

Ministério da Saúde

FIOCRUZ

Fundação Oswaldo Cruz

FUNDAÇÃO OSWALDO CRUZ  
INSTITUTO NACIONAL DE INFECTOLOGIA EVANDRO CHAGAS  
MESTRADO PROFISSIONAL EM PESQUISA CLÍNICA

PATRÍCIA COSTA DOS SANTOS

**IMPLEMENTAÇÃO DO *DATA MART* NA PESQUISA CLÍNICA: ESTUDO DE  
CASO NO LABORATÓRIO DE PESQUISA CLÍNICA E VIGILÂNCIA EM  
LEISHMANIOSES**

Rio de Janeiro

2020

PATRÍCIA COSTA DOS SANTOS

**IMPLEMENTAÇÃO DO *DATA MART* NA PESQUISA CLÍNICA: ESTUDO DE CASO NO LABORATÓRIO DE PESQUISA CÍNICA E VIGILÂNCIA EM LEISHMANIOSES**

Dissertação apresentada ao Curso de Mestrado Profissional em Pesquisa Clínica do Instituto Nacional de Infectologia Evandro Chagas, para obtenção do grau de Mestre em Pesquisa Clínica.

Orientadora: Dra. Aline Fagundes da Silva  
Co-orientadora: Me. Mary Lucy Ribeiro Pinto

Rio de Janeiro

2020

PATRICIA COSTA DOS SANTOS

**IMPLEMENTAÇÃO DO *DATA MART* NA PESQUISA CLÍNICA: ESTUDO DE CASO NO LABORATÓRIO DE PESQUISA CLÍNICA E VIGILÂNCIA EM LEISHMANIOSES**

Dissertação apresentada ao Curso de Mestrado Profissional em Pesquisa Clínica do Instituto Nacional de Infectologia Evandro Chagas para obtenção do grau de Mestre.

Orientadora: Dra. Aline Fagundes da Silva  
Aprovada em \_\_\_ / \_\_\_ / \_\_\_\_

BANCA EXAMINADORA

---

**Dra. Samanta Cristina das Chagas Xavier Azeredo (Presidente e Revisora)**

Doutora em Biologia Parasitária  
Instituto Oswaldo Cruz - Fiocruz

---

**Dr. Ronaldo Imerio Moreira (Membro)**

Doutor em Epidemiologia  
Instituto Nacional de Infectologia Evandro Chagas - Fiocruz

---

**Dra. Cristina Maria Giordano Dias (Membro)**

Doutora em Medicina Tropical  
Secretaria de Estado de Saúde do Rio de Janeiro

---

**Dra. Andreza Paim Marcelino (Suplente)**

Doutora em Medicina Veterinária Preventiva  
Instituto Nacional de Infectologia Evandro Chagas – Fiocruz

## AGRADECIMENTOS

À Deus por ter me dado força e coragem para ultrapassar todos os obstáculos e atingir meus objetivos.

Ao meu esposo, Marco, pela paciência diária, pela parceria nas dificuldades técnicas encontradas e o conforto nos momentos difíceis dessa jornada.

Aos meus filhos, Lucas e Maria Eduarda, que sempre me apoiaram e torciam durante a realização desse trabalho.

À minha coordenadora e amiga, Mary Lucy, que me ajudou nos momentos mais difíceis dessa jornada e com toda dedicação. Muito obrigada!

À Dra. Aline Fagundes, minha orientadora, pelo incentivo que me foi dado durante a elaboração desse trabalho.

À minha amiga de trabalho e de turma, Luciene, por todo apoio durante o curso e na realização da dissertação.

Aos meus amigos de trabalho, Diogo e Thiago, pelo conhecimento e atividades realizadas para a implantação do *Data Mart*.

Aos profissionais do INI participantes do estudo que contribuíram com os dados coletados durante as entrevistas.

Santos, Patrícia Costa. **Implementação do *Data Mart* na Pesquisa Clínica: Estudo de caso no Laboratório de Pesquisa Clínica e Vigilância em Leishmanioses.** Rio de Janeiro, 2020. Dissertação de Mestrado Profissional em Pesquisa Clínica – Instituto Nacional de Infectologia Evandro Chagas.

## RESUMO

**Introdução:** A produção da informação com a introdução de novas tecnologias influencia e modifica a produção compartilhada do conhecimento, costumes e práticas na área de saúde. Este trabalho demonstra a implementação de um *Data Mart* com a finalidade de apoiar o processamento, análise, consultas e relatórios das informações pelos pesquisadores do Laboratório de Pesquisa Clínica e Vigilância em Leishmanioses do Instituto Nacional de Infectologia da Fiocruz. As pesquisas científicas realizadas por este laboratório geram um grande volume de dados que necessitam de tratamento e análise. Tal processo, demanda um grande tempo de execução, podendo levar meses. **Objetivo:** Identificar e descrever os processos de coleta e armazenamento dos dados do Laboratório de Pesquisa Clínica e Vigilância em Leishmanioses sob a perspectiva de implantação de um *Data Mart* (repositório de dados) de modo a apoiar à tomada de decisão dos pesquisadores em um menor tempo com uma melhor qualidade na análise dos dados da pesquisa. Como objetivos secundários temos: (i) Analisar o processo de coleta, armazenamento e processamento dos dados do Laboratório de Pesquisa Clínica e Vigilância em Leishmanioses, (ii) Construir e implementar o *Data Mart* e (iii) Desenvolver um material de apoio aos usuários do *Data Mart*. **Métodos:** No processo de criação do *Data Mart* foi realizada a extração e gerenciamento dos dados provenientes de diversas fontes em um só local, que permitiu aos pesquisadores realizar inferências em um banco de dados modelado de forma multidimensional, tornando mais rápido, seguro e eficiente o acesso às consultas a qualquer momento através de ferramentas *Online Analytical Processing*. A partir da técnica aplicada foi possível coletar as informações operacionais, filtrá-las, padronizá-las até que as mesmas atingissem o formato e o nível de detalhes desejados, carregá-las em um repositório de dados e gerar relatórios de forma dinâmica. **Resultados:** O *Data Mart* foi criado integrando os dados oriundos das diversas fontes de dados de modo que todas as perguntas elaboradas no processo de análise dos pacientes fossem respondidas. O tempo na obtenção da informação foi minimizado através do uso de alguns recursos que realizam a extração e gerenciamento dos dados, melhorando assim a qualidade de análise dos mesmos na pesquisa. **Conclusões:** O *Data Mart* eliminou o desencontro dos dados, reduziu a dependência dos pesquisadores quanto aos profissionais de TI e reduziu o tempo gasto na obtenção de informações utilizadas na análise dos dados da pesquisa clínica.

Palavras-chave: *Data Mart*. *Online Analytical Processing*. Leishmanioses.

Santos, Patrícia Costa. **Implementation of Data Mart in Clinical Research: Case study in the Laboratory of Clinical Research and Surveillance in Leishmaniasis.** Rio de Janeiro, 2020. Dissertation Professional Master's in Clinical Research – National Institute of Infectious Diseases Evandro Chagas.

## ABSTRACT

**Introduction:** The production of information with the introduction of new technologies influences and modifies the shared production of knowledge, customs and practices in the health area. This project demonstrates the implementation of a Data Mart with the purpose of supporting the processing, analysis and discovery of information by researchers from the Laboratory of Clinical Research and Surveillance in Leishmaniasis of the National Institute of Infectology at Fiocruz. The scientific research carried out by this laboratory generates a large volume of data that needs treatment and analysis. Such a process takes a long time to complete and can take months.

**Objective:** To identify and describe the data collection and storage processes of the Clinical Research and Surveillance Laboratory in Leishmaniasis from the perspective of implementing a Data Mart (data repository) in order to support researchers' decision-making in a shorter time with better quality in the analysis of research data. Secondary objectives are: (i) Analyze the process of collecting, storing and processing data from the Leishmaniasis Clinical Research and Surveillance Laboratory, (ii) Building and implementing the Data Mart and (iii) Develop support material for Data Mart users.

**Methods:** In the process of creating the Data Mart, it was performed the extraction and management of this data from different sources in a single location, which allowed researchers to make inferences in a multidimensional modeled database, making it faster, safer and more efficient access to consultations at any time through Online Analytical Processing tools. Based on the applied technique, it was possible to collect operational information, filter it, standardize it until it reached the desired format and level of details, upload it to a data repository and generate reports dynamically.

**Results:** The Data Mart was created by integrating data from different data sources so that all questions elaborated in the Patient Analysis process were answered. The time in obtaining the information was minimized through the use of some resources that perform the extraction and management of the data, thus improving the quality of their analysis in the research.

**Conclusions:** Data Mart eliminated the mismatch of data, reduced the dependence of researchers on IT professionals and reduced the time spent in obtaining information used in the analysis of clinical research data.

Key words: *Data Mart. Online Analytical Processing.* Leishmaniasis.

## LISTA DE FIGURAS

|   |    |
|---|----|
| Figura 1 – Ambiente <i>Data Warehouse</i>   | 16 |
| Figura 2 – Exemplo de dados baseados em assuntos de negócio                       | 17 |
| Figura 3 – Não volátil  | 17 |
| Figura 4 – Exemplo de integração de dados   | 18 |
| Figura 5 – Modelo Estrela ( <i>Star schema</i> )                                  | 18 |
| Figura 6 – Modelo Floco de Neve ( <i>Snowflake schema</i> )                       | 19 |
| Figura 7 – Exemplo de um Cubo   | 20 |
| Figura 8 – Exemplo de uma operação <i>roll up</i>                                 | 22 |
| Figura 9 – <i>Data Warehouse</i> e <i>Data Mart</i>                               | 22 |
| Figura 10 – Processo de Análise dos Pacientes do INI                              | 31 |
| Figura 11 – Modelo estrela Laboratório Leishmaniose                               | 34 |
| Figura 12 – Planilha referente aos dias da semana                                 | 36 |
| Figura 13 – Planilha referente aos meses do ano                                   | 37 |
| Figura 14 – Banco de dados Leishmania   | 37 |
| Figura 15 – Exemplo do <i>Execution Results</i> do extrator dos dados do paciente | 38 |
| Figura 16 – Esquema de funcionamento do Extrator dimensão Tempo                   | 39 |
| Figura 17 – Esquema de funcionamento do Extrator dimensão Encaminhamento          | 39 |
| Figura 18 – Esquema de funcionamento do Extrator dimensão Paciente                | 40 |
| Figura 19 – Esquema de funcionamento do Extrator dimensão Exames                  | 40 |
| Figura 20 – Esquema de funcionamento do Extrator Fato                             | 41 |
| Figura 21 – <i>Job</i> executor dos extratores das dimensões                      | 41 |
| Figura 22 – <i>Job</i> executor do extrator da fato                               | 42 |
| Figura 23 – <i>Job</i> principal  | 42 |
| Figura 24 – Visões criadas para a pesquisadora Aline Fagundes                     | 45 |
| Figura 25 – Operações permitidas para o perfil de acesso <i>Administrator</i>     | 46 |
| Figura 26 – Operações permitidas para o perfil de acesso <i>Power User</i>        | 46 |

## LISTA DE QUADROS

|   |    |
|---|----|
| Quadro 1 – Relatório extraído do cubo LaPClinVigiLeish contendo a quantidade de pacientes confirmados e não confirmados com leishmanioses, no período de 2010 a 2019.                             | 53 |
| Quadro 2 – Relatório extraído do cubo LaPClinVigiLeish contendo os encaminhamentos realizados para uma especialidade dos pacientes não confirmados para leishmanioses, no período de 2010 a 2019. | 54 |
| Quadro 3 – Relatório extraído do cubo LaPClinVigiLeish contendo os encaminhamentos realizados para um laboratório dos pacientes não confirmados para leishmanioses, no período de 2010 a 2019.    | 55 |
| Quadro 4 – Relatório extraído do cubo LaPClinVigiLeish contendo os encaminhamentos realizados para uma instituição dos pacientes não confirmados para leishmanioses, no período de 2010 a 2019.   | 55 |
| Quadro 5 – Relatório extraído do cubo LaPClinVigiLeish contendo a quantidade de pacientes não confirmados que não foram encaminhados evoluindo para óbito, nos anos de 2012, 2013, 2014 e 2017.   | 56 |
| Quadro 6 – Relatório extraído do cubo LaPClinVigiLeish contendo a quantidade de pacientes não confirmados, não encaminhados e que não foram a óbito, no período de 2010 a 2019.                   | 56 |



## LISTA DE SIGLAS

|                  |   |
|------------------|---|
| BI               | Business Intelligence   |
| CEP              | Comitê de Ética em Pesquisa                                   |
| CID              | Classificação Internacional de Doenças                        |
| DM               | Data Mart   |
| DW               | Data Warehouse  |
| ERP              | Planejamento de Recursos Empresariais                         |
| ETC              | Extração, Transformação e Carga                               |
| FDA              | do inglês, Food and Drug Administration                       |
| HOLAP            | Hybrid Online Analytical Processing                           |
| INI              | Instituto Nacional de Infectologia                            |
| LaPClinVigiLeish | Laboratório de Pesquisa Clínica e Vigilância de Leishmanioses |
| LT               | Leishmaniose Tegumentar                                       |
| LTA              | Leishmaniose Tegumentar Americana                             |
| MOLAP            | Multidimensional Online Analytical Processing                 |
| NICHHD           | National Institute of Child Health and Human Development      |
| NIH              | National Institute of Health                                  |
| OLAP             | Online Analytical Processing                                  |
| OLTP             | Online Transactional Processing                               |
| OMS              | Organização Mundial da Saúde                                  |
| POSIC            | Política de Segurança da Informação e Comunicações            |
| ROLAP            | Relational Online Analytical Processing                       |
| SETIC            | Serviço de Tecnologia da Informação e Comunicação             |
| SGBD             | Sistemas Gerenciadores de Banco de Dados                      |
| SINAN            | Sistema de Informação de Agravos de Notificação               |
| SIPEC            | Sistema de Prontuário Eletrônico do INI                       |
| TCLE             | Termo de Consentimento Livre e Esclarecido                    |
| TI               | Tecnologia da Informação                                      |

## SUMÁRIO

|   |           |
|---|-----------|
| <b>1 INTRODUÇÃO E REFERENCIAL TEÓRICO .....</b>   | <b>12</b> |
| <b>2 JUSTIFICATIVA .....</b>  | <b>25</b> |
| <b>3 OBJETIVOS .....</b>  | <b>27</b> |
| 3.1 OBJETIVO GERAL.....   | 27        |
| 3.2 OBJETIVOS ESPECÍFICOS.....  | 27        |
| <b>3.2.1 Analisar o processo de coleta, armazenamento e processamento dos dados do Laboratório de Pesquisa Clínica e Vigilância em Leishmanioses;</b> ..... | 27        |
| <b>3.2.2 Desenvolver um material de apoio aos usuários do <i>Data Mart</i>.</b> .....   | 27        |
| <b>4 METODOLOGIA.....</b>   | <b>28</b> |
| 4.1 DESENHO DO ESTUDO.....  | 28        |
| 4.2 CASUÍSTICA .....  | 28        |
| <b>4.2.1 Objeto do Estudo</b> .....   | 28        |
| <b>4.2.2 Critérios de inclusão e de exclusão</b> .....  | 28        |
| <b>4.2.3 Amostra</b> .....  | 28        |
| 4.3 MATERIAIS, PROCEDIMENTOS E TÉCNICAS .....   | 29        |
| <b>4.3.1 Instalação dos <i>softwares</i> da <i>Suíte Pentaho</i></b> .....  | 29        |
| <b>4.3.2 Coleta de Dados</b> .....  | 30        |
| <b>4.3.3 Modelagem de dados</b> .....   | 31        |
| <b>4.3.4 Extração dos dados</b> .....   | 35        |
| <b>4.3.5 Modelagem dos Dados no modelo Multidimensional</b> .....   | 43        |
| <b>4.3.6 Visualização dos dados</b> .....   | 44        |
| <b>4.3.7 Migração para o servidor de Produção</b> .....   | 44        |
| <b>4.3.8 Análise dos Resultados</b> .....   | 44        |
| <b>4.3.9 Criação de Visões Pré-definidas</b> .....  | 45        |
| <b>4.4.0 Definição da Segurança de Informação</b> .....   | 45        |
| 4.4 RISCOS E BENEFÍCIOS.....  | 47        |
| 4.5 MATERIAL DE APOIO.....  | 48        |
| 4.6 PLANO DE GESTÃO DE DADOS .....  | 48        |
| <b>4.6.1 Dados e metadados produzidos pelo projeto</b> .....  | 48        |
| <b>4.6.2 Questões éticas e Legais</b> .....   | 50        |
| <b>4.6.3 Armazenamento e Backup</b> .....   | 50        |
| <b>5 RESULTADOS .....</b>   | <b>50</b> |
| <b>6 DISCUSSÃO .....</b>  | <b>56</b> |

|   |           |
|---|-----------|
| <b>7 CONCLUSÃO.....</b>   | <b>59</b> |
| <b>REFERÊNCIAS BIBLIOGRÁFICAS.....</b>                                      | <b>60</b> |
| <b>APÊNDICE A - TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO (TCLE).....</b>  | <b>62</b> |
| <b>APÊNDICE B - QUESTIONÁRIO DE REQUISITOS SOBRE BANCO DE DADOS.....</b>    | <b>65</b> |
| <b>APÊNDICE C – TERMO DE COMPROMISSO E CONFIDENCIALIDADE .....</b>          | <b>67</b> |
| <b>APÊNDICE D – MANUAL DE USO DO <i>PENTAHO USER CONSOLE</i> .....</b>      | <b>68</b> |
| <b>APÊNDICE E – DICIONÁRIO DE DADOS .....</b>                               | <b>73</b> |
| <b>APÊNDICE F – CÓDIGO DE CRIAÇÃO DOS OBJETOS NOS BANCOS DE DADOS .....</b> | <b>77</b> |

## 1 INTRODUÇÃO E REFERENCIAL TEÓRICO

O atual Laboratório de Pesquisa Clínica e Vigilância em Leishmanioses (LaPClinVigiLeish ) teve seu início como decorrência das atividades de pesquisa do Dr. Mauro Marzochi iniciadas no departamento de Protozoologia do Instituto Oswaldo Cruz (IOC) em 1978 e continuadas no departamento de Ciências Biológicas da Escola Nacional de Saúde Pública (ENSP), a partir de 1985. No ano de 2000 esse grupo de profissionais chegou ao Instituto de Pesquisa Evandro Chagas (IPEC) sendo vinculado ao Laboratório de Parasitologia, e em 2007 com a implantação do novo organograma do IPEC, o grupo de profissionais passou a constituir o LaPClinVigiLeish que hoje é formado por 32 profissionais. Este, integra a Rede de Referências da Fiocruz junto ao Ministério da Saúde (MS) em nível nacional para Leishmaniose Tegumentar (LT) e possui como missão desenvolver pesquisa científica e operacional, ensino, atenção clínica e laboratorial de referência visando subsidiar as ações de vigilância e controle das leishmanioses, atuando conjuntamente à Rede de Referências da Fiocruz.

De acordo com o Manual de Vigilância da Leishmaniose Tegumentar (Brasil, 2017), a leishmaniose é uma doença infecciosa, não contagiosa, que é ocasionada por espécies de protozoários do gênero *Leishmania* que acomete tanto o ser humano como animais silvestres e domésticos, podendo manifestar-se sob a forma cutânea e mucosa, sendo transmitida pela picada de insetos infectados denominados flebotomíneos, conhecidos popularmente, de acordo com a localização geográfica, como mosquito-palha, tatuquira, birigui, entre outros. Além do homem, que representa um hospedeiro acidental da doença por entrar em contato com o ciclo de transmissão, algumas espécies de roedores, marsupiais, edentados, quirópteros e canídeos silvestres foram registrados como hospedeiros e possíveis reservatórios naturais, os animais domésticos, assim como o homem, são considerados hospedeiros acidentais pela não existência de evidências científicas que comprovem que esses sejam reservatórios das espécies de leishmânias. É uma doença de notificação compulsória, sendo esta realizada no Sistema de Informação de Agravos de Notificação (Sinan).

Conforme o Manual de Vigilância da Leishmaniose Tegumentar (BRASIL, 2017), a leishmaniose tegumentar (LT) é tida como um problema de saúde pública em 85 países. Devido a sua capacidade de produzir deformidades e pelo seu alto

coeficiente de detecção, a LT é considerada pela Organização Mundial da Saúde (OMS) como uma das seis mais importantes doenças infecciosas, ainda de acordo com a OMS, no Brasil a LT é denominada Leishmaniose Tegumentar Americana (LTA) e encontra-se distribuída por todo o território brasileiro.

O tratamento é complicado, com o uso de poucas drogas disponíveis e com muitos efeitos colaterais, sendo gratuito fornecido pelo governo (BRASIL, 2017).

As pesquisas realizadas no LaPClinVigiLeish geram um enorme volume de dados e estes estão armazenados em diferentes locais.

Os dados são fatos que podem ser armazenados e que possuem um significado subentendido como por exemplo nomes, números de telefones e endereços (ELMASRI; NAVATHE, 2005).

Um banco de dados é uma coleção de dados relacionados, que representa algum aspecto do mundo real com uma finalidade específica. Eles são de grande importância para gerenciamento dos projetos de pesquisa, sendo através deles que informações valiosas são geradas (ELMASRI; NAVATHE, 2005).

A utilização dos bancos de dados é feita através de *softwares* específicos chamados de Sistemas Gerenciadores de Banco de Dados (SGBD), estes são um grupo de programas que manipulam o banco de dados, é o responsável pela manutenção e armazenamento dos dados de um banco de dados, além de fornecer uma interface entre o usuário e o banco de dados ou entre aplicativos e o banco de dados (RASLAN; CALAZANS, 2014).

Para simplificar e resolver o problema de organização dos dados e fazer com que a capacidade total do banco de dados seja utilizada, surgiram as técnicas de modelagem de dados que permitem a construção de abstrações do mundo real para relacionar dados de maneira eficiente. Através delas é possível entender melhor os processos de relacionamento entre os objetos em estudo. É uma técnica que tem como objetivo o entendimento do problema em estudo de modo que os dados relativos a esse problema possam ser organizados e armazenados de maneira segura no banco de dados por meio do SGBD, permitindo assim o acesso à informação na tomada de decisões (RASLAN; CALAZANS, 2014).

A geração das informações por meio dos bancos de dados pode ser feita através dos sistemas transacionais ou por sistemas analíticos.

Os sistemas transacionais tem como objetivo processar e registrar as transações que são realizadas na organização. Se caracterizam por terem uma alta taxa de atualização nos bancos de dados, pelo grande volume de dados e acessos pontuais e por pesquisas onde os resultados tenham um volume pequeno de informações. Esses sistemas também são conhecidos como Oltp - *Online Transactional Processing*. São usados no dia a dia da organização e desenvolvidos de acordo com o negócio da mesma (BEUREN; MARTINS, 2001).

Os sistemas analíticos, também conhecidos como Olap - *Online Analytical Processing* - caracterizam-se por fornecer o apoio à tomada de decisão a partir de análises realizadas em bases de dados históricas contendo um grande volume de dados; proporcionar um ambiente favorável para o desempenho do planejamento e a otimização e monitoramento das tarefas executadas; utilizar recursos gráficos na apresentação dos resultados; permitir a recuperação rápida da informação; e criar um ambiente único de trabalho (BEUREN; MARTINS, 2001).

Com o avanço da Tecnologia da Informação (TI), novas ferramentas foram criadas para facilitar a manipulação desse grande volume de dados que estão armazenados em diferentes locais, tornando mais simples o processo de geração de conhecimento e informação.

Sendo assim, para que esses dados gerados sejam consultados de forma produtiva pelos pesquisadores do LaPClinVigiLeish , os mesmos devem estar organizados em um banco de dados único, tendo uma interface com acesso fácil e rápido, no qual através dela, o pesquisador possa obter informações consolidadas, tendo a possibilidade de decidir estrategicamente o que deve ser feito.

