UNIVERSIDADE DE SÃO PAULO
ESCOLA DE ENGENHARIA DE SÃO CARLOS
FACULDADE DE MEDICINA DE RIBEIRÃO PRETO
INSTITUTO DE QUÍMICA DE SÃO CARLOS

VINÍCIUS COSTA LIMA

Arquitetura e recursos computacionais para coleta, gestão e interoperabilidade de dados de tuberculose

São Carlos
2022

VINÍCIUS COSTA LIMA

VERSÃO CORRIGIDA

Arquitetura e recursos computacionais para coleta, gestão e interoperabilidade de dados de tuberculose

Tese apresentada ao Programa de Pós-Graduação Interunidades em Bioengenharia da Escola de Engenharia de São Carlos – Faculdade de Medicina de Ribeirão Preto e Instituto de Química de São Carlos da Universidade de São Paulo, como requisito para a obtenção do Título de Doutor em Ciências.

Orientador: Prof. Dr. Domingos Alves
Coorientador: Prof. Dr. Rui Pedro Charters Lopes Rijo

São Carlos
2022

AUTORIZO A REPRODUÇÃO TOTAL OU PARCIAL DESTE TRABALHO, POR QUALQUER MEIO CONVENCIONAL OU ELETRÔNICO, PARA FINS DE ESTUDO E PESQUISA, DESDE QUE CITADA A FONTE.

Ficha catalográfica elaborada pela Biblioteca Prof. Dr. Sérgio Rodrigues Fontes da EESC/USP com os dados inseridos pelo(a) autor(a).

Costa Lima, Vinícius
C732a Arquitetura e recursos computacionais para coleta, gestão e interoperabilidade de dados de tuberculose / Vinícius Costa Lima; orientador Domingos Alves; coorientador Rui Pedro Charters Lopes Rijo. São Carlos, 2022.

Tese (Doutorado) - Programa de Pós-Graduação Interunidades em Bioengenharia e Área de Concentração em Bioengenharia -- Escola de Engenharia de São Carlos; Faculdade de Medicina de Ribeirão Preto; Instituto de Química de São Carlos, da Universidade de São Paulo, 2022.

1. Interoperabilidade. 2. Sistemas de Informação em Saúde. 3. Web Semântica. 4. Dados em Saúde. 5. Gerenciamento de Dados. I. Título.

Eduardo Graziosi Silva - CRB - 8/8907

FOLHA DE JULGAMENTO

Candidato(a): Vinícius Costa Lima

TÍTULO: “Arquitetura e recursos computacionais para coleta, gestão e interoperabilidade de dados de tuberculose”

Data da defesa: 23/03/2022

Comissão Julgadora	Resultado
Prof(a). Dr(a). Domingos Alves FMRP/USP -- Orientador	Não votante
Prof(a). Dr(a). Antônio Ruffino Neto FMRP/USP	Aprovado
Prof(a). Dr(a). Valdes Roberto Bollela FMRP/USP	Aprovado
Prof(a). Dr(a). Rafael Mello Galliez UFRJ	Aprovado
Prof(a). Dr(a). Ricardo João Cruz-Correia Universidade do Porto	Aprovado
Prof(a). Dr(a). Paulo Mazzoncini de Azevedo Marques FMRP/USP	Aprovado

Presidente da Comissão de Pós-Graduação: Prof. Dr. Adair Roberto Aguiar

Dedicado aos meus pais, Matilde e Guilherme, ao meu querido irmão Felipe, à minha esposa Raquel, à minha filha Maria Clara e à minha sogra Margarida.

AGRADECIMENTOS

À minha esposa Raquel, pelos *insights*, pelas sábias orientações sobre a prática da pesquisa científica e todo o apoio incondicional;

Aos meus pais, pela educação proporcionada, da qual me orgulho fortemente, e pelo sempre efetivo apoio à continuidade dos meus estudos.

Aos meus grandes amigos Felipe Pellison, Filipe Bernardi e Francisco Júnior, que caminharam comigo durante todo o tempo e que, sem eles, os conceitos de colaboração e trabalho em equipe teriam de ser reinventados.

Aos profissionais Rafael Galliez e Karla Gram, pelas dicas, ensinamentos e trocas de experiências que ajudaram a nortear as atividades desenvolvidas.

Ao meu caríssimo coorientador, Prof. Rui Rijo, pelo acolhimento em todos os momentos e por estar sempre de braços abertos.

Ao meu orientador, Prof. Domingos Alves, pela disponibilidade, pela supervisão recebida e por, principalmente, acreditar no potencial de seus alunos.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), pela atribuição da bolsa de Desenvolvimento Tecnológico e Industrial (DTI), Nível A, que permitiu a execução de atividades relacionadas a este projeto de pesquisa.

Eu não tenho ídolos.

Tenho admiração por trabalho, dedicação e competência.

Ayrton Senna (1960-1994)

RESUMO

LIMA, V. C. **Arquitetura e recursos computacionais para coleta, gestão e interoperabilidade de dados de tuberculose**. 2022. Tese (Doutorado) – Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, 2022.

A Tuberculose é uma doença infectocontagiosa bacteriana que afeta principalmente os pulmões e permanece como um dos maiores problemas mundiais de saúde pública. Em 2015, tornou-se a maior causa de morte por agente infeccioso no mundo. Devido à sua incidência, um grande volume de dados heterogêneos é produzido continuamente. Por isso, gestores de saúde são obrigados a interpretar dados complexos obtidos de fontes dispersas, com baixa (ou nenhuma) integração, precisão variada e frequentemente caracterizados como conjuntos de dados isolados, acessíveis apenas por sistemas de informação específicos ou em um contexto definido, resultando na redução da qualidade e integridade dos dados. Isso geralmente impede a extração de conhecimento e seu uso como referência por profissionais e gestores de saúde em processos operacionais e administrativos, afetando diretamente as ações de serviços de saúde. Para reverter esse cenário, a interoperabilidade em Sistemas de Informação em Saúde se apresenta como uma característica essencial para estabelecer a capacidade de comunicação, troca e reuso de informação. A integração de bases de dados e a transmissão de informação com valor semântico agregado contribui com a redução da complexidade dos dados. Dada a importância da enfermidade e a necessidade de dados de qualidade, este projeto tem como objetivo propor uma arquitetura de software interoperável composta por um conjunto de ferramentas para permitir a integração, a coleta e o gerenciamento de dados de saúde. A solução permite consumir, integrar e disponibilizar dados de Tuberculose obtidos a partir de bases de dados heterogêneas, sistemas de informação em saúde isolados e projetos de pesquisa, de modo a suportar o monitoramento e o tratamento da doença e facilitar a condução de pesquisas clínicas.

