



INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DO  
AMAZONAS  
DEPARTAMENTO DE ENSINO SUPERIOR  
CURSO DE TECNOLOGIA EM ANÁLISE E DESENVOLVIMENTO DE SISTEMAS

ÉRICLE LUNA COSTA

TB BI: SOLUÇÃO DE INTEGRAÇÃO DE DADOS DOS SISTEMAS E-SUS E SINAN PARA APOIAR O  
CONTROLE DE CASOS DE TUBERCULOSE

MANAUS, AMAZONAS  
2023

ÉRICLE LUNA COSTA

TB BI: SOLUÇÃO DE INTEGRAÇÃO DE DADOS DOS SISTEMAS E-SUS E SINAN PARA APOIAR O  
CONTROLE DE CASOS DE TUBERCULOSE

Trabalho de Conclusão de Curso apresentado ao Curso Superior de Tecnologia em Análise e Desenvolvimento de Sistemas do Instituto Federal de Educação, Ciências e Tecnologia do Amazonas (IFAM), Campus Manaus Centro, como requisito para a conclusão do curso.

Orientador: Prof. Esp. Rogério Luiz Araújo Carminé,

Coorientador: Prof. Me. Miguel Bonafé

---

**Biblioteca do IFAM – Campus Manaus Centro**

---

C837t Costa, Éricle Luna.

TB BI: solução de integração de dados dos sistemas e-SUS e SINAN para apoiar o controle de casos de tuberculose / Éricle Luna Costa. – Manaus, 2023.

46 p. : il. color.

Monografia (Tecnologia em Análise e Desenvolvimento de Sistemas). – Instituto Federal de Educação, Ciência e Tecnologia do Amazonas, *Campus Manaus Centro*, 2023.

Orientador: Prof. Esp. Rogério Luiz Araújo Carminé.

Coorientador: Prof. Me. Miguel Bonafé.

1. Desenvolvimento de sistemas. 2. Sistema SINAN. 3. Sistema e-SUS APS. 4. Integração de dados. I. Carminé, Rogério Luiz Araújo. (Orient.) II. Bonafé, Miguel. (Coorient.). III. Instituto Federal de Educação, Ciência e Tecnologia do Amazonas. IV. Título.

CDD 005.3

---

## RESUMO

A tuberculose é um dos grandes problemas de saúde mundial, uma das doenças infecciosas mais mortais do planeta. No Brasil, foi registrado um aumento no número de casos nos anos de 2020 e 2021, que passou de 66.819 para 68.271 pessoas com a doença. O tratamento da tuberculose é oferecido pela rede pública de saúde e para gerenciar essa situação o Ministério da Saúde oferece alguns sistemas de informação, no nível municipal é utilizado o Sistema de Informação de Agravos de Notificação (SINAN), que é responsável pela notificação dos casos, e o sistema e-SUS APS, pelos atendimentos clínicos nas unidades de saúde. Esses sistemas não estão integrados, o que dificulta o controle mais amplo e completo das ações de saúde. O TB-BI é a solução proposta que visa integrar esses dados, fornecendo uma base de dados integrada e dashboard para geração da informação. Neste trabalho, é considerado o desenvolvimento de uma versão inicial (Prova de Conceito) a ser desenvolvida no escopo do município de Manaus, utilizando recursos tecnológicos amplamente utilizados no mercado, como linguagem PHP e SQL. A solução pode contribuir com o fornecimento de informação diferenciada, ampla e de qualidade sobre os casos de tuberculose e os atendimentos relacionados, para que possa impactar positivamente na gestão municipal de saúde e, conseqüentemente, na qualidade de vida da população.

**Palavras-chave:** Informações de tuberculose, Sistema SINAN, Sistema e-SUS APS, Integração de Dados, Gestão Municipal de Saúde.

## **ABSTRACT**

Tuberculosis is one of the biggest global health problems, one of the deadliest infectious diseases on the planet. In Brazil, an increase in the number of cases was recorded in 2020 and 2021, from 66,819 to 68,271 people with the disease. Tuberculosis treatment is offered by the public health network and to manage this situation the Ministry of Health offers some information systems, at the municipal level Notifiable Diseases Information System (SINAN) is used, which is responsible for notifying cases, and the e-SUS APS system, for clinical care in health units. These systems are not integrated, which makes broader and more complete control of health actions difficult. TB-BI is the proposed solution that aims to integrate this data, providing an integrated database and dashboard for generating information. In this work, the development of an initial version (Proof of Concept) to be developed within the scope of the municipality of Manaus is considered, using technological resources widely used in the market, such as PHP and SQL languages. The solution can contribute to providing differentiated, broad and quality information about tuberculosis cases and related care, so that it can positively impact municipal health management and, consequently, the population's quality of life.

**Keywords:** Tuberculosis information, SINAN System, e-SUS APS System, Data Integration, Municipal Health Management.

## LISTA DE FIGURAS

Figura 1 Elementos básicos do data warehouse.....	14
Figura 2 Arquitetura de alto nível do BI.....	16
Figura 3 Dashboard da gestão de segurança do trabalho.....	16
Figura 4 Tela de atendimento de um paciente no e-SUS PEC.....	22
Figura 5 Fluxo e instrumentos de registro utilizados na vigilância epidemiológica da tuberculose no Brasil .....	24
Figura 6 Tela principal do SINAN NET versão 5.0 .....	25
Figura 7 Esquema Conceitual da Solução "TB BI" .....	27
Figura 8 Diagrama das etapas do processo de integração .....	28
Figura 9 Tabela tb_tuberculose_sinan .....	29
Figura 10 Tabela tb_tuberculose_esus .....	29
Figura 11 Diagrama UML da ferramenta TB-BI.....	33
Figura 12 Diagrama Entidade-Relacionamento.....	35
Figura 13 Tela inicial da ferramenta TB-BI.....	36
Figura 14 Tela de Configuração.....	37
Figura 15 Tela de Parâmetros de Conexão .....	37
Figura 16 Tela de Integração .....	39
Figura 17 Exemplo de tela do painel de monitoramento com dados fictícios .....	40

## **LISTA DE SIGLAS E ABREVIações**

APS - Atenção Primária em Saúde

BI - Business Intelligence

CadSUS Web - Sistema de Cadastro Nacional de Usuários do Sistema Único de Saúde

CBHPM - Classificação Brasileira Hierarquizada de Procedimentos Médicos

CDS - Coleta de Dados Simplificado

CIAP - Classificação Internacional de Atenção Primária

CID - Classificação Internacional de Doenças

CNS - Cartão Nacional de Saúde

DATASUS - Departamento de Informática do SUS

DESF - Departamento de Saúde da Família

e-SUS APS - Sistema e-SUS Atenção Primária em Saúde

ETL - Extract-Transformation-Load

KPI - Key Performance Indicators

MS - Ministério da Saúde

PEC - Prontuário Eletrônico do Cidadão

PNCT - Programa Nacional de Controle da Tuberculose

SAPS - Secretaria de Atenção Primária à Saúde

SCNES - Sistema de Cadastro Nacional de Estabelecimentos de Saúde

SIASUS - Sistema de Informação Ambulatorial do SUS

SIGTAP - Tabela de Procedimentos do Sistema Único de Saúde

SIM - Sistema de Informação Sobre Mortalidade

SINAN - Sistema de Informação de Agravos de Notificação

SINASC - Sistema de Informações sobre Nascidos Vivos

SNOMED CT - Systematized Nomenclature of Medicine Clinical Terms

TB - Tuberculose

TI – Tecnologia da Informação

TUSS - Terminologia Unificada da Saúde Suplementar

UBS - Unidade Básica de Saúde

UFSC - Universidade Federal de Santa Catarina

# Sumário

<b>RESUMO</b> .....	3
<b>ABSTRACT</b> .....	4
<b>LISTA DE FIGURAS</b> .....	5
<b>LISTA DE SIGLAS E ABREVIACÕES</b> .....	6
<b>1. INTRODUÇÃO</b> .....	9
1.1. Contextualização do Problema.....	9
1.2. Problematização.....	10
1.3. Justificativa.....	11
1.4 Objetivos.....	11
1.4.1 Objetivo Geral.....	11
1.4.2 Objetivo Específico.....	12
1.5 Metodologia.....	12
1.7 Organização do documento.....	13
<b>2. FUNDAMENTAÇÃO TEÓRICA</b> .....	13
2.1 <i>Data Warehouse</i> .....	13
2.2 <i>Business Intelligence</i> .....	15
2.3 <i>Dashboard</i> .....	16
2.4 Similaridade de Dados Textuais.....	17
2.5 Padrões de Códigos de Doenças.....	18
2.5.1 Classificação Internacional de Doenças e Problemas Relacionados à Saúde (CID) .....	18
2.5.2 Classificação Internacional de Atenção Primária (CIAP).....	19
2.6 Sistemas de Informação em Saúde.....	19
2.7 Sistema e-SUS Atenção Primária à Saúde.....	21
2.8 Sistema de Informação de Agravos de Notificação.....	23
2.9 CadSUS WEB.....	25
2.10 Indicadores de Saúde.....	26
<b>3. Desenvolvimento da Solução</b> .....	26
3.1 Descrição da Solução.....	26
3.1.1 Processo para Integração das Bases.....	28
3.3 Tecnologias Utilizadas.....	31
3.4. Modelagem da Ferramenta.....	32
3.4.1. Modelo de Caso de Uso.....	32
As descrições de alguns casos de uso definidos no diagrama (Figura 11) estão disponíveis a seguir.....	33
3.4.2. Modelo de Dados.....	35

3.5 Implementação da Solução .....	36
3.5 Demonstração de Aplicação dos Dados Integrados .....	39
<b>4. CONSIDERAÇÕES FINAIS</b> .....	<b>41</b>
<b>5. REFERÊNCIA BIBLIOGRÁFICA</b> .....	<b>42</b>
ANEXO A - FICHAS DE NOTIFICAÇÃO DE TUBERCULOSE DO SINAN .....	45

## 1. INTRODUÇÃO

### 1.1. Contextualização do Problema

No cenário mundial, cerca de 1,5 milhão de pessoas morreram devido à tuberculose em 2020, sendo observado um aumento pela primeira vez em mais de uma década, resultado atribuído à sobrecarga dos sistemas de saúde em 2020 pela pandemia de COVID-19. Existe uma previsão, segundo a Organização Mundial da Saúde (OMS), que esse número continue a aumentar nos próximos anos (OMS, 2021).

No Brasil, o Ministério da Saúde (MS) também registrou um aumento no número de casos novos da doença nos anos de 2020 e 2021, passando de 66.819 para 68.271. O crescimento também ocorreu no número de óbitos, que passou de 4.532 em 2019 para 4.543 em 2020, resultando em uma taxa de mortalidade 2,1 por 100 mil habitantes (Brasil, 2021a, 2022a).

Em 2021, o Amazonas apresentou a maior taxa de incidência da doença no país, com 3.065 casos novos registrados. Desse total, 2.223 casos foram registrados em Manaus (72,53%) e 842 (27,47%) no interior do Estado (Brasil, 2022a).

O controle da tuberculose compreende uma série de atividades relacionadas a atendimentos clínicos, planejamento dos serviços, contato com outras áreas, além do uso de sistemas de informação para o atendimento e controle de casos da doença. O Programa Nacional de Controle da Tuberculose (PNCT), estabelecido pelo MS, é responsável por definir instruções para notificação e controle de casos, além do atendimento aos pacientes, oferecido no âmbito do Sistema Único de Saúde (SUS) (Brasil, 2019).

Segundo o PNCT, a notificação deve ser realizada por meio do Sistema de Informação de Agravos de Notificação (SINAN), que abrange os procedimentos necessários para investigação e acompanhamento de casos. O registro é feito por profissionais em diversos níveis do sistema de saúde, desde as unidades básicas de saúde (UBS), secretarias municipais e estaduais, até o Ministério da Saúde (Rocha et al., 2020).