Raslan e Calazans (2014) mencionaram que mesmo com os bancos de dados mantendo e recuperando uma grande quantidade de dados, ainda se observaram inúmeras dificuldades na organização destes, de modo a se conseguir utilizar a capacidade total dos mesmos.

Grande quantidade de dados e consultas complexas levam a um aumento inaceitável do tempo de resposta e enfatizam a necessidade de criar um mecanismo eficiente para responder rapidamente a essas consultas (HUANG; DUY; FANG, 2014).

Com toda essa necessidade chegou-se ao conceito de *Data Warehouse* (DW) ou simplesmente Depósito de Dados que é uma coleção de dados orientados por assuntos, integrados, variáveis com o tempo e não voláteis, que possuem a função

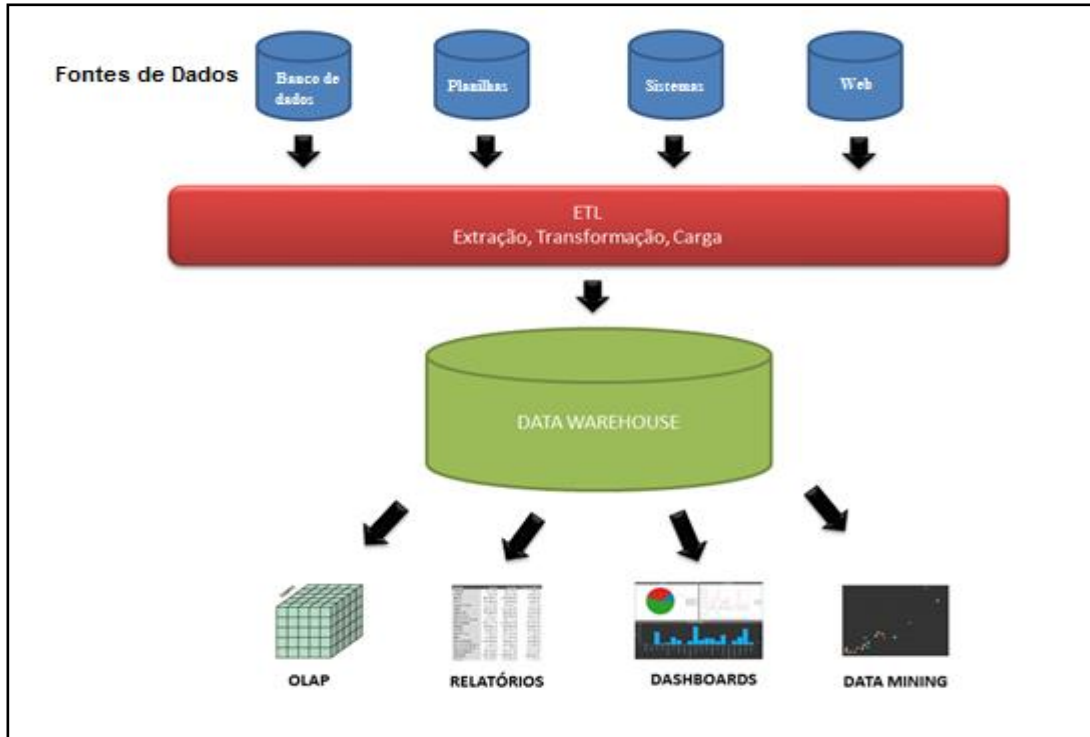
de dar suporte ao processo de tomada de decisões de maneira fácil e rápida (INMON, 1997).

Para Silberschatz, Korth e Sudarshan (1999), um *Data Warehouse* é um repositório de dados obtidos de diversas fontes, armazenados sob um esquema de banco de dados comum, em um só local. Após serem coletados, os dados são armazenados por um período longo, permitindo acesso a dados históricos.

O *Data Warehouse* possui algumas diferenças quando comparado ao banco de dados operacional, entre essas diferenças está a integração das informações fornecidas por vários bancos de dados que é realizada pelo *Data Warehouse* e também este é usado para fins de análise, geralmente para processamento analítico *on-line* (Olap), em que consultas *ad hoc* complexas que computam valores agregados em grandes quantidades de dados ocorrem com frequência, enquanto os usuários nunca iniciam transações de atualização. Dessa forma, as características de acesso a dados observadas nos *Data Warehouses*, diferentes daquelas em bancos de dados operacionais, no qual se tem transações curtas (geralmente pré-definidas e parametrizadas) que exigem alto desempenho e tempos de resposta curtos (LECHTENBÖRGER; VOSSSEN, 2003).

É certa a grande importância da análise de dados no mundo atual baseado em informação e conhecimento no qual se tem o Olap como uma das tecnologias envolvidas na análise de dados empregados por *Data Warehouses* (MANSMANN et al., 2014).

O DW é um ambiente que possibilita realizar consultas, análises e gerar informações, tendo como finalidade extrair, integrar, limpar e dar consistência aos diversos dados provenientes de locais distintos. Ele consolida e dimensiona os dados a partir da organização destes, sendo um local seguro que permite que dados confiáveis sejam publicados, auxiliando a tomada de decisão (KIMBALL; ROSS, 2002). Esse ambiente pode ser visualizado através da figura 1.

Figura 1 – Ambiente *Data Warehouse*

Fonte: Elaborada pela autora.

O desenvolvimento de um DW permite com que seja realizada a extração de informações importantes dos dados dispersos em alguns sistemas de informação em um armazenamento integrado e centralizado, apoiando a necessidade de um histórico de dados. Estes dados integrados podem ser utilizados para atividades de entrega de informações que podem ser vistas de várias dimensões, além de permitir definir o nível de detalhe desejado (SANTOSO; YULIA, 2017).

A extração é o primeiro passo no processo de inserção dos dados no ambiente do DW, isto é, leitura e compreensão dos dados de origem e cópia dos dados necessários para o DW para futura manipulação. Após a extração, ocorre a transformação na qual os dados são limpos, isto é, o momento no qual ocorre a correção da falta de alguns elementos, problemas de conflito, dados duplicados, etc. O último passo é o processo no qual a carga dos dados é feita no DW (KIMBALL; ROSS, 2002).

Os dados usados no processo de extração, transformação e carga (ETC) podem vir de uma variedade de fontes, incluindo aplicativos de planejamento de

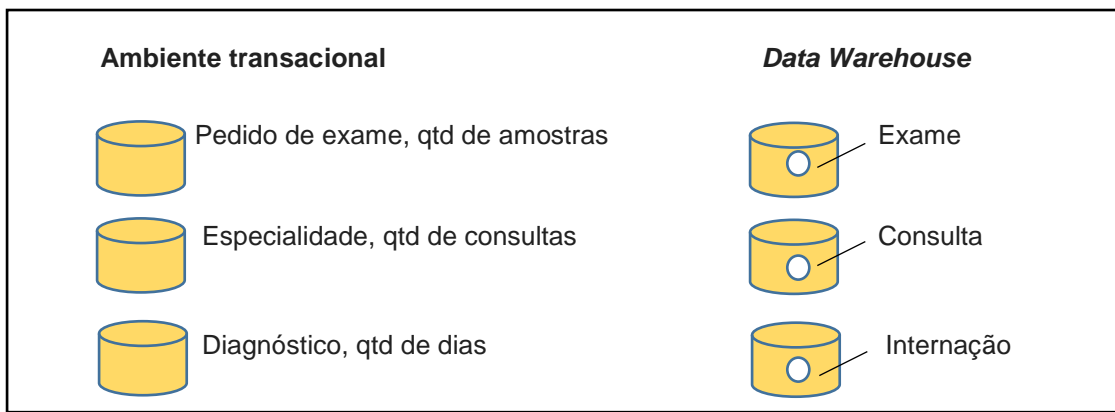


recursos empresariais (ERP), arquivos simples e planilhas eletrônicas (SANTOSO; YULIA, 2017).

De acordo com Nery (2012), as principais características de um DW são:

- Orientação por assunto – Significa que os dados são armazenados de forma agrupada por assuntos mais importantes, diferentemente dos sistemas operacionais, onde estes são orientados a processos desenvolvidos para manter as transações realizadas diariamente conforme mostrado na figura 2.

Figura 2 – Exemplo de dados baseados em assuntos de negócio

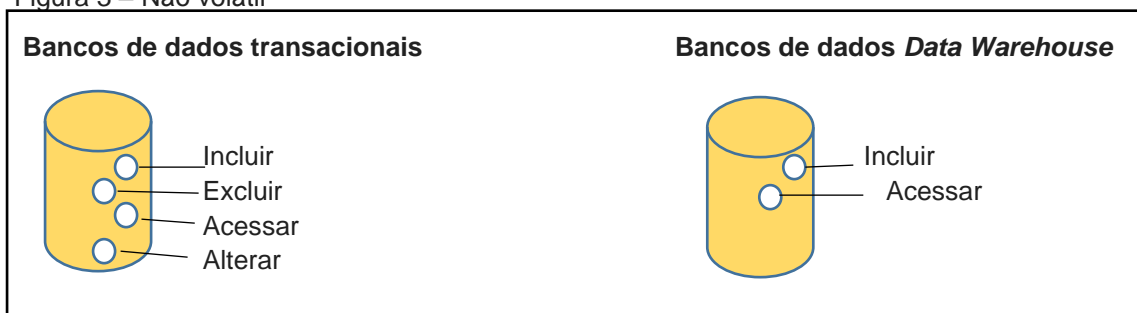


Fonte adaptada (Nery, 2012, p.27).

- Variação de tempo – Os dados são precisos em relação ao tempo, representando resultados operacionais em um determinado momento de tempo em que foram capturados, são como fotografias de um conjunto estático de registros de uma ou mais tabelas capturadas em um momento de tempo determinado.

- Não volátil – Somente duas operações básicas são realizadas, a carga dos dados (inicial e incremental) e o acesso a esses em modo de leitura, conforme mostrado na figura 3.

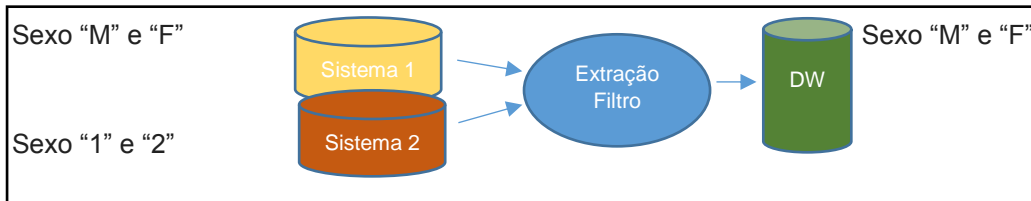
Figura 3 – Não volátil



Fonte: (Nery, 2012, p.29).

- Integração – Essa característica é fundamental para termos a unicidade de dados. A existência de fontes de dados com diferentes padrões de codificação quando carregados no DW são solucionados pelo processo de filtragem e agregação conforme mostrado na figura 4.

Figura 4 – Exemplo de integração de dados

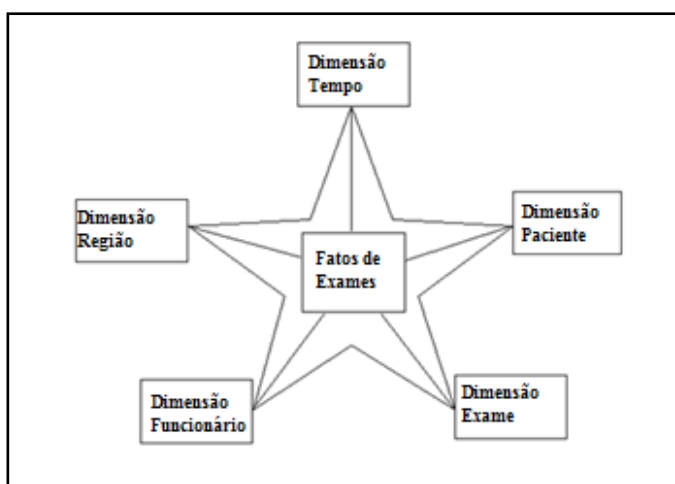


Fonte: (Nery, 2012, p.30).

A utilização do DW permite armazenar, consultar e atualizar uma grande variedade de dados sem a necessidade de realizar mudanças na estrutura principal do banco de dados (KOEHLER et al., 2005).

Os dados no DW são modelados de forma diferenciada, isto é, estão em um modelo multidimensional composto de uma tabela de fato central e um conjunto de tabelas de dimensões ao redor, cada uma correspondendo a um dos componentes da tabela de fatos, formando assim o modelo estrela ou *Star schema*, conforme mostrado na figura 5. O modelo estrela pode ser refinado em modelo floco de neve ou *Snowflake schema*, isto é, permite que as tabelas de dimensões tenham tabelas subdimensionadas (LEVENE; LOIZOU, 2003).

Figura 5 – Modelo Estrela (*Star schema*)

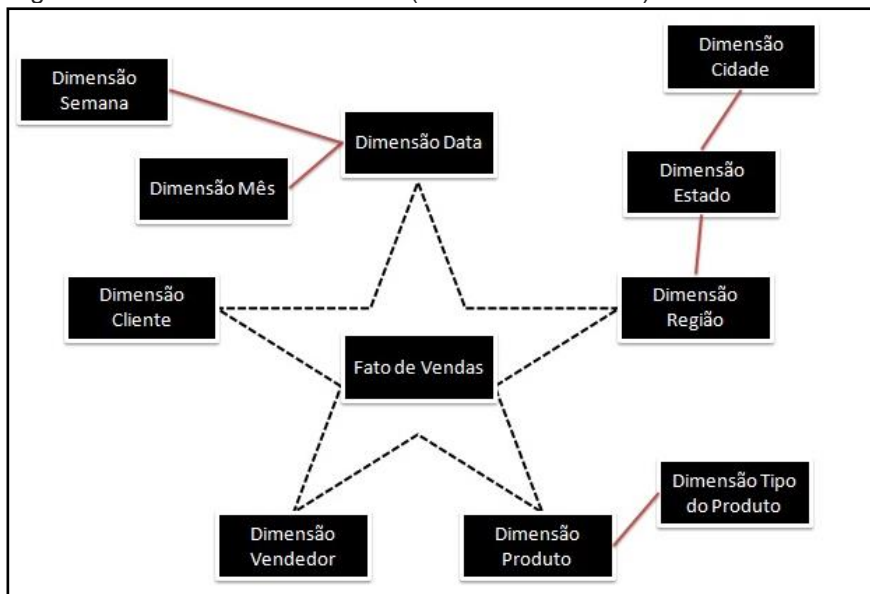


Fonte adaptada: (Nery, 2012, p.92).

Os relacionamentos entre a tabela fato e as dimensões são simples ligações em que cada tabela de dimensão está associada à tabela de fatos em um relacionamento de um para muitos, isto é, eu posso ter uma tabela fato ligada a várias dimensões (NERY, 2012).

O modelo floco de neve, exemplificado na figura 6, é mais complexo que o modelo estrela. Nele, os dados das dimensões são agrupados em várias tabelas, visando eliminar a redundância existente no modelo estrela. Este modelo busca uma economia de espaço, porém como consequência, o resultado das consultas se torna mais complexo e com um baixo desempenho (RASLAN; CALAZANS, 2014).

Figura 6 – Modelo Floco de Neve (*Snowflake schema*)



Fonte: (Nery, 2012, p.94).

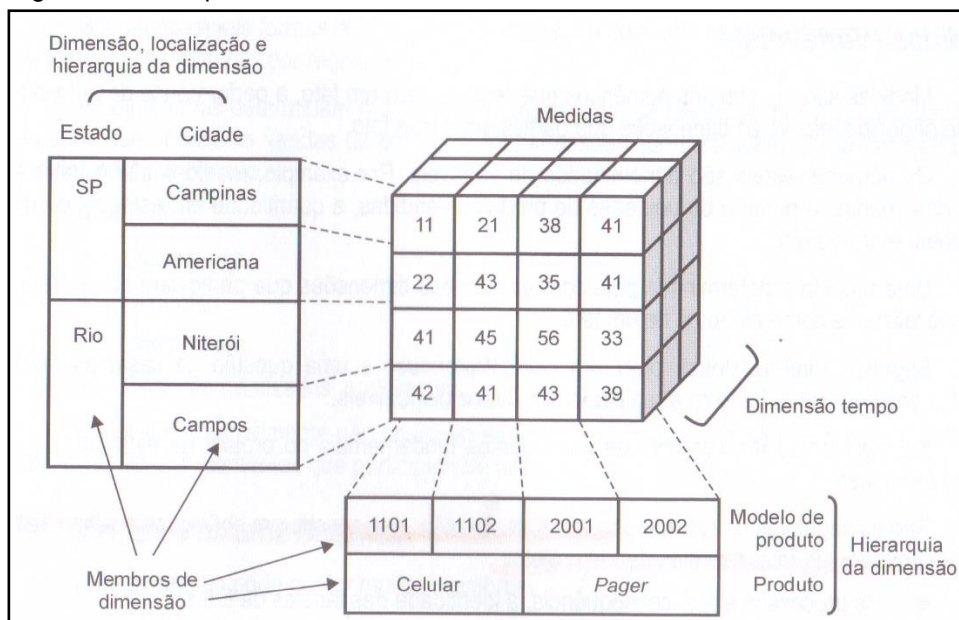
Os fatos representam sujeitos de análise (por exemplo, "vendas") e são descritos usando medidas (por exemplo, "preço"). As dimensões representam perspectivas de análise (por exemplo, "produtos"). Eles são organizados em diferentes hierarquias de dimensão, que são compostas de níveis de agregação relacionados, permitindo a visualização de indicadores em diferentes granularidades. Os indicadores (por exemplo, "volume de negócios") são calculados aplicando funções agregadas (por exemplo, Soma, Média, Mínimo e Máximo) para medir valores (BOULIL; BIMONTE; PINET, 2015).

Os dados no *Data Warehouse* estão na forma de cubos de dados, isso permite que os mesmos sejam modelados e visualizados em várias dimensões (HUANG; DUY; FANG, 2014).

O cubo de dados é uma matriz multidimensional que permite aos usuários explorar e analisar uma coleção de dados de muitas perspectivas diferentes e geralmente de várias dimensões de uma só vez (HUANG; CHOU; SENG, 2007).

De acordo com Nery (2012), a palavra cubo é apenas uma metáfora que realiza uma aproximação da forma como os dados estão organizados, e não a verdadeira expressão da realidade conforme exemplificado na figura 7.

Figura 7 – Exemplo de um Cubo



Fonte: (Nery, 2012, p.82).

Interpretando o exemplo da figura 7, pode-se verificar que as medidas informam o volume de vendas realizado, isso é determinado através da combinação das três dimensões (localização, produto e tempo). Por exemplo, em um determinado período do tempo, na cidade de Niterói foram vendidos 41 celulares do modelo 1101.

Os *Data Warehouses* suportam cubos dimensionais onde os dados podem ser combinados para fornecer resultados previamente desconhecidos. Os cubos de dados são explorados com operadores Olap, estes permitem uma navegação dentro de hierarquias das dimensões e dados agregados (BERRAHOU et al., 2015).

O objetivo principal do Olap é fornecer respostas em um menor tempo possível. Por exemplo, em um banco de dados que contém dados de desempenho de

vendedores, pode-se calcular *on-line* a quantidade de vendas feitas nos últimos 10 dias, incluindo apenas vendedores com dois ou mais anos de experiência. Usando um banco de dados relacional contendo informações de vendas, tal cálculo pode ser muito demorado. Usando o Olap, entretanto, o cálculo é tipicamente feito *on-line*.

De acordo com Kaser e Lemire (2006) existem três tipos de mecanismos Olap:

- Olap relacional (Rolap - *Relational Online Analytical Processing*);
- Olap multidimensional (Molap - *Multidimensional Online Analytical Processing*);
- Olap híbrido (Holap - *Hybrid Online Analytical Processing*).

No Rolap, os dados são armazenados em um banco de dados relacional, enquanto que com o Molap, um grande array multidimensional é construído com os dados. Muitos fornecedores como *Speedware*, *Hyperion*, IBM e *Microsoft* utilizam o Olap híbrido (Holap), armazenando regiões densas do cubo usando Molap e armazenando o restante usando uma abordagem Rolap (KASER; LEMIRE, 2006).

Song et al. (2015) relatam as diferenças entre os três tipos de Olap da seguinte forma:

1 Os servidores Molap suportam diretamente a visualização multidimensional de dados através de um mecanismo de armazenamento que usa a abstração de matriz multidimensional, enquanto em Rolap, o modelo multidimensional e suas operações precisam ser mapeados em relacionamento e consultas SQL;

2 - No Molap, normalmente calcula previamente grandes cubos de dados para acelerar o processamento de consultas, enquanto o Rolap confia nas técnicas de armazenamento de dados para acelerar o processamento de consultas relacionais;

3 - Molap sofre de má utilização de armazenamento, especialmente quando o conjunto de dados é esparsos;

4 - Como o Rolap depende mais do banco de dados para realizar cálculos, ele tem mais limitação na função especializada que pode usar;

5 - Holap combina Rolap e Molap, dividindo o armazenamento de dados em um Molap e um armazenamento relacional.

As consultas Olap são criadas para explorar e executar análises complexas, gerando interativamente diferentes visualizações desses dados de diferentes perspectivas. As operações Olap têm como objetivo selecionar planos a partir do DW e alterar a granularidade na qual os dados são observados. As principais operações Olap exemplificadas na figura 8 são: “roll-up”, para passar para um nível mais abstrato em uma dimensão; “drill-down”, o inverso do “roll-up”, para mover para dados mais

detalhados; “slice and dice”, para focar em dados relacionados a subconjuntos de valores de dimensão; e “pivot”, para girar os eixos nos quais os dados são exibidos para fornecer uma exibição alternativa dos dados (BOUADI et al., 2017).

Figura 8 – Exemplo de uma operação roll up

| Volume de Internações |            | 2017   |        |        |        |
|-----------------------|------------|--------|--------|--------|--------|
|                       |            | Trim.1 | Trim.2 | Trim.3 | Trim.4 |
| Tipos de Leito        | CTI        | 30     | 22     | 15     | 20     |
|                       | Enfermaria | 90     | 78     | 88     | 97     |

*Roll Up*

| Volume de Internações |            | Trimestre 1 |           |       |
|-----------------------|------------|-------------|-----------|-------|
|                       |            | Janeiro     | Fevereiro | Março |
| Tipos de Leito        | CTI        | 10          | 8         | 12    |
|                       | Enfermaria | 35          | 28        | 27    |

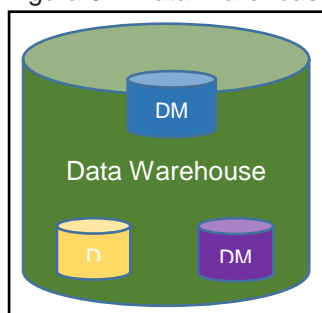
Fonte: Elaborada pela autora.

Os DW podem armazenar grandes quantidades de informação e são projetados para tamanhos de centenas de gigabytes até terabytes e podem ser divididos em unidades lógicas menores que são chamadas de *Data Marts* (DM) (RASLAN; CALAZANS, 2014).

Segundo Nery (2012), o *Data Mart*, visualizado na figura 9, é um subconjunto de dados do DW criados para um departamento ou área específica tendo como principal vantagem a possibilidade de retorno rápido, com um grande envolvimento do usuário final, capaz de avaliar os benefícios extraídos de seu investimento.

Pelo fato do *Data Mart* ser um subconjunto do *Data Warehouse*, todas as definições e atribuições existentes para o *Data Warehouse* podem ser atribuídas ao *Data Mart*.

Figura 9 – *Data Warehouse* e *Data Mart*



Fonte: Elaborada pela autora.

A aplicabilidade do DW é grande: sua utilização pode ser feita por qualquer instituição que deseje organizar seus dados de maneira integrada, com uma única versão da verdade, em um histórico variável no tempo, gerando uma única fonte de dados.