Palavras-chave: Interoperabilidade, Sistemas de Informação em Saúde, Web Semântica, Dados em Saúde, Gerenciamento de Dados.

ABSTRACT

LIMA, V. C. **A computational architecture and resources for collection, management and interoperability of tuberculosis data.** 2022. Tese (Doutorado) – São Carlos School of Engineering, University of São Paulo, 2022.

Tuberculosis is a bacterial infectious disease that mainly affects the lungs and remains as one of the biggest public health problems in the world. In 2015, it became the leading cause of death from an infectious agent in the world. Due to its incidence, a large volume of heterogeneous data is continuously produced. Therefore, health managers are forced to interpret complex data obtained from dispersed sources, with low (or none) integration, varied precision and often characterized as isolated data sets, accessible only by specific information systems or in a defined context, reducing the quality and integrity of data. This usually prevents the extraction of knowledge and its use as a reference by health professionals and managers in operational and administrative processes, directly affecting the actions of health services. To reverse this scenario, interoperability in Health Information Systems is presented as an essential feature to establish the ability to communicate, exchange and reuse information. The integration of databases and the transmission of information with added semantic value contributes to the reduction of data complexity. Given the importance of the disease and the need for quality data, this project aims to propose an interoperable software architecture composed of a set of tools to allow the integration, collection and management of health data. The solution allows consuming, integrating and making available Tuberculosis data obtained from heterogeneous databases, isolated health information systems and research projects, in order to support the monitoring and the treatment of the disease and facilitate the execution of clinical research.

Keywords: Interoperability, Health Information Systems, Semantic Web, Health Data, Data Management.

LISTA DE ILUSTRAÇÕES

Figura 1. Ciclo de vida de dados de pesquisas científicas 11

Figura 2 - Estrutura da Web Semântica - Semantic Layer Cake 13

Figura 3. Diagrama PRISMA 25

Figura 4. Representação do método científico 37

Figura 5. Modelo de integração de dados da TB 39

Figura 6. Arquitetura de software 46

Figura 7. Modelo Entidade-Relacionamento – Módulo de Gerenciamento de Dados, integração REDCap/KoBoToolbox e Rede Blockchain da TB 47

Figura 8. Modelo Entidade-Relacionamento – Módulo de Segurança 48

Figura 9. *Interface* de configuração dos metadados de um projeto do REDCap. 48

Figura 10. Consumo de dados do CADSUS através do REDCap 50

Figura 11. Marcação manual do SISTB 51

Figura 12. Marcação automatizada através do D2R Server 51

Figura 13. Visão macro da arquitetura de interoperabilidade do SISTB 52

Figura 14. Conversor de formulários 55

Figura 15. Processo de criptografia e descriptografia, utilizando Criptografia Híbrida 58

Figura 16 - Software de Administração da Arquitetura de Interoperabilidade 61

Figura 17 - Resposta enviada para (a) sistemas com nível amarelo e vermelho e (b) para sistemas com nível verde. 61

Figura 18 - Estrutura da camada de segurança e os mecanismos implementados 63

Figura 19. Fluxo de comunicação 66

Figura 20. Dados no formato JSON armazenados em uma transação no Tangle. 66

Figura 21. Painel de validação de campos de instrumentos de coleta do REDCap 69

Figura 22. Agenda de visitas 70

Figura 23. Gerador de Plano de Gerenciamento de Dados 71

Figura 24. Dados desidentificados no formato JSON-LD..... 77

Figura 25. Detalhes da transação Ethereum 79

Figura 26. Fluxo de informação e funcionamento da rede Blockchain..... 80

Figura 27. Integrador da TB do Projeto Avaliação do LPA..... 83

Figura 28. Interface do software para cálculo de indicadores laboratoriais da TB 84

Figura 29. Modelo conceitual de DSS para TB..... 91

Figura 30. Protótipo Funcional para Android - (a) tela principal; (b) calculadora de depuração de creatinina..... 92

LISTA DE TABELAS

Tabela 1. Sumarização dos resultados da revisão da literatura 28

Tabela 2 - Comparação entre as principais DLT 35

Tabela 3. Sistemas de informação e bases de dados de saúde governamentais. 38

Tabela 4. Exemplo de correspondência entre instrumentos de coleta e ontologias 57

LISTA DE ABREVIATURAS E SIGLAS

ABI - Application Binary Interface

AES - Advanced Encryption Standard

BaaS - Blockchain as a Service

BPC - Boas Práticas Clínicas

CID - Confidencialidade, Integridade e Disponibilidade (Tríade)

CNES - Cadastro Nacional de Estabelecimentos de Saúde

CNS - Cartão Nacional de Saúde

CPF - Cadastro de Pessoa Física

CRF - Case Report Form

CSV - Comma-separated values

DAG - Directed Acyclic Graph

DATASUS - Departamento de Informática do SUS

DDI - Data Documentation Initiative

DLT - Distributed Ledger Technol

DR-TB - Tuberculose resistente a medicamentos

DSS - Decision Support System

DSS-TB - Decision Support System for Tuberculosis

EDC - Electronic Data Capture System

FAPESP - Fundação de Amparo à Pesquisa do Estado de São Paulo

FMRP - Faculdade de Medicina de Ribeirão Preto

FMT-HVD - Fundação de Medicina Tropical Doutor Heitor Vieira Dourado

GAL - Gerenciador de Ambiente Laboratorial

HXL - Humanitarian Exchange Language

IoT - Internet of Things

JSON - JavaSCript Object Notation

JSON-LD - JavaSCript Object Notation for Linking Data

LGPD - Lei Geral de Proteção de Dados

LIS - Laboratório de Informática em Sáude

LOD - Linked Open Data

LPA - Line Probe Assay

M2M - Machine-to-Machine

MAM - Masked Authenticated Messages

MD5 - Message-Digest 5 (algoritmo)

MDR-TB - Tuberculose multiressistente

MIT - Massachusetts Institute of Technology

NIC.BR - Núcleo de Informação e Coordenação do Ponto BR

OMS - Organização Mundial de Saúde

OWL - Web Ontology Language

P2P - Ponto-a-ponto (peer-to-peer)

