2023** - Nº da Revista: **2746**

Rua Mayrink Veiga, 9 - Centro - RJ - CEP: 20090-910



APÊNDICE I – REGISTRO DO PROV-HEALTH NO INPI**INPI** INSTITUTO
NACIONAL DA PROPRIEDADE
INDUSTRIAL
Assinado
Digitalmente

REPÚBLICA FEDERATIVA DO BRASIL
MINISTÉRIO DO DESENVOLVIMENTO, INDÚSTRIA, COMÉRCIO E SERVIÇOS
INSTITUTO NACIONAL DA PROPRIEDADE INDUSTRIAL
DIRETORIA DE PATENTES, PROGRAMAS DE COMPUTADOR E TOPOGRAFIAS DE CIRCUITOS

Certificado de Registro de Programa de Computador**Processo Nº: BR512023001481-8**

O Instituto Nacional da Propriedade Industrial expede o presente certificado de registro de programa de computador, válido por 50 anos a partir de 1º de janeiro subsequente à data de 01/01/2023, em conformidade com o §2º, art. 2º da Lei 9.609, de 19 de Fevereiro de 1998.

Título: PROV-HEALTH**Data de publicação:** 01/01/2023**Data de criação:** 01/01/2023**Titular(es):** UNIVERSIDADE FEDERAL DE SANTA CATARINA**Autor(es):** DOUGLAS DYLLON JERONIMO DE MACEDO; ALEXANDRE AUGUSTO GIMENES MARQUEZ FILHO; MÁRCIO JOSÉ SEMBAY**Linguagem:** SQL; PYTHON; JSON**Campo de aplicação:** IF-07; IN-02; TC-01**Tipo de programa:** AP-03; CD-01; GI-02; GI-08; TI-01**Algoritmo hash:** SHA-512**Resumo digital hash:**

74be54a374f8b24eb15bb0d97cfa376881c7fd3cf03d7d1f2791a75936c51d1672a646bef73376af387a9d315d06f39a0be8ed72de2fa0c9b5d9e0ae5ca5609

Expedido em: 06/06/2023**Aprovado por:**Carlos Alexandre Fernandes Silva
Chefe da DIPTO