Fischer et al. (2006) construíram um sistema de *Data Warehouse* para analisar famílias de proteínas. Este sistema foi projetado para comparar várias famílias de proteínas para inferir especificidades do substrato e para orientar a engenharia de enzimas otimizadas.

Zhou et al. (2010) abordaram o desenvolvimento de um sistema de DW clínico para ser utilizado na Medicina Tradicional Chinesa, esta é uma disciplina científica que trata da fisiologia e patologia humanas, do diagnóstico, tratamento e prevenção das doenças através da prática clínica. Este sistema foi desenvolvido contendo dados estruturados do registro médico eletrônico de modo que possibilitasse a descoberta do conhecimento médico e suporte a decisão clínica. Com a construção deste sistema foi possível identificar o conhecimento médico comum e repetitivo e estratégias clínicas eficazes.

A utilização do DW na área da saúde apresenta-se como um grande potencial de benefícios para as comunidades científicas e para a sociedade em geral.

A reutilização de dados de saúde na pesquisa em saúde pública tem sido de grande valia. Pesquisas derivadas da aplicação de tecnologias de informação e comunicação em medicina, operam em um contexto que envolve a globalização da coleta, compartilhamento, armazenamento, transferência e reutilização de dados de saúde. A informatização dos dados de saúde dentro dos Sistemas de Informação Clínica deve permitir a reutilização dos mesmos para fins de pesquisa clínica e saúde pública. Um dos objetos que permitem a integração de sistemas de informações de pesquisa e assistência médica é a utilização de um DW (LAMAS et al., 2015).

A Pediatrix Medical Group, desenvolveu um sistema de *Data Warehouse* que foi patentado como BabySteps, esse desenvolvimento se deu por conta de uma grande quantidade de pacientes neonatos em unidades de terapia intensiva nos EUA. A empresa reconheceu que poderia realizar a extração de dados significativos para medir os resultados, assim como analisar como as várias práticas feitas poderiam afetar esses resultados. Acredita-se que o BabySteps seja um dos maiores repositórios de dados sobre o neonato no mundo, contendo informações sobre mais de 1.025.000 bebês e aproximadamente 18.600.000 pacientes-dia. Devido à grande

quantidade dos dados coletados, o BabySteps foi utilizado não apenas dentro da organização para novas observações de pesquisas, mas também pela *Food and Drug Administration* (FDA), pelo *National Institute of Health* (NIH), pelo *National Institute of Child Health and Human Development* (NICHD), e numerosos programas acadêmicos de neonatologia (SPITZER; ELLSBURY; CLARK, 2015).

Assim, este projeto propõe a implementação de um *Data Mart* para gerenciamento dos dados obtidos através da pesquisa, incluindo a criação de um modelo que fará a extração, transformação e carga dos dados em um local seguro criado com o auxílio de ferramentas *Open Source*, que possibilitem obter gráficos, relatórios e pesquisas customizadas sobre a informação desejada.



## 2 JUSTIFICATIVA

A medida que se caminha para uma sociedade cada vez mais orientada para a informação, a forma de como os dados são organizados com o objetivo de maximizar sua utilização torna-se uma tarefa muito importante, uma vez que a correta manipulação destes, além de tornar mais dinâmico os diversos setores de uma organização, pode também fornecer indicadores que auxiliam a gestão estratégica da mesma.

Uma grande quantidade de relatórios utilizados hoje pelo Laboratório de Pesquisa Clínica e Vigilância de Leishmanioses (LaPClinVigiLeish ), é obtida através dos bancos de dados dos sistemas transacionais, sistemas de suporte para atividades do dia a dia da organização utilizados no nível operacional, existentes no Instituto de Pesquisa, que proporcionam consultas pré-definidas e foram desenvolvidos para gerar um padrão estático, dessa forma, os pesquisadores podem apenas obter os dados no formato pré-definido, ou seja, com colunas e linhas imutáveis, além disso, esses relatórios não possuem capacidade para retornar dados históricos referentes a longos períodos.

O LaPClinVigiLeish utiliza dados operacionais provenientes do Sistema de Prontuário Eletrônico do INI (SIPEC), planilhas eletrônicas e banco de dados Access®, porém, devido à falta de padronização e integração dos dados existentes nas diversas fontes, tornava-se difícil a elaboração de relatórios que contenham dados integrados, padronizados e relevantes à pesquisa, tornando mais complexa a análise dos mesmos e conseqüentemente afeta o processo de tomada de decisão. Organizar estes dados tornou-se um grande desafio, eles representavam um papel crítico para o pesquisador que lida com eles.

A implantação de um *Data Mart* fez uma sólida e concisa integração dos dados, ele é responsável por integrar e consolidar os dados de todas as fontes de dados selecionadas previamente que são heterogêneas, agrupando, filtrando e limpando esses dados, tornando-os aptos para análise e suporte à tomada de decisão. Acrescenta-se ainda a vantagem de que a obtenção da informação tornou-se mais rápida a partir de um grande volume de dados operacionais, sendo possível realizar o cruzamento de informações de modo mais dinâmico e intuitivo, além de proporcionar uma estrutura que suporta esse grande volume de dados.

Os pesquisadores ao realizarem a análise dos dados através das ferramentas de *Data Warehouse* (DW) podem decidir quais serão as linhas e quais serão as colunas que farão parte da sua análise, podendo mudar totalmente a visualização dos dados na planilha com um mínimo de tempo e esforço, foi disponibilizada também a opção de gerar gráficos a partir destes dados sem a necessidade de realizar nenhuma programação no código.

Diante do exposto, tornou-se necessária a implantação de um *Data Mart* que faz a integração dos dados existentes nas diversas fontes utilizadas hoje pelo LaPClinVigiLeish , contribuindo diretamente para a otimização dos recursos disponíveis aumentando a efetividade do processo de tomada de decisões e descoberta das informações de maneira fácil e rápida, e indiretamente na redução da morbidade, deformidades e óbitos em pacientes com leishmanioses que vem sendo realizada através das pesquisas por este laboratório.

### 3 OBJETIVOS

#### 3.1 OBJETIVO GERAL

Construir e implementar o *Data Mart* de modo a apoiar à tomada de decisão dos pesquisadores em um menor tempo com uma melhor qualidade na análise dos dados da pesquisa.

#### 3.2 OBJETIVOS ESPECÍFICOS

**3.2.1 Identificar e descrever os processos de coleta, armazenamento e processamento dos dados do Laboratório de Pesquisa Clínica e Vigilância em Leishmanioses;**

**3.2.2 Desenvolver um material de apoio aos usuários do *Data Mart*.**

## **4 METODOLOGIA**

### **4.1 DESENHO DO ESTUDO**

Trata-se de um estudo descritivo cuja técnica utilizada foi estudo de caso envolvendo o Laboratório de Pesquisa Clínica e Vigilância em Leishmanioses com a aplicação de uma metodologia da informação envolvendo análise de dados secundários.

### **4.2 CASUÍSTICA**

#### **4.2.1 Objeto do Estudo**

O objeto de estudo foi constituído por dados provenientes do Sistema de Prontuário Eletrônico do INI, planilhas eletrônicas e banco de dados Access®.

#### **4.2.2 Critérios de inclusão e de exclusão**

Na construção do *Data Mart* foram utilizados somente dados estruturados, isto é, dados que são organizados em linhas e colunas.

Não foram utilizados dados textuais por necessitarem de um complexo processo de extração, não sendo viável realizá-lo no período de desenvolvimento deste projeto.

#### **4.2.3 Amostra**

A amostra utilizada foi composta por 2 planilhas eletrônicas, 11 tabelas do banco de dados onde estão armazenadas as informações do Prontuário Eletrônico dos pacientes e 2 tabelas do banco de dados Access® utilizado pelo LaPClinVigiLeish. Esses dados foram obtidos no intervalo de tempo de 9 anos, de 1 de janeiro de 2010 a 31 de dezembro de 2019.

## 4.3 MATERIAIS, PROCEDIMENTOS E TÉCNICAS

A metodologia descrita a seguir tem como objetivo a orientação para o desenvolvimento evolutivo do *Data Mart* para LaPClinVigiLeish. Ela está dividida em etapas bem caracterizadas pelo agrupamento dos principais processos, descrevendo a finalidade e as principais técnicas de cada fase, identificando, assim, os pontos críticos.

### 4.3.1 Instalação dos softwares da *Suíte Pentaho*

O primeiro procedimento realizado foi a instalação e configuração dos softwares da *Suíte Pentaho* em um servidor de teste localizado no Setor de Informática para posterior migração para o servidor que contém o ambiente de Produção.

Foram utilizados o Pentaho *Data Integration*, o Pentaho *User Console*, o Pentaho *Schema Workbench* e o Pentaho *Server*, todos na versão 8.0 extraídos do endereço <https://sourceforge.net/projects/pentaho/files/Pentaho%208.0/>.

O Pentaho *Data Integration* é o componente da *suíte Pentaho* usado para criar os processos de extração, transformação e carga (ETC) que alimentam o banco de dados. Este é formado por duas categorias de artefatos, Jobs e Transformações, e estes artefatos são construídos por meio de sua interface gráfica, o *Spoon*, este facilita na concepção de rotinas e lógica ETC.

O Pentaho *Schema Workbench* é uma interface de designer que permitiu criar e testar modelos de cubos Olap que utilizam tabelas Fato e Dimensão existentes no banco Relacional. Não requer que um cubo físico real seja construído ou mantido, apenas o modelo de metadados é criado.

O Pentaho *User Console* é um ambiente de design baseado na Web, onde os dados podem ser analisados e os relatórios criados de modo interativo.

O Pentaho *Server* é responsável por toda a camada servidora da plataforma do Pentaho, fornece a arquitetura e infra-estrutura necessária para a construção de soluções para *business intelligence* (BI).

A escolha da *Suíte Pentaho* foi baseada no alto desempenho apresentado por esta ferramenta e por ser *Open Source*, além de outras facilidades como geração de

gráficos de forma dinâmica e permitir a exportação dos dados para outros formatos como “xls” e “pdf”.

### **4.3.2 Coleta de Dados**

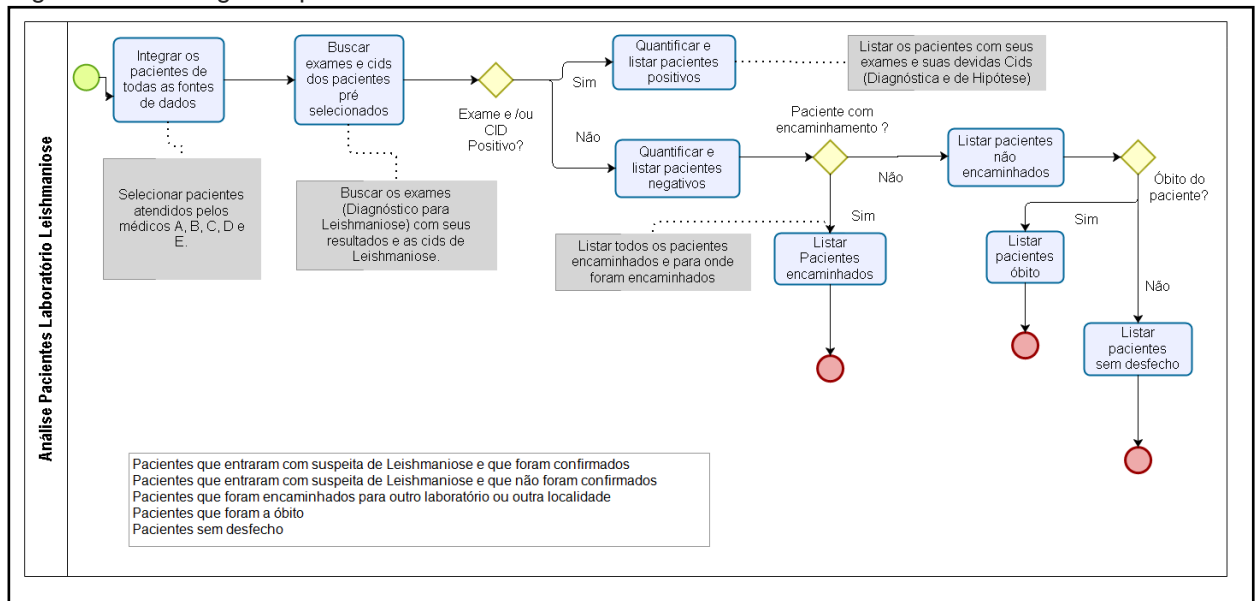
A etapa de coleta de dados define e documenta as necessidades dos pesquisadores para alcançar os objetivos do projeto. Nesta etapa foi confeccionado um questionário (APÊNDICE B) que foi aplicado aos pesquisadores previamente selecionados pela pesquisadora responsável pelo estudo, essa seleção foi realizada de acordo com o perfil de atuação do profissional dentro do laboratório.

#### **4.3.2.1 Análise das fontes de informação e Identificação dos dados para extração**

Como instrumento de coleta de dados foram realizadas 02 (duas) entrevistas de forma presencial com 5 pesquisadores onde foi aplicado um questionário contendo 13 perguntas sobre os relatórios, dados e análises que são realizadas rotineiramente, visando identificar e selecionar as fontes de dados que podiam conter informações pertinentes a serem integradas ao modelo inicialmente proposto pela pesquisadora responsável pelo projeto. As entrevistas foram realizadas de forma individualizada, na sala do LaPClinVigiLeish ,LaPClinVigiLeish, com tempo de 1 hora e foram agendadas de acordo com a conveniência e disponibilidade de cada pesquisador. A análise de documentos foi uma das fontes utilizadas neste projeto, a fim de permitir o entendimento de situações omitidas ou não questionadas.

Após a coleta de dados ter sido realizada, foi desenhado um processo visando alcançar algumas respostas através da análise dos dados da pesquisa conforme mostrado na figura 10. Essa análise foi montada sob a ótica de garimpar informações de pacientes que realizaram exames específicos para o diagnóstico das leishmanioses, auditando os pacientes já existentes no banco de dados e identificando novos pacientes.

Figura 10 – Fluxograma para a Análise dos Pacientes do Laboratório de Leishmanioses do INI



Fonte: Elaborada pela autora.

A análise dos dados foi iniciada pela pesquisa em tabelas contidas no banco de dados onde estão armazenadas as informações do Sistema de Prontuário Eletrônico do INI (SIPEC) e no banco de dados Access® do LaPClinVigiLeish.

A análise das bases de dados foi mais complexa, pois nem todas continham documentação e nem mesmo seguiam os padrões de normalização de dados, isto é, alguns dados contidos nessas bases estão armazenados de forma redundante.

Foi realizada uma análise dos processos do LaPClinVigiLeish, onde foram observadas as regras de negócio, estas, originaram as transformações na fase de extração de dados. Em geral, as transformações que ocorreram, foram por questões técnicas, mudanças de formatos ou uma visão de negócio.

#### 4.3.3 Modelagem de dados

Após terem sido identificados os dados dos sistemas transacionais, foi iniciada a modelagem dos mesmos para armazenamento no *Data Mart*.

O *Data Mart* construído é um grande repositório de dados contendo os dados previamente selecionados.

#### **4.3.3.1 Definição da granularidade dos dados**

Segundo Nery (2012), na modelagem física de dados, um dos aspectos mais importantes é a definição da granularidade dos dados, pelo fato dela afetar profundamente o volume de dados que terá o *Data Mart* assim como o tipo de consulta que será realizada.

Nesta etapa foi definido o nível de sumarização dos elementos e de detalhes disponíveis nos dados previamente selecionados, sendo possível visualizar a maior quantidade de detalhes existentes, tornando menor o nível de granularidade. Esta característica, dos dados estarem na menor granularidade, permite uma melhor flexibilidade nas consultas a serem efetuadas, bem como fornece um meio simplificado de extração e disponibilização dos dados exportados, sendo possível identificar o paciente em estudo.

#### **4.3.3.2 Construção do Modelo Estrela**

O modelo estrela foi o modelo de dados escolhido devido a sua maior velocidade no processamento das consultas, este apresenta uma grande entidade central (tabela fato) e um conjunto de entidades menores (dimensões) arranjadas ao redor da entidade central. O relacionamento dos fatos com as dimensões são simples ligações entre as duas entidades em um relacionamento de um para muitos no sentido da dimensão para o fato. Os elementos das tabelas de dimensão estarão relacionados com vários itens da tabela de fato, caracterizando assim a necessidade de agregação.

De acordo com Nery (2012), um fato é um grupo de itens de dados, composta de dados de medidas e de contexto, onde cada um representa um item, uma transação ou um evento de negócio possuindo como característica a representação por valores numéricos e implementado em tabelas denominadas “tabelas de fato”. As dimensões são os elementos que participam de um fato, são as formas de visualizar os “por” dos dados: “por mês”, “por país”, etc. É o momento em que são respondidas as quatro perguntas:

Quando? (Período de tempo referente à análise), O quê? (O principal objeto de análise), Onde? (Onde aconteceu o fato) e Quem? (Quem executou o fato).

As medidas são os valores numéricos que representam um fato. São o valor em reais do custo com os exames para diagnóstico de leishmanioses, a quantidade



de pacientes notificados ao ano, entre outros. Foram definidas as primeiras dimensões, a tabela fato e a medida contida nela.

Os sistemas transacionais diferem do *Data Mart* no que diz respeito ao tempo, sendo o primeiro um *status* instantâneo, atualizado sempre à medida que as transações são concretizadas. No *Data Mart* os dados possuem uma marca de tempo ao serem armazenados, onde são observadas as mudanças de comportamento, seja de dados dos pacientes ou das amostras biológicas. Ao ter sido criada a tabela de fato no *Data Mart* pode ser observada a necessidade do atributo de tempo, dessa forma, a dimensão “tempo” teve grande importância para o modelo de dados do *Data Mart*. Modelar os dados dos pacientes de forma que seja possível observar as mudanças do mesmo, ao longo do tempo, é fundamental para atender a esta finalidade.

Existiu uma grande preocupação com relação ao armazenamento dos dados no *Data Mart*, de modo que fosse garantida uma alta performance e um baixo custo de espaço para o armazenamento. Foi realizado o dimensionamento dos tipos de dados das chaves das dimensões e através deste que o espaço necessário para o armazenamento foi determinado.

A partir das dimensões e fato criadas, foi gerado o modelo estrela utilizado neste projeto que permite realizar consultas pertinentes aos pacientes internos do INI como buscar os exames e CID's (Classificação Internacional de Doenças), quantificar e listar os pacientes positivos e negativos, verificar os encaminhamentos realizados e identificar os pacientes que não possuem desfecho. O modelo estrela criado pode ser observado na figura 11.

A dimensão “Tempo” (tbl\_dmns\_tmp) fornece uma identidade temporal aos fatos aos quais a mesma está relacionada.

A dimensão “Consulta” (tbl\_dmns\_consulta) contém todas as consultas realizadas pelos médicos previamente selecionados no período de 1/1/2010 a 31/1/2019.

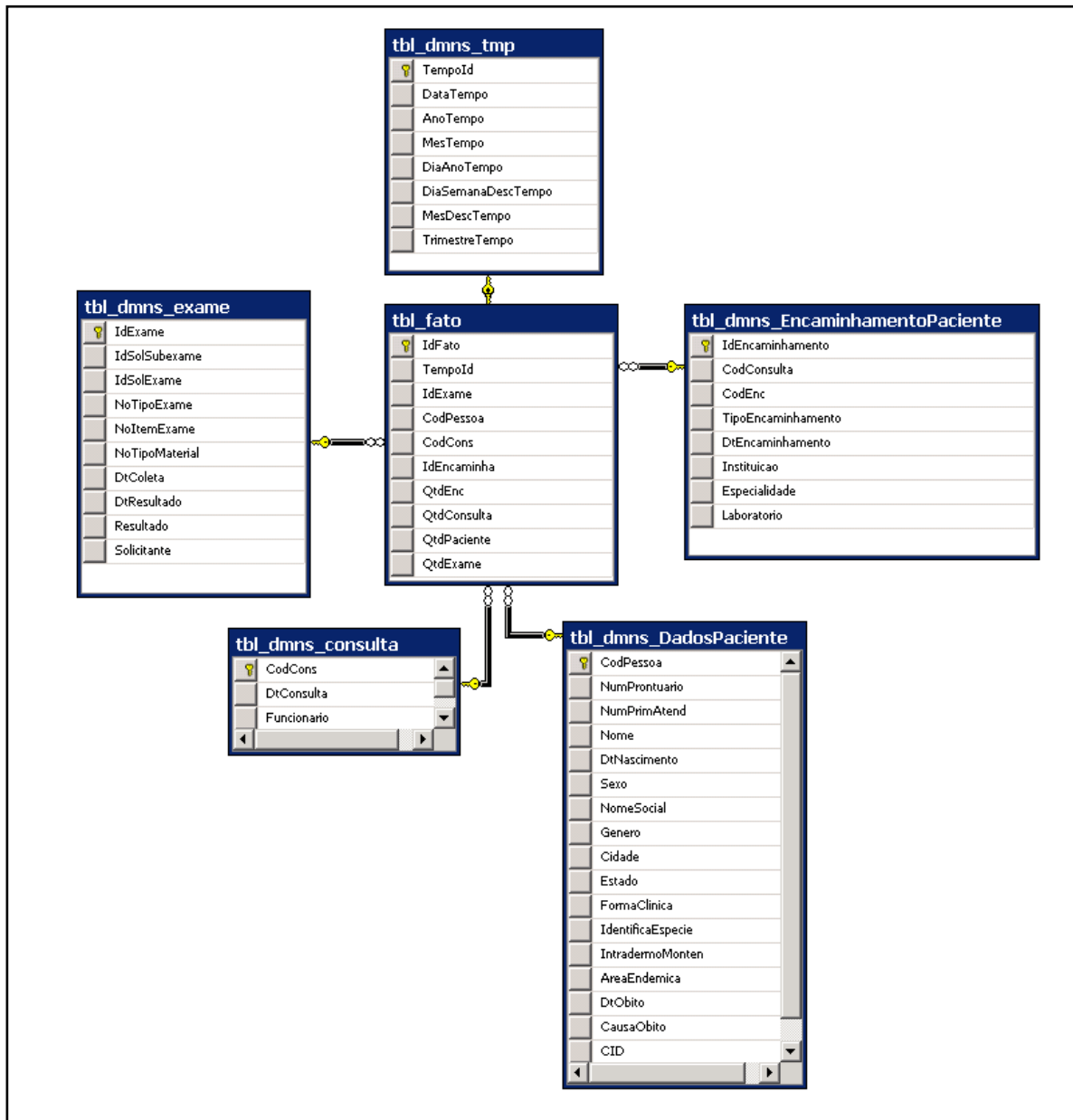
A dimensão “Paciente” (tbl\_dmns\_dadospaciente) caracteriza os pacientes internos que tenham sido consultados conforme os critérios da dimensão Consulta.

A dimensão “Encaminhamento” (tbl\_dmns\_encaminhamentopaciente) representa todos os encaminhamentos realizados para os pacientes existentes na dimensão Paciente.

A dimensão “Exame” (tbl\_dmns\_exame) é composta pelos exames finalizados de diagnóstico para leishmanioses dos pacientes contidos na dimensão Paciente.

A fato leishmaniose (tbl\_fato) representa a quantidade de pacientes, encaminhamentos, consultas e exames em um determinado tempo. Esta tabela de fato é composta por cinco dimensões: tempo, encaminhamento, paciente, consulta e exame.

Figura 11 – Modelo estrela Laboratório Leishmaniose



Fonte: Elaborada pela autora.

#### 4.3.4 Extração dos dados

Após ter sido realizada a primeira modelagem e definida a granularidade dos dados selecionados, foi realizado o processo de extração destes, sendo esta fase a mais trabalhosa.

Esta etapa foi subdividida em duas: a extração e a importação dos dados, onde esta última é responsável pelas transformações dos dados.