PGD - Plano de Gerenciamento de Dados

PNCT - Programa Nacional de Controle da Tuberculose

PRIMSA - Preferred Reporting Items for Systematic Review and Meta-Analysis

RDF - Resource Description Framework

REDCap - Research Electronic Data Capture

RLP - Recursive Length Prefix

RNDS - Rede Nacional de Dados em Saúde

RR-TB - Tuberculose resistente à Rifampicina

RSA - Rivest–Shamir–Adleman (algoritmo)

SHA - Secure Hash Algorithm

SIM - Sistema de Informações sobre Mortalidade

SINAN - Sistema de Informação de Agravos de Notificação

SISTB - Sistema da Tuberculose

SITETB - Sistema de Informação de Tratamentos Especiais de Tuberculose

SPARQL - Simple Protocol and RDF Query Language

SUS - Sistema Único de Saúde

TB - Tuberculose

TBWeb - Sistema de Controle de Pacientes com Tuberculose

TCLE - Termo de Consentimento Livre Esclarecido

TDO - Tratamento Diretamente Observado

TI - Tecnologia da Informação

TRM-TB - Teste Rápido Molecular para Tuberculose

TSA - Teste de Sensibilidade aos Antimicrobianos

USP - Universidade de São Paulo

W3C - World Wide Web Consortium

WWW - World Wide Web

XDR-TB - Tuberculose extensivamente resistente

SUMÁRIO

1. INTRODUÇÃO 1
1.1. Contextualização 1
1.1. Objetivos 3
1.2. Organização do documento 3
2. A TUBERCULOSE E O SISTEMA ÚNICO DE SAÚDE 4
3. DADOS EM SAÚDE 7
3.1. Heterogeneidade e Semântica 7
3.2. Indicadores de saúde 8
3.3. Interoperabilidade e Integração de Sistemas de Informação em Saúde 9
3.4. Dados em pesquisas científicas 10
4. WEB SEMÂNTICA 13
4.1. A Web Semântica aplicada à dados abertos e sensíveis 14
4.2. O uso da *Web* Semântica na área da saúde 15
5. SEGURANÇA DA INFORMAÇÃO 21
5.1. Métodos de criptografia 22
5.2. Abordagens de segurança para manipulação de dados eletrônicos de saúde por meio da Web Semântica: uma revisão de escopo 23
5.3. Blockchain 33
6. MATERIAIS E MÉTODOS 37
6.1. Integração de sistemas de informação em saúde de Tuberculose 37
6.1.1. Modelo de integração de dados 37
6.1.2. Justificativa da necessidade de integração 40
6.1.3. Acesso aos sistemas de informação e bases de saúde governamentais 40
6.2. *Framework* de segurança 41
6.2.1. Pilares e propriedades da segurança da informação 41
6.2.2. Aplicabilidades 42
6.3. Ferramentas para coleta, gerenciamento e compartilhamento de dados de pesquisas científicas de Tuberculose 42
6.3.1. Sistemas de captura eletrônica de dados 42
6.3.2. Plataforma Blockchain 43
6.3.3. Anotação semântica 44
7. RESULTADOS 46
7.1. Arquitetura de *software* 46
7.2. Módulo de Interoperabilidade e Integração 48
7.2.1. Integração dos sistemas da TB 49
7.2.2. Coleta e gerenciamento de dados de pesquisa de TB 54

7.3. Módulo de Segurança 57
7.3.1. Mecanismos de segurança aplicado à arquitetura de interoperabilidade semântica do SISTB 57
7.3.2. Verificação adicional de integridade dos dados do SISTB 64
7.4. Módulo de Gerenciamento de Dados de Pesquisas Científicas 67
7.5. Rede Blockchain para compartilhamento de dados de pesquisa de Tuberculose 72
7.6. Caso de Uso – Projeto Avaliação do LPA 81
7.7. Resultados complementares 83
7.7.1. Cálculo de Indicadores Laboratoriais de Tuberculose 83
7.7.2. Divulgação da relevância das estratégias digitais de saúde para apoiar os serviços de TB no contexto da pandemia de COVID-19 85
7.7.3. Sistemas de suporte à decisão para TB 86
7.7.4. Guia rápido para profissionais de saúde – aplicativo para apoio ao cuidado do paciente de TB 91
8. CONSIDERAÇÕES FINAIS 93
8.1. Limitações do trabalho realizado 94
8.2. Desenvolvimentos subsequentes e proposta de trabalhos futuros 94
REFERÊNCIAS 96
APÊNDICE A – MANUAL DE UTILIZAÇÃO DO *FRAMEWORK* SEGURANÇA 112
APÊNDICE B - MODELO DE PLANO DE GERENCIAMENTO DE DADOS 121
APÊNDICE C - PUBLICAÇÕES E APRESENTAÇÕES DE TRABALHOS COMO PRIMEIRO AUTOR 125
APÊNDICE D - PUBLICAÇÕES E APRESENTAÇÕES DE TRABALHOS COMO COAUTOR 135

1. INTRODUÇÃO

1.1. Contextualização

Com base nos princípios de universalidade, igualdade e integralidade, o Sistema Único de Saúde (SUS) se constitui como uma rede regionalizada e hierarquizada que reconhece a interdependência entre municípios e entre níveis de governo para garantir a oferta dos serviços. Ou seja, a União e os estados compartilham com os municípios as responsabilidades quanto à garantia do direito constitucional de acesso aos bens e serviços de saúde no país (JACCOUD; VIEIRA, 2018).

Com a regulamentação do SUS pela Lei nº 8.080 de 1990 (BRASIL, 1990), os Sistemas de Informação em Saúde (SIS) foram valorizados, atribuindo à União, Estados e Distrito Federal a organização e coordenação dos sistemas. Do ponto de vista tecnológico, o processo de informatização contribuiu (e continua a contribuir) com a ampliação das redes de saúde regionalizadas, o acesso aos serviços de saúde e a continuidade do cuidado em diferentes níveis de complexidade.

Entretanto, observa-se que os SIS são subutilizados para os fins a que se destinam, devido à falta de articulação entre os próprios sistemas e os processos de planejamento e gestão da saúde, resultando, em geral, em dificuldades de acesso e uso dos dados existentes por gestores e profissionais de saúde (OLIVEIRA; PASSOS; TOLFO, 2017). Ainda, nota-se uma grande quantidade de SIS de base nacional. De acordo com (CARDOSO; NETO; CHIORO, 2021), 54 sistemas de base nacional coexistiram no período de 2010 a 2018. Além disso, os autores refletem sobre a complexidade, da frágil integração e pouca transparência de como tem sido historicamente conduzida a política de tecnologia de informação (TI) do Ministério da Saúde.