Com relação ao atendimento de saúde, as pessoas que contraíram a doença

são atendidas nas unidades básicas de saúde (UBS), abrangendo a realização de consultas com profissionais de saúde, fornecimento de medicamentos para o tratamento, além de outros cuidados necessários. Esse processo de atendimento é realizado por meio do sistema e-SUS Atenção Primária (e-SUS APS), que possui o Prontuário Eletrônico do Cidadão (PEC), onde são registradas as informações clínicas e administrativas, possibilitando o compartilhamento de conduta entre profissionais (Brasil, 2021b).

Esses dois sistemas compartilham um elemento comum: o usuário do SUS. Todo usuário para ter acesso aos serviços do SUS precisa ter o número do Cartão Nacional de Saúde (CNS), seja para tratamento dos casos de tuberculose ou para qualquer outro serviço. Esse cadastro se torna a principal referência para informações de identificação dos usuários na rede de saúde, podendo auxiliar na identificação do usuário e na validação dos dados.

## 1.2. Problematização

Os sistemas SINAN e e-SUS APS contemplam processos de trabalho diferentes. No primeiro, o foco é a vigilância e controle de casos de doenças; e no segundo, o atendimento de saúde dos pacientes. Eles não possuem uma integração nativa para possibilitar um relacionamento de dados sobre casos de tuberculose e pacientes em tratamento.

No âmbito municipal, esses sistemas são gerenciados normalmente por unidades organizacionais diferentes, dificultando a integração e o acesso à informação para tomada de decisão, conforme aponta Lima (2022). Além disso, é possível observar situações de redundância e inconsistência de dados, seguido de retrabalho para manutenção dos cadastros de usuários da saúde pública (Coelho Neto; Andreazza; Chioro, 2021).

O reflexo dessa descentralização pode ser observado durante o registro dos atendimentos de saúde nas unidades. O profissional precisa registrar dados clínicos do cidadão nos dois sistemas (SINAN e e-SUS APS). Desta forma, a gestão municipal de saúde possui dois sistemas como fonte de dados da doença, mas que não se comunicam nativamente, tornando o acesso aos dados dividido, diminuindo a disponibilidade e a qualidade das informações para a tomada de decisão.

A integração de sistemas de informação de saúde, segundo Lima (2022), facilita o acesso à informação, melhora a comunicação, cooperação e coordenação

entre as partes envolvidas. Os dados desses dois sistemas, se integrados, podem permitir a geração de informação diferenciada para o monitoramento de casos de tuberculose relacionados com os atendimentos realizados nas unidades de saúde, possibilitando um tratamento com maior efetividade, além da melhoria nos processos de trabalho e na aplicação de recursos públicos.

### 1.3. Justificativa

Com o aumento crescente do número de casos de tuberculose no Brasil em 2020 e 2021, é importante investir em soluções que apoiem os processos de monitoramento e controle de casos da doença, além do tratamento dos pacientes no âmbito da gestão municipal.

Segundo Lima (2022), os sistemas de informação em saúde devem fornecer benefícios operacionais para as instituições de saúde e gerar informações estratégicas, ampliando as capacidades e fornecendo novos tipos de ferramentas de suporte à decisão. Os sistemas SINAN e e-SUS podem gerar informações mais abrangentes para o controle da doença, se os dados necessários fossem integrados e atualizados em tempo oportuno para tomada de decisão, porém não é a realidade na maioria das secretarias municipais de saúde.

Portanto, é necessária uma solução que contemple a integração de dados de notificação de casos (SINAN) e atendimento clínico (e-SUS) referentes a tuberculose para geração de informação diferenciada, disponibilizada em tempo oportuno aos gestores de saúde municipais para apoiar no monitoramento e controle de casos relacionado com as ações de cuidado dos pacientes com esta patologia na rede de saúde.

### 1.4 Objetivos

#### 1.4.1 Objetivo Geral

Desenvolver uma solução computacional inicial para integrar dados dos sistemas SINAN e e-SUS APS para apoiar o monitoramento dos casos de tuberculose e oferecer um suporte mais eficaz à tomada de decisão dos gestores municipais de saúde.

### 1.4.2 Objetivo Específico

- Identificar funcionalidades que agreguem valor à solução deste trabalho;
- Identificar tecnologias e componentes de software necessários para o desenvolvimento da solução;
- Identificar as principais chaves de integração entre os bancos de dados relacionados.
- Desenvolver a solução considerando as funcionalidades, tecnologias e estrutura de dados definida;
- Demonstrar o uso de dados integrados em um instrumento para geração de informação.

### 1.5 Metodologia

Para alcançar os objetivos deste trabalho, adotou-se uma pesquisa de natureza aplicada, utilizando-se do conhecimento disponível para elaborar a solução proposta. Empregou-se um tipo de pesquisa exploratória quanto aos objetivos, utilizando procedimentos que incluíram pesquisas bibliográficas e uma abordagem qualitativa. Quanto ao método científico, fundamentou-se em uma investigação indutiva.

Para o desenvolvimento da ferramenta, foram consideradas as seguintes atividades:

- Modelagem Conceitual da Solução: Contempla as definições de requisitos funcionais e de qualidade do projeto para realização da modelagem conceitual da solução;
- Definição de Informação Necessária: Consiste na identificação, análise e validação de indicadores necessários para solução;
- Identificação das Fontes de Dados: Consiste na identificação e análise das fontes de dados disponíveis na rede de saúde municipal para definição da estrutura de dados que contemple a geração de informação definida;
- Pesquisa Bibliográfica: Consiste na pesquisa bibliográfica sobre os conceitos necessários para o desenvolvimento do trabalho, além da identificação de trabalhos relacionados;

- Modelagem do Software: Contempla a realização dos principais artefatos de modelagem de software na linguagem UML para contribuir com a implementação da solução;
- Implementação do Software: Consiste na codificação do software e realização de testes unitários relacionados;
- Demonstração: Consiste na construção de uma aplicação para demonstrar o uso dos dados integrados.

## 1.7 Organização do documento

Este documento seguirá a estrutura mostrada a seguir:

O Capítulo 2 apresenta uma revisão bibliográfica sobre os temas de *data warehouse*, *business intelligence*, e as tecnologias consideradas para o desenvolvimento da solução.

O Capítulo 3 apresenta a solução, contendo a análise e modelagem do software;

O Capítulo 4 apresenta as considerações finais sobre o que foi observado no desenvolvimento da solução.

## 2. FUNDAMENTAÇÃO TEÓRICA

A revisão teórica abrangeu tópicos que enriquecem o conhecimento e compreensão dos objetivos deste estudo. No domínio da ciência de dados, foram examinadas as tecnologias de *data warehouse*, *business intelligence*, *dashboards* e técnicas de junção por similaridade textual. Adicionalmente, foram explorados conceitos pertinentes à área da saúde, incluindo sistemas de informação, indicadores e códigos de padronização de doenças.

### 2.1 Data Warehouse

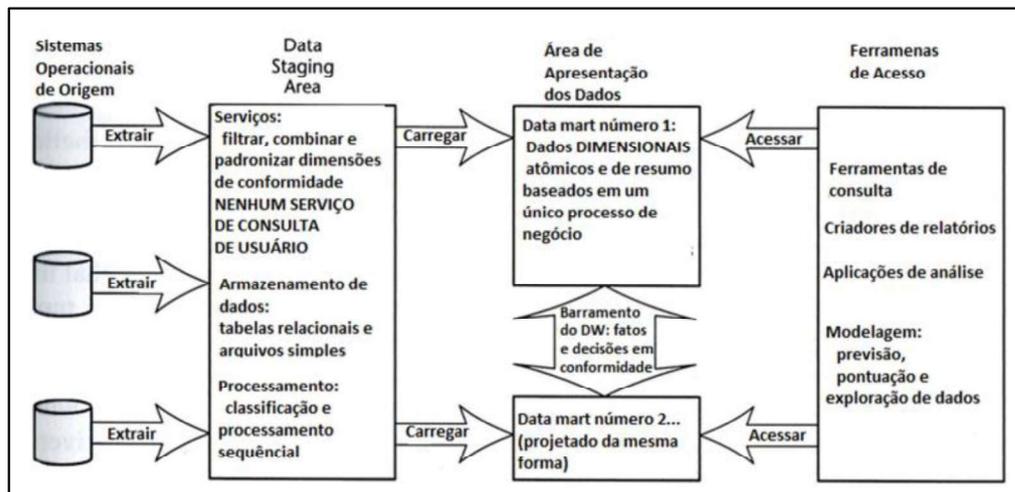
Um *data warehouse* é definido como um conjunto de dados integrado, orientado a assunto, com o principal objetivo de propiciar suporte à tomada de decisões (Inmon, 1992).

Os *data warehouses*, assim como os bancos de dados tradicionais, são uma coleção de informações, porém são otimizados para recuperação de dados que servirão principalmente para aplicações de apoio à decisão (Elmasri; Navathe, 2011).

Segundo Kimball e Ross (2002), como mostrado na Figura 1, é preciso considerar um ambiente de data warehouse formado por quatro componentes: sistemas operacionais de origem, *data staging area*, área de apresentação de dados e ferramentas de acesso a dados.

Sistemas operacionais de origem são os sistemas de registro que capturam as transações do negócio, que normalmente são sistemas independentes que não compartilham dados comuns e serão as fontes de dados que irão compor o *data warehouse* (Kimball; Ross, 2002). As fontes de dados podem ser de origens diversas como arquivos texto, planilhas, banco de dados relacionais e demais ambientes operacionais.

FIGURA 1 – ELEMENTOS BÁSICOS DO DATA WAREHOUSE



Fonte: Kimball; Ross, 2002

*Data Staging Area* é uma área de armazenamento intermediária que possui um conjunto de processos denominado *Extract-Transformation-Load* (ETL). A extração é a primeira etapa do processo que envolve a leitura e a compreensão de dados de origem e cópia dos dados necessários. Em seguida, na etapa de transformação podem ocorrer muitas modificações, como filtragem dos dados, relacionamento de dados de várias origens, cancelamento de dados duplicados, entre outras. Após as transformações, os dados serão carregados na área de apresentação do *data warehouse* (Kimball; Ross, 2002).

A área de apresentação dos dados é o local onde os dados ficam organizados e disponíveis para serem consultados por usuários, técnicos e outras aplicações de análise (Kimball; Ross, 2002).

A consulta aos dados é o ponto principal do uso do *data warehouse*. No componente ferramentas de acesso a dados é por onde todas as ferramentas poderão fazer consultas aos dados da área de apresentação do *data warehouse*, disponibilizando ao usuário de negócio recursos que podem contribuir nas ações de tomada de decisões (Kimball; Ross 2002).