##### **4.3.4.1 Definição dos extratores e Importação dos dados para o *Data Mart***

A execução da etapa 'Coleta de Dados' foi a base para a definição dos extratores e importações dos dados. No momento da definição dos extratores foi realizada a definição da seleção dos dados, parâmetros da interface, critérios de seleção, objetos (programas, arquivos, conectores) que deveriam ser criados, formato de saída e o local de armazenamento dos objetos.

Os dados foram extraídos com a utilização do *Spoon*, ferramenta de Extração, Transformação e Carga (ETC), este, foi usado para realizar a integração dos ambientes transacionais e o *Data Mart*, facilitando assim o processo de ETC no *Data Mart*.

Os bancos de dados Cenlab e Ambulatório, armazenados em um servidor de banco de dados no Serviço de Tecnologia da Informação e Comunicação (SETIC), armazenam as informações do Sistema de Prontuário Eletrônico do INI, estes, foram utilizados na obtenção de alguns dados de origem. As tabelas e *view* pertencentes a esses bancos utilizados durante o processo de ETC estão listados abaixo.

- ambulatorio.pessoa
- ambulatorio.consulta
- ambulatorio.agendamento
- ambulatorio.paciente
- ambulatorio.cidade
- ambulatorio.estado
- ambulatorio.funcionario
- ambulatorio.encaminhamento
- ambulatorio.especialidade

- ambulatorio.instituicao
- ambulatorio.projeto
- ambulatorio.linhapesq
- ambulatorio.pepes
- cenlab.vw\_ultimosresultadoitemexame\_por\_solexame2
- cenlab.SolltemExame
- cenlab.Situacao\_SolltemExame
- cenlab.SolExame

O banco de dados Access® (LTA30jan19.accdb) utilizado pelos pesquisadores do LaPClinVigiLeish forneceu algumas informações que não existem nos bancos de dados que armazenam as informações do SIPEC, as seguintes tabelas foram utilizadas:

- Cadastro
- Investigação Diagnóstica 1

As planilhas Excel® abaixo listadas foram usadas de modo a padronizar as informações semanais (DiasSemanasDesc.xls) e mensais (MesDesc.xls) conforme figuras 12 e 13 respectivamente.

Figura 12 – Planilha referente aos dias da semana

| A     | B             | C          |
|-------|---------------|------------|
| DiaNr | DiaDesc       | DiaDescAbv |
| 2     | SEGUNDA-FEIRA | SEG        |
| 3     | TERÇA-FEIRA   | TER        |
| 4     | QUARTA-FEIRA  | QUA        |
| 5     | QUINTA-FEIRA  | QUI        |
| 6     | SEXTA-FEIRA   | SEX        |
| 7     | SÁBADO        | SAB        |
| 1     | DOMINGO       | DOM        |

Fonte: Elaborada pela autora.

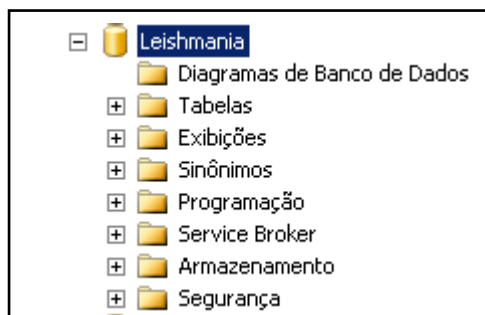
Figura 13 – Planilha referente aos meses do ano

| A     | B         | C           |
|-------|-----------|-------------|
| MesNr | MesDesc   | MesDescAbrv |
| 1     | JANEIRO   | JAN         |
| 2     | FEVEREIRO | FEV         |
| 3     | MARÇO     | MAR         |
| 4     | ABRIL     | ABR         |
| 5     | MAIO      | MAI         |
| 6     | JUNHO     | JUN         |
| 7     | JULHO     | JUL         |
| 8     | AGOSTO    | AGO         |
| 9     | SETEMBRO  | SET         |
| 10    | OUTUBRO   | OUT         |
| 11    | NOVEMBRO  | NOV         |
| 12    | DEZEMBRO  | DEZ         |

Fonte: Elaborada pela autora.

A partir das tabelas, *view* e planilhas Excel®, listadas acima, foi criado um novo banco de dados, Leishmania, que reúne todas as informações extraídas em um só local. Este, representado na figura 14, está armazenado nos mesmos servidores de banco de dados já existente no SETIC. O primeiro servidor contém os dados acessados pelos pesquisadores e o outro servidor é utilizado como backup dos dados armazenados no servidor principal.

Figura 14 – Bando de dados Leishmania



Fonte: Elaborada pela autora.

Para que os dados fossem extraídos obedecendo às regras de negócio, foi necessária a criação de alguns objetos, estes estão listados a seguir e seus códigos de criação disponíveis no APÊNDICE F.

- vw\_resultadoexame\_leishmaniose;
- tb\_resultadoexame\_leishmaniose;
- dta\_index\_tb\_resultadoexame\_leishmaniose;
- fn\_ret\_diag\_leishmaniose;

- fn\_ret\_dt\_primeiro\_exame;
- ret\_todos\_cids\_paciente\_leishmania;
- ret\_todos\_cids\_paciente\_leishmania\_com\_exame.

As falhas no processo de integração com o *Data Mart* são monitoradas através do *Execution Results* existente no *Spoon*, através dele é possível visualizar detalhes da execução dos extratores, as transformações executadas, incluindo a data e a hora em que foram executadas, seu status e resultados conforme mostrado na figura 15.

Figura 15 – Exemplo do *Execution Results* do extrator dos dados do paciente

| Execution Results   |                                |        |       |         |       |        |         |          |        |        |          |             |              |
|---|--------------------------------|--------|-------|---------|-------|--------|---------|----------|--------|--------|----------|-------------|--------------|
| Logging Execution History Step Metrics Performance Graph Metrics Preview data |                                |        |       |         |       |        |         |          |        |        |          |             |              |
| #   | Stepname                       | Copynr | Read  | Written | Input | Output | Updated | Rejected | Errors | Active | Time     | Speed (r/s) | input/output |
| 1   | Microsoft Access Invest Diag 1 | 0      | 0     | 1028    | 1028  | 0      | 0       | 0        | 0      | 0      | Finished | 0.1s        | 14.898       |
| 2   | Microsoft Access Cadastro      | 0      | 0     | 1020    | 1020  | 0      | 0       | 0        | 0      | 0      | Finished | 0.1s        | 16.190       |
| 3   | SQL Ambulatorio                | 0      | 0     | 7579    | 7579  | 0      | 0       | 0        | 0      | 0      | Finished | 1mn 28s     | 86           |
| 4   | If field value is null 2 2     | 0      | 7579  | 7579    | 0     | 0      | 0       | 0        | 0      | 0      | Finished | 1mn 28s     | 86           |
| 5   | Sexo                           | 0      | 7579  | 7579    | 0     | 0      | 0       | 0        | 0      | 0      | Finished | 1mn 28s     | 86           |
| 6   | Genero                         | 0      | 7579  | 15158   | 0     | 0      | 0       | 0        | 0      | 0      | Finished | 1mn 28s     | 172          |
| 7   | Select values                  | 0      | 7579  | 7579    | 0     | 0      | 0       | 0        | 0      | 0      | Finished | 1mn 28s     | 86           |
| 8   | Stream lookup 2                | 0      | 8607  | 7579    | 0     | 0      | 0       | 0        | 0      | 0      | Finished | 1mn 28s     | 98           |
| 9   | Identificacao Especie          | 0      | 7579  | 7579    | 0     | 0      | 0       | 0        | 0      | 0      | Finished | 1mn 28s     | 86           |
| 10  | Forma Clinica                  | 0      | 7579  | 7579    | 0     | 0      | 0       | 0        | 0      | 0      | Finished | 1mn 28s     | 86           |
| 11  | If field value is null         | 0      | 7579  | 7579    | 0     | 0      | 0       | 0        | 0      | 0      | Finished | 1mn 28s     | 86           |
| 12  | Stream lookup                  | 0      | 8599  | 7579    | 0     | 0      | 0       | 0        | 0      | 0      | Finished | 1mn 28s     | 98           |
| 13  | If field value is null 2       | 0      | 7579  | 7579    | 0     | 0      | 0       | 0        | 0      | 0      | Finished | 1mn 28s     | 86           |
| 14  | Merge Join                     | 0      | 15158 | 7579    | 0     | 0      | 0       | 0        | 0      | 0      | Finished | 1mn 28s     | 171          |
| 15  | Insert / Update                | 0      | 7579  | 7579    | 7579  | 7579   | 0       | 0        | 0      | 0      | Finished | 1mn 49s     | 69           |

Fonte: Elaborada pela autora.

A frequência com que a extração dos dados é realizada, bem como a quantidade de registros movimentada varia de acordo com os dados a serem extraídos. Os dados das dimensões e da fato foram definidos para serem extraídos mensalmente.

Algumas regras de negócio foram definidas para atender o processo de análise dos pacientes do Laboratório de Leishmaniose do INI, definido no item 4.3.2.1, são elas:

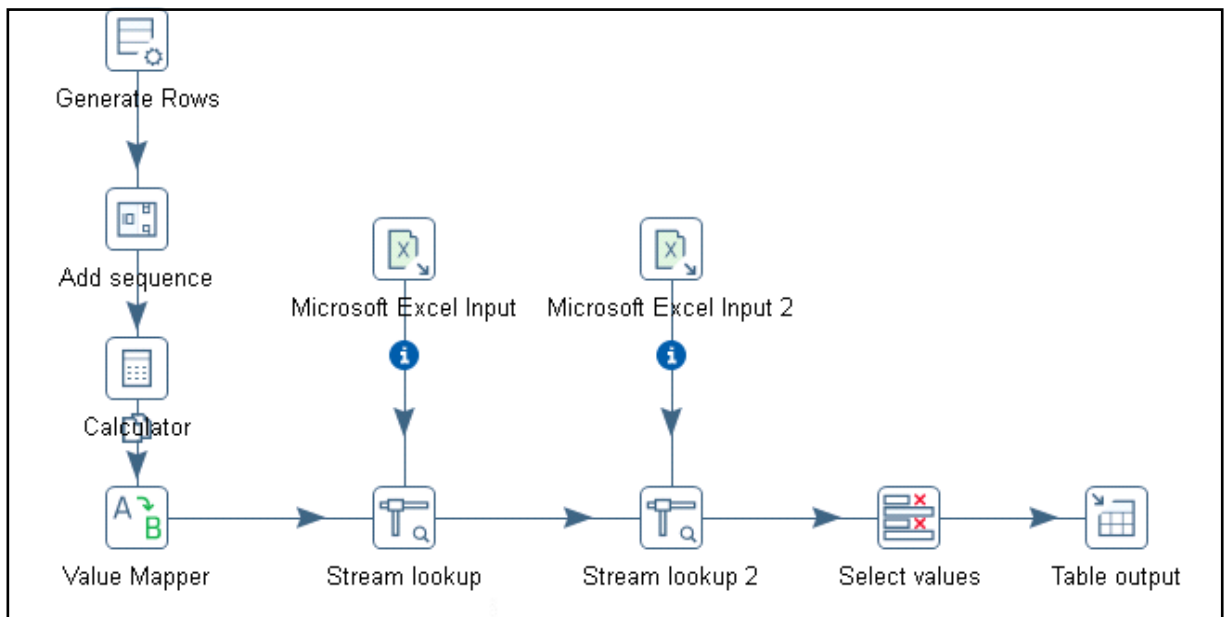
- Todos os dados foram extraídos a partir de 01-01-2010;
- Os pacientes selecionados foram os que passaram por uma consulta com os médicos do INI, denominados A, B, C, D e E;
- Os pacientes foram considerados positivos quando possuíam CID para leishmanioses e/ou exames positivos para o diagnóstico da doença;

- Os pacientes foram considerados negativos quando não possuíam CID para Leishmaniose e os exames eram negativos para o diagnóstico da doença.

Abaixo são apresentados todos os extratores desenvolvidos para atender o modelo estrela definido na etapa 'Construção do Modelo Estrela'. A definição de cada dimensão e da fato estão descritas acima no item '4.3.3 Modelagem de dados'.

Para popular a dimensão (tbl\_dmns\_tmp), onde estão armazenadas as informações de tempo, foi criado o extrator apresentado na figura 16.

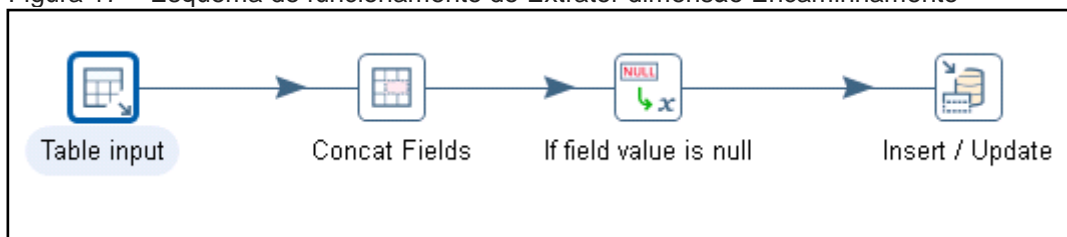
Figura 16 – Esquema de funcionamento do Extrator dimensão Tempo



Fonte: Elaborada pela autora.

Para popular a dimensão (tbl\_dmns\_encaminhamentopaciente), onde estão armazenadas as informações dos encaminhamentos realizados, foi criado o extrator apresentado na figura 17.

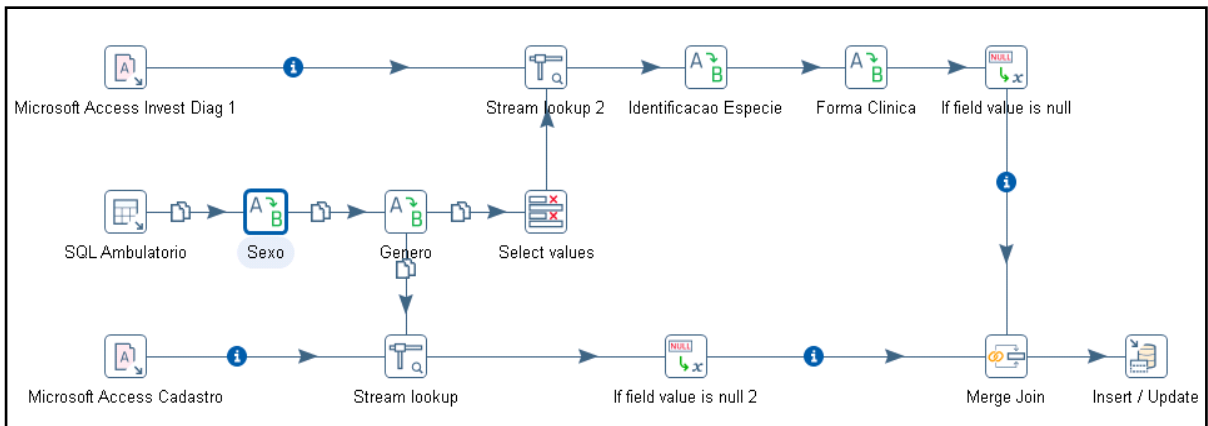
Figura 17 – Esquema de funcionamento do Extrator dimensão Encaminhamento



Fonte: Elaborada pela autora.

Para popular a dimensão (tbl\_dmns\_dadospaciente), onde estão armazenadas as informações dos pacientes atendidos, criou-se o extrator e a transformação da informação referente ao sexo, gênero, identificação da espécie e forma clínica armazenada no banco conforme apresentado na figura 18.

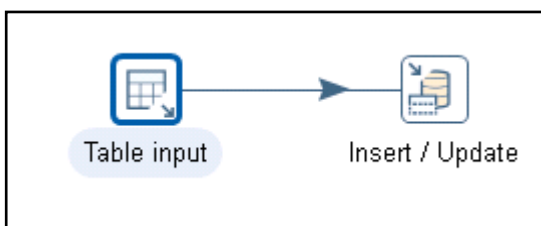
Figura 18 – Esquema de funcionamento do Extrator dimensão Paciente



Fonte: Elaborada pela autora.

Para popular a dimensão (tbl\_dmns\_exame), onde estão armazenadas as informações dos exames de diagnóstico de leishmanioses com status finalizado realizados para os pacientes do INI, foi criado o extrator apresentado na figura 19.

Figura 19 – Esquema de funcionamento do Extrator dimensão Exames

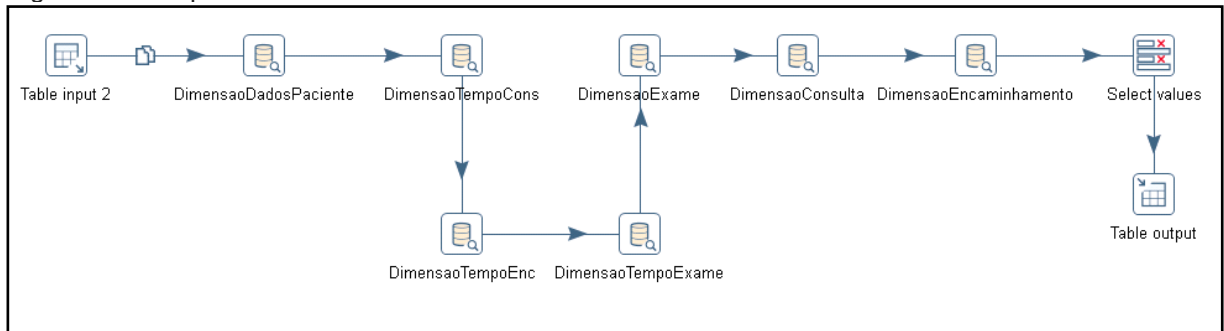


Fonte: Elaborada pela autora.

Para popular a fato, (tbl\_fato\_exame\_leish), onde estão armazenadas as informações que quantificam o número de pacientes atendidos pelos médicos previamente definidos, que realizaram os exames de diagnóstico de leishmanioses em um determinado momento, foi criado o extrator apresentado na figura 20.



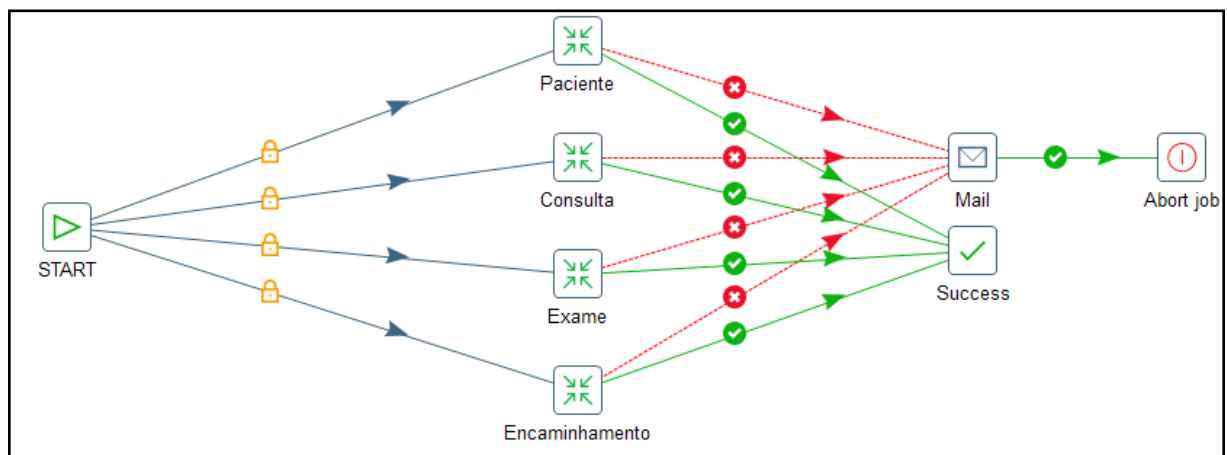
Figura 20 – Esquema de funcionamento do Extrator Fato



Fonte: Elaborada pela autora.

Nesta etapa foram construídos os *Jobs* que encadeiam as chamadas de execução dos extratores, podendo resultar em sucesso ou em falha, e no caso de falha, é enviado um e-mail contendo o erro para a pessoa responsável do setor de TI pelo *Data Mart*.

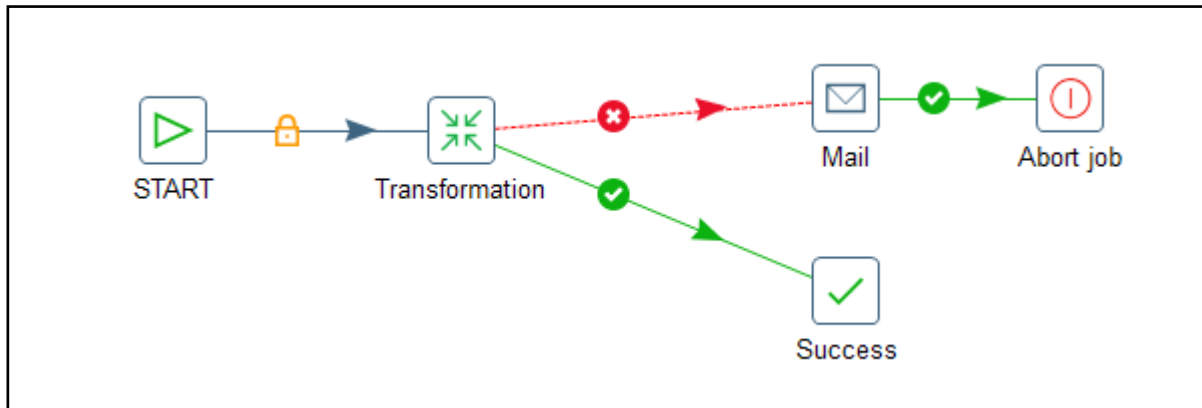
O primeiro *Job* criado pode ser observado na figura 21, ele é responsável por executar os extratores das dimensões listadas acima.

Figura 21 – Esquema de funcionamento do *Job* executor dos extratores das dimensões

Fonte: Elaborada pela autora.

Após a criação do *Job* responsável por executar os extratores das dimensões, foi criado o *Job* responsável por executar o extrator da fato, mencionada acima, que está exibido na figura 22.

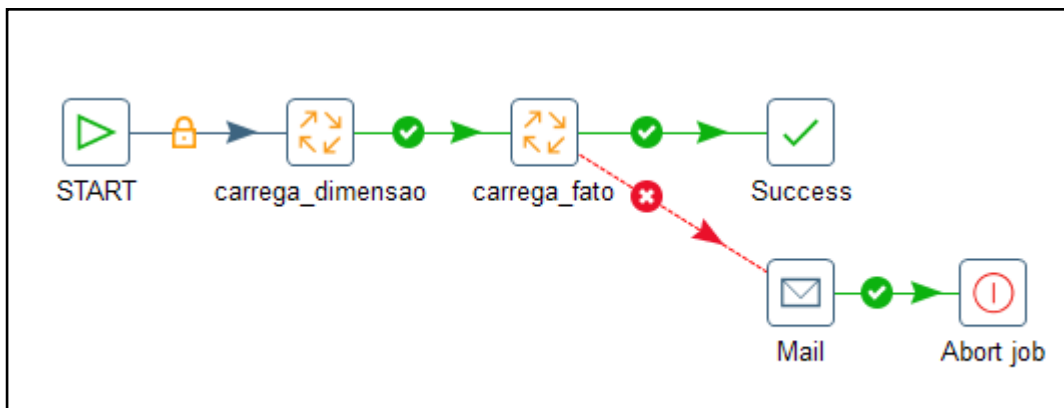
Figura 22 – Esquema de funcionamento do *Job* executor do extrator da fato



Fonte: Elaborada pela autora.

Para orquestrar a execução dos *Jobs* criados, para as dimensões e para a fato, foi criado um *Job* principal que realiza esse trabalho. Foi programado para rodar todo dia primeiro de cada mês às 9h da manhã. Esse *Job* pode ser visualizado na figura 23.

Figura 23 – Esquema de funcionamento do *Job* principal



Fonte: Elaborada pela autora.