A Tuberculose (TB) é uma doença infectocontagiosa bacteriana que afeta principalmente os pulmões. Apesar de ter sido declarada como uma emergência global em 1993 (NAKAJIMA, 1993), ainda permanece como um dos maiores problemas mundiais de saúde pública (SÁ et al., 2011; WORLD HEALTH ORGANIZATION, 2021). No Brasil, o diagnóstico e o tratamento da TB são realizados pelo SUS, que, em diversas ocasiões, sofre com a escassez de recursos para o combate à doença (ISABEL et al., 2008a; PEREIRA; MARINHO, 2009; ALICE; MENEZES; TRAJMAN, 2012; MINISTÉRIO DA SAÚDE, 2019b; REZENDE et al., 2020).

No âmbito da TB são utilizados diversos SIS, variando de acordo com o tipo de

informação, a fase/tipo de tratamento e a localização geográfica do paciente. Entretanto, os mesmos problemas se repetem. O suporte à tomada de decisão para gestores e profissionais de saúde é prejudicado devido às dificuldades de acesso e interpretação das informações armazenadas nos SIS. Como os SIS são subutilizados, não-interoperáveis (ou seja, não se comunicam e não trocam informações entre si) e não possuem semântica (significado) vinculada aos dados, notam-se grave problemas como a baixa qualidade do preenchimento das informações (completude) e a grande quantidade de informações repetidas entre os sistemas (retrabalho). Esse cenário pode, em última instância, contribuir com a ocorrência de erros humanos e/ou decisões equivocadas, impactando diretamente no bem-estar e estado de saúde do paciente (ANDRADE, 2008; BRASIL, 2009; RINNER et al., 2016; CARDOSO; NETO; CHIORO, 2021).

A falta de conhecimento da importância desses sistemas, além da precariedade de pessoal e de estrutura de TI dos serviços de saúde também são fatores que contribuem para o uso ineficaz dos SIS (OLIVEIRA; PASSOS; TOLFO, 2017). Além de impactar diretamente na qualidade do atendimento prestado ao paciente, em diferentes níveis de complexidade, o processo de geração de conhecimento é igualmente afetado. Como exemplo, pesquisas clínicas que requerem levantamento de dados históricos (estudos retrospectivos) ou aquelas que realizam o acompanhamento de pacientes (estudos prospectivos) dependem, geralmente, da extração de dados armazenados nos SIS. Nesses casos, portanto, pode-se notar que a análise de desfecho de uma determinada pesquisa será diretamente afetada pela qualidade dos dados disponíveis nos SIS.

Preocupações relacionadas à segurança da informação são, geralmente, desincentivos para uma efetiva troca de informações entre sistemas. O uso inadequado e a divulgação de informações pessoais de saúde de um indivíduo pode causar potenciais danos econômicos, sociais e psicológicos. Assim, é imprescindível a implementação de protocolos e mecanismos de segurança para garantir uma comunicação segura entre SIS, de modo a permitir acesso somente às entidades autorizadas.

Nesse contexto, este trabalho visa propor uma arquitetura e um conjunto de recursos computacionais para abordar questões relativas à interoperabilidade, semântica e segurança de dados de saúde, bem como aperfeiçoar a coleta e gerenciamento de dados de pesquisas científicas. As soluções são aplicáveis, especialmente, em contextos com potencial baixo nível de recursos financeiros, humanos e tecnológicos, como é o caso da TB.

As subseções a seguir apresentam os objetivos deste trabalho, a justificativa, e a forma de organização deste documento.

1.1. Objetivos

Este projeto de pesquisa tem por finalidade disponibilizar recursos computacionais para a área da TB, de forma a operacionalizar a interoperabilidade e integração de dados, aprimorar a interpretação e o compartilhamento de dados com semântica (com origem em sistemas de informação ou projetos de pesquisa), contribuir com a diminuição do estigma e receio da troca de dados de saúde por meio do uso de mecanismos de segurança, e, finalmente, oferecer ferramentas integradas para coleta, gerenciamento e compartilhamento de dados em pesquisas científicas.

Os objetivos específicos são:

- Propor o uso da Web Semântica, aliada a técnicas tradicionais de comunicação e troca de informações, para promover a interoperabilidade técnica e semântica de sistemas de informação em saúde e a integração de bases de dados para facilitar o acesso e o uso de dados da TB;
- Propor um conjunto de mecanismos de segurança abrangente;
- Fornecer ferramentas de coleta, gerenciamento e compartilhamento de dados em pesquisas científicas;
- Propor o uso de Blockchain para suportar o compartilhamento e reuso de dados coletados em pesquisas de TB;
- Integrar e disponibilizar dados de TB, provenientes de diferentes sistemas de informação, para facilitar o acesso, a visualização e o reuso de dados em pesquisas clínicas;
- Propor um modelo de sistema de suporte à decisão para TB baseado na Web Semântica.

1.2. Organização do documento

A presente tese está organizada da seguinte forma: no Capítulo 1 são apresentados o contexto e os objetivos do trabalho; do Capítulo 2 ao Capítulo 5, os fundamentos teóricos e os conceitos computacionais e tecnológicos essenciais para o desenvolvimento deste trabalho são introduzidos, a saber: a Tuberculose, os dados em saúde, a Web Semântica, e a segurança da informação; no Capítulo 6, a metodologia aplicada é detalhada; no Capítulo 7, os resultados primários e complementares obtidos são descritos; e, finalmente, no Capítulo 8, são apresentadas as considerações finais.

2. A TUBERCULOSE E O SISTEMA ÚNICO DE SAÚDE

A TB é uma doença antiga causada pelo agente patogênico Mycobacterium tuberculosis, podendo ser primária ou pós-primária (RUSSELL, 2001; FITZGERALD; STERLING; HAAS, 2005). Em 2020, ocorreram no mundo cerca de 10 milhões de novos casos e 1,3 milhão de óbitos por TB. Nesse mesmo ano, no Brasil a incidência de TB foi de 96 mil casos, além de 11 mil casos de TB associada ao HIV, 914 casos de TB resistente à rifampicina (RR-TB) ou multirresistente (MDR-TB), e 33 casos de TB extensivamente resistente (XDR-TB). O número de mortos no país é estimado em 5700 para indivíduos HIV negativos e em 2000 para HIV positivos (WORLD HEALTH ORGANIZATION, 2021).