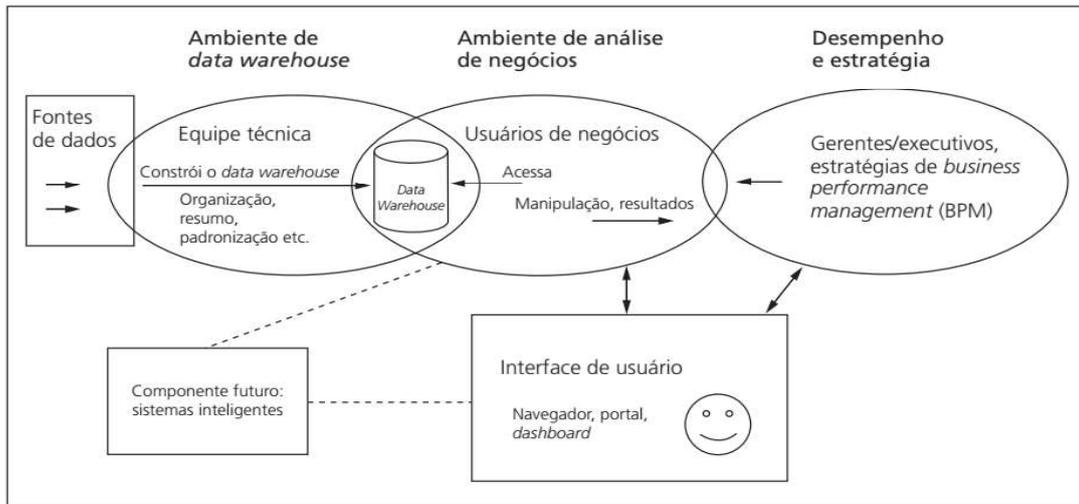
## 2.2 Business Intelligence

*Business intelligence* (BI) é um termo abrangente que pode incluir arquiteturas, bancos de dados, aplicações e metodologias. O principal objetivo do BI é permitir o acesso aos dados, proporcionando a manipulação desses dados e fornecendo aos gerentes e equipe técnica a possibilidade de realizar análises adequadas. Com a análise desses dados, os profissionais podem perceber importantes insights que servirão como base para decisões melhores e mais informadas. O processo do BI está baseado numa sequência encadeada, que é a transformação de dados em informações, incentivando decisões e finalmente causando ações (Turban, 2009).

Antes de se tornar um recurso específico oferecido por profissionais de BI com soluções dependentes do serviço de TI, o *business intelligence* surgiu na década de 1960 como um sistema para compartilhar informações nas organizações. Desenvolveu-se nos anos 80 em paralelo com as ferramentas computacionais para auxiliar no acesso às informações e tomada de decisões. As soluções de BI modernas priorizam o acesso rápido à informação, análise através de ferramentas facilitadoras e autonomia dos usuários (Tableau Software, [s.d.]).

Segundo Turban (2009), como mostrado na Figura 2, uma arquitetura de alto nível de BI tem quatro grandes componentes: um *data warehouse* com seus dados-fonte; análise de negócios com uma coleção de ferramentas para manipular e analisar os dados; *business performance management* (BPM) para monitoria e análise do desempenho; e uma interface de usuário (como o dashboard).

FIGURA 1 – ARQUITETURA DE ALTO NÍVEL DO BI

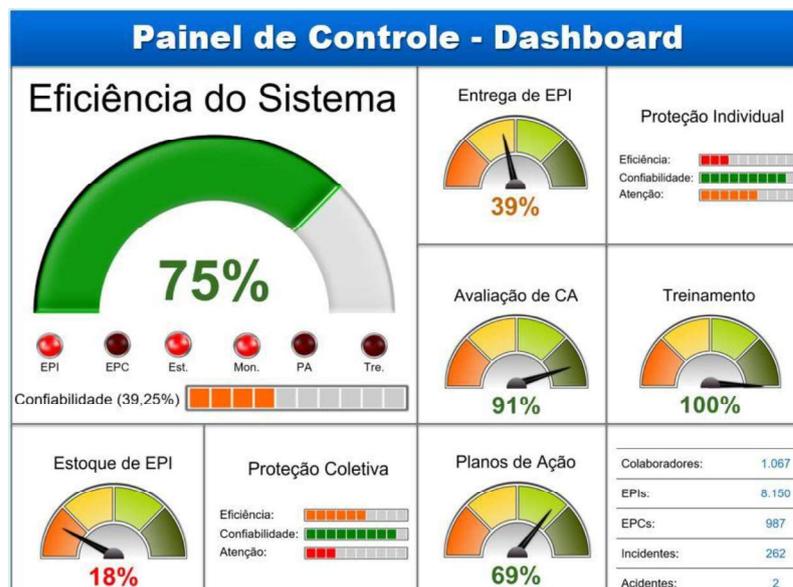


Fonte: Turban, 2009

## 2.3 Dashboard

O *dashboard* ou painel de bordo é uma ferramenta de visualização gráfica, normalmente composta por gráficos ou tabelas, sendo expostos os indicadores que permitem monitorar todo o processo organizacional. Em termos gerais, é construído para que os gestores e os responsáveis possam ter acesso de forma fácil, rápida e constante à informação mais relevante sobre o desempenho da organização (Caldeira, 2010). Na Figura 4 é apresentado um exemplo de *dashboard* de gestão de segurança do trabalho.

FIGURA 2 – DASHBOARD DA GESTÃO DE SEGURANÇA DO TRABALHO



Fonte: "CM Center", 2017

As métricas usadas em *dashboards* são conhecidas como *Key Performance Indicators* (KPIs), ou indicadores chave de performance, e medem o desempenho, objetivos e alvos pré-estabelecidos no planejamento da organização.

Normalmente, os indicadores demonstram medidas quantitativas sobre algum acontecimento. Para melhorar o entendimento dos números, estes precisam ser colocados em um contexto. Isso pode ser feito comparando os números a uma meta de referência ou alvo, indicando se os números estão bons ou ruins, e usando mecanismos especializados de visualização que conseguem estabelecer comparação e evolução (Turban, 2009).

## 2.4 Similaridade de Dados Textuais

Na área de integração de dados, principalmente quando se utilizam fontes de dados distintas, uma operação altamente indicada é a junção por similaridade. Neste trabalho, esse conceito é importante para relacionar registros que não possuem uma chave bem definida.

Junção por similaridade é uma operação que retorna pares de objetos que possuam um grau de semelhança igual ou acima de um limite estabelecido (Sidney, 2014).

Na similaridade textual, os dados são representados por cadeia de caracteres, ou seja, strings. Duas strings podem não ser iguais, mas ainda assim representar uma mesma entidade. Para compará-las, é almejado um método que determine o grau de similaridade entre elas. Vários algoritmos foram criados com este objetivo e são conhecidos como funções de similaridade (Munaretti, 2008).

Agrupar funções de similaridade textual envolve categorizar diferentes métodos ou abordagens que medem a semelhança entre textos com base em características específicas, as funções podem ser agrupadas em três eixos principais (Dezembro, 2019):

1. Funções que se baseiam em tokenização, a comparação das strings é feita a partir da comparação dos tokens das mesmas, nesse contexto, tokens podem ser entendidos como os segmentos atômicos que formam uma string, por exemplo, Jaccard e TF-IDF (Munaretti, 2008).

2. Funções que se baseiam em distância de edição, a similaridade é determinada pelo menor número de operações de edições elementares necessárias para transformar uma sequência de caracteres na outra que está sendo comparada, exemplo: Jaro Wrinkler e Levenshtein (Santos, 2012).
3. As medidas de similaridade híbrida incorporam técnicas tanto das medidas baseadas em distância de edição quanto das medidas de similaridade baseadas em token, exemplo: Monte Elkan e Soft TF-IDF (Dezembro, 2019).

## 2.5 Padrões de Códigos de Doenças

Na esfera das ciências da saúde, a atenção à padronização de processos, procedimentos e serviços relacionados ao atendimento e cuidado ao paciente é uma constante. A ausência de uma padronização adequada de normas e técnicas poderia complicar a identificação dos processos de trabalho, resultando em riscos para a execução de tarefas e aumentando significativamente as possibilidades de patologias decorrentes de equívocos médicos (Sales; Bentes Pinto, 2019).

No cenário atual da saúde, diversos padrões são empregados, abrangendo comunicação, vocabulário, representação de dados clínicos, segurança, imagens, conteúdo e estrutura. Alguns exemplos são: *Systematized Nomenclature of Medicine Clinical Terms* (SNOMED CT); Classificação Internacional de Doenças e Problemas Relacionados à Saúde (CID); Classificação Internacional de Atenção Primária (CIAP); Tabela de Procedimentos do Sistema Único de Saúde (SIGTAP); Classificação Brasileira Hierarquizada de Procedimentos Médicos (CBHPM); Terminologia Unificada da Saúde Suplementar (TUSS) (Sales; Bentes Pinto, 2019).

### 2.5.1 Classificação Internacional de Doenças e Problemas Relacionados à Saúde (CID)

A Classificação Internacional de Doenças (CID) é um documento elaborado pela Organização Mundial de Saúde (OMS). Esse documento padroniza e organiza de maneira sistemática uma ampla gama de informações médicas, incluindo doenças, lesões, sintomas, queixas, circunstâncias sociais, causas externas e até mesmo as causas de óbito que foram identificadas ao longo da história da medicina (Amplimed, 2023).

Esse recurso é essencial para detectar e reconhecer a ocorrência de doenças e epidemias. Ao empregar adequadamente a tabela CID de doenças, torna-se possível identificar aumentos ou diminuições de doenças, bem como elaborar um perfil da população. Isso facilita a determinação dos recursos necessários para abordar os problemas de saúde específicos de uma determinada região (Amplimed, 2023).

A CID-10 corresponde à décima edição desse documento, aprovada em 1994 e adotada no Brasil dois anos após sua aprovação. A CID-10 permaneceu em vigor até 1º de janeiro de 2022, atualmente está sendo sucedida pela décima-primeira versão (Amplimed, 2023).

### 2.5.2 Classificação Internacional de Atenção Primária (CIAP)

A CIAP é uma classificação que representa a distribuição e os conteúdos característicos da atenção primária, sendo empregada para categorizar três componentes essenciais de uma consulta: os motivos que levaram à marcação da consulta (MC), os diagnósticos ou problemas, e os procedimentos. Ela também abrange informações relacionadas aos avanços recentes na base conceitual da medicina de família e comunidade, os quais, em grande parte, emergiram da adoção de uma classificação específica para essa área (Gusso, 2009).

A CIAP-2, segunda edição, foi lançada principalmente por dois motivos: para estabelecer uma conexão com a 10ª edição da CID, a CID-10, publicada pela OMS em 1992, e para assegurar estabilidade e coesão, sendo introduzidas poucas modificações na classificação (Gusso, 2009).

O CIAP é comumente empregado por profissionais de saúde na atenção primária para registrar e codificar as razões que levam os pacientes a buscar cuidados, além de documentar diagnósticos e intervenções associadas. Sua utilização auxilia na organização das informações relacionadas à saúde do paciente, contribuindo para a gestão eficaz dos cuidados primários.

## 2.6 Sistemas de Informação em Saúde

Os Sistemas de informação em saúde (SIS) podem ser definidos como ferramentas especializadas que coletam, armazenam e distribuem a informação para

auxiliar a organização e apoiar tomada de decisão na área da saúde. Assim, permitem sustentar o planejamento, o aperfeiçoamento e o processo decisório dos múltiplos profissionais da área da saúde envolvidos no atendimento aos pacientes e usuários do sistema de saúde (Marin, 2010).

O Prontuário Eletrônico do Paciente (PEP) é uma solução informatizada apoiada num modelo de sistema de saúde que busca uma maior integração da informação. É um registro que reside em um sistema especificamente desenhado para apoiar os usuários, fornecendo acesso a um completo conjunto de dados. Pressupõe-se por integração, o compartilhamento da informação por profissionais da saúde, unidades de saúde, laboratórios e demais setores envolvidos, de acordo com direito de acesso de cada um (Marin, 2010).

Os sistemas de informação públicos são desenvolvidos pelo Sistema Único de Saúde (SUS) com o propósito de dar suporte às necessidades da informação para gestão, controle de produtividade e repasse de recursos financeiros, além de monitoramento de situações de risco. São regulados por políticas de saúde, estratégias de gestão e normas administrativas (Bittar et al., 2018).