#### **4.3.4.2 Definição da duração histórica do *Data Mart***

Após as cargas de dados terem sido efetuadas, foi definido que o *Data Mart* armazenará os dados dos últimos 15 anos, esse tempo foi baseado na duração dos projetos de Pesquisa Clínica que são desenvolvidos no LaPClinVigiLeish .

#### 4.3.5 Modelagem dos Dados no modelo Multidimensional

De acordo com Nery (2012), a modelagem multidimensional é utilizada para sumarizar e reestruturar dados apresentando-os em visões que sejam capazes de suportar a análise dos valores desses dados, além disso, é uma técnica de concepção e visualização de um modelo de dados de um conjunto de medidas que descrevem aspectos comuns de negócios.

Um modelo multidimensional é formado por três elementos básicos: Fatos, Dimensões e Medidas (variáveis).

Foram criadas as medidas de QTDPaciente, QTDConsulta, QTDEncaminhamento e QTDExame que permitem quantificar os pacientes, consultas, encaminhamentos e exames existentes no *Data Mart* criado.

A modelagem multidimensional é responsável por formar uma camada intermediária entre a base de dados e as ferramentas de consulta de dados. Nesta fase foram utilizadas as ferramentas *Online Analytical Processing* (Olap) que possibilitaram:

- A exploração dos dados no *Data Mart*, permitindo que os pesquisadores sejam capazes de construir seus relatórios transformando os dados em informação estratégica.

- A análise dos dados por diferentes visões, isto é, diferentes ângulos, permitindo aos pesquisadores investigarem os dados, validando suposições e confrontando diferentes aspectos do assunto.

- Realizar a agregação das informações, que fez com que as mesmas estejam disponíveis imediatamente, independente da complexidade da consulta e da quantidade de dados armazenados. Elas calculam (agregam) todos os totais e combinações possíveis de forma simplista.

Para a criação dos cubos foi utilizada a ferramenta Pentaho *Schema Workbench*, que é uma interface de designer, onde foi permitido criar e testar visualmente o cubo Olap.

Na etapa de modelagem multidimensional foi desenvolvido o cubo LaPClinVigiLeish, conforme o modelo estrela apresentado na etapa 'Construção do Modelo Estrela'.

Algumas regras de negócio, identificadas na coleta de dados, estão representadas por meio de fórmulas na estrutura dos cubos.

#### **4.3.6 Visualização dos dados**

A visualização dos dados oriundos do *Data Mart* é efetuada através do uso da ferramenta *Pentaho User Console*. Esta ferramenta possibilita a elaboração de algumas consultas pré-estabelecidas sendo bastante flexível na parametrização das mesmas (esta parametrização permite a inserção de novos agrupamentos e colunas bem como a remoção dos mesmos).

Algumas visões mais rotineiras foram pré-definidas para facilitar o uso enquanto outras ficam completamente a cargo do pesquisador, podendo este criar novas consultas no momento em que houver necessidade e de acordo com as necessidades emergentes.

#### **4.3.7 Migração para o servidor de Produção**

Nesta etapa, o *Data Mart* foi migrado para o servidor de Produção, este com 16 processadores Intel® Xeon® CPU E5-2660 v2 @2.20Ghz, disco rígido de 2 terabytes e memória RAM de 16 gigabytes e contém o servidor de banco de dados SQL Server 2008 com o Sistema Gerenciador de Banco de dados (SGBD) da Microsoft®.

#### **4.3.8 Análise dos Resultados**

Nesta etapa os dados foram disponibilizados no *Data Mart* e validados pelo pesquisador responsável pelo projeto com o apoio do analista de TI, foi realizada uma comparação dos dados contidos nas fontes de origem com os registros criados no *Data Mart*. Neste ponto alguns ajustes foram realizados no processo de extração.

Alguns acontecimentos como adequações à legislação, fatos externos e mudanças organizacionais poderão afetar de forma direta ou indiretamente o *Data Mart*, dessa forma, analisar os dados contidos no mesmo será uma tarefa constante, mesmo após a sua implantação.

### 4.3.9 Criação de Visões Pré-definidas

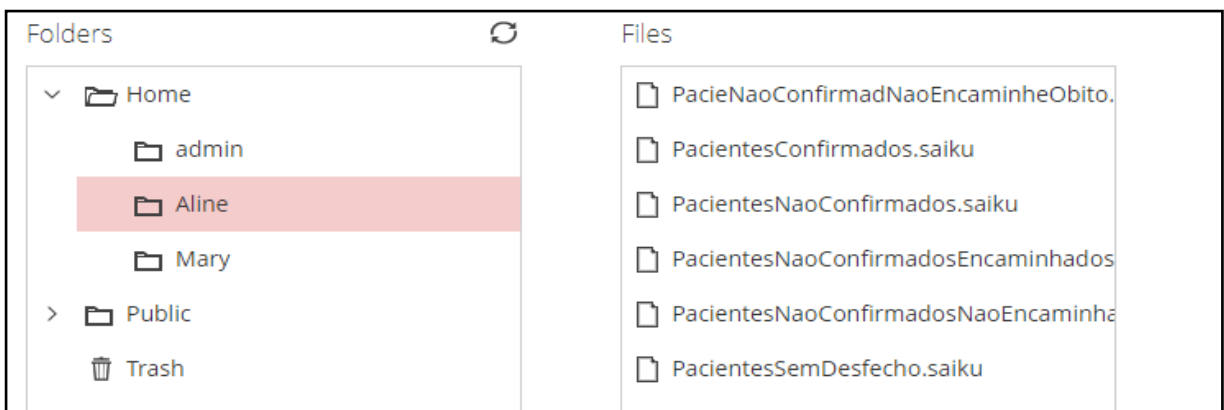
Esta etapa do projeto abordou a criação de visões (“relatórios”) direcionadas, isto é, visões pré-definidas que são frequentemente utilizadas pelos pesquisadores.

Algumas visões foram criadas de forma orientada, onde os pesquisadores podem “navegar” através dos relatórios obtendo o detalhamento necessário para suas análises.

O analista de TI responsável pelo projeto auxiliou a pesquisadora responsável pelo projeto a desenvolver algumas consultas mais simples no início da utilização da ferramenta Olap e as consultas com um alto grau de complexidade foram desenvolvidas pelo analista de TI responsável pelo projeto.

Na figura 24 estão ilustradas as visões criadas para a pesquisadora Aline Fagundes.

Figura 24 – Visões criadas para a pesquisadora Aline



Fonte: Elaborada pela autora.

### 4.4.0 Definição da Segurança de Informação

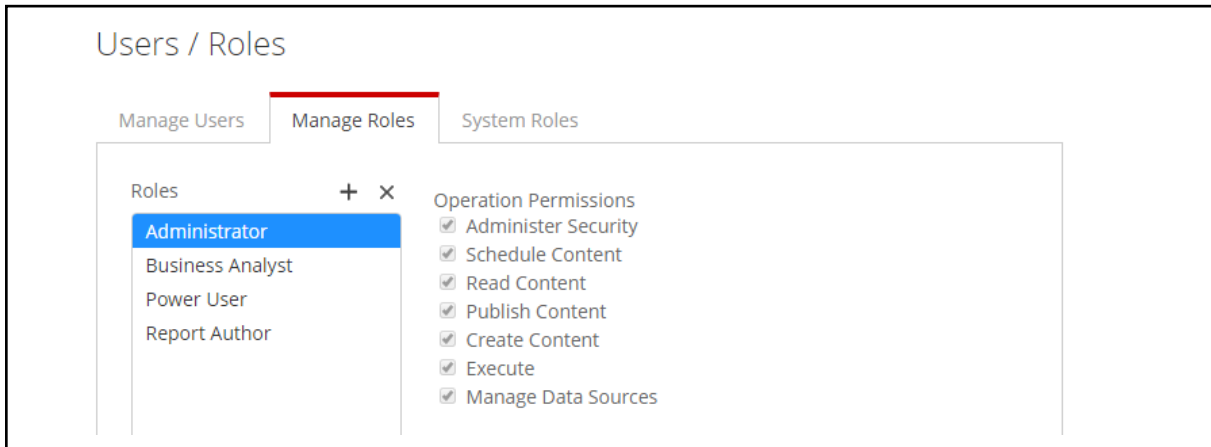
As informações disponibilizadas no *Data Mart* são referentes aos dados dos projetos de pesquisa clínica do LaPClinVigiLeish, fazem parte destes dados informações dos pacientes que participam das pesquisas, portanto o acesso a estas informações é extremamente restrito.

Esta fase representa a segurança da informação onde foi identificado quem pode, como pode e como as informações são consultadas. Nenhum dado de identificação do paciente é divulgado.

Foram criados dois níveis de acesso para a utilização da ferramenta Pentaho:

- O primeiro nível de acesso criado foi o de *Administrator*, que possui privilégio em realizar operações de *Administer Security*, *Manage Data Sources*, entre outras. Somente os Analistas de TI possuem esse nível de acesso. Todas as operações permitidas podem ser visualizadas na figura 25.

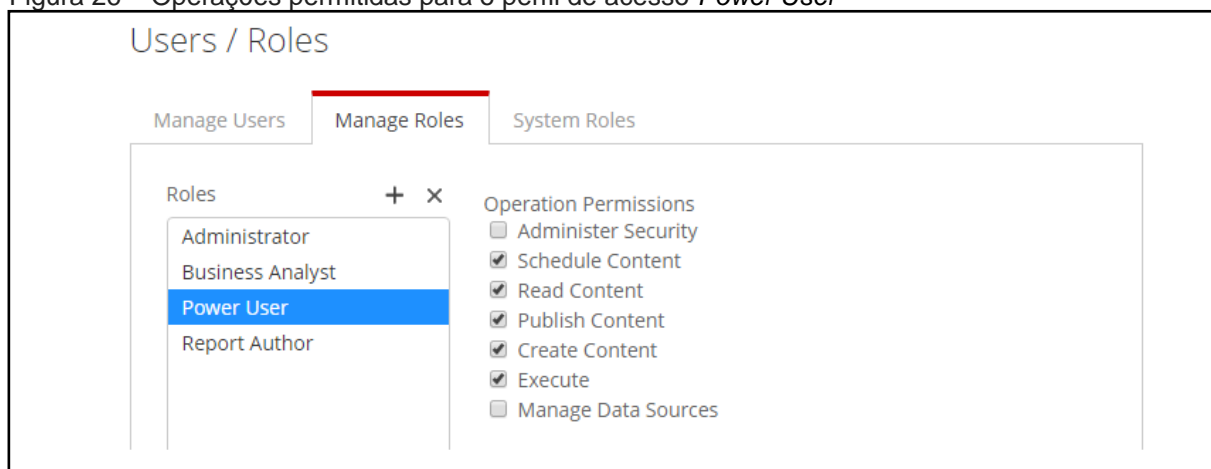
Figura 25 – Operações permitidas para o perfil de acesso *Administrator*



Fonte: Elaborada pela autora.

- O segundo nível de acesso foi o *Power User*, onde operações como *Read Content*, *Publish Content*, entre outras são possíveis. Este nível de acesso foi concedido aos pesquisadores envolvidos no projeto. Todas as operações permitidas podem ser visualizadas na figura 26.

Figura 26 – Operações permitidas para o perfil de acesso *Power User*



Fonte: Elaborada pela autora.

Os dados obtidos através das diversas fontes de dados foram armazenados em um banco de dados único dentro de um servidor já existente localizado no SETIC definido pela equipe de Redes seguindo os critérios de Segurança da Informação conforme a Política de Segurança da Informação e Comunicações (Posic) Fiocruz.

Todos os profissionais envolvidos no projeto assinaram o Termo de Confidencialidade, fornecido pelo serviço de Gestão da Qualidade do INI/Fiocruz.

#### 4.4 RISCOS E BENEFÍCIOS

Os principais riscos estão relacionados a problemas tecnológicos, tais como, confidencialidade, indisponibilidade, o acesso por uma pessoa não autorizada, ou manipulação e/ou interpretação equivocada. Esses problemas poderiam trazer prejuízos ao Laboratório devido as relevantes informações disponíveis neste ambiente.

Para minimizar os riscos foram tomadas as seguintes medidas:

- Os dados oriundos do SIPEC são mantidos em servidores em locais específicos, com entrada e saída autorizadas somente a pessoas do SETIC;
- Os dados oriundos de planilhas eletrônicas e banco de dados Access® são mantidos em um diretório específico na rede onde o acesso foi concedido ao responsável pelo desenvolvimento da pesquisa e a pesquisadora responsável pelo projeto;
- O acesso ao servidor que armazena o *Data Mart* foi concedido ao responsável pelo desenvolvimento da pesquisa, a pesquisadora responsável pelo projeto e a equipe de Rede do SETIC;
- Sistema de controle de acesso à sala dos servidores por meio de mecanismo de abertura da porta através de biometria ou senha, *firewall* e sistemas de detecção de intrusão são utilizados para controlar o acesso e manter os dados seguros;
- A liberação do uso do *Data Mart* foi concedida mediante a um treinamento na ferramenta de modo que os usuários possam desenvolver os trabalhos e gerar benefícios ao INI;
- A segurança da informação é realizada pela equipe de Rede do SETIC.

São inúmeros os benefícios obtidos através desse projeto, dentre eles pode-se destacar a possibilidade de obter informações detalhadas, com maior rapidez e facilidade.

O tempo gasto no processo de busca e análise da informação foi invertido, antes se gastava muito tempo para conseguir a informação e tinha-se pouco tempo para analisá-la, com a implantação do *Data Mart*, os pesquisadores podem dedicar menos tempo para a montagem das consultas e uma parcela maior de tempo para a análise das informações e a tomada de decisão, o que proporcionará uma maior probabilidade de acerto sob a decisão tomada.

No *Data Mart* o usuário pode gerar a informação, analisar e modificar o modo de apresentação. Além do mais, o usuário tem a flexibilidade de detalhar ou consolidar dados, facilitando o entendimento para a análise.

A utilização do *Data Mart* por si só, não garante vantagem, mas sim a maneira eficiente do modo que será utilizado.

Os participantes da pesquisa não tiveram nenhum benefício direto com a realização deste projeto, este será indiretamente através do conhecimento procedente da pesquisa que contribuirá na redução da morbidade, deformidades e óbitos em pacientes com Leishmaniose que vem sendo realizada através das pesquisas por este laboratório.

#### 4.5 MATERIAL DE APOIO

Foi elaborado um material de apoio (APÊNDICE D) aos usuários do *Data Mart* contendo o passo a passo para a criação de relatórios e/ou acesso aos dados utilizando o *plugin Saiku Analytics* da ferramenta Pentaho de modo a proporcionar uma forma fácil de utilização da ferramenta.

Um Dicionário de dados (APÊNDICE E) foi criado de modo que as informações sobre o conjunto de dados estejam centralizadas, melhorando a comunicação entre todos os usuários do *Data Mart*. Este é um documento feito com informações sobre os dados, tais como: significado e o tipo dos dados.

#### 4.6 PLANO DE GESTÃO DE DADOS

##### 4.6.1 Dados e metadados produzidos pelo projeto

O *Data Mart* integra dados estruturados provenientes do Sistema de Prontuário Eletrônico do INI (SIPEC), planilhas eletrônicas e banco de dados Access®.



Os dados provenientes do SIPEC estão armazenados em um banco de dados localizado em um servidor no Serviço de Tecnologia da Informação e Comunicação (SETIC) do Instituto Nacional de Infectologia (INI) seguindo os critérios de Segurança da Informação conforme a Posic Fiocruz.

As planilhas eletrônicas e os dados do banco Access® são obtidos de um diretório específico localizado na rede do INI.

Somente foram utilizados dados organizados em linhas e colunas devido ao período de tempo em que o projeto foi realizado, dessa forma, dados textuais não foram utilizados neste projeto.

O *Data Mart* gerado foi armazenado em um banco de dados SQL Server 2008 com um Sistema Gerenciador de Banco de dados (SGBD) da Microsoft® dentro de um servidor com 16 processadores Intel® Xeon® CPU E5-2660 v2 @2.20Ghz, disco rígido de 2 terabytes e memória RAM de 16 gigabytes localizado no SETIC seguindo os critérios de Segurança da Informação adotados para os servidores já existentes no local.

Será realizada mensalmente a avaliação quantitativa e a avaliação qualitativa da qualidade dos dados.

A avaliação quantitativa será feita do processo de Extração, Transformação e Carga. Os dados extraídos da origem que não tiverem sido carregados no *Data Mart* serão gravados em um arquivo texto contendo o motivo da sua não inserção para posterior análise e solução.

A avaliação qualitativa da qualidade será realizada pelo pesquisador responsável pelo projeto.

Foram elaborados dois tipos de metadados:

**O técnico** - Este contém o *Data Mart* incluindo informações sobre a estrutura dos dados, as fontes de dados, estruturas de registro, restrições de atributos, movimentações dos dados, filtros, transformações e agregações, as rotinas de extração dos dados e visões definidas em diferentes bases de dados.

**O semântico** – O modelo de dados e seus relacionamentos, o modelo de dados multidimensional contém informações sobre as dimensões, cubos de dados e níveis de agregação dos dados.

Todos os documentos produzidos ao longo do projeto foram armazenados em repositório único no SETIC.

#### 4.6.2 Questões éticas e Legais

Para os dados obtidos não foi necessária aprovação dos participantes das diversas pesquisas realizadas no laboratório de Pesquisa Clínica e Vigilância em Leishmanioses, pois estes já assinaram um Termo de Consentimento Livre e Esclarecido (TCLE) relativos aos projetos de pesquisa que direcionaram a coleta dos dados a serem analisados, onde os objetivos destas já constam nos diferentes projetos já aprovados.

O projeto desta dissertação está aprovado pelo Comitê de Ética em Pesquisa (CEP) do INI, parecer 3.146.209 datado de 14 de fevereiro de 2019, CAAE: 99366918.3.0000.5262, cuja pesquisadora responsável é Aline Fagundes Silva.

Foi realizada a assinatura do TCC (APÊNDICE C) pelos pesquisadores com acesso aos dados pessoais dos participantes dos diversos projetos e os pesquisadores que serão entrevistados assinarão o TCLE (APÊNDICE A) no momento da entrevista.

Os dados de identificação dos pacientes são mantidos sob guarda dos pesquisadores coordenadores do projeto e sob a guarda institucional do INI, obedecendo toda a legislação acerca do assunto. Foram seguidos todos os protocolos de sigilo e a proteção das cópias de segurança dos dados, de acordo com o que já é realizado rotineiramente pelos Serviços de Documentação e SETIC do INI.

Foi seguida toda a política e regulamentação da Fiocruz acerca da propriedade intelectual, e nenhum dado foi licenciado para patenteamento ou reutilização.

#### 4.6.3 Armazenamento e Backup

É realizado o *backup* dos dados de modo incremental diariamente às 19h com uma retenção de 90 dias e de modo completo todo primeiro sábado do mês às 0h.

O *backup* dos dados em caso de perda é restaurado em outro servidor, sendo validado e transportado posteriormente para o ambiente de produção.

Os administradores da rede do SETIC são os responsáveis por realizar o *backup* e a recuperação dos dados, caso necessário.

O armazenamento do *backup* da rotina diária é feito em um servidor localizado no SETIC do INI e da rotina mensal na sala cofre da Fiocruz.

## 5 RESULTADOS

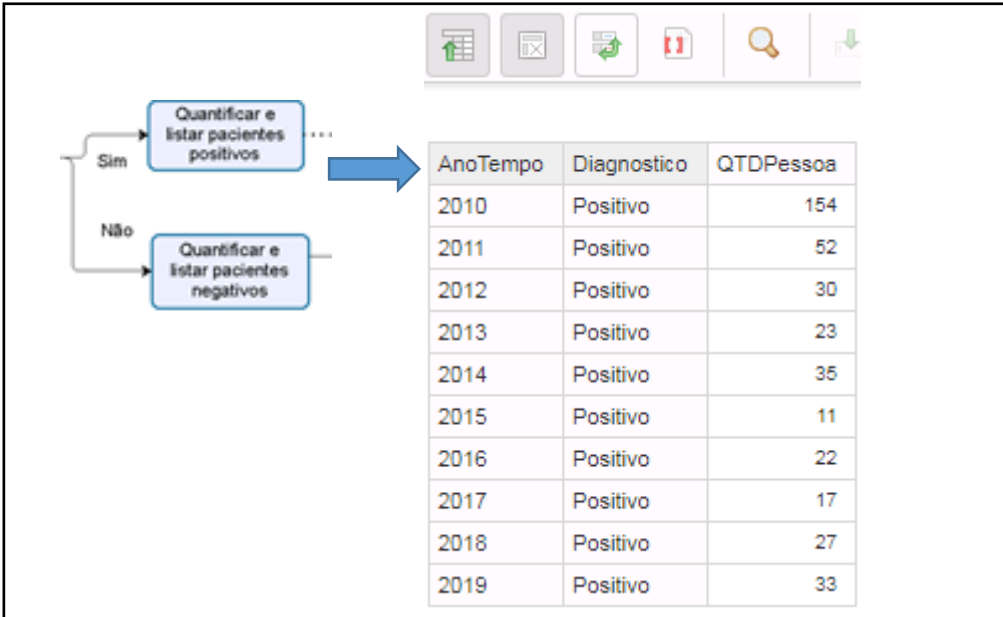
A utilização dos conceitos de BI junto às ferramentas de apoio *SQL Server* e a *Suit Pentaho* criaram o ambiente de desenvolvimento que fornece dados relevantes às análises de forma simples e organizada.

Com a utilização de componentes tais como tabelas e gráficos de diversos estilos, os dados foram agrupados de forma com que respondessem às questões levantadas no processo de análise dos pacientes do Laboratório de Leishmaniose. Assim, as visões, o cubo e as medidas que formam o resultado final deste projeto foram criadas de acordo com esse processo.

O *Saiku Analytics* permite a geração de relatórios *ad-hoc*, com isso foi possível analisar o cubo LaPClinVigiLeish.

Visando uma melhor apresentação final dos resultados, foram gerados alguns relatórios com dados a partir de 01-01-2010 até 31-12-2019. Dois modos de exibição dos relatórios foram gerados com o objetivo de demonstrar as opções disponíveis na ferramenta.

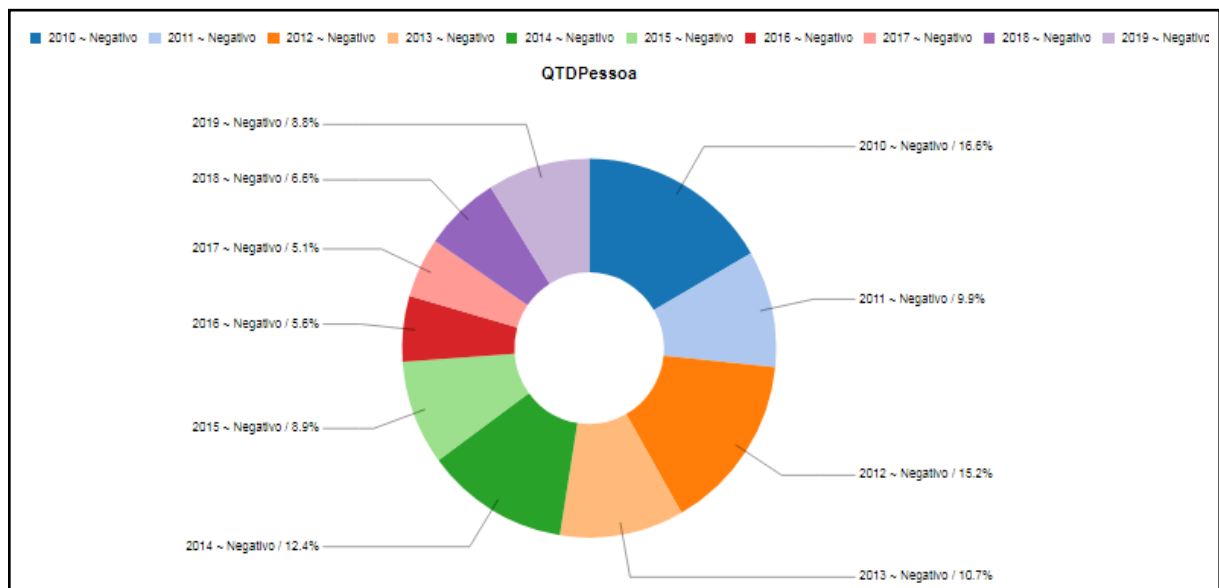
O início do nosso processo de análise se deu pela integração dos dados oriundos das diversas fontes seguido pela busca dos exames para diagnóstico de leishmanioses, bem como seus resultados, e as CIDS de Leishmaniose atribuídas aos pacientes atendidos pelos médicos previamente selecionados. Seguindo o processo, quantificamos os pacientes que entraram com suspeita de leishmaniose e foram confirmados, estes representados no Quadro 1 e os que não foram confirmados, no gráfico 1.



| AnoTempo | Diagnostico | QTD Pessoa |
|----------|-------------|------------|
| 2010     | Positivo    | 154        |
| 2011     | Positivo    | 52         |
| 2012     | Positivo    | 30         |
| 2013     | Positivo    | 23         |
| 2014     | Positivo    | 35         |
| 2015     | Positivo    | 11         |
| 2016     | Positivo    | 22         |
| 2017     | Positivo    | 17         |
| 2018     | Positivo    | 27         |
| 2019     | Positivo    | 33         |

Fonte: Elaborada pela autora.