Diferentes técnicas diagnósticas estão disponíveis, como o radiograma torácico, o teste tuberculínico, o exame e cultura de escarro e o Teste Rápido Molecular para TB (TRM-TB). O Xpert MTB RIF é atualmente o teste rápido recomendado pela Organização Mundial de Saúde (OMS), capaz de detectar o DNA do agente patogênico e verificar formas resistentes à medicamentos (WORLD HEALTH ORGANISATION, 2013).

Rifampicina, Isoniazida, Pirazinamida e Etambutol são utilizados como fármacos de primeira linha e são administrados de acordo com esquemas de tratamento recomendados pela OMS e atualmente em uso no Brasil. Entretanto, o resultado de um Teste de Sensibilidade aos Antimicrobianos (TSA) pode influenciar o esquema medicamentoso, o qual deverá ser ajustado de acordo com a resistência apresentada, incluindo-se, portanto, fármacos de segunda linha no tratamento do paciente (BRASIL, 2011a).

No Brasil, a TB é considerada um agravo de notificação compulsória (BRASIL, 2010) (BRASIL, 2016). Obrigatoriamente, todo caso diagnosticado deve ser informado aos órgãos governamentais competentes. A notificação é feita através do preenchimento de instrumentos específicos e/ou pelo uso de sistemas informatizados.

O acompanhamento do paciente infectado com TB sensitiva às drogas é, preferencialmente, realizado através da estratégia de Tratamento Diretamente Observado (TDO) – do inglês, *Directly Observed Therapy* –, preconizado pela Organização Mundial de Saúde (OMS) (WORLD HEALTH ORGANIZATION, 1999). Essa estratégia envolve o monitoramento do paciente com fornecimento regular de drogas e o uso de um sistema de informação para registro e avaliação do tratamento. O TDO permite o estabelecimento de um vínculo entre profissional de saúde e a pessoa com tuberculose, que deve ser orientada, de forma clara, quanto às características da doença e do tratamento. O TDO também busca a observação da ingestão dos medicamentos pelo profissional de saúde para mitigar os riscos de abandono de

tratamento. Preferencialmente, o acompanhamento deve ser realizado em todos os dias úteis da semana e pode ser feito presencialmente ou à distância, por meio de recursos tecnológicos (SÁ et al., 2011; ALBANO DOS SANTOS et al., 2019). O tratamento dura, no mínimo, seis meses, e deve ser completado para garantir a cura ao final do esquema terapêutico. No Brasil, o tratamento é gratuito e disponibilizado pelo SUS.

A OMS estabeleceu uma estratégia global para prevenção, tratamento e controle da tuberculose, denominada *End TB Strategy*, a qual visa deter a epidemia de TB até 2035 (WORLD HEALTH ORGANIZATION, 2014). A estratégia baseia-se em três pilares, nomeadamente: Cuidados e prevenção integrados e centrados no paciente; Políticas fortes e sistemas de apoio; e Intensificação da investigação científica e inovação tecnológica. Esses pilares referem-se às intervenções críticas para garantir que todas as pessoas com TB tenham acesso equitativo a um diagnóstico e tratamento de alta qualidade, além de mitigar os custos associados e as repercussões sociais.

O primeiro pilar concentra-se nas ações de cuidado e prevenção ao paciente, atendendo às suas necessidades de saúde específicas. O segundo pilar estimula a participação conjunta de entidades governamentais, comunidades e parceiros privados na promoção de uma cobertura de cuidados universal, levando-se em consideração as condições sociais dos grupos mais afetados pela doença (pessoas de baixa renda, indivíduos HIV positivos, imigrantes, refugiados e prisioneiros). O terceiro pilar tem como objetivo apoiar fortemente a redução da carga global de TB através do estabelecimento de novos algoritmos diagnósticos, medicamentos e vacinas, bem como formas inovadoras de distribuição e acesso aos novos recursos. Países com alta incidência de TB podem, por exemplo, conduzir investigações por meio de seus planos nacionais de pesquisa e de controle da tuberculose e colaborações com parceiros nacionais e internacionais (WORLD HEALTH ORGANIZATION, 2015a).

Seguindo as diretrizes da estratégia da OMS, o Brasil lançou, em 2017, por meio do Programa Nacional de Controle da Tuberculose (PNCT), o Plano Nacional pelo Fim da Tuberculose como Problema de Saúde Pública até o ano de 2035. O plano define objetivos e estratégias associados aos pilares da OMS que podem ser usados como suporte pelos programas de controle da tuberculose, nas três esferas de governo, na construção de seus planos de trabalho (MINISTÉRIO DA SAÚDE BRASIL, 2017).

Especificamente, os objetivos relacionados ao terceiro pilar visam:

- Estabelecer parcerias para fomentar a realização de pesquisas no País em temas de interesse para saúde pública;
- Promover a incorporação de iniciativas inovadoras para aprimorar o controle da

tuberculose.

Tais metas reconhecem a importância da intensificação das investigações científicas e da inovação tecnológica e da divulgação e disseminação dos resultados para posterior utilização pelo SUS. Entende-se, então, que pesquisas para incorporação de novos medicamentos, formas de tratamento e técnicas de diagnóstico são cruciais na redução da carga de TB no Brasil. Entretanto, a condução de estudos de qualidade requer planejamento sólido e a disponibilidade de recursos financeiros, humanos e tecnológicos para a coleta, gerenciamento e análise de dados.

Cabe destacar que o desempenho dos serviços de saúde de TB pode ser gravemente afetado se afetado pela falta de recursos financeiros, humanos e tecnológicos para a entrega de tratamento e para a condução de pesquisas científicas (ISABEL et al., 2008b). Além disso, segundo (REZENDE et al., 2020), a relevância epidemiológica da tuberculose está diretamente relacionada ao perfil socioeconômico dos países, o que agrava o cenário nacional, uma vez que a expressividade da doença tende a aumentar, enquanto pode não haver recursos para combatê-la a curto e/ou a longo prazo.

3. DADOS EM SAÚDE

Dados em saúde referem-se a qualquer informação sobre a condição física ou mental de uma pessoa, à prestação de serviços de saúde a um indivíduo, epidemiologia, ou aspectos gerenciais e operacionais relacionados à execução de cuidados de saúde e serviços correlacionados (KITSOS; PAPPA, 2015; DURMUŞ; UYDACI, 2020).