O SUS é um dos grandes sistemas de saúde pública do mundo, compreendendo desde atendimentos básicos, até transplante de órgãos, garantindo acesso integral, universal para toda a população do país (“Sistema Único de Saúde (SUS)”, [s.d.]). De acordo com o catálogo de produtos de 2019 do Departamento de Informática do SUS (DATASUS) existem mais de 250 sistemas e aplicativos, dentre os mais conhecidos destacam-se: Sistema de Informação Ambulatorial do SUS (SIASUS), Sistema de Informação Sobre Mortalidade (SIM), Sistema de Informação de Agravos de Notificação (SINAN), Sistema de Informações sobre Nascidos Vivos (SINASC), Sistema e-SUS Atenção Primária à Saúde (e-SUS APS), Sistema de Cadastro Nacional de Estabelecimentos de Saúde (SCNES), etc.

Mesmo com essa ampla lista de sistemas desenvolvidos pelo DATASUS, existem municípios e estados em que seus gestores optam por utilizar sistemas terceirizados. Os motivos podem ser diversos, mas dentre eles, talvez a busca por maior integração e acesso às informações, diminuição da quantidade de sistemas e maior proximidade com serviços de estrutura e suporte sejam os principais. Porém,

esses sistemas devem ser capazes de se comunicar com as bases de dados do SUS, pois o envio da informação é muitas vezes obrigatório.

## 2.7 Sistema e-SUS Atenção Primária à Saúde

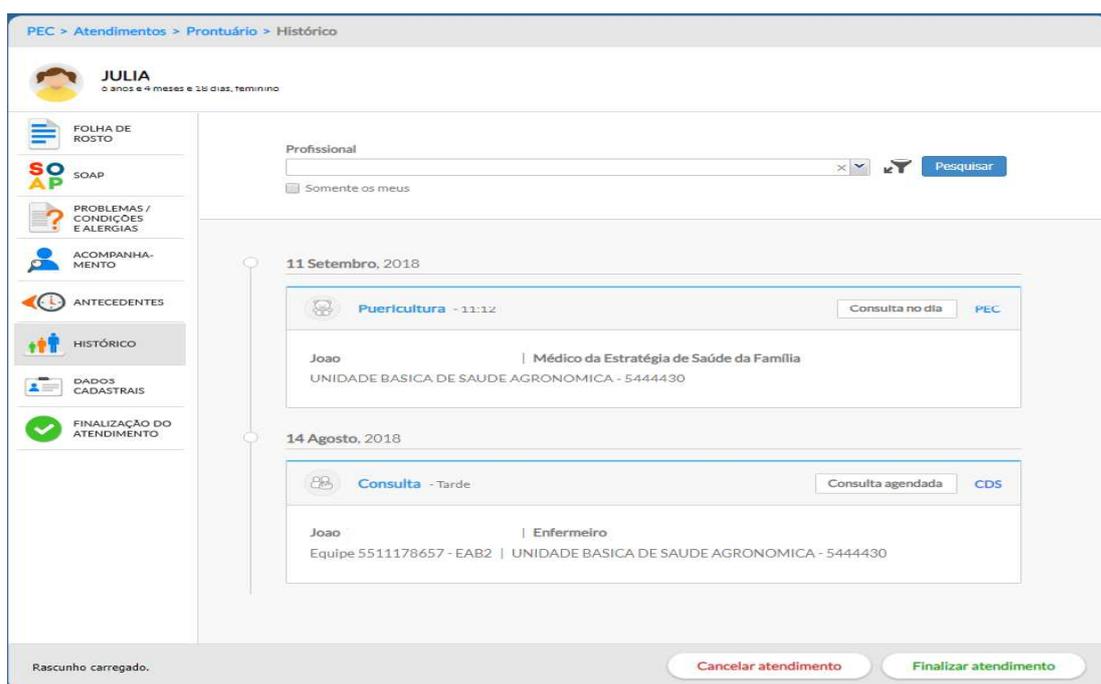
O sistema e-SUS Atenção Primária em Saúde é um sistema que compõe a estratégia e-SUS APS criada pelo SUS para reestruturar as informações da Atenção Primária à Saúde (APS) em nível nacional (Brasil, 2021).

É um software público que considera as necessidades de modernização dos atendimentos em saúde nas unidades de Atenção Primária à Saúde (APS). A responsabilidade administrativa é do Departamento de Saúde da Família (DESF) e da Secretaria de Atenção Primária à Saúde (SAPS) e o desenvolvimento é realizado em cooperação com a Universidade Federal de Santa Catarina (UFSC) (Brasil, 2021b).

O sistema e-SUS APS possui dois componentes de software: O Sistema de Coleta de Dados Simplificada (CDS), baseado no preenchimento de fichas e processo de digitação delas, sistema de caráter transitório e de contingência; e o Sistema com Prontuário Eletrônico do Cidadão (PEC), sistema com prontuário eletrônico, que apoia o processo de informatização das Unidades Básicas de Saúde (UBS) (Brasil, 2021b).

O sistema auxilia, entre outras coisas, nas ações de: gerenciamento do trabalho nas unidades de Atenção Primária à Saúde (APS); organização da agenda de pacientes; registro e organização, em formato de prontuário eletrônico, para os profissionais de saúde. Na Figura 4, é apresentada a tela de atendimento inicial de um paciente no Sistema e-SUS PEC (Brasil, 2021b).

FIGURA 3 – TELA DE ATENDIMENTO DE UM PACIENTE NO E-SUS PEC



Fonte: Brasil, 2021b

O sistema tem características particulares de uma aplicação web, utiliza um servidor web e um banco de dados PostgreSQL, que vem configurado por padrão na instalação do sistema. Pode ser utilizado em uma UBS como servidor local ou a partir de um servidor único no município, funcionando como a instância única de prontuário no ambiente municipal (Brasil, 2021b).

No e-SUS PEC para registros dos atendimentos de TB são utilizados códigos que estão relacionados à doença, sendo eles:

- Códigos de Classificação Internacional de Doenças e Problemas Relacionados à Saúde (CID-10): A15-A16 (Tuberculose respiratória); A15.0-A15.3, A16.0 - A16.3 (Tuberculose pulmonar); A15.4-A15.9, A16.4-A16.9 (Outras tuberculoses respiratórias); A17-A19 (Outras tuberculoses); A17 (Tuberculose do sistema nervoso); A18.3 (Tuberculose do intestino, do peritônio e dos gânglios mesentéricos); A18.0 (Tuberculose óssea e das articulações); A18.1 (Tuberculose do aparelho geniturinário); A19 (Tuberculose miliar); A18.2, A18.4-A18.8 (Restante de outras tuberculoses).

- Código de Classificação Internacional de Atenção Primária (CIAP2): A70 (Tuberculose);

- Código de Procedimentos do Sistema de Gerenciamento da Tabela de Procedimentos, Medicamentos e OPM do SUS (SIGTAP): 02.02.08.006-4 (Baciloscopia Direta p/ BAAR Tuberculose (CONTROLE); 02.02.08.004-8 (Baciloscopia Direta p/ BAAR Tuberculose - Diagnóstica); 02.13.01.069-0 (Teste de Sensibilidade a Drogas Contra a Tuberculose); 03.01.01.001-3 (Consulta ao Paciente Curado de Tuberculose - Tratamento Supervisionado); 03.01.01.002-1 (Consulta com Identificação de Casos Novos de Tuberculose) e 03.03.01.021-5 (Tratamento De Tuberculose - A15 A A19).

## 2.8 Sistema de Informação de Agravos de Notificação

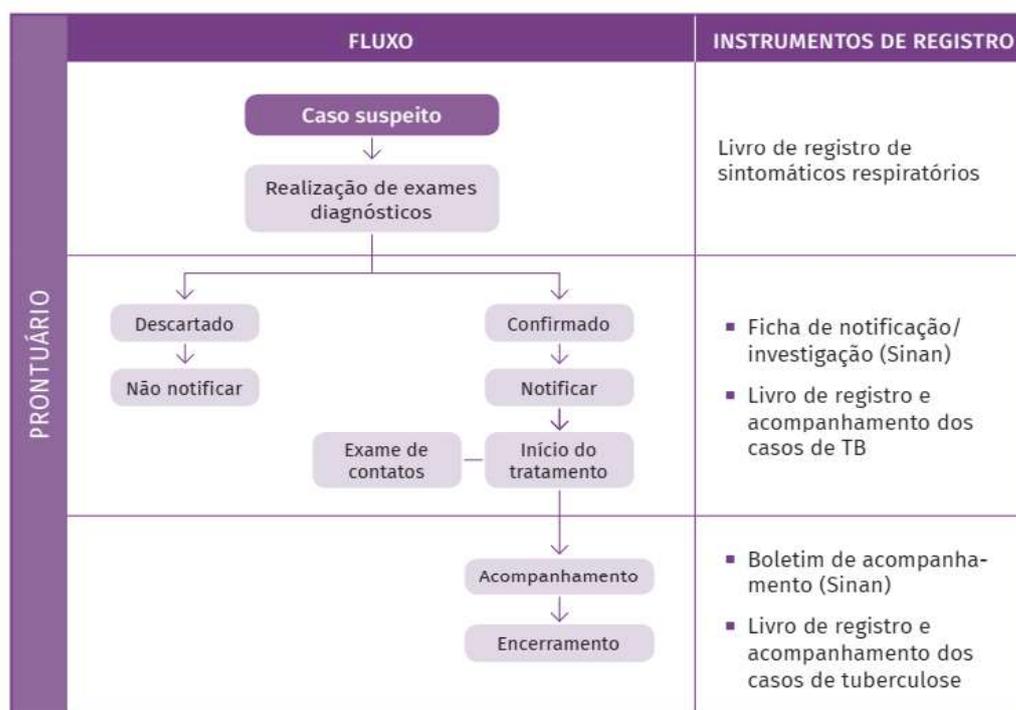
O Sistema de Informação de Agravos de Notificação (SINAN) tem como objetivo gerenciar informações do Sistema de Vigilância Epidemiológica para apoiar o processo de investigação e fornecer elementos para análise das informações de vigilância epidemiológica das doenças de notificação compulsória (Brasil, 2014).

O SINAN é abastecido, principalmente, pela notificação e investigação de casos de doenças e agravos que constam da lista nacional de doenças de notificação compulsória (Portaria GM/MS nº 420, de 2 de março de 2022), porém estados e municípios podem incluir outros problemas de saúde relevantes em sua região (Brasil, 2022b).

A tuberculose (TB) é uma doença infecciosa de notificação obrigatória, causada por bactérias que contaminam principalmente os pulmões. Sua transmissão ocorre usualmente por via aérea, iniciando com a inalação de gotículas contendo bacilos expelidos pela tosse, fala ou espirro da pessoa doente (BRASIL, 2019).

A notificação é feita por meio da confirmação de um caso de tuberculose, onde a unidade de saúde (pública ou privada) que detecta o caso é responsável pelo preenchimento da “Ficha de Notificação/Investigação de Tuberculose”, conforme está apresentado no Anexo - A, onde estão contempladas informações de identificação do paciente, do serviço, dados clínicos e epidemiológicos (BRASIL, 2019). Foram desenvolvidos instrumentos de registro para todas as etapas das ações dos profissionais de saúde na vigilância epidemiológica da TB, conforme a Figura 5.

FIGURA 4 – FLUXO E INSTRUMENTOS DE REGISTRO UTILIZADOS NA VIGILÂNCIA EPIDEMIOLÓGICA DA TUBERCULOSE NO BRASIL



Fonte: Brasil, 2019

Para que o tratamento da TB seja efetivo se faz necessário o cumprimento de algumas ações programáticas, dentre elas: consultas, teste de HIV e exame de BAAR controle (baciloscopia direta para bacilos álcool-ácido resistentes).