Gráfico 1 – Relatório extraído do cubo LaPClinVigiLeish contendo a quantidade de pacientes não confirmados com leishmanioses, no período de 2010 a 2019.



Fonte: Elaborada pela autora.

Dos pacientes não confirmados, verificamos quais encaminhamentos foram realizados para os mesmos. Esses encaminhamentos podem ter sido realizados para outra especialidade, laboratório ou outra localidade e estão representados nos quadros 2, 3 e 4 respectivamente.

Quadro 2 – Relatório extraído do cubo LaPClinVigiLeish contendo os encaminhamentos realizados para uma especialidade dos pacientes não confirmados para leishmanioses, no período de 2010 a 2019.



| AnoTempo - AnoTempo            | 2010       | 2011       | 2012       | 2013       | 2014       | 2015       | 2016       | 2017       | 2018       | 2019       |
|--------------------------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|
| Diagnostico - Diagnostico      | Negativo   | Negativo   | Negativo   | Negativo   | Negativo   | Negativo   | Negativo   | Negativo   | Negativo   | Negativo   |
| Especialidade                  | QTD Pessoa | QTD Pessoa | QTD Pessoa | QTD Pessoa | QTD Pessoa | QTD Pessoa | QTD Pessoa | QTD Pessoa | QTD Pessoa | QTD Pessoa |
| ASSISTENTE SOCIAL              | -          | -          | -          | -          | -          | -          | 1          | -          | -          | -          |
| CARDIOLOGIA                    | -          | -          | 1          | -          | -          | -          | -          | -          | -          | -          |
| CIRURGIÃO                      | -          | -          | -          | -          | -          | -          | -          | -          | -          | 1          |
| DERMATOLOGIA                   | 1          | 3          | 3          | 6          | 45         | 38         | 29         | 26         | 45         | 39         |
| ENDOCRINOLOGIA                 | -          | -          | -          | -          | 1          | -          | -          | -          | -          | -          |
| ENFERMAGEM                     | 10         | 22         | 21         | 12         | 13         | 2          | 1          | 2          | 2          | -          |
| EPIDEMIOLOGIA                  | -          | -          | -          | 1          | -          | -          | -          | -          | -          | -          |
| FARMACÊUTICO                   | -          | -          | 1          | -          | 1          | -          | -          | -          | -          | -          |
| FONOAUDILOGIA                  | 1          | 1          | 1          | -          | -          | 1          | -          | -          | 1          | -          |
| GASTROENTEROLOGIA              | -          | -          | -          | -          | -          | 1          | -          | -          | -          | -          |
| HEMATOLOGIA                    | -          | -          | -          | -          | 1          | -          | -          | -          | -          | -          |
| HOSPITAL-DIA                   | -          | 2          | 1          | -          | -          | -          | -          | -          | -          | -          |
| INFECTOLOGIA                   | -          | 1          | 1          | 4          | 7          | 10         | 2          | 6          | 8          | 8          |
| INFECTOLOGIA - IMUNIZAÇÕES     | -          | 1          | 1          | -          | 1          | 1          | -          | -          | -          | -          |
| NEUROLOGIA                     | 1          | -          | -          | -          | -          | -          | -          | -          | -          | -          |
| NUTRIÇÃO                       | 1          | -          | -          | -          | -          | -          | -          | -          | -          | -          |
| OFTALMOLOGIA                   | 5          | -          | 1          | 1          | 1          | -          | -          | 1          | -          | 1          |
| OTORRINOLARINGOLOGIA           | 14         | 5          | 4          | 1          | 12         | 21         | 3          | 1          | 4          | 23         |
| PNEUMOLOGIA                    | -          | -          | -          | -          | -          | 1          | -          | -          | -          | -          |
| PRIM.ATEND./ ATEND.INTERCORRÉ. | 1          | -          | -          | -          | -          | -          | -          | -          | -          | -          |
| PSIQUIATRIA                    | 1          | -          | -          | -          | -          | -          | -          | -          | -          | -          |

Fonte: Elaborada pela autora.

Quadro 3 – Relatório extraído do cubo LaPClinVigiLeish contendo os encaminhamentos realizados para um laboratório dos pacientes não confirmados para leishmanioses, no período de 2010 a 2019.

| AnoTempo - AnoTempo                 | 2010       | 2011       | 2012       | 2014       | 2015       | 2018       | 2019       |
|-------------------------------------|------------|------------|------------|------------|------------|------------|------------|
| Diagnostico - Diagnostico           | Negativo   | Negativo   | Negativo   | Negativo   | Negativo   | Negativo   | Negativo   |
| Laboratorio                         | QTD Pessoa | QTD Pessoa | QTD Pessoa | QTD Pessoa | QTD Pessoa | QTD Pessoa | QTD Pessoa |
| DOENÇAS FEBRIS AGUDAS / DENGUE      | 1          | -          | -          | -          | -          | -          | -          |
| DOENÇAS INFECCIOSAS EM DERMATOLOGIA | -          | -          | -          | -          | -          | -          | 2          |
| LEISHMANIOSES                       | -          | -          | -          | 1          | 4          | 2          | -          |
| MICOSES                             | -          | -          | -          | -          | 1          | -          | -          |
| OUTRAS DIPS                         | 1          | -          | -          | -          | -          | -          | -          |
| TUBERCULOSE E OUTRAS MICOBACTÉRIAS  | -          | 1          | 2          | 1          | -          | -          | -          |

Fonte: Elaborada pela autora.

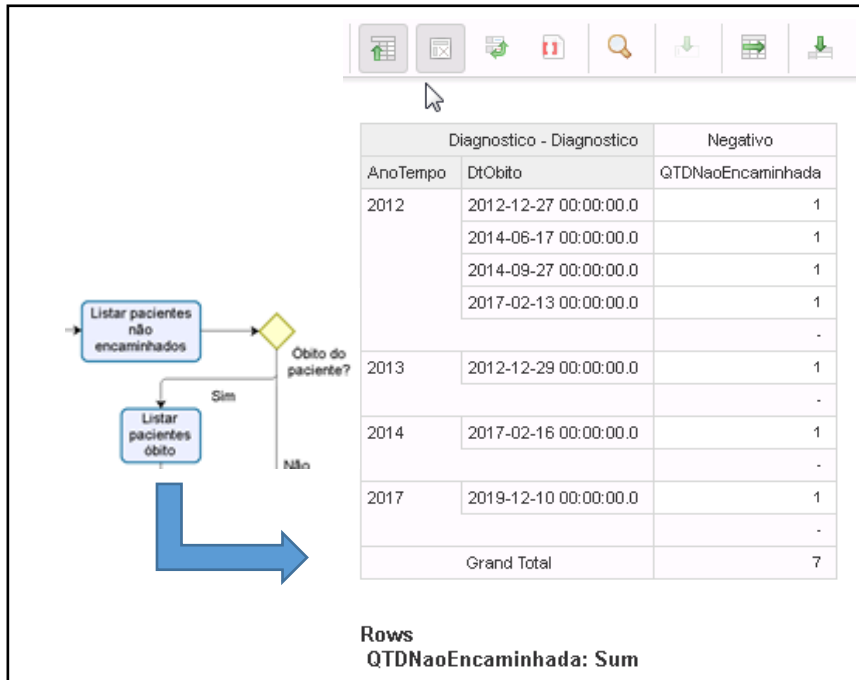
Quadro 4 – Relatório extraído do cubo LaPClinVigiLeish contendo os encaminhamentos realizados para uma instituição dos pacientes não confirmados para leishmanioses, no período de 2010 a 2019.

| AnoTempo - AnoTempo   | 2010       | 2011       | 2012       | 2013       | 2014       | 2015       | 2016       | 2017       | 2018       | 2019       |
|---|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|
| Diagnostico - Diagnostico                                       | Negativo   | Negativo   | Negativo   | Negativo   | Negativo   | Negativo   | Negativo   | Negativo   | Negativo   | Negativo   |
| Instituicao   | QTD Pessoa | QTD Pessoa | QTD Pessoa | QTD Pessoa | QTD Pessoa | QTD Pessoa | QTD Pessoa | QTD Pessoa | QTD Pessoa | QTD Pessoa |
| AMBULATÓRIO SOUZAARAÚJO - ASA                                   | 2          | -          | 1          | 3          | -          | 1          | 1          | -          | -          | 4          |
| CENTRO MUNICIPAL DE SAÚDE                                       | -          | -          | -          | -          | 1          | -          | -          | -          | -          | -          |
| HOSPITAL DOS SERVIDORES DO ESTADO                               | 1          | -          | -          | -          | -          | -          | -          | -          | -          | -          |
| HOSPITAL FEDERAL DO ANDARAÍ                                     | -          | -          | -          | 1          | -          | -          | -          | -          | -          | 1          |
| HOSPITAL GERAL DE BONSUCESSO                                    | 2          | 2          | 3          | -          | 1          | 1          | -          | -          | -          | -          |
| HOSPITAL PRIVADO  | -          | -          | -          | 1          | -          | 1          | -          | -          | -          | -          |
| HOSPITAL PÚBLICO  | 1          | 3          | 2          | -          | 2          | -          | -          | -          | -          | -          |
| HOSPITAL UNIVERSITÁRIO CLEMENTINO FRAGA FILHO                   | 1          | -          | -          | -          | -          | -          | 1          | -          | -          | -          |
| HOSPITAL UNIVERSITÁRIO GAFFREE E GUINLE                         | -          | -          | -          | 1          | -          | -          | -          | -          | -          | -          |
| HOSPITAL UNIVERSITÁRIO PEDRO ERNESTO                            | 1          | -          | -          | -          | -          | -          | -          | -          | -          | 1          |
| INSTITUTO ESTADUAL DE HEMATOLOGIA ARTHUR DE SIQUEIRA CAVALCANTI | -          | 1          | -          | -          | -          | -          | -          | -          | -          | -          |
| INSTITUTO FERNANDES FIGUEIRA - IFF                              | -          | -          | -          | -          | 1          | -          | -          | -          | -          | -          |
| POSTO DE SAÚDE  | 6          | 2          | 1          | 6          | 2          | 3          | -          | 2          | 4          | 1          |
| SERVIÇO DE AMBULATÓRIO - CECLIN                                 | 1          | 1          | 1          | -          | -          | 1          | 1          | 1          | 6          | 6          |

Fonte: Elaborada pela autora.

Dando continuidade ao nosso processo de análise, quantificamos os pacientes que entraram com suspeita de leishmanioses, que não foram confirmados, que não foram encaminhados e que evoluíram para óbito, sendo esse resultado representado no quadro 5.

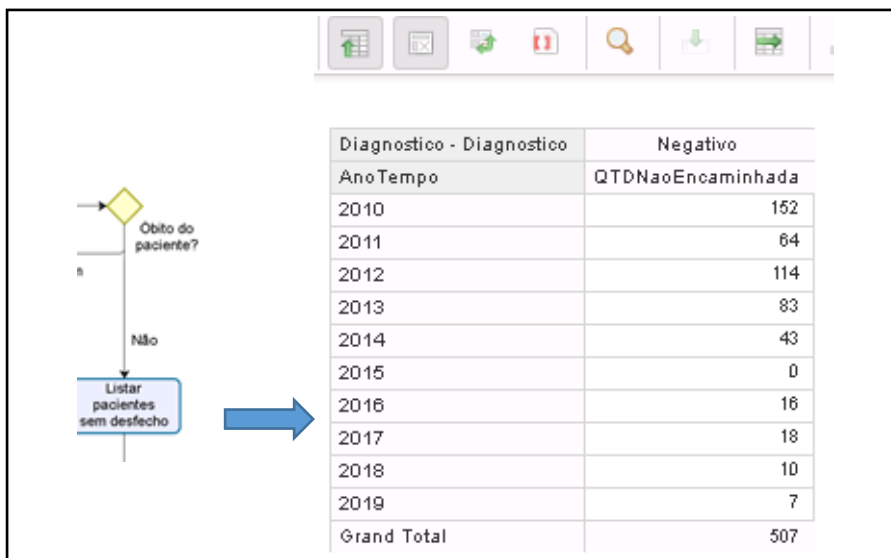
Quadro 5 – Relatório extraído do cubo LaPClinVigiLeish contendo a quantidade de pacientes não confirmados que não foram encaminhados evoluindo para óbito, nos anos de 2012, 2013, 2014 e 2017.



Fonte: Elaborada pela autora.

Para fechar o processo de análise foi quantificado os pacientes não confirmados, não encaminhados e que não foram a óbito, sendo este resultado representado no quadro 6.

Quadro 6 – Relatório extraído do cubo LaPClinVigiLeish contendo a quantidade de pacientes não confirmados, não encaminhados e que não foram a óbito, no período de 2010 a 2019.



Fonte: Elaborada pela autora.

## 6 DISCUSSÃO

A falta de informações para análise e suporte à tomada de decisão vem a cada dia se tornando um dos fatores mais críticos relacionados ao sucesso das pesquisas. Decisões equivocadas tomadas com base em informações errôneas geradas a partir de fontes duvidosas e imprecisas.

Antes da implementação do *Data Mart*, as informações relevantes para as pesquisas realizadas pelo Laboratório de Pesquisa Clínica e Vigilância em Leishmanioses eram obtidas através de diversas planilhas eletrônicas, informações oriundas do Sistema de Prontuário Eletrônico do INI, entre outras. Todas essas informações eram reunidas em uma só planilha o que demandava muito tempo e retrabalho devido à grande quantidade de unidades e fatores. Após a elaboração da planilha final, a análise e busca de informações eram realizadas pelos pesquisadores.

Diante deste contexto, este trabalho propôs a criação e implementação de um *Data Mart* visando suprir tanto a necessidade de informações quanto a qualidade e confiabilidade das mesmas. Permitiu que os dados fossem extraídos de diversas fontes com mais segurança e agilidade, proporcionando uma base de dados única e sólida para auxiliar nos processos decisórios, sendo este processo de extrema importância para a melhora da qualidade e confiabilidade dos dados.

A *Suit Pentaho* mostrou-se eficiente, sendo capaz de atender as necessidades levantadas no início do projeto. Ela pode ser facilmente customizada, porém necessita de um profissional da área de TI com conhecimento em *Data Warehouse*. Por ser uma ferramenta *Open Source*, não existem despesas com sua aquisição ou licenciamento, tornando-se uma alternativa viável.

Para a utilização do *Pentaho User Console* não se fez necessário instalar nenhum programa adicional nas máquinas dos usuários, o acesso é feito via web.

A *Suit Pentaho* possui uma flexibilidade para acessar qualquer base de dados, bastando existir um driver de conexão JDBC.

Este trabalho, em conformidade com estudos prévios de Bréant et al.(2005), onde obtiveram ganho em velocidade na geração de um grande volume de dados, assim como o cruzamento de informações, o que garantiu uma melhora na análise e suporte à tomada de decisão, mostrou que foi possível a implementação da Tecnologia de *Business Intelligence (BI)* em Laboratórios de Pesquisa Clínica.



A construção e implementação do *Data Mart* utilizando a *Suit Pentaho* levou 5 meses, com ela foi possível transferir todos os dados em apenas 2 minutos.

Foi necessária a criação de algumas transformações para corrigir pequenos problemas, como informações incorretas, que estavam sendo capturadas no banco de dados de origem e arrumar esses dados para que fossem gravados e exibidos de forma correta no destino.

Após a inclusão dos dados no *Data Mart* realizamos a conferência dos mesmos, verificando se todos os dados na origem estavam no destino.

A falta de documentação técnica das fontes de dados utilizadas mostrou-se um grande problema no momento do levantamento dos mesmos, sendo gasto 20% a mais do tempo estimado inicialmente para a realização desse processo. Para resolver esse problema, foram realizadas algumas simulações no Sistema de Prontuário Eletrônico, ambiente de desenvolvimento, a fim de descobrir as informações necessárias para esse projeto.

Através do uso da ferramenta foi possível observar a quantidade de pacientes que entraram no Instituto com suspeita de leishmaniose e foram confirmados, assim como a quantidade dos que não foram. Além da quantidade, através do modelo construído, é possível identificar a moradia, área endêmica, a espécie de leishmania encontrada, a CID atribuída, a cidade, o estado, a data de nascimento, nome, gênero, nome social, sexo, número de primeiro atendimento e número de prontuário do paciente.

Realizamos uma pesquisa com os pacientes que não tiveram o diagnóstico confirmado para leishmanioses para sabermos se foram encaminhados para alguma instituição, especialidade ou outro laboratório de pesquisa. Essa pesquisa nos retornou um quantitativo dos encaminhamentos realizados, sendo possível visualizarmos a data em que ocorreu o encaminhamento, bem como o tipo (interno ou externo) e a consulta que originou esse encaminhamento.

Descobrimos que alguns pacientes que entraram com suspeita de leishmaniose foram a óbito, porém ao realizarmos o cruzamento dos dados vimos que esses pacientes não tiveram o diagnóstico confirmado para leishmaniose e não tiveram nenhum tipo de encaminhamento.

Existe ainda um número de pacientes a serem analisados que não se enquadraram em nenhuma das condições criadas no processo de análise para o

Laboratório de Leishmaniose, esses pacientes foram classificados como “pacientes sem desfecho”.

Após a implantação do *Data Mart* no ambiente de produção, os resultados foram avaliados pelo pesquisador responsável pelo projeto e algumas alterações necessárias foram realizadas de modo a ajustar os dados extraídos.

Não houve tempo hábil para a realização do treinamento para a utilização da ferramenta com os pesquisadores do laboratório. Houve a necessidade de utilização desse tempo para a realização de algumas alterações e ajustes necessários no *Data Mart*. O treinamento dos pesquisadores será realizado após a defesa desse projeto.

Espera-se que após a utilização do *Data Mart* pelos pesquisadores seja possível melhorar a análise dos dados da pesquisa clínica do LaPClinVigiLeish, da mesma forma mencionada por Roelofs et al. (2013), após a implantação do DW na radioterapia, onde após agregar várias fontes de dados em um DW combinado com ferramentas para extração de parâmetros relevantes melhorou a qualidade da análise dos mesmos.

Da mesma forma que Bréant et al. (2005) apresentaram a utilização do DW em um Hospital Universitário de Geneva onde o propósito da implementação do DW foi facilitar o acesso a uma visão integrada e coerente do atendimento médico de modo a apoiar atividades analíticas, como estatísticas médicas, estudos clínicos, recuperação de casos semelhantes e processos de Data Mining, acredita-se que o *Data Mart* implantado traga benefícios semelhantes a estes mencionados.

Não foram exibidas informações que pudessem identificar o paciente e cujo o conteúdo do tipo texto fosse muito extenso, devido a limitação de espaço neste documento para exibição da mesma.

O *BI* era uma tecnologia, até o momento, inexistente no Instituto Nacional de Infectologia Evandro Chagas (INI), bem como a quebra do paradigma da desnormalização do modelo de dados. Estes, foram fatores que aumentaram a complexidade e as dificuldades no desenvolvimento deste projeto.

Este *Data Mart* poderá ser adaptado para a utilização em outros laboratórios de pesquisa ou até mesmo em outros setores do INI. A facilidade em adaptar a ferramenta a novas necessidades ou até diferentes contextos constitui-se em um dos pontos positivos do projeto apresentado.

## 7 CONCLUSÃO

Com a implantação do *Data Mart* e a consequente disponibilização de informações, foi eliminado o desencontro de dados e, a problemática da centralização da informação, que eram fatores de maior relevância e motivação para o desenvolvimento deste projeto.

A redução da dependência dos pesquisadores quanto aos profissionais de Tecnologia da Informação (TI) na geração de relatórios também é um fator importante, visto que, antes, para cada visualização e/ou combinação dos dados, era necessário o desenvolvimento de novos programas, muitas vezes, utilizados somente em situações momentâneas. Tais consultas passaram a ser efetuadas diretamente pelo pesquisador, proporcionando grande flexibilidade na combinação, visualização e composição dos dados, além de permitir um acompanhamento mais eficiente, visto que os mesmos, passaram a estar disponíveis diariamente para a elaboração das consultas necessárias, tornando assim mais ágil o processo de tomada de decisão.

Um grande benefício obtido foi a redução do tempo gasto para a obtenção de informações utilizadas no processo de análise da pesquisa clínica e cálculo dos indicadores do LaPClinVigiLeish, pois o tempo, que hoje é utilizado para a elaboração das planilhas passará a ser utilizado em análises mais efetivas que desencadearão controles e atitudes mais eficientes.

O aumento na rapidez da análise dos dados já pode ser observado ao analisar os resultados de consultas realizadas na ferramenta.

## REFERÊNCIAS BIBLIOGRÁFICAS

- BERRAHOU, L. et al. A quality-aware spatial data warehouse for querying hydroecological data. **Computers & Geosciences**, v. 85, p. 126–135, 1 dez. 2015.
- BEUREN, I. M.; MARTINS, L. W. Sistema de informações executivas: suas características e reflexões sobre sua aplicação no processo de gestão. **Revista Contabilidade & Finanças**, v. 12, n. 26, p. 6–24, ago. 2001.
- BOUADI, T. et al. A data warehouse to explore multidimensional simulated data from a spatially distributed agro-hydrological model to improve catchment nitrogen management. **Environmental Modelling & Software**, v. 97, p. 229–242, 1 nov. 2017.
- BOULIL, K.; BIMONTE, S.; PINET, F. Conceptual model for spatial data cubes: A UML profile and its automatic implementation. **Computer Standards & Interfaces**, v. 38, p. 113–132, 1 fev. 2015.
- BRASIL, 2017. Manual de Vigilância da Leishmaniose Tegumentar Americana, 1. ed. Ministério da Saúde, Brasília (DF).
- BRÉANT, C. et al. Design of a Multi Dimensional Database for the Archimed DataWarehouse. **Stud Health Technol Inform**, v. 116, p. 169–74, set. 2005.
- ELMASRI, R.; NAVATHE, S. B. **Sistemas de banco de dados**. 4. ed. São Paulo: Pearson Education do Brasil, 2005.
- FISCHER, M. et al. DWARF--a data warehouse system for analyzing protein families. **BMC Bioinformatics**, v. 7, p. 495–495, nov. 2006.
- HUANG, S.-M.; CHOU, T.-H.; SENG, J.-L. Data warehouse enhancement: A semantic cube model approach. **Information Sciences**, v. 177, n. 11, p. 2238–2254, 1 jun. 2007.
- HUANG, Y.-S.; DUY, D.; FANG, C.-C. Efficient maintenance of basic statistical functions in data warehouses. **Decision Support Systems**, v. 57, p. 94–104, 1 jan. 2014.
- INMON, W. H. **Como construir o data warehouse**. 2. ed. Rio de Janeiro: Campus, 1997.
- KASER, O.; LEMIRE, D. Attribute value reordering for efficient hybrid OLAP. **Information Sciences**, v. 176, n. 16, p. 2304–2336, 22 ago. 2006.
- KIMBALL, R.; ROSS, M. **The data warehouse toolkit: the complete guide to dimensional modeling**. 2. ed., [Nachdr.] ed. New York, NY: Wiley, 2002.
- KOEHLER, J. et al. Linking experimental results, biological networks and sequence analysis methods using Ontologies and Generalised Data Structures. **In Silico Biol**, v. 5, n. 1, p. 33–44, jun. 2005.