Atividades como o controle e a organização de programas de saúde pública, de instalações de saúde, de insumos e da qualidade de atendimento podem ser prejudicadas pela falta de instrumentos para apoiar a definição de parâmetros gerenciais. Entre os principais fatores que influenciam negativamente o processo de tomada de decisão, destacam-se a baixa qualidade, disponibilidade e integração de dados populacionais de saúde. Apesar da existência de tais dados, existem razões que inviabilizam o acesso por gestores e profissionais de saúde, como a não informatização de processos, a heterogeneidade e a duplicidade de dados em sistemas de informação e a existência de uma grande quantidade de dados isolados e acessíveis apenas em um determinado contexto. Além disso, grande parte dos dados estão em formato de texto não estruturado, o que dificulta a extração de informação. Tais fatores podem ocasionar problemas de qualidade de informação, tornando difícil sua coordenação e avaliação, tendo em vista que, apesar do intenso volume, as informações ainda permanecem descentralizadas sem que seja possível auxiliar o processo de tomada de decisão (BRASIL, 2009; RINNER et al., 2016).

O modelo do sistema de saúde brasileiro instaurado pela Constituição de 1988, visa garantir acesso à saúde para toda a população através dos princípios da equidade, universalidade e descentralização, além de estabelecer a participação da sociedade através do controle social para garantir a formulação, implantação controle e avaliação de políticas públicas (BRASIL, 1988). Devido à abrangência e características particulares, a gestão de dados, de conhecimentos e de serviços de saúde deve ser cuidadosamente executada, de modo que um atendimento de qualidade possa ser oferecido em todos os níveis de atenção. As práticas envolvem, além de aspectos políticos, administrativos e financeiros, a garantia da correta execução de diagnósticos, tratamentos e procedimentos médicos, os quais impactam diretamente na saúde do paciente, através do fornecimento de dados confiáveis.

3.1. Heterogeneidade e Semântica

É bastante comum encontrar ambientes computacionais heterogêneos em instituições de saúde e empresas, tanto públicas quanto privadas. Em geral, os sistemas são desenvolvidos em

diferentes linguagens, executam em diferentes plataformas e possuem bases de dados em formatos distintos (CUNHA; JÚNIOR; ALMEIDA, 2005; BRASIL. MINISTÉRIO DA SAÚDE, 2009; MADURO-ABREU et al., 2020).

A heterogeneidade dos dados está intrinsecamente conectada ao tipo de informação gerada pelos serviços de saúde, que, por si só, é considerada diversificada e complexa. A natureza altamente heterogênea, e às vezes ambígua, da linguagem médica e sua constante evolução, a enorme quantidade de dados gerados constantemente pela automação de processos e pelo surgimento de novas tecnologias e a necessidade de processar, analisar e tomar decisões com base nessas informações (KONOPKA, 2015) constituem o alicerce para a inevitável informatização da saúde.

Esse cenário complicado certamente dificulta a interpretação dos dados necessários para a produção e gestão de conhecimento na área da saúde, tanto nos processos executados por humanos quanto naqueles controlados por máquinas e sistemas computacionais. Assim, passa a ser desejável fornecer informações adicionais sobre esses dados (metadados), de forma a agregar valor semântico aos mesmos. Em outras palavras, trata-se da expressão explícita do verdadeiro significado de determinado termo, por exemplo, baseado em vocabulários formais previamente definidos, como as ontologias e terminologias de saúde.

3.2. Indicadores de saúde

Indicadores medem vários aspectos de saúde dentro de uma população em uma região demográfica e em um período de tempo definido (SHABAN-NEJAD et al., 2017). São medidas-síntese que possuem informações relevantes sobre determinados atributos e dimensões do estado de saúde, assim como o desempenho do sistema de saúde. Quando vistos em conjunto, devem mostrar a situação sanitária de uma população e ser usados para a vigilância de condições de saúde. Indicadores também são calculados para avaliar a qualidade de processos laboratoriais.

Os indicadores são de extrema importância para o estabelecimento de políticas e prioridades melhor ajustadas às necessidades da população. Para a tomada de decisões e para a programação de ações de saúde, é necessária a disponibilidade de informação apoiada em dados válidos e confiáveis. Os indicadores são instrumentos valiosos para a gestão e avaliação de saúde se gerados de forma regular e manejados em um sistema dinâmico. A disponibilidade de um conjunto de indicadores tende a facilitar o monitoramento de objetos e metas em saúde e promover o desenvolvimento de sistemas de informação intercomunicados. A seleção do conjunto básico de indicadores deve ajustar-se à disponibilidade de sistemas de informação,

fonte de dados, recursos, prioridades e necessidades específicas em cada região (REDE INTERAGENCIAL DE INFORMAÇÃO PARA A SAÚDE, 2008; PERES RONDELLI, 2017).

3.3. Interoperabilidade e Integração de Sistemas de Informação em Saúde

Os Sistemas de Informação em Saúde representam um conjunto integrado de ferramentas, processos e atores para coletar, processar, relatar e usar informações e conhecimentos de saúde para influenciar a formulação de políticas e ações e pesquisas de saúde (ABOUZAHR; BOERMA, 2005). Segundo a OMS, um SIS é um mecanismo de coleta, processamento, análise e transmissão da informação necessária para se organizar e operar os serviços de saúde e, também, para a investigação e o planejamento com vistas ao controle de doenças (SHIFERAW et al., 2017).

Os SIS também podem ser utilizados para auxiliar no processo de gestão, como uma forma de contribuir com o cuidado de alta qualidade e eficiente ao paciente (HAUX, 2006), além de suprirem vários outros objetivos dentro do sistema de saúde, como fornecer benefícios operacionais para as instituições de saúde e gerar informações estratégicas (ANDRADE, 2008), ampliando as capacidades e fornecendo novos tipos de ferramentas de suporte à decisão.

A capacidade de dois ou mais sistemas trocarem informação e a utilizarem de forma transparente é definida como interoperabilidade (GERACI et al., 1991). Para isso, existem padrões, linguagens e protocolos que devem ser adotados, dependendo do tipo de interoperabilidade que se deseja alcançar.