As consultas devem ser mensais e objetivam avaliar a evolução e/ou regressão da doença após o início do tratamento, monitorar o peso, bem como conhecer possíveis reações adversas. O teste de HIV deve ser oferecido a toda pessoa com diagnóstico de TB e tem como intuito diagnosticar precocemente infecção pelo HIV. Já o exame de BAAR controle deve ser realizado mensalmente nos casos de TB pulmonar para o monitoramento da efetividade do tratamento (Brasil, 2019).

Quando analisadas em conjunto, essas variáveis apresentam informações relevantes sobre o tratamento do paciente com tuberculose na rede de APS do município de Manaus.

A versão atual do SINAN chama-se SINAN NET versão 5.0, que trouxe melhorias em relação à versão anterior e rotinas pela internet de atualização de tabelas, transferências de dados e fluxo de retorno. É um sistema desktop que utiliza o banco de dados PostgreSQL. Na Figura 6 é mostrada a tela principal do SINAN NET.

O SINAN On Line é uma versão web, foi criado para trabalhar em paralelo com o Sinan Net até o momento em que todos os agravos estiverem disponíveis na versão online. O sistema permite a exportação de dados, em formato DBASE, para tabulação dos dados. Atualmente contempla apenas os agravos de Dengue e Febre de Chikungunya (Brasil, 2022b).

FIGURA 5 – TELA PRINCIPAL DO SINAN NET VERSÃO 5.0



Fonte: Brasil, 2014

## 2.9 CadSUS WEB

CadSUS WEB é o sistema de Cadastro Nacional de Usuários do Sistema Único de Saúde, ferramenta on-line responsável pelo cadastramento de cidadãos na base de cadastros do SUS, provendo assim o acesso do usuário ao Cartão Nacional de Saúde (CNS) e conseqüentemente a utilização de toda rede de atendimento no país. Permite que profissionais das unidades de saúde que atendem pelo SUS, tenham o controle total de todos os procedimentos realizados dentro do ambiente do SUS, como cadastros de novos cartões, pesquisa de consultas e dados do cidadão, alteração de informações e impressão de documentos (Brasil, 2022c).

Para ter acesso à base de dados do CadSUS é possível utilizar o Web Service do Cartão Nacional de Saúde que foi desenvolvido para operacionalizar a troca de informações cadastrais dos usuários do SUS, permitindo a integração entre sistemas de informação estaduais, municipais, e outros clientes que necessitem e tenham autorização para acessar (Brasil, 2018).

## 2.10 Indicadores de Saúde

Indicadores de saúde são medidas que procuram sintetizar informações sobre determinada situação, objetivo, meta ou o próprio desempenho de um sistema de saúde. Em geral, devem proporcionar a mensuração da situação sanitária de uma população e viabilizar o monitoramento das condições de saúde. A construção de um indicador pode ser um processo complexo, pois pode calcular desde a quantidade de atendimentos de uma equipe de saúde, até proporções, taxas ou índices mais sofisticados (Rede Interagencial de Informação para a Saúde, 2008).

Um dos principais motivos para o uso de indicadores é o empenho para melhorar a gestão e a qualidade do serviço oferecido. O monitoramento dos resultados desses instrumentos pode evidenciar o desenvolvimento eficiente de uma gestão. Organizações têm buscado instrumentos confiáveis que forneçam informações relevantes para um processo de tomada de decisão (De Soárez; Padovan; Ciconelli, 2005).

Na escolha dos indicadores é de grande importância o conhecimento teórico sobre o qual eles estão baseados, a definição dos critérios a serem utilizados na seleção e a indicação de um esclarecimento evidente de como os mesmos foram calculados (De Soárez; Padovan; Ciconelli, 2005). Espera-se que a análise e a interpretação dos indicadores aconteçam facilmente, e que sejam compreensíveis pelos usuários que desejam a informação, especialmente gerentes, gestores e outros profissionais do sistema de saúde (Rede Interagencial de Informação para a Saúde, 2008).

## 3. Desenvolvimento da Solução

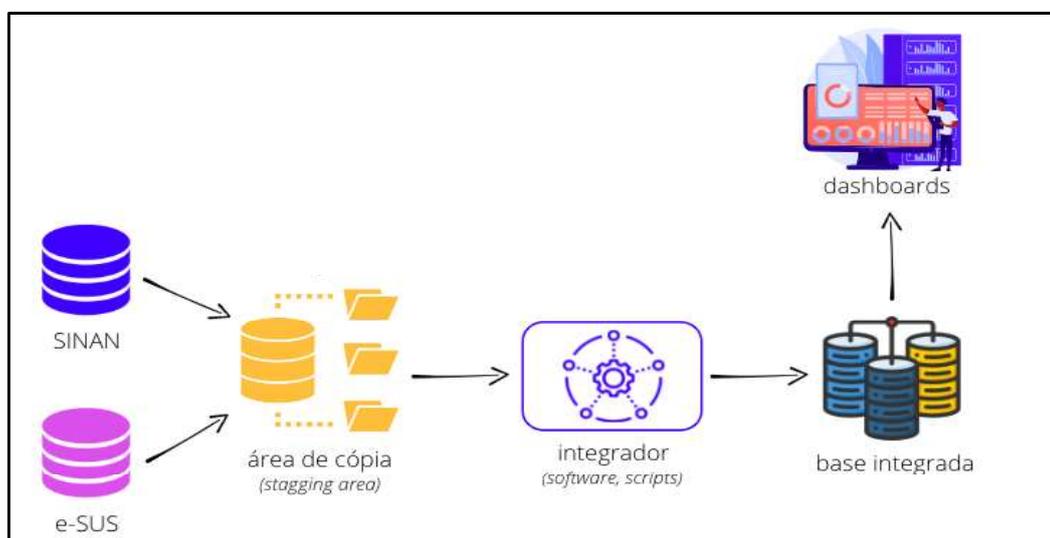
Este capítulo apresenta a solução para o problema apresentado neste trabalho, enunciando principais características, modelagem, funcionamento e funcionalidades planejadas.

### 3.1 Descrição da Solução

A solução TB BI integra dados de notificação de agravos no sistema SINAN com os dados de atendimento clínicos disponíveis no sistema e-SUS, com a finalidade

de gerar informação diferenciada para o monitoramento de casos de tuberculose na gestão municipal de saúde.

FIGURA 6 – ESQUEMA CONCEITUAL DA SOLUÇÃO "TB BI"



Fonte: Elaborado pelo autor, 2022

Conforme o esquema conceitual apresentado na Figura 7, a solução é composta dos seguintes elementos: fontes de dados, ferramenta de integração, base de dados integrada. O primeiro elemento refere-se às bases de dados dos sistemas SINAN e e-SUS; o segundo refere-se ao software que realiza a integração dos dados conforme uma estrutura definida; por fim, o último elemento é a base de dados integrados. Os painéis de monitoramento presentes na figura demonstram o uso dos dados integrados para geração de informação necessária na gestão de saúde.

Com relação às fontes de dados, é realizada a cópia dos dados originais dos sistemas SINAN e e-SUS, que ficam disponíveis em uma área de armazenamento temporária (*staging area*) para iniciar o processo de integração de dados.

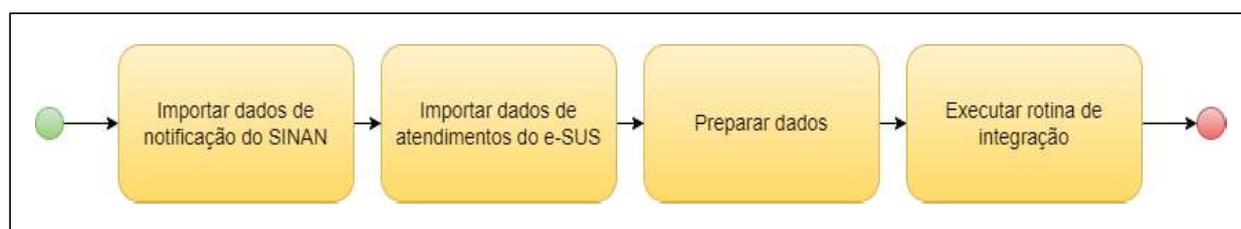
A ferramenta de integração realiza o processo de ETL (do inglês: *extraction, transform and load*), que extrai dados relevantes dos sistemas SINAN e e-SUS, depois transforma esses dados conforme uma estrutura definida, e por fim, realiza a carga desses dados transformados em uma base que pode ser conectada aos painéis de monitoramento (*dashboards*).

A base de dados integrada fornece dados preparados para geração de informação que não seria possível se os bancos de dados fossem usados isoladamente. É considerado que essa ferramenta seja capaz de prover uma grande quantidade de dados com desempenho satisfatório para o elemento de geração de informação (dashboards).

### 3.1.1 Processo para Integração das Bases

Nesta seção, serão descritas as atividades do processo de integração das bases, que é composto de 5 atividades principais, conforme mostrado na figura 8. Cada uma dessas atividades possui tarefas que são realizadas nos sistemas de origem dos dados e também no sistema integrador. Nas subseções abaixo, são descritas as atividades com mais detalhes.

FIGURA 7 – DIAGRAMA DAS ETAPAS DO PROCESSO DE INTEGRAÇÃO



Fonte: Elaborado pelo autor, 2023

#### 3.1.1.1 Importar dados de notificação do SINAN

O início do processo de integração acontece com a busca pelos dados dos casos de tuberculose nos bancos de dados dos sistemas SINAN e e-SUS APS, para em seguida serem exportados, tratados e processados no banco de integração.

No SINAN, são selecionadas as informações de notificação e investigação dos agravos de tuberculose, sendo exportadas para base de integração como uma tabela chamada `tb_tuberculose_sinan`, demonstrada na figura 9, que conterà os dados dos pacientes, de notificações e acompanhamentos da doença.

FIGURA 8 – TABELA TB\_TUBERCULOSE\_SINAN

Column Name	#	Data type	Identity	Collation	Not Null	Default
abc nu_notificacao	1	varchar(50)		default	[ ]	
abc dt_notificacao	2	varchar(50)		default	[ ]	
abc co_municipio_notifi...	3	varchar(50)		default	[ ]	
abc co_cnes	4	varchar(50)		default	[ ]	
abc cnes_atual	5	varchar(50)		default	[ ]	
abc ds_estabelecimento	6	varchar(150)		default	[ ]	
abc ds_estab_atual	7	varchar(150)		default	[ ]	
abc tp_notificacao	8	varchar(10)		default	[ ]	
abc dt_diagnostico_sint...	9	varchar(50)		default	[ ]	
abc co_cid	10	varchar(50)		default	[ ]	
abc dt_notificacao_atual	11	varchar(50)		default	[ ]	
abc nome_pac	12	varchar(150)		default	[ ]	
abc nome_mae	13	varchar(150)		default	[ ]	
abc dt_nascimento	14	varchar(50)		default	[ ]	
abc tp_sexo	15	varchar(10)		default	[ ]	

Fonte: Elaborado pelo autor, 2023

### 3.1.1.2 Importar dados de atendimento do e-SUS

No e-SUS são selecionados os dados dos pacientes que tiveram atendimentos identificados com tuberculose, através dos códigos CID-10 e CIAP2 mencionados no item 2.5 deste trabalho, que são exportados para base de integração como a tabela `tb_tuberculose_esus`, apresentada na figura 10. As atividades clínicas do e-SUS (p.ex. consultas, exames de HIV, baciloscopia e radiografia de tórax) são exportadas para a base de integração, na tabela `tb_atividades_esus`.