- LAMAS, E. et al. Ethical, Legal and Social Issues related to the health data-warehouses: re-using health data in the research and public health research. **Stud Health Technol Inform**, v. 210, p. 719–23, maio 2015.
- LECHTENBÖRGER, J.; VOSSEN, G. Multidimensional normal forms for data warehouse design. **Information Systems**, v. 28, n. 5, p. 415–434, 1 jul. 2003.
- LEVENE, M.; LOIZOU, G. Why is the snowflake schema a good data warehouse design? **Information Systems**, v. 28, n. 3, p. 225–240, 1 maio 2003.
- MANSMANN, S. et al. Discovering OLAP dimensions in semi-structured data. **Information Systems**, v. 44, p. 120–133, 1 ago. 2014.
- NERY, F. **Tecnologia e Projeto de Data Warehouse**. 5. ed. São Paulo: Érica, 2012.
- RASLAN, D. A.; CALAZANS, A. T. S. **Data Warehouse: conceitos e aplicações** - DOI: 10.5102/un.gti.v4i1.2612. **Universitas: Gestão e TI**, v. 4, n. 1, 4 ago. 2014.
- ROELOFS, E. et al. Benefits of a clinical data warehouse with data mining tools to collect data for a radiotherapy trial. **Radiother Oncol**, v.108, n.1, p. 9-174, fev. 2013.
- SANTOSO, L. W.; YULIA. Data Warehouse with Big Data Technology for Higher Education. **Procedia Computer Science**, 4th Information Systems International Conference 2017, ISICO 2017, 6-8 November 2017, Bali, Indonesia. v. 124, p. 93–99, 1 jan. 2017.
- SILBERSCHATZ, A.; KORTH, H.; SUDARSHAN, S. **Sistema de Banco de Dados**. São Paulo: Makron Books, 1999.
- SONG, J. et al. HaOlap: A Hadoop based OLAP system for big data. **Journal of Systems and Software**, v. 102, p. 167–181, 1 abr. 2015.
- SPITZER, A. R.; ELLSBURY, D.; CLARK, R. H. The Pediatrix BabySteps® Data Warehouse — A Unique National Resource for Improving Outcomes for Neonates. **The Indian Journal of Pediatrics**, v. 82, n. 1, p. 71–79, 1 jan. 2015.
- ZHOU, X. et al. Development of traditional Chinese medicine clinical data warehouse for medical knowledge discovery and decision support. **Artificial Intelligence in Medicine**, v. 48, n. 2–3, p. 139–152, fev. 2010.

## APÊNDICE A - TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO (TCLE)

### Termo de Consentimento Livre e Esclarecido (TCLE)

**\*Versão 01 de 21/05/2018**

Prezado participante, você está sendo convidado a participar da pesquisa “Implementação do *Data Mart* na Pesquisa Clínica: Estudo de Caso no Laboratório de Pesquisa Clínica e Vigilância em Leishmanioses (LaPClinVigiLeish)”, desenvolvida por Patrícia Costa dos Santos, discente do Mestrado Profissional em Pesquisa Clínica do INI sob orientação da Doutora Aline Fagundes.

O objetivo desta pesquisa é Implementar um *Data Mart* no LaPClinVigiLeish de modo a apoiar à tomada de decisão dos pesquisadores em um menor tempo possível com uma melhor qualidade na análise dos dados da pesquisa.

Para alcançar o objetivo será realizada a coleta de informações sobre os dados referentes as pesquisas realizadas no LaPClinVigiLeish. Estes dados serão integrados e transformados em informação pertinente para fins de suporte à decisão, e disponibilizados em um ambiente seguro, de acesso rápido e com qualidade denominado *Data Mart*. Este, é um banco de dados com capacidade de organizar um grande volume de dados retirados de diversas fontes (Sistema de Prontuário Eletrônico do INI (SIPEC), planilhas eletrônicas e banco de dados Access®) e armazená-los em um único local, que permitirá aos pesquisadores o acesso as consultas a qualquer momento através de ferramentas que possuem uma interface gráfica amigável.

Após a implementação deste projeto os pesquisadores do LaPClinVigiLeish terão o benefício de obter os dados de maneira mais rápida a partir de um grande volume de dados operacionais, sendo possível realizar o cruzamento de informações de modo mais dinâmico e intuitivo.

Será possível mudar totalmente a visualização dos dados na ferramenta com um mínimo de tempo e esforço, será disponibilizada também a opção de gerar gráficos a partir destes dados sem a necessidade de realizar nenhuma programação no código.

Este estudo incluirá dados das pesquisas do LaPClinVigiLeish obtidos no intervalo de tempo de 9 anos, de 1 de janeiro de 2010 a 31 de dezembro de 2019.

Serão realizadas 02 (duas) entrevistas de forma presencial onde será aplicado um questionário contendo algumas perguntas sobre os relatórios, dados e análises que são realizadas rotineiramente. As entrevistas serão realizadas de forma

individualizada, em local adequado, com tempo estimado de 1 hora e será agendada de acordo com a sua conveniência e disponibilidade.

O convite à sua participação se deve ao fato de você atuar como pesquisador do LaPClinVigiLeish, utilizar o Sistema de Prontuário Eletrônico do INI (SIPEC) e planilhas eletrônicas para registro das pesquisas realizadas no LaPClinVigiLeish. Sua participação é inteiramente voluntária, entretanto, se quiser, poderá deixar de responder a qualquer pergunta durante a entrevista, ou deixar de participar da pesquisa a qualquer momento. Contudo ela é muito importante para a execução da pesquisa. Nenhuma remuneração será oferecida, não havendo nenhum custo para sua participação.

A qualquer momento, durante a pesquisa, ou posteriormente, você poderá solicitar dos pesquisadores, informações sobre sua participação e/ou sobre a pesquisa, o que poderá ser feito através dos meios de contato explicitados neste Termo.

Os riscos quanto à confidencialidade serão minimizados com o sigilo das informações. Qualquer dado que possa identificá-lo será omitido na divulgação dos resultados da pesquisa, e o material será armazenado em local seguro. Todos os dados serão digitados e armazenados em arquivos digitais. Somente terão acesso aos dados da entrevista a discente responsável pelo desenvolvimento da pesquisa e sua orientadora.

Os resultados desta pesquisa serão divulgados posteriormente para o INI e em eventos e/ou revistas científicas.

Em caso de dúvida quanto à condução ética do estudo, entre em contato com o comitê de ética em pesquisa responsável pela análise e aprovação do estudo.

### **Comitê de Ética em Pesquisa do INI-FIOCRUZ**

Endereço: Av. Brasil, 4365, Manguinhos – RJ. CEP 21040-900

Tel.: 3865-9585 / 3865-9107

E-mail: cep@ini.fiocruz.br

Caso você queira tirar dúvidas sobre o projeto e sua participação, entre em contato com a pesquisadora principal e a discente responsável pelo desenvolvimento da pesquisa.

### **Pesquisadora Responsável**

Orientadora

Dra. Aline Fagundes da Silva - Coordenadora do  
Laboratório de Pesquisa Clínica e Vigilância em Leishmanioses do INI/FIOCRUZ  
Endereço: Av. Brasil, 4365, Manguinhos – RJ. CEP 21040-900  
Tel.: (21)3865-9541  
E-mail: aline.fagundes@ini.fiocruz.br

**Responsável pelo desenvolvimento da pesquisa**

Mestranda

Patrícia Costa dos Santos  
Endereço: Av. Brasil, 4365, Manguinhos – RJ. CEP 21040-900  
Tel.: (21)3865-9556  
E-mail: patricia.costa@ini.fiocruz.br

Declaro estar ciente e que compreendi as informações constantes neste “Termo de Consentimento Livre e Esclarecido” e concordo em participar da pesquisa.

---

Nome do participante da pesquisa

\_\_\_\_\_ Data: \_\_\_\_/\_\_\_\_/\_\_\_\_

Assinatura do participante da pesquisa

---

Nome do responsável pela aplicação do TCLE

\_\_\_\_\_ Data: \_\_\_\_/\_\_\_\_/\_\_\_\_

Assinatura do responsável pela aplicação do TCLE

Este termo contém 3 páginas e redigido em duas vias, sendo uma para o participante e outra para o responsável pelo desenvolvimento da pesquisa. Todas as páginas deverão ser rubricadas pelo participante da pesquisa e pelo responsável pela aplicação do TCLE.



## APÊNDICE B - QUESTIONÁRIO DE REQUISITOS SOBRE BANCO DE DADOS

### QUESTIONÁRIO DE REQUISITOS SOBRE BANCO DE DADOS

\*Versão 01 de 21/05/2018

Data da entrevista: \_\_\_\_/\_\_\_\_/\_\_\_\_

| OBJETIVO   |
|--|
| <p>Identificar e descrever processo de coleta e armazenamento dos dados do Laboratório de Pesquisa Clínica e Vigilância em Leishmaniose no Instituto Nacional de Infectologia Evandro Chagas (INI), sob a perspectiva de implantação de um <i>Data Mart</i> (repositório de dados) que viabilize agilidade no acesso à informação, potencial produção compartilhada de conhecimento e melhor qualidade na análise dos dados para a pesquisa clínica.</p> |

| IDENTIFICAÇÃO DO USUÁRIO |
|--------------------------|
| Nome:                    |
| Especialidade:           |
| Serviço/Setor:           |

| LOCAL DE OBTENÇÃO DAS INFORMAÇÕES UTILIZADAS NA ELABORAÇÃO DOS RELATÓRIOS   |
|---|
| <input type="checkbox"/> Planilhas eletrônicas <input type="checkbox"/> SIPEC <input type="checkbox"/> Banco de dados próprio |
| Outras fontes:<br>_____<br>_____  |
| Nome do relatório/arquivo:  |
| Extensão do relatório/arquivo:  |

|   |
|---|
| Quem manipula?  |
| Com que periodicidade é utilizado?  |
| É atualizado? Se sim, com que frequência?   |
| Quais dados/colunas são utilizados?   |
| Onde é aplicada essa informação?  |
| Quem solicita essa análise?   |
| Utiliza algum dado de um outro relatório/arquivo para complementar o relatório primário? Se sim, qual dado e relatório/arquivo é utilizado? |
| O relatório utilizado possui todos os dados que você necessita? Se não, quais são eles? Onde podem ser obtidos?                             |
| Caso o relatório seja manipulado, quanto tempo é gasto para a conclusão do mesmo?   |
| Qual o período de informação histórica é necessário?  |

Este questionário contém 2 páginas e redigido em duas vias, sendo uma para o participante e outra para o responsável pelo desenvolvimento da pesquisa. Todas as páginas deverão ser rubricadas pelo participante da pesquisa e pelo responsável pela aplicação do questionário.

**APÊNDICE C – TERMO DE COMPROMISSO E CONFIDENCIALIDADE****TERMO DE COMPROMISSO E RESPONSABILIDADE**

**Projeto:** “Implementação do *Data Mart* na Pesquisa Clínica: Estudo de caso no Laboratório de Pesquisa Clínica e Vigilância em Leishmanioses”

**Pesquisador:** nome do participante da pesquisa

**Instituição:** Instituto Nacional de Infectologia Evandro Chagas – FIOCRUZ

Parecer CEP/INI 3.061.965

Eu, (nome do participante da pesquisa), (cargo) do projeto de pesquisa acima intitulado, comprometo-me a manter a confidencialidade e a privacidade dos sujeitos da pesquisa. A identidade dos participantes, assim como os resultados obtidos com este projeto serão mantidos em um banco de dados sob a minha responsabilidade. Os resultados serão divulgados em comunicações científicas mantendo o anonimato dos participantes e o material utilizado não será empregado em outras pesquisas, a não ser quando abertos novos protocolos.

Rio de Janeiro, (dia) de (mês) de 2019.

---

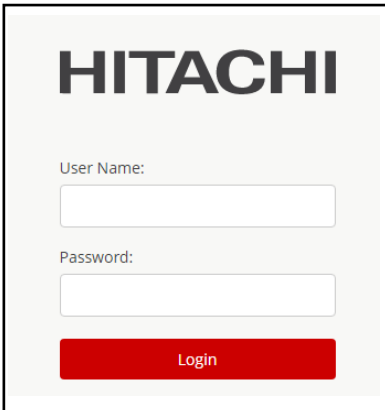
Assinatura do participante da pesquisa

## APÊNDICE D – MANUAL DE USO DO *PENTAHO USER CONSOLE*

### Usando a Pentaho *User Console* (PUC)

1. Para acessar a *Pentaho User Console* abra o seu web-browser (Internet Explorer/Firefox/Chrome/Safari) e digite no campo URL o seguinte endereço, <https://pentaho/login>.

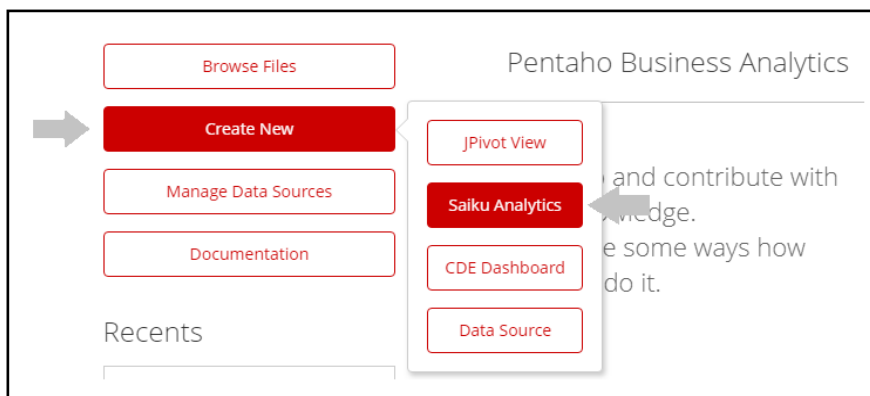
Figura 1 – Tela de login para a *Pentaho User Console*



Fonte: Elaborada pela autora.

2. Para conectar a *Pentaho User Console* entre com o seu *login* e sua *password* e clique no botão *Login* conforme a figura 1 abaixo.

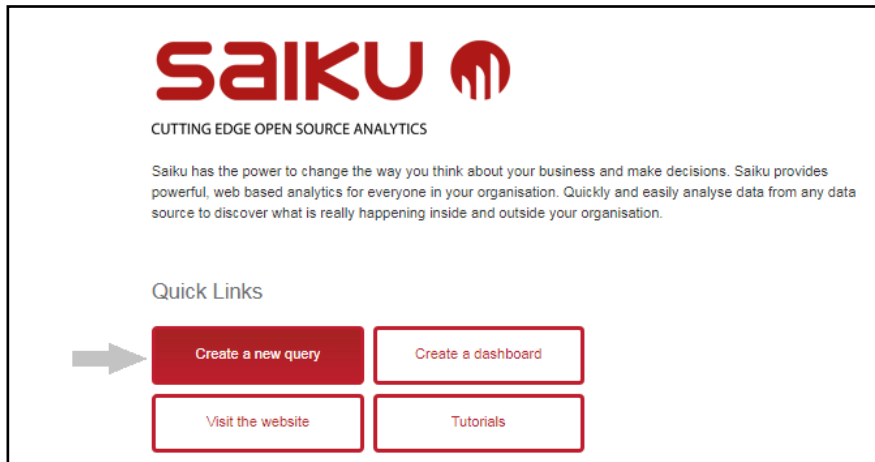
Figura 2 – Tela para abrir o *Saiku Analytics*



Fonte: Elaborada pela autora.

3. Para criar um relatório no *Saiku Analytics*, clique no botão *Create New*, depois no botão *Saiku Analytics* conforme a figura 2 e clique no botão *Create a new query* conforme a figura 3.

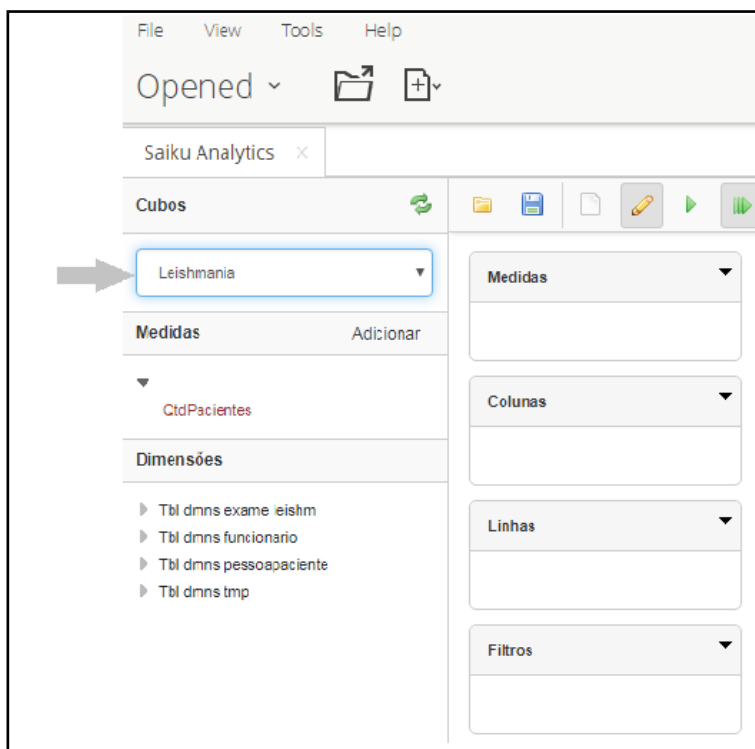
Figura 3 – Tela de entrada do *Saiku Analytics*



Fonte: Elaborada pela autora.

4. Na tela de consultas do *Saiku Analytics*, clique na caixa de combinação Cubos e escolha o cubo Leishmania conforme a figura 4.

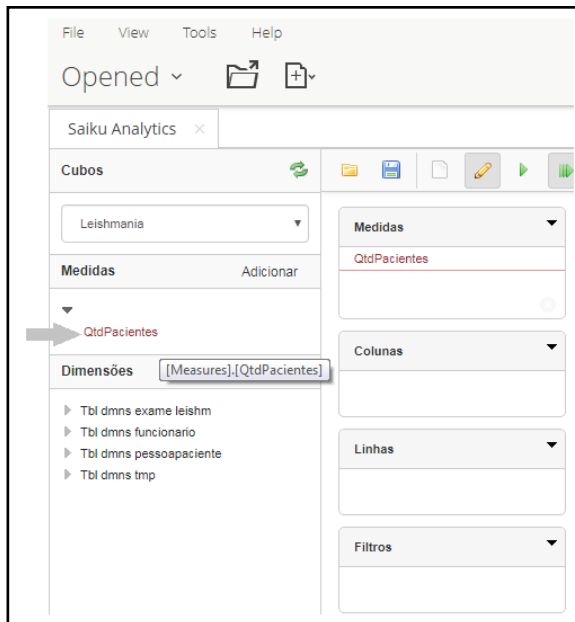
Figura 4 – Tela de consultas do *Saiku Analytics*



Fonte: Elaborada pela autora.

5. Para elaborar o relatório, selecione a medida QtdPacientes clicando em cima do nome conforme mostrado na figura 5.

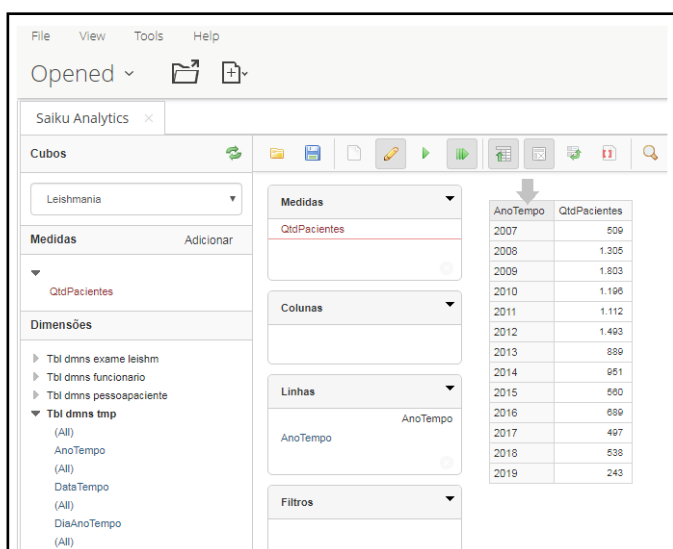
Figura 5 – Tela de consultas do Saiku Analytics



Fonte: Elaborada pela autora.

6. Após a medida ter sido selecionada, é necessário selecionar as dimensões e os respectivos campos que servirão de base para a construção dos relatórios. Na figura 6, foi selecionado somente o campo AnoTempo da dimensão Tempo. A quantidade de pacientes é exibida para cada ano na tela.

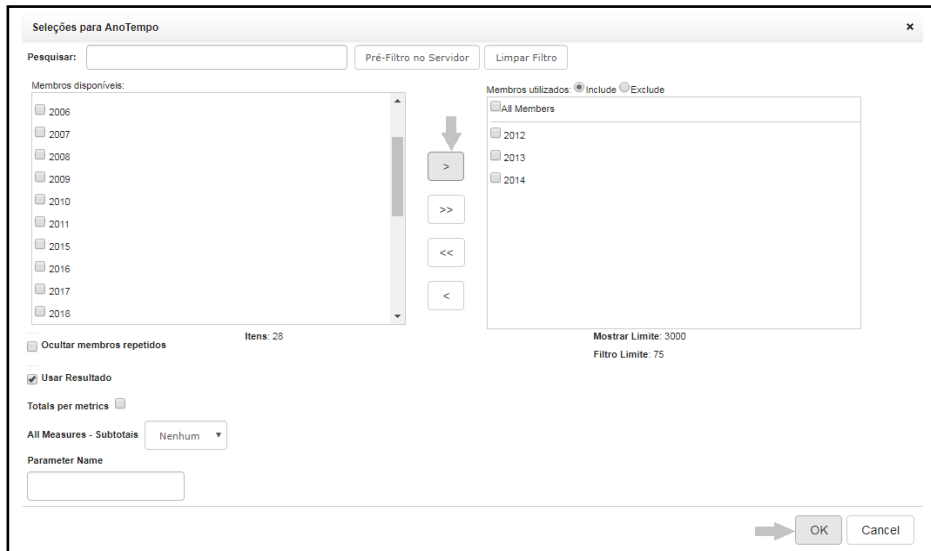
Figura 6 – Seleção do nível do ano na dimensão Tempo



Fonte: Elaborada pela autora.

7. Ao clicar no campo AnoTempo, abaixo de Linhas, o usuário pode criar um filtro para este campo conforme figura 7. Então seleciona-se os anos no painel à esquerda da figura 7, clicando no botão “>” e os anos selecionados serão movidos para o painel à direita. Então clique no botão “Ok” e os resultados serão exibidos somente para os anos selecionados.