Em nível de sistemas, a interoperabilidade pode ser categorizada em:

- Interoperabilidade funcional/técnica: está relacionada com o conjunto de regras definidas que serão respeitadas durante o processo de comunicação e troca de informação entre sistemas (FERREIRA; FELIX, 2009). A disponibilização de *webservices* e *Application Programming Interface* (APIs) são técnicas consolidadas para se alcançar esse tipo de interoperabilidade.
- Interoperabilidade semântica: é a capacidade de sistemas compartilharem informações compreendidas através de uma definição de conceitos (FERREIRA; FELIX, 2009), utilizando-se vocabulários específicos para determinado domínio de conhecimento, garantindo-se a transmissão do correto significado de um determinado dado.

No cenário brasileiro, a Portaria nº 2073/2011 do Ministério da Saúde regulamenta o uso de padrões de interoperabilidade no âmbito do SUS e do setor de saúde suplementar

(BRASIL, 2011b), de modo a garantir a interoperabilidade técnica e semântica em SIS em uso no país. Além disso, a Portaria nº 1434/2020 instituiu a Rede Nacional de Dados em Saúde (RNDS), que apresenta-se como uma plataforma de integração de dados em saúde (saúde pública e suplementar) para ampliar o acesso às informações de saúde, oferecer um melhor acompanhamento ao paciente, melhorar a eficiência da gestão de recursos públicos e apoiar a inovação em saúde (MINISTÉRIO DA SAÚDE, 2020). A RNDS também recomenda o uso de padrões de interoperabilidade para a troca de informações em saúde.

Por integração entende-se a facilitação do acesso à informação e uma consequente melhora no processo de comunicação, cooperação e coordenação entre as partes envolvidas, de forma que elas possam ser vistas como um "todo" integrado (VERNADAT, 1996; ESCOLA NACIONAL DE ADMINISTRAÇÃO PÚBLICA, 2015).

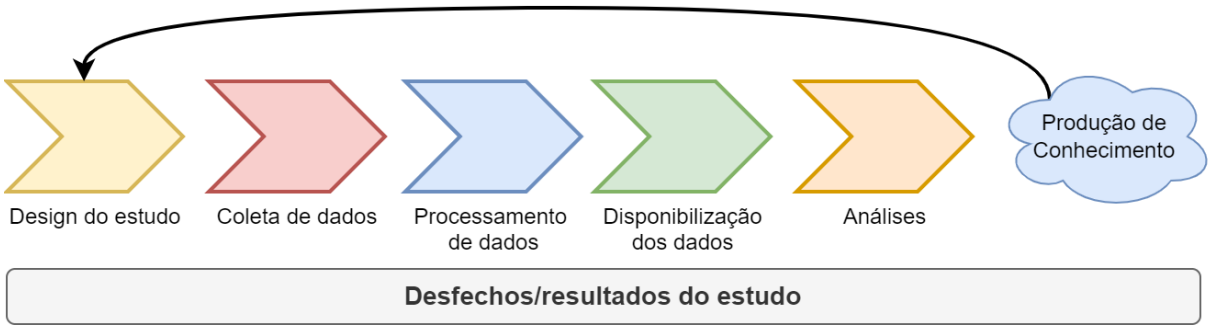
Uma vez que a interoperabilidade técnica e semântica entre SIS passa a ser considerada, viabiliza-se a comunicação integrada, a troca de informações e a conservação do valor semântico de um dado, permitindo que certa informação seja corretamente compreendida por outro sistema ou aplicação que não tenha sido inicialmente desenvolvida para o mesmo propósito (EUROPEAN COMISSION, 2008).

Observa-se que um sistema de informação pertencente a um conjunto de sistemas altamente integrados terá à sua disposição dados confiáveis que poderão ser utilizados para a construção de relacionamentos complexos entre múltiplos conceitos, obtidos com diferente acurácia, a partir de fontes de dados heterogêneas (SHABAN-NEJAD et al., 2017), que irão contribuir com a construção de conhecimento útil, permitindo que usuários como profissionais e gestores de saúde possam visualizar informações relevantes para, por exemplo, direcionar tratamentos de pacientes, investimentos em saúde e ações voltadas para a saúde pública, além de reduzir a perda de informações, aumentar a qualidade dos dados e facilitar a o cruzamento de informações de saúde entre diferentes instituições.

3.4. Dados em pesquisas científicas

O ciclo de vida de dados de pesquisas científicas representa o curso de um sistema maior, como o processo de pesquisa em si, por meio de uma série de etapas ou fases relacionadas sequencialmente nas quais informação é produzida ou manipulada, fornecendo uma visão sobre as relações entre as etapas e atividades da pesquisa (KOWALCZYK, ; HUMPHREY, 2005; ANDREAS, 2021). A Figura 1 apresenta o modelo de ciclo de vida.

Figura 1. Ciclo de vida de dados de pesquisas científicas



Fonte: Adaptado de (DDI WORKING GROUP, 2009)

Cada estágio consiste em processos, muitos dos quais geram informações. Os pontos de transição de fase são os locais mais vulneráveis à perda de informações, além de representarem conexões cruciais para outras atividades importantes de pesquisa, pois muitos fluxos de atividades fluem ao longo do ciclo de vida da pesquisa, incluindo administração de projetos, licitações, gerenciamento de dados, criação de conhecimento, julgamentos éticos, supervisão de propriedade intelectual e gerenciamento de tecnologia (HUMPHREY, 2005).

A adoção de novos métodos, ferramentas e fontes de dados mudou a forma como as pesquisas são conduzidas. No entanto, novos desafios surgiram, exigindo abordagens inovadoras para coletar, gerenciar e publicar dados. Dados bem gerenciados são mais fáceis de usar e analisar para a confirmação de uma hipótese de pesquisa. Além disso, a reutilização de dados em estudos posteriores é aprimorada. Em outras palavras, estimula uma maior colaboração entre pesquisadores e maximiza o investimento dos financiadores (STRASSER, 2015). O sucesso de uma pesquisa clínica depende diretamente da definição correta do protocolo de pesquisa, da estratégia de coleta de dados e do Plano de Gerenciamento de Dados (PGD) (LYNCH, 2008). Esses elementos impulsionam a qualidade e a confiabilidade dos dados.

A coleta de dados é um dos momentos mais cruciais em todos os tipos de projetos de pesquisa e pode levar um projeto ao sucesso ou ao fracasso. A falta de qualidade dos dados às vezes é notada quando a fase de coleta termina ou está quase no fim. Para evitar isso, além de um coletor de dados treinado (LAW et al., 2009), é imprescindível a utilização de um sistema de captura de dados confiável.