FIGURA 9 – TABELA TB\_TUBERCULOSE\_ESUS

Column Name	#	Data type	Identity	Collation	Not Null	Default
123 co_seq_fat_cidadao...	1	int4			[ ]	
abc no_cidadao	2	varchar(150)		default	[ ]	
abc nome_mae	3	varchar(150)		default	[ ]	
abc dt_nasc	4	varchar(150)		default	[ ]	
abc nu_cns	5	varchar(20)		default	[ ]	
abc nu_cpf_cidadao	6	varchar(20)		default	[ ]	
abc no_end	7	varchar(150)		default	[ ]	
abc nu_numero	8	varchar(150)		default	[ ]	
abc bairro_desc	9	varchar(150)		default	[ ]	
abc ds_cep	10	varchar(10)		default	[ ]	
abc nu_telefone_celular	11	varchar(150)		default	[ ]	
abc nu_telefone_residen...	12	varchar(150)		default	[ ]	
abc nu_telefone_contato	13	varchar(150)		default	[ ]	
abc nu_cnes	14	varchar(10)		default	[ ]	
abc no_unidade_saude	15	varchar(150)		default	[ ]	

Fonte: Elaborado pelo autor, 2023

### 3.1.1.3 Preparar dados

A integração dos dados acontece identificando o paciente e fazendo a relação dele entre os dois sistemas. Porém, principalmente no SINAN, os dados do paciente carecem de um atributo de chave forte para fazer este relacionamento, ou seja, não possui um campo como número do CPF ou número do Cartão SUS exigido como obrigatório, além de não ter uma validação do que é informado nesses campos.

Para fazer o relacionamento de integração das bases, foi definido como chave os atributos do paciente de cada tabela que contenham os dados do nome do paciente, data de nascimento e nome da mãe. Considerando que se trata de campos abertos do tipo textual, para os atributos nome do paciente e nome da mãe, será adotada a técnica de junção por similaridade para realização do relacionamento, enquanto a data de nascimento será uma relação de igualdade.

O Sistema Gerenciador de Banco de Dados (SGBD) PostgreSQL é empregado para armazenar e administrar a base de integração. Este SGBD apresenta funcionalidades para avaliar a similaridade de dados com base em trigramas, permitindo calcular a similaridade entre duas strings. Esse cálculo é expresso como um número entre 0 e 1, sendo que quanto mais próximo de 1, mais similares são os atributos (“(15) Trigramas, Índices e PostgreSQL | LinkedIn”, [s.d.]

As tabelas `tb_tuberculose_sinan` e `tb_tuberculose_esus` são empregadas para fazer o relacionamento de integração. No entanto, antes de relacioná-las é necessário fazer uma padronização dos atributos nome do paciente e nome da mãe para que a função de similaridade seja executada de maneira satisfatória. Serão feitas cópias destes campos na própria tabela de origem aplicando-se um padrão que consiste em retirar os espaços em branco e converter para maiúsculos todos os caracteres.

Na tabela `tb_tuberculose_esus`, devido à possibilidade de existir mais de um registro cadastral para um mesmo paciente, será necessário criar um atributo de *status* que indicará qual é o registro válido para integração, identificado através de uma regra que consiste em detectar qual o registro mais completo e atual deste paciente.

#### 3.1.1.4 Executar rotina de integração

No fluxo da relação de integração são considerados os registros da tabela `tb_tuberculose_esus` percorrendo a `tb_tuberculose_sinan` para encontrar os pacientes correspondentes, observando que no processo real o paciente é normalmente identificado primeiro no atendimento clínico no e-SUS, para em seguida ser notificado no SINAN.

Uma rotina para realização do relacionamento de integração será executada, ela pesquisará conforme os atributos-chave definidos, verificando a correspondência dos registros da tabela `tb_tuberculose_esus` com a `tb_tuberculose_sinan`. Para as situações que tiverem uma combinação positiva serão adicionados à tabela `tb_integração`, copiando os atributos chaves que servirão para relacionar as duas tabelas. Foram adotados índices para os atributos que utilizaram função de similaridade no relacionamento, para o nome do paciente foi admitido um índice acima de 60% e para o nome da mãe um índice acima de 50%.

### 3.3 Tecnologias Utilizadas

Para o desenvolvimento da solução, foram utilizadas tecnologias e componentes de software disponíveis no mercado. Devido a questões de limitação de custos, e complexidade na aquisição de equipamentos e recursos de infraestrutura de TI em órgãos governamentais (contexto de operação) a solução foi idealizada com softwares livres ou de código aberto, e também com o uso de recursos de nuvem.

Com relação ao uso de recursos de nuvem, embora eles tenham um custo mensal envolvido, foram observados os seguintes pontos positivos que favorecem sua escolha para esse trabalho que visa desenvolver uma versão inicial da solução para validar o conceito (PoC - Proof of Concept):

- Maior rapidez para implantação e configuração dos softwares de base e de apoio em comparação aos recursos físicos de TI;
- Custo menor se comparado à aquisição de materiais permanentes, links de comunicação e contratação de mão de obra especializada de TI;
- A facilidade de uso (usabilidade) das ferramentas de nuvem se comparada a ferramentas livres ou de código aberto pode ser considerada superior, com

relação, principalmente, aos softwares para gerenciamento de dashboards (p.ex. Power BI, Looker Studio e Tableau).

Para o desenvolvimento, foram utilizadas as seguintes tecnologias conforme os elementos da solução:

- Fontes de Dados:
  - Cópia dos dados dos bancos relacionais através scripts na linguagem SQL;
- Desenvolvimento da Ferramenta Integradora: Linguagem PHP e Banco de Dados PostgreSQL;
  - Tratamento dos Dados: Linguagem SQL para criação de scripts de manipulação de dados nas bases de dados;
- Banco de Dados Integrado: PostgreSQL, devido à robustez de armazenar uma grande quantidade de dados e alto desempenho na entrega de resultados, além de possui funções para tratar similaridade;
- Painéis de Monitoramento: Power BI devido à facilidade de uso e possibilidade de produção de instrumentos de informação (indicadores) com aparência atrativa para o público-alvo (gestores e analistas de informação).

### 3.4. Modelagem da Ferramenta

Esta seção apresenta os principais artefatos de modelagem da solução que foram desenvolvidos: modelo de caso de uso e o modelo de dados. Para o primeiro, é apresentado o diagrama com a descrição de caso uso associada, e para o segundo o diagrama lógico de banco de dados.

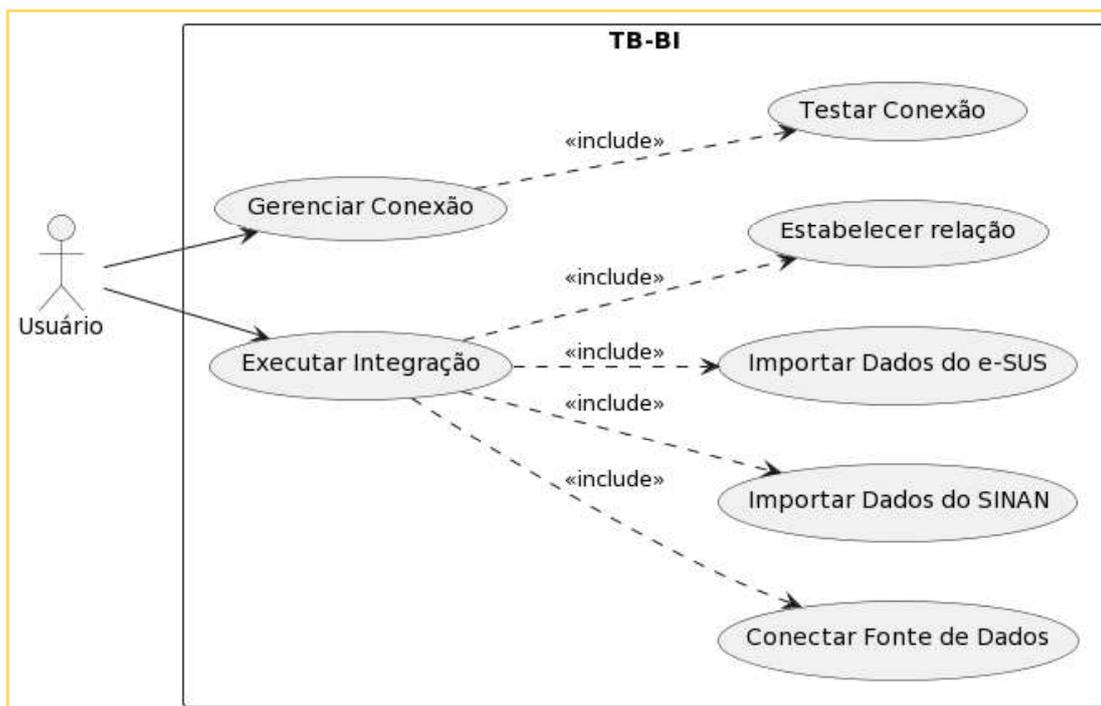
#### 3.4.1. Modelo de Caso de Uso

No diagrama de caso de uso (Figura 11). O usuário interage com os casos de uso base: **Gerenciar Conexão** e **Executar Integração**. No primeiro, o usuário pode definir as conexões, informando dados de servidores, portas, usuário e senha de

acesso) com os bancos de dados que serão usados no processo (SINAN, e-SUS, Base Integrada); o caso de uso **Testar Conexão** é incluído nele (relação *include*). Já no segundo, é onde é realizada integração dos dados, e são incluídos (relação *include*) os casos de uso específicos para realização do processo:

1. **Conectar Fontes de Dados** (realizar conexão com os bancos de dados necessários);
2. **Importar Dados do e-SUS** (importar os dados relacionados do e-SUS para uma área de trabalho - banco de dados integrado);
3. **Importar Dados do SINAN** (importar os dados relacionados do SINAN para o banco de dados integrado);
4. **Estabelecer Relação** (executar o procedimento de relacionamento entre as bases utilizando o conceito de similaridade).

FIGURA 10 – DIAGRAMA UML DA FERRAMENTA TB-BI



Fonte: Elaborado pelo autor, 2023

As descrições de alguns casos de uso definidos no diagrama (Figura 11) estão disponíveis a seguir.

### **Caso de Uso: Gerenciar Conexão**

Cenário Principal:

1. O usuário seleciona a opção Gerenciar Conexão na barra de menu;
2. O sistema lista todas as bases de dados utilizadas;
3. O usuário seleciona a base de dados que deseja configurar;
4. O sistema exibe os campos dos parâmetros de conexão;
5. O usuário preenche os parâmetros de conexão;
6. O usuário seleciona opção para salvar parâmetros de conexão;
7. O sistema salva os parâmetros na base de dados e atualiza o status da conexão para 'Configurado'.

Fluxo Alternativo:

- 5a. O usuário seleciona excluir dados de conexão.

### **Caso de Uso: Testar Conexão**

Pré-Condição: Usuário se encontra na tela de Gerenciar conexão.

Cenário Principal:

1. Usuário seleciona base de dados a qual deseja testar a conexão;
2. O sistema verifica os parâmetros da conexão e faz um teste de conexão com o banco de dados;
3. O sistema consegue validar a conexão, atualiza o status da conexão para 'Conectado' e atualiza data e hora da última verificação.

Fluxo Alternativo:

- 3a. O sistema não consegue validar a conexão, atualiza o status da conexão para 'Não Conectado' e atualiza data e hora da última verificação.