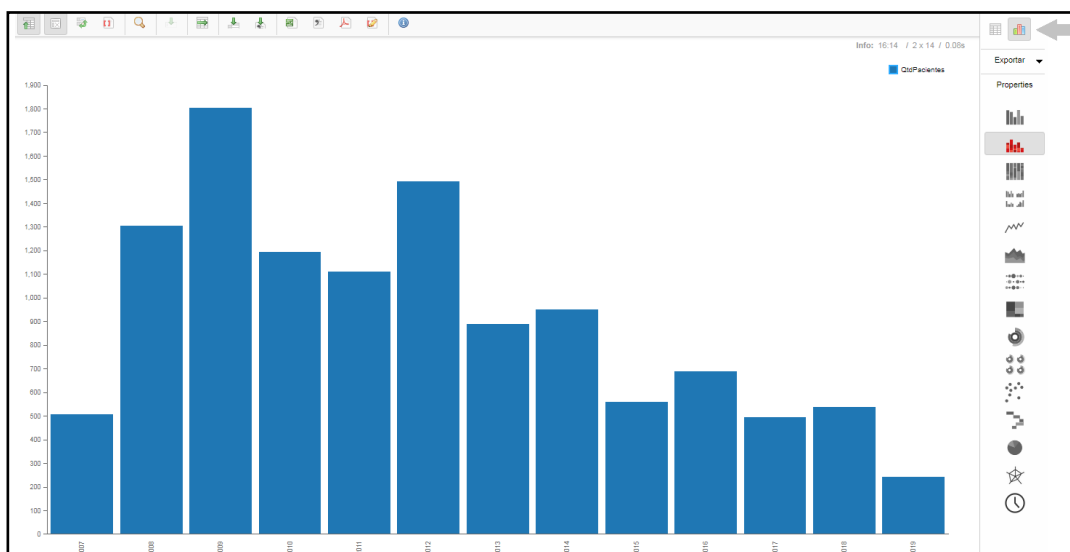
Figura 7 – Seleção de filtros para o campo AnoTempo



Fonte: Elaborada pela autora.

8. Ao clicar no botão indicado pela seta cinza na figura 8, o usuário pode escolher uma variedade de gráficos exibidos no lado direito da tela. Neste exemplo, os resultados são mostrados em um gráfico de barras empilhadas.

Figura 8 – Modo gráfico



Fonte: Elaborada pela autora.

9. Os resultados podem também ser exportados para outros formatos. Por exemplo, ao clicar no botão indicado pelas setas cinzas na figura 9, o usuário pode exportar os resultados para um arquivo do Excel ou PDF.

Figura 9 – Outros formatos de visualização

| AnoTempo | QtyPacientes |
|----------|--------------|
| 2007     | 509          |
| 2008     | 1.305        |
| 2009     | 1.803        |
| 2010     | 1.196        |
| 2011     | 1.112        |
| 2012     | 1.493        |
| 2013     | 889          |
| 2014     | 951          |
| 2015     | 560          |
| 2016     | 689          |
| 2017     | 497          |
| 2018     | 538          |
| 2019     | 243          |

Fonte: Elaborada pela autora.

10. Os resultados podem também ser visualizados de uma outra maneira. Por exemplo, ao clicar no botão indicado pela seta cinza na figura 10, o que é linha vira coluna e vice-versa.

Figura 10 – Outras formas de visualização, trocar linha por coluna

| Nome Exame - Nome Exame | CULTURA PARA LEISHMANIOSE | DIRETO PARA LEISHMANIOSE |
|-------------------------|---------------------------|--------------------------|
| AnoTempo                | QtyPacientes              | QtyPacientes             |
| 2007                    | -                         | -                        |
| 2008                    | -                         | -                        |
| 2009                    | -                         | -                        |
| 2010                    | -                         | -                        |
| 2011                    | 21                        | 8                        |
| 2012                    | 92                        | 47                       |
| 2013                    | 80                        | 49                       |
| 2014                    | 62                        | 65                       |
| 2015                    | 112                       | 2                        |
| 2016                    | 74                        | 2                        |
| 2017                    | 72                        | 8                        |
| 2018                    | 87                        | 49                       |
| 2019                    | 23                        | 21                       |

Fonte: Elaborada pela autora.



## APÊNDICE E – DICIONÁRIO DE DADOS

### Data Mart Leishmaniose - Dicionário de dados

#### Dimensão: Tbl\_dmns\_consulta

| Variável       | Descrição                            | Tipo         |
|----------------|--------------------------------------|--------------|
| <b>CodCons</b> | Código da consulta do paciente       | Inteiro      |
| DtConsulta     | Data em que ocorreu a consulta       | Datetime     |
| Funcionario    | Especialista que realizou a consulta | Varchar(250) |

#### Dimensão: Tbl\_dmns\_dadospaciente

| Variável          | Descrição   | Tipo         |
|-------------------|---|--------------|
| <b>CodPessoa</b>  | Código de identificação da pessoa                     | Inteiro      |
| NumProntuario     | Número do prontuário                                  | Varchar(20)  |
| NumPrimAtend      | Número do primeiro atendimento                        | Varchar(20)  |
| Nome              | Nome do paciente                                      | Varchar(250) |
| DtNascimento      | Data de nascimento                                    | Datetime     |
| Sexo              | Sexo  | Varchar(50)  |
| NomeSocial        | Nome Social   | Varchar(250) |
| Genero            | Gênero  | Varchar(50)  |
| Cidade            | Cidade  | Varchar(400) |
| Estado            | Estado  | Varchar(50)  |
| FormaClinica      | Forma Clinica da doença                               | Varchar(70)  |
| IdentificaEspecie | Nome da espécie                                       | Varchar(70)  |
| IntradermoMonten  | Resultado do teste de Intradermorreação de montenegro | Inteiro      |

|              |   |              |
|--------------|---|--------------|
| AreaEndemica | Área Endêmica   | Varchar(250) |
| Dtbito       | Data do óbito   | Datetime     |
| CausaObito   | Causa do óbito  | CausaObito   |
| CID          | Classificação Estatística Internacional de Doenças e Problemas Relacionados à Saúde | Varchar(400) |

**Dimensão: Tbl\_dmns\_encaminhamentopaciente**

| Variável                       | Descrição                                    | Tipo         |
|--------------------------------|--|--------------|
| IdEncaminham <sup>ento</sup>   | Código de identificação do encaminhamento    | Inteiro      |
| CodConsulta                    | Código da consulta do paciente               | Inteiro      |
| CodEnc                         | Código do encaminhamento                     | Inteiro      |
| TipoEncaminham <sup>ento</sup> | Tipo do encaminhamento feito                 | Varchar(50)  |
| DtEncaminham <sup>ento</sup>   | Data do encaminhamento                       | Datetime     |
| Instituicao                    | Instituição que o paciente foi encaminhado   | Varchar(250) |
| Especialidade                  | Especialidade que o paciente foi encaminhado | Varchar(50)  |
| Laboratorio                    | Laboratório que o paciente foi encaminhado   | Varchar(100) |

**Dimensão: Tbl\_dmns\_exame**

| Variável       | Descrição                         | Tipo         |
|----------------|-----------------------------------|--------------|
| IdExame        | Código de identificação do exame  | Inteiro      |
| IdSolSubexame  | Código da solicitação do subexame | Inteiro      |
| IdSolExame     | Código da solicitação do exame    | Inteiro      |
| NoTipoExame    | Nome do tipo do exame             | Varchar(50)  |
| NoItemExame    | Nome do item do exame             | Varchar(100) |
| NoTipoMaterial | Nome do material usado no exame   | Varchar(50)  |

|             |                             |              |
|-------------|-----------------------------|--------------|
| DtColeta    | Data da coleta do material  | Datetime     |
| DtResultado | Data do resultado do exame  | Datetime     |
| Resultado   | Resultado do exame          | Varchar(50)  |
| Solicitante | Médico solicitante do exame | Varchar(250) |

**Dimensão: Tbl\_dmns\_tmp**

| Variável           | Descrição                        | Tipo        |
|--------------------|----------------------------------|-------------|
| Tempold            | Código de identificação do tempo | Inteiro     |
| DataTempo          | Data contendo dia, mês e ano     | Datetime    |
| AnoTempo           | Ano                              | Inteiro     |
| MesTempo           | Mês em numeral                   | Inteiro     |
| DiaAnoTempo        | Dia                              | Inteiro     |
| DiaSemanaDescTempo | Dia da semana descritivo         | Varchar(30) |
| MesDescTempo       | Mês descritivo                   | Varchar(30) |
| TrimestreTempo     | Trimestre                        | Varchar(5)  |

**Fato: Tbl\_fato**

| Variável    | Descrição                                 | Tipo    |
|-------------|---|---------|
| IdFato      | Código de identificação                   | Inteiro |
| Tempold     | Código de identificação do tempo          | Inteiro |
| IdExame     | Código de identificação do exame          | Inteiro |
| CodPessoa   | Código de identificação da pessoa         | Inteiro |
| CodCons     | Código de identificação da consulta       | Inteiro |
| IdEncaminha | Código de identificação do encaminhamento | Inteiro |

|             |                              |         |
|-------------|------------------------------|---------|
| QtdEnc      | Quantidade de encaminhamento | Inteiro |
| QtdConsulta | Quantidade de consulta       | Inteiro |
| QtdExame    | Quantidade de exame          | Inteiro |
| QtdPaciente | Quantidade de paciente       | Inteiro |

## APÊNDICE F – CÓDIGO DE CRIAÇÃO DOS OBJETOS NOS BANCOS DE DADOS

### Código fonte da criação dos objetos nos bancos de dados

#### 1 – Criação da view vw\_resultadoexame\_leishmaniose

```

drop view vw_resultadoexame_leishmaniose
go
USE [cenlab]
GO

SET ANSI_NULLS ON
GO

SET QUOTED_IDENTIFIER ON
GO

create view [dbo].[vw_resultadoexame_leishmaniose] as SELECT TOP (100)
PERCENT
dbo.tb_ValorTipoResultado.de_valor_tipo_resultado,
dbo.tb_ValorTipoResultado.id_valor_tipo_resultado,
dbo.tb_ValorTipoResultado.fk_id_tipo_resultado,
dbo.tb_Situacao.no_situacao,
dbo.tb_SolExame.id_sol_exame,
SollItem.id_sol_item_exame,
dbo.tb_Resultado.dt_resultado,
dbo.tb_PerguntaResultado.id_pergunta_resultado,
dbo.tb_TipoExame.id_tipo_exame
FROM      dbo.tb_ItemKit_KitExame_ItemExame RIGHT OUTER JOIN
          dbo.tb_Result_TpResult_ItemExame INNER JOIN
          dbo.tb_TpResult_ItemExame ON
dbo.tb_Result_TpResult_ItemExame.fk_id_tpresult_itemexame =
dbo.tb_TpResult_ItemExame.id_tpresult_itemexame INNER JOIN
          dbo.tb_PerguntaResultado ON
dbo.tb_TpResult_ItemExame.fk_id_pergunta_resultado =
dbo.tb_PerguntaResultado.id_pergunta_resultado RIGHT OUTER JOIN
          dbo.tb_TipoMaterial INNER JOIN
          dbo.tb_Laboratorio INNER JOIN
          dbo.tb_Situacao INNER JOIN
          dbo.tb_Motivo_Situacao_ItemExame INNER JOIN
          dbo.tb_SolExame INNER JOIN
          dbo.tb_SollItemExame AS SollItem ON dbo.tb_SolExame.id_sol_exame
= SollItem.fk_id_sol_exame INNER JOIN
          dbo.tb_TipoExame INNER JOIN
          dbo.tb_ItemExame ON dbo.tb_TipoExame.id_tipo_exame =
dbo.tb_ItemExame.fk_id_tipo_exame ON

```

```

        SollItem.fk_id_item_exame = dbo.tb_ItemExame.id_item_exame ON
dbo.tb_Motivo_Situacao_ItemExame.fk_id_item_exame = SollItem.fk_id_item_exame
AND    dbo.tb_Motivo_Situacao_ItemExame.fk_id_situacao =
SollItem.fk_ultima_situacao ON
        dbo.tb_Situacao.id_situacao =
dbo.tb_Motivo_Situacao_ItemExame.fk_id_situacao INNER JOIN
        dbo.tb_Resultado ON SollItem.dt_ultima_situacao =
dbo.tb_Resultado.dt_sistema AND SollItem.id_sol_item_exame =
dbo.tb_Resultado.fk_id_sol_item_exame ON
        dbo.tb_Laboratorio.id_laboratorio = SollItem.fk_id_laboratorio ON
dbo.tb_TipoMaterial.id_tipo_material = SollItem.fk_id_tipo_material INNER JOIN
        dbo.tb_AmostraMaterial ON dbo.tb_TipoMaterial.id_tipo_material =
dbo.tb_AmostraMaterial.fk_id_tipo_material AND
        dbo.tb_SolExame.id_sol_exame =
dbo.tb_AmostraMaterial.fk_id_sol_exame ON
        dbo.tb_TpResult_ItemExame.fk_id_item_exame =
dbo.tb_ItemExame.id_item_exame AND
        dbo.tb_Result_TpResult_ItemExame.fk_id_resultado =
dbo.tb_Resultado.id_resultado LEFT OUTER JOIN
        dbo.tb_ValorTipoResultado INNER JOIN
        dbo.tb_tiporesultado_valor ON
        dbo.tb_ValorTipoResultado.fk_id_tipo_resultado =
        dbo.tb_tiporesultado_valor.fk_id_tipo_resultado AND
        dbo.tb_ValorTipoResultado.id_valor_tipo_resultado =
        dbo.tb_tiporesultado_valor.fk_id_valor ON
        dbo.tb_Result_TpResult_ItemExame.fk_id_tpresult_itemexame =
        dbo.tb_tiporesultado_valor.fk_id_tpresult_itemexame AND
        dbo.tb_Result_TpResult_ItemExame.fk_id_resultado =
        dbo.tb_tiporesultado_valor.fk_id_resultado LEFT OUTER JOIN
        dbo.tb_Antibiotico INNER JOIN
        dbo.tb_tiporesultado_antibiotico ON dbo.tb_Antibiotico.id_Antibiotico =
        dbo.tb_tiporesultado_antibiotico.fk_id_antibiotico ON
        dbo.tb_Result_TpResult_ItemExame.fk_id_tpresult_itemexame =
        dbo.tb_tiporesultado_antibiotico.fk_id_tpresult_itemexame AND
        dbo.tb_Result_TpResult_ItemExame.fk_id_resultado =
        dbo.tb_tiporesultado_antibiotico.fk_id_resultado ON
        dbo.tb_ItemKit_KitExame_ItemExame.fk_id_item_exame =
        dbo.tb_Resultado.fk_id_item_exame AND
        dbo.tb_ItemKit_KitExame_ItemExame.fk_id_kit_exame =
        dbo.tb_Resultado.fk_id_kit_exame AND
        dbo.tb_ItemKit_KitExame_ItemExame.fk_id_item_kit =
        dbo.tb_Resultado.fk_id_item_kit LEFT OUTER JOIN
        dbo.tb_Bacteria INNER JOIN
        dbo.tb_tiporesultado_bacteria ON dbo.tb_Bacteria.id_bacteria =
        dbo.tb_tiporesultado_bacteria.fk_id_bacteria ON
        dbo.tb_Result_TpResult_ItemExame.fk_id_tpresult_itemexame =
        dbo.tb_tiporesultado_bacteria.fk_id_tpresult_itemexame AND
        dbo.tb_Result_TpResult_ItemExame.fk_id_resultado =
        dbo.tb_tiporesultado_bacteria.fk_id_resultado LEFT OUTER JOIN

```

```

        dbo.tb_Situacao_SolltemExame AS TbSituacao ON
        Solltem.id_sol_item_exame = TbSituacao.fk_id_sol_item_exame AND
        TbSituacao.fk_id_situacao = 2 AND
        TbSituacao.fk_id_motivo = 1 AND TbSituacao.fk_id_item_exame =
        dbo.tb_ItemExame.id_item_exame LEFT OUTER JOIN
        dbo.tb_Maquina AS Maquina ON Maquina.id_maquina =
        dbo.tb_Resultado.equipamento LEFT OUTER JOIN
        dbo.tb_Maquina AS M ON M.id_maquina =
        dbo.tb_Resultado.equipamento LEFT OUTER JOIN
        dbo.tb_ValorTipoResultado AS VR ON VR.id_valor_tipo_resultado =
        dbo.tb_Resultado.metodologia AND VR.fk_id_tipo_resultado = 289
        where dbo.tb_TipoExame.id_tipo_exame in(52,54,172,201)
        and dbo.tb_PerguntaResultado.id_pergunta_resultado in(59,184,185)
        and dbo.tb_ValorTipoResultado.fk_id_tipo_resultado
        in(129,224,30,240,238,344)
        and dbo.tb_Situacao.no_situacao = 'FINALIZADO'
        and dbo.tb_Resultado.dt_resultado >= '20100101'

```

GO

## **2 – Criação da tabela tb\_resultadoexame leishmaniose**

```
drop table tb_resultadoexame_leishmaniose
```

GO

```

SELECT * INTO tb_resultadoexame_leishmaniose FROM
vw_resultadoexame_leishmaniose
go

```

## **3 – Criação do índice**

```

CREATE NONCLUSTERED INDEX [_dta_index_tb_resultadoexame_leishmaniose]
ON [dbo].[tb_resultadoexame_leishmaniose]
(
[id_sol_exame] ASC,
[de_valor_tipo_resultado] ASC
)
INCLUDE ( [dt_resultado]) WITH (SORT_IN_TEMPDB = OFF, IGNORE_DUP_KEY
= OFF, DROP_EXISTING = OFF, ONLINE = OFF) ON [PRIMARY]

```

## **4 – Criação da função fn\_ret\_diag leishmaniose**

```

USE [cenlab]
GO

```

```

SET ANSI_NULLS ON
GO

```

```

SET QUOTED_IDENTIFIER ON

```

GO

```
Create function [dbo].[fn_ret_diag_leishmaniose](@codpessoa int)
returns varchar(20)
```

```
as
begin
declare
```

```
@resultpos varchar(50),@resultneg varchar(50), @resultfinal varchar(20), @cid
varchar(500), @dtprimexame datetime , @anomesprimexame varchar(10), @mes
int, @ano int, @anomesresult varchar(10)
```

```
-----Busco o mes e ano do primeiro exame -----
```

```
Select @dtprimexame = cenlab.dbo.fn_ret_dt_primeiro_exame(@codpessoa)
Set @mes = month(@dtprimexame)
Set @ano = year(@dtprimexame)
--Set @anomesprimexame = Convert(Varchar(2),@mes) +
Convert(Varchar(4),@ano)
Set @anomesprimexame = ltrim( str(@mes)) + '/01/' + ltrim( str(@ano))
```

```
--select @anomesprimexame
```

```
If @dtprimexame is null
Begin
```

```
Select top 1 @cid =
ambulatorio.dbo.ret_todos_cids_paciente_leishmania(@codpessoa)
IF (@cid is not Null and @cid <>")
Begin
set @resultfinal = 'Positivo'
end
End
Else
Begin
```

```
-----Verifico CID com exame-----
```

```
-- Tem a data de Exame, testa o resultado pois a CID ainda não foi lançada ou não
tem
```

```
select @cid =
ambulatorio.dbo.ret_todos_cids_paciente_leishmania_com_exame(@codpessoa,
@anomesprimexame)
```

```
-- Positivo se ele tem exame com CID e tem data de exame igual a data da CID
```

```
IF (@cid is not Null and @cid <>")
```

```
Begin
set @resultfinal = 'Positivo'
end
Else
Begin
```

```
set @resultpos = 'N'
```



```

select distinct top 1
    @resultpos = vws.de_valor_tipo_resultado,
    @mes = month(vws.dt_resultado),
    @ano = year(vws.dt_resultado)
From tb_resultadoexame_leishmaniose vws
Join dbo.tb_SolExame SE on SE.id_sol_exame = vws.id_sol_exame
Join ambulatorio.dbo.cectb49_pessoa PE on PE.CodPessoa =
se.fk_cd_pessoa_dona_material
where PE.CodPessoa = @codpessoa
and (vws.de_valor_tipo_resultado = 'POSITIVO' OR vws.de_valor_tipo_resultado like
'Reator%')

```

```

set @anomesresult = ltrim( str(@mes)) + '/01/' + ltrim( str(@ano))
End

```

```

-- Teste para os exames Negativos
If @resultfinal = 'Positivo' or (@resultpos <> 'N' and (@anomesresult =
@anomesprimexame ))
begin
set @resultfinal = 'Positivo'
end
else
set @resultfinal = 'Negativo'
end

```

```

return @resultfinal;
end

```

```
GO
```

### **5 – Criação da função fn\_ret\_dt\_primeiro\_exame**

```

USE [cenlab]
GO

```

```

SET ANSI_NULLS OFF
GO

```

```

SET QUOTED_IDENTIFIER OFF
GO

```

```

CREATE function [dbo].[fn_ret_dt_primeiro_exame](@codpessoa int)
returns datetime
as
begin
declare @ret datetime;

```

```

select @ret = MIN(vws.dt_resultado)

From tb_resultadoexame_leishmaniose vws
Join dbo.tb_SolExame SE on SE.id_sol_exame = vws.id_sol_exame

where SE.fk_cd_pessoa_dona_material = @codpessoa
return @ret;
end
GO

```

## **6 – Criação da função ret\_todos\_cids\_paciente\_leishmania**

```

USE [ambulatorio]
GO

SET ANSI_NULLS ON
GO

SET QUOTED_IDENTIFIER ON
GO

CREATE function [dbo].[ret_todos_cids_paciente_leishmania](@codpessoa int)
returns nvarchar(max)
as
begin
declare
@ret nvarchar(max),
@CodCid varchar(100);
set @ret="";

DECLARE Vetor
CURSOR FOR

select distinct tb08.NOME
from dbo.cectb10_cid_x_paciente tb10
join dbo.cectb08_cid tb08 ON tb10.CodCid = tb08.CodCid
where CodPessoa = @codpessoa
and tb10.CodCid in('B55','B550','B551','B552','B559')
order by 1 asc;

OPEN vetor
--Atribuindo valores do select nas variáveis
FETCH NEXT FROM vetor INTO @CodCid;
--Iniciando laço
WHILE @@FETCH_STATUS = 0
BEGIN
set @ret = @ret + ltrim(@CodCid) + '|';

```

```
FETCH NEXT FROM vetor INTO @CodCid;
```

```
END
```

```
CLOSE vetor;
DEALLOCATE vetor;
```

```
return @ret;
```

```
end
```

```
GO
```

7 – Criação da função ret\_todos\_cids\_paciente\_leishmania\_com\_exame

```
USE [ambulatorio]
GO
```

```
SET ANSI_NULLS ON
GO
```

```
SET QUOTED_IDENTIFIER ON
GO
```

Create function

```
[dbo].[ret_todos_cids_paciente_leishmania_com_exame](@codpessoa int,
@anomespriexame varchar(10))
```

```
returns nvarchar(max)
```

```
as
```

```
begin
```

```
declare
```

```
@ret nvarchar(max),
```

```
@CodCid varchar(100),
```

```
@mes int, @ano int, @anomesprimcid date ,@dtprimcid datetime
```

```
set @ret="";
```

```
select top 1 @ret = tb08.NOME,
```

```
@dtprimcid = tb10.DtInc
```

```
from dbo.cectb10_cid_x_paciente tb10
```

```
join dbo.cectb08_cid tb08 ON tb10.CodCid = tb08.CodCid
```

```
where CodPessoa = @codpessoa
```

```
and tb10.CodCid in('B55','B550','B551','B552','B559')
```

```
--select @dtprimcid
```

```
-----Busco o mes e ano da inclusao da CID -----
```

```
Set @mes = month(@dtprimcid)
```

```
Set @ano = year(@dtprimcid)
set @anomesprimcid = ltrim( str(@mes)) + '/01/' + ltrim( str(@ano))
--Set @anomesprimcid = Convert(int,Convert(Varchar(2),@mes) +
Convert(Varchar(4),@ano))
--declare @teste int
--set @teste = Convert(int, @anomesprimcid)
-----

--select @anomesprimcid

If (@anomesprimcid >= @anomespriexame) and (@ret is not Null and @ret <>")
set @ret = 1
else
set @ret = "
return @ret;

End
```