### **Caso de Uso: Executar Integração**

Cenário Principal:

1. O usuário seleciona a opção 'Integração' na barra de menu;
2. O sistema abre a tela de executar integração;
3. O usuário seleciona a opção executar integração;
4. O sistema informa a data e hora de início;
4. Caso de Uso Conectar Fontes de Dados;
5. Caso de Uso Importar Dados do SINAN;
6. Caso de Uso Importar Dados do e-SUS;

7. Caso de Uso Estabelecer Relação;

8. O sistema informa a data e hora do final da execução.

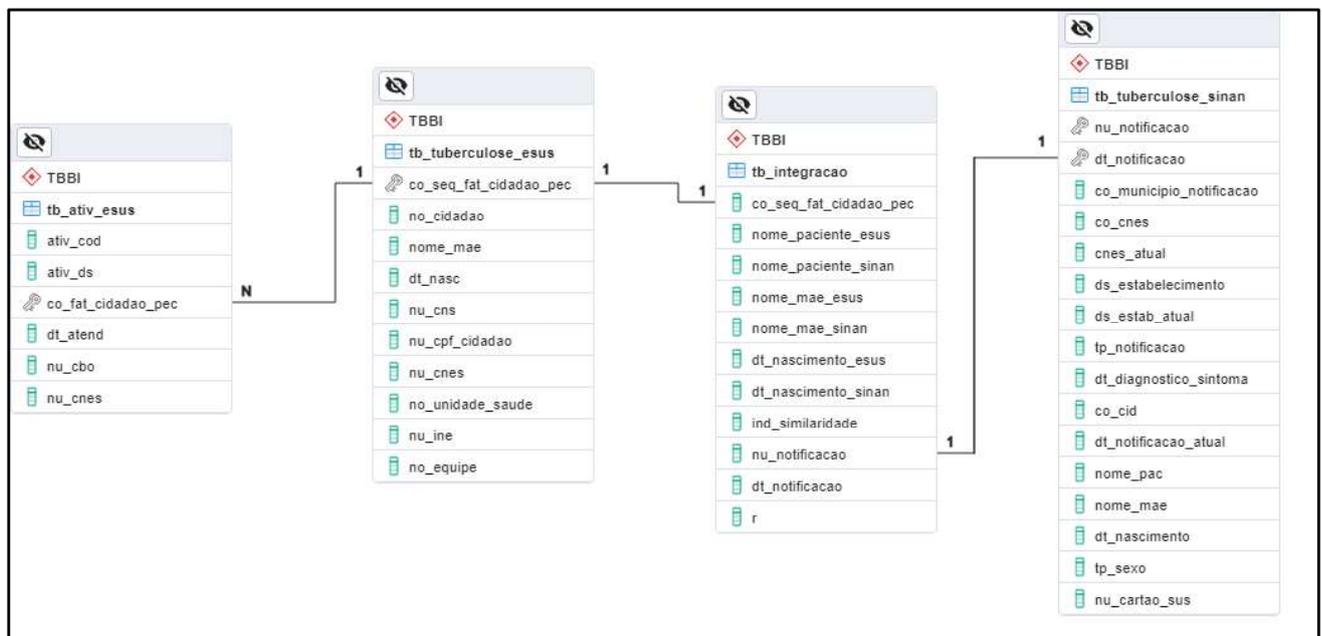
Fluxo Alternativo:

(4-7) Em caso de erro de qualquer uma das etapas o sistema interrompe a execução.

### 3.4.2. Modelo de Dados

O diagrama da Figura 12 mostra as principais tabelas de dados envolvidas no processo, que contém cópias dos dados originais: `tb_ativ_esus`, `tb_tuberculose_esus`, `tb_tuberculose_sinan`, `tb_integracao`. A primeira refere-se aos procedimentos médicos e de laboratório que estão associados ao paciente cadastrado no e-sus referente a tuberculose, a segunda contém os dados do paciente do e-sus, a terceira armazena os dados de notificação do sistema SINAN sobre casos de tuberculose, e por fim a última contém os dados do relacionamento dos pacientes nos dois sistemas (e-SUS e SINAN) após a execução da integração.

FIGURA 11 – DIAGRAMA ENTIDADE-RELACIONAMENTO

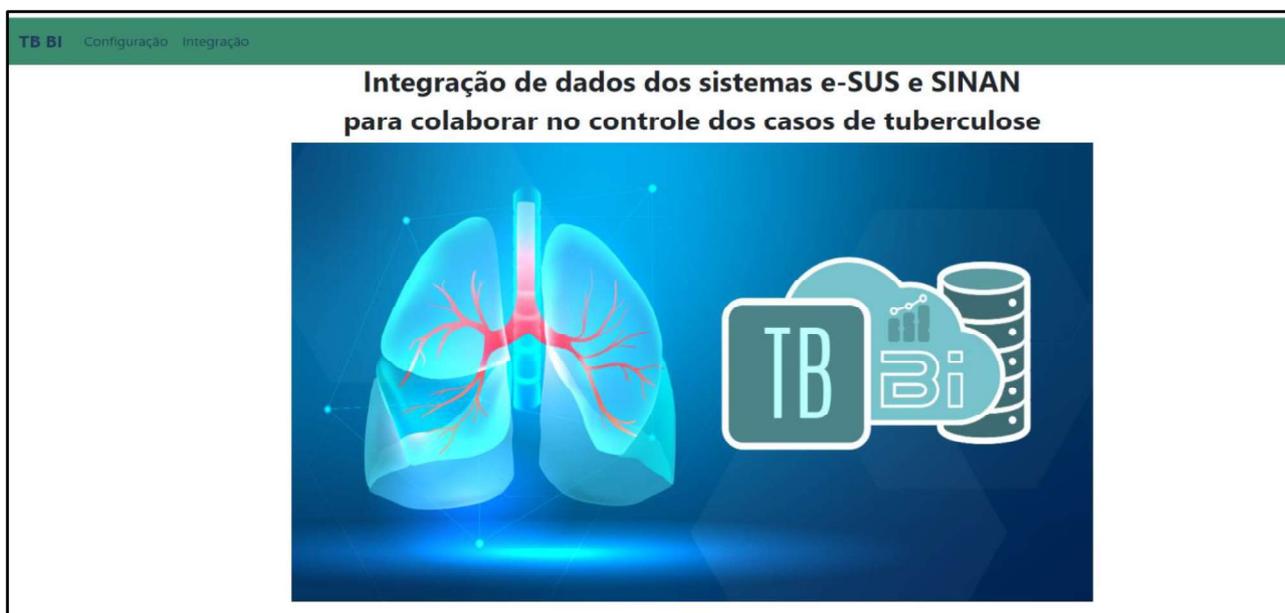


Fonte: Elaborado pelo autor, 2023

### 3.5 Implementação da Solução

A ferramenta de integração foi desenvolvida com a finalidade de reunir dados dos sistemas SINAN e e-SUS relacionados à tuberculose e consolidá-los numa base de dados integrada. Possui duas funcionalidades principais, a primeira é **Gerenciar Configurações** consiste em controlar e verificar todas as conexões com as fontes de dados, a segunda é a **Execução da Rotina de Integração** que fornece a execução das ações de extração, tratamento e carga dos dados na base integrada, na figura 13 é exibida a tela inicial.

FIGURA 12 – TELA INICIAL DA FERRAMENTA TB-BI



Fonte: Elaborado pelo autor, 2023

Na opção **Configuração** da barra de menu é feito o acesso à tela de configuração das conexões, mostrada na figura 14. São listadas as bases de dados utilizadas pela ferramenta, com os botões de ações, status da conexão e a última verificação de cada uma delas.

FIGURA 13 – TELA DE CONFIGURAÇÃO

Base de Dados	Ações	Status Conexão	Verificação
SINAN	Configurar Testar	Conectado	29-11-2023 08:31:26
eSUS APS	Configurar Testar	Conectado	29-11-2023 08:41:13
Base de Integração	Configurar Testar	Conectado	29-11-2023 08:37:06

Fonte: Elaborado pelo autor, 2023

Para gerenciar os parâmetros de conexão de uma base de dados pode-se pressionar o botão 'Configurar' que levará para tela de parâmetros de conexão, mostrada na figura 15, onde podem ser alterados os parâmetros da base selecionada. Pode-se inserir ou alterar os valores dos campos, o parâmetro **servidor** corresponde ao endereço IP da máquina onde o banco de dados está instalado, **banco de dados** refere-se ao nome do próprio banco que será utilizado, **porta** é o número da porta de acesso, **usuário** é o nome do usuário para acesso ao banco de dados e **senha** é a senha de acesso. Os botões permitem salvar as alterações, excluir todas as informações ou voltar à tela de Configuração.

FIGURA 14 – TELA DE PARÂMETROS DE CONEXÃO

Paramêtros de Conexão

SINAN

Servidor:  Banco de Dados:  Porta:  Usuário:  Senha:

Salvar Excluir Voltar

Fonte: Elaborado pelo autor, 2023

Voltando à tela de **Configuração**, o botão **Testar** possibilita checar as condições de conexão com a base de dados, atualizando o status da conexão e o último teste realizado.

A funcionalidade para execução da rotina de integração é acessada através da opção **Integração** na barra de menu, que abre a referida tela mostrada na figura 16. As etapas de integração serão executadas ao pressionar o botão **Executar Integração**, e os passos que estiverem ocorrendo serão mostrados na tela.

A execução inicia informando a data e hora do início do processo, que serve para observar quanto tempo será gasto para completar toda a execução.

Em seguida, verifica e estabelece todas as conexões com as bases de dados informadas, certificando-se que poderá utilizar os bancos neste processo, caso alguma das conexões apresentem problema, será informado um erro e o processo será interrompido.

A próxima etapa prepara os dados do banco do SINAN, executando os scripts sql que buscarão as informações de notificação de tuberculose nas tabelas de origem. Em seguida os dados selecionados serão importados para a base de integração, informando na tela a conclusão desta atividade.

Logo após, faz a preparação dos dados do banco do e-SUS, executando as consultas sql elaboradas para buscar as informações dos pacientes identificados com tuberculose e as atividades clínicas relacionadas. Depois, faz a importação dos dados selecionados para a base integrada, informando na tela a conclusão desta tarefa.

Em seguida, é feita a preparação dos dados importados para base integrada, é realizada a padronização dos campos que serão utilizados no processo de integração. O processo consiste em relacionar os pacientes notificados com tuberculose no SINAN com os pacientes identificados para tratamento de tuberculose no e-SUS utilizando um método de similaridade textual. Esta etapa busca todos os registros que conseguiu relacionar e importa para a base de integração, informando na tela a conclusão do processo.

FIGURA 15 – TELA DE INTEGRAÇÃO



Fonte: Elaborado pelo autor, 2023

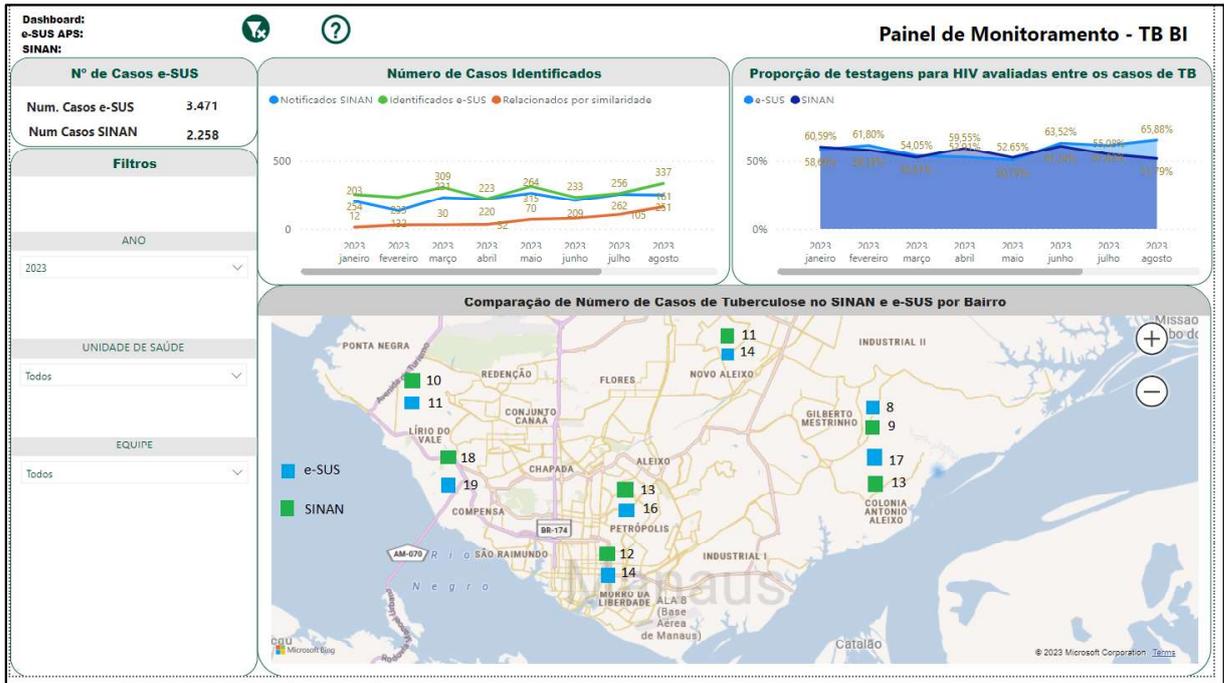
### 3.5 Demonstração de Aplicação dos Dados Integrados

Os painéis de monitoramento irão exibir medidas calculadas baseadas nos dados integrados. A Figura 17 mostra uma tela de um painel de monitoramento (*dashboard*) contendo métricas de gestão para demonstrar a potencialidade de geração de informação. Os dados apresentados são fictícios devido à necessidade de controle de exposição por questões de política institucional. Com relação às métricas, foram definidos os seguintes:

- Número de casos identificados de tuberculose, compara o número de casos notificados no SINAN, identificados no e-SUS e os casos que foram associados por similaridade, possibilitando analisar mensalmente esses valores;
- Proporção de testes de HIV avaliados entre os casos de tuberculose no e-SUS e SINAN, observando que o teste de HIV é um dos procedimentos obrigatórios no tratamento da doença;
- Comparação do número de casos notificados e atendidos por bairro, proporcionando uma análise geográfica dos casos.

Há a possibilidade da aplicação de filtros no painel, além da visualização geral do município, pode-se filtrar as informações por unidade e equipe de saúde auxiliando melhor na análise de cada situação.

FIGURA 16 – EXEMPLO DE TELA DO PAINEL DE MONITORAMENTO COM DADOS FICTÍCIOS



Fonte: Elaborado pelo autor, 2023

## 4. CONSIDERAÇÕES FINAIS

Com o crescente aumento de casos de tuberculose no Brasil nos anos de 2020 e 2021, é importante que os gestores tenham ferramentas para geração de informação que apoie no monitoramento dessa situação. A solução desenvolvida neste trabalho integra dados de casos notificados da doença com os registros de atendimentos clínicos para gerar informação diferenciada aos gestores de saúde pública nos municípios.

Para delimitar o escopo deste trabalho, foi considerado o desenvolvimento de uma versão inicial da solução, no formato de PoC (*Proof of Concept*), para validar e avaliar o investimento em versões futuras. Com relação ao escopo de dados, contribuindo para modelagem e teste da ferramenta, foram utilizados dados anônimos sobre casos da doença, seguindo os cuidados necessários definidos na LGPD.

Pode ser considerado que os objetivos deste trabalho foram alcançados, pois foi desenvolvida uma versão inicial da solução proposta para integração dos dados referente à tuberculose nos dois sistemas (SINAN e e-SUS), e também os objetivos específicos referente ao processo de construção do software, como identificação de necessidades e tecnologias, além da implementação do software, enfim chegando a uma demonstração do uso dos dados integrados em um painel de monitoramento institucional.

Por fim, considera-se que a solução proposta tem um grande potencial para apoiar nos processos de monitoramento dos casos de tuberculose e gestão clínica desses pacientes nos municípios do Brasil devido à carência desse tipo de solução no serviço público municipal, podendo impactar positivamente na saúde da população.

## 5. REFERÊNCIA BIBLIOGRÁFICA

(15) **Trigramas, Índices e PostgreSQL | LinkedIn**. Disponível em:

<<https://www.linkedin.com/pulse/trigramas-%C3%ADndices-e-postgresql-alfredo-henrique-martins-lopes/?originalSubdomain=pt>>. Acesso em: 4 dez. 2023.

AMPLIMED. **Por que a CID-10 não deve ser mais usada? | Amplimed**. , 11 out. 2023.

Disponível em: <<https://www.amplimed.com.br/cid-10/>>. Acesso em: 4 dez. 2023

BITTAR, O. J. N. et al. Sistemas de informação em saúde e sua complexidade. **Revista de Administração em Saúde**, v. 18, n. 70, 12 jan. 2018.

BRASIL. MINISTÉRIO DA SAÚDE. **Cartão Nacional de Saúde - Catálogo de APIs governamentais**. Disponível em: <<https://www.gov.br/conecta/catalogo/apis/cadsus-cadastro-de-usuarios-do-sus>>. Acesso em: 7 dez. 2022.

BRASIL. MINISTÉRIO DA SAÚDE. **SINAN**. Disponível em: <<http://portalsinan.saude.gov.br/o-sinan>>. Acesso em: 6 dez. 2022.

BRASIL. MINISTÉRIO DA SAÚDE. **Manual de apoio ao Cadastro Nacional de Usuários do Sistema Único de Saúde - CadSUS Web versão 7.0**. Brasília:Ministério da Saúde, , 2022c. Disponível em: <<https://cadastro.saude.gov.br/segcartao/manual-cadsusweb.pdf>>. Acesso em: 24 out. 2022

BRASIL. MINISTÉRIO DA SAÚDE. SECRETARIA DE ATENÇÃO PRIMÁRIA À SAÚDE. **Manual do Sistema com Prontuário Eletrônico do Cidadão PEC – Versão 5.0 [recurso eletrônico]**. Brasília:Ministério da Saúde, 2021b. Disponível em: <<https://cgiap-saps.github.io/Manual-eSUS-APS/>>. Acesso em: 25 out. 2022

BRASIL. MINISTÉRIO DA SAÚDE. SECRETARIA DE VIGILÂNCIA. **Sistema de Informação de Agravos de Notificação – Sinan Net. Manual**. Brasília:Ministério da Saúde, 2014.

BRASIL. MINISTÉRIO DA SAÚDE. SECRETARIA DE VIGILÂNCIA. **Manual de Recomendações para o controle da tuberculose no Brasil**. Brasília: Ministério da Saúde, 2019.

BRASIL. MINISTÉRIO DA SAÚDE. SECRETARIA DE VIGILÂNCIA. **Boletim Epidemiológico de Tuberculose 2021**. Brasília:Ministério da Saúde, 2021a. Disponível em: [https://www.gov.br/saude/pt-br/centrais-de-conteudo/publicacoes/boletins/epidemiologicos/especiais/2021/boletim-tuberculose-2021\\_24.03](https://www.gov.br/saude/pt-br/centrais-de-conteudo/publicacoes/boletins/epidemiologicos/especiais/2021/boletim-tuberculose-2021_24.03)

BRASIL. MINISTÉRIO DA SAÚDE. SECRETARIA DE VIGILÂNCIA. **Boletim Epidemiológico de Tuberculose 2022**. Brasília:Ministério da Saúde, 2022a. Disponível em: <<https://www.gov.br/saude/pt-br/centrais-de-conteudo/publicacoes/boletins/epidemiologicos/especiais/2022/boletim-epidemiologico-de-tuberculose-numero-especial-marco-2022.pdf>>. Acesso em: 24 out. 2022

CALDEIRA, J. **Dashboards. Comunicar Eficazmente a Informação de Gestão**. 1ª ed. [s.l.] Almedina, 2010. **CM Center**. Disponível em: <<https://cmcenter.com.br/pt-br/blog/gestao-de-seguranca-do-trabalho-painel-de-control/>>. Acesso em: 5 dez. 2022.

COELHO NETO, G. C.; ANDREAZZA, R.; CHIORO, A. Integração entre os sistemas nacionais de informação em saúde: o caso do e-SUS Atenção Básica. **Revista de Saúde Pública**, v. 55, p. 93, 1 dez. 2021.

DE SOÁREZ, P. C.; PADOVAN, J. L.; CICONELLI, R. M. Indicadores de saúde no Brasil: um processo em construção. **Revista de Administração em Saúde**, v. 7, p. 27, 2005.

DEZEMBRO, D. G. **Uma medida de similaridade híbrida para correspondência aproximada de múltiplos padrões**. Mestrado em Computação Aplicada—Ribeirão Preto: Universidade de São Paulo, 15 jul. 2019.

ELMASRI, R.; NAVATHE, S. B. **Sistemas de banco de dados**. 6. ed ed. São Paulo (SP): Pearson Addison Wesley, 2011.

GUSSO, G. D. F. **Classificação Internacional De Atenção Primária (ciap 2)**. [s.l.] Sbmfc, 2009.

INMON, W. H. **Building the data warehouse**. Boston: QED Technical Pub. Group, 1992.

KIMBALL, R.; ROSS, M. **The data warehouse toolkit: the complete guide to dimensional modeling**. 2nd ed ed. New York: Wiley, 2002.

LIMA, V. C. **Arquitetura e recursos computacionais para coleta, gestão e interoperabilidade de dados de tuberculose**. Doutorado Direto em Bioengenharia—São Carlos: Universidade de São Paulo, 23 mar. 2022.

MARIN, H. DE F. Sistemas de informação em saúde: considerações gerais. **Journal of Health Informatics**, v. 2, n. 1, 31 mar. 2010.

MUNARETTI, R. B. Qualificador Lattes : uma ferramenta para a padronização de dados utilizando junção por similaridade textual. 2008.

OMS. **Global tuberculosis report 2021**. Geneva: Organização Mundial de Saúde, 2021.

REDE INTERAGENCIAL DE INFORMAÇÃO PARA A SAÚDE. **Indicadores básicos para a saúde no Brasil: conceitos e aplicações**. Brasília: Organização Pan-Americana da Saúde, , 2008. Disponível em: <http://tabnet.datasus.gov.br/tabdata/livroidb/2ed/indicadores.pdf>

ROCHA, M. S. et al. Sistema de Informação de Agravos de Notificação (Sinan): principais características da notificação e da análise de dados relacionada à tuberculose. **Epidemiologia e Serviços de Saúde**, v. 29, n. 1, mar. 2020.

SALES, O. M. M.; BENTES PINTO, V. Tecnologias digitais de informação para a saúde: revisando os padrões de metadados com foco na interoperabilidade. **Revista Eletrônica de Comunicação, Informação e Inovação em Saúde**, v. 13, n. 1, 29 mar. 2019.

SANTOS, L. F. D. **Explorando variedade em consultas por similaridade**. Mestrado em Ciências de Computação e Matemática Computacional—São Carlos: Universidade de São Paulo, 26 out. 2012.

SIDNEY, C. F. **Predição de desempenho para junções por similaridade baseadas em conjuntos**. 2014. 91 p. Dissertação (Mestrado em Ciência da Computação) - Universidade Federal de Lavras, Lavras, 2014.

**Sistema Único de Saúde (SUS): Estrutura, princípios e como funciona (Brasil)**. Disponível em: <<https://www.gov.br/saude/pt-br/assuntos/saude-de-a-a-z/s/sus-estrutura-principios-e-como-funciona>>. Acesso em: 24 out. 2022.

TABLEAU SOFTWARE. **Business intelligence: o que é e por que é importante**. Disponível em: <<https://www.tableau.com/pt-br/learn/articles/business-intelligence>>. Acesso em: 1 dez. 2022.

TURBAN, E. **Business intelligence: um enfoque gerencial para a inteligência do negócio**. [s.l.] Grupo A - Bookman, 2009.