O uso de um Sistema de Captura Eletrônica de Dados (do inglês, *Electronic Data Capture System* - EDC) reduz o risco de armazenamento de dados potencialmente sensíveis em papel e ajuda a garantir a conformidade com as diretrizes de privacidade, segurança e regulamentações de dados médicos, melhorando a qualidade dos dados, capacidade de

gerenciamento e reduzindo tempo e custos (BART, 2003). Um sistema EDC deve ser capaz de funcionar independentemente de um sistema operativo ou protocolos proprietários e ser interoperável, ou seja, capaz de se comunicar com outros sistemas de forma transparente e consistente (IROJU et al., 2013).

A elaboração de um PGD é essencial para controlar as fases do ciclo de vida dos dados, de modo a evitar a perda de informações e o monitoramento da qualidade dos dados durante a execução da pesquisa. Um PGD é um documento que descreve como os dados serão tratados durante e após o término do projeto. Esses planos normalmente cobrem todo ou partes do ciclo de vida dos dados, incluindo a descoberta, coleta, organização, documentação, acesso, preservação, compartilhamento, segurança e privacidade de dados (MICHENER, 2015).

O PGD impacta diretamente os resultados da pesquisa e, portanto, deve ser o mais completo e robusto possível. As medidas e estratégias de controle de qualidade e monitoria dos dados devem estar previstas no plano. A monitoria dos estudos clínicos pretende verificar se a condução do protocolo de pesquisa está de acordo com as Boas Práticas Clínicas (BPC) para garantir a segurança dos sujeitos de pesquisa e a qualidade dos dados (MICHENER, 2015).

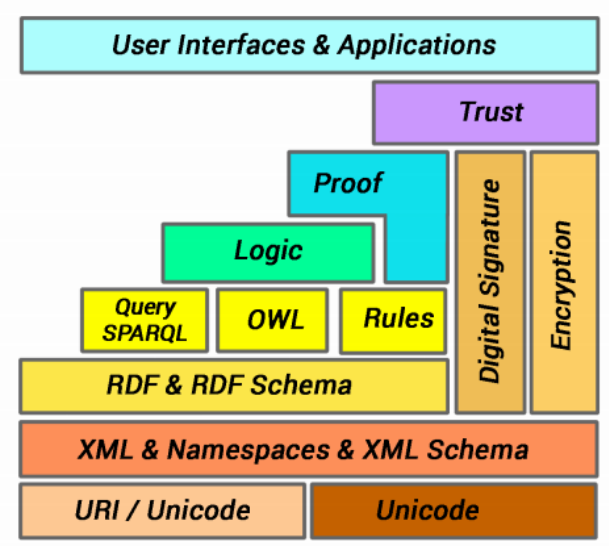
Além disso, na pesquisa em saúde, os pesquisadores precisam trabalhar em diferentes contextos. De instalações com dispositivos de última geração disponíveis até aquelas com baixa disponibilidade de recursos, como conexão de internet instável ou mesmo sem energia elétrica confiável. Na TB, os recursos para pesquisa podem ser raros e os custos para usar um EDC podem ser uma limitação. Esses aspectos se destacam como barreiras para a coleta de dados na pesquisa em TB e, portanto, disponibilizar dados para futuros estudos baseados em dados é fundamental para apoiar o desenvolvimento de novas ferramentas de tomada de decisão baseadas em evidências.

4. WEB SEMÂNTICA

Na *web* atual, a maioria do conteúdo disponível é adequado para interpretação humana, não sendo, portanto, facilmente acessível por outras máquinas e sistemas. A *Web* Semântica, definida por Tim Berners-Lee em (BERNERS-LEE; HENDLER; LASSILA, 2001), pode ser interpretada como uma extensão da *web* atual, com o objetivo de adicionar lógica ao conteúdo para expressar o significado de uma informação, suas propriedades e os relacionamentos complexos existentes entre diferentes tipos de dados, de modo que seja possível interpretar o significado de um dado sem se preocupar com sua forma de representação (ROBU; ROBU; THIRION, 2006). A ideia geral é criar uma maneira eficiente para representar dados na *World Wide Web* (WWW) para construir um banco global de dados conectados (BRASIL, 2015a), através da marcação semântica de páginas *web* e da adaptação de bases de dados relacionais já existentes, por exemplo.

Os blocos básicos que definem a *Web* Semântica são: um modelo de dados padrão, um protocolo padrão de consulta e um conjunto de vocabulários de referência. Padrões e definições recomendadas pela *World Wide Web Consortium* (W3C), como o (i) *Resource Description Framework* (RDF), (ii) *Simple Protocol and RDF Query Language* (SPARQL), (iii) ontologias e *Web Ontology Language* (OWL), (iv) *Microdata*/RDFa e (v) *Uniform Resource Identifier* (URI) são geralmente considerados para adicionar e recuperar valor semântico de um conjunto de dados. A Figura 2 apresenta a estrutura da *web* semântica.

Figura 2 - Estrutura da Web Semântica - Semantic Layer Cake



Fonte: (GEZER; BERGWEILER, 2016)

O *Resource Description Framework* (RDF) é uma linguagem de descrição e um modelo

para troca de dados na *web*, permitindo a combinação, exposição e compartilhamento de dados estruturados ou semiestruturados (W3C, 2014).

O SPARQL *Protocol and RDF Query Language* é semelhante à linguagem de consulta estruturada (SQL), é o protocolo usado para a obtenção de dados nativamente armazenados ou visualizados em RDF (W3C, 2014).

As Ontologias e *Web Ontology Language* (OWL) são outros dois pilares relevantes. As ontologias podem ser definidas como uma representação formal de um conhecimento em um domínio específico (LIYANAGE; KRAUSE; LUSIGNAN, 2015). Por sua vez, OWL é uma linguagem para publicação e compartilhamento de ontologias, projetada para descrever classes e relacionamentos entre elas (LIMA; CARVALHO, 2005).

A Microdata/RDFa representa a extensão de atributos HTML para realizar marcações semânticas em documentos HTML, com a finalidade de agregar valor semântico a determinado dado com base em uma ontologia específica (W3C, 2017). Essas informações podem ser posteriormente extraídas por outros sistemas e armazenadas em RDF.

O URI (*Uniform Resource Identifiers*) é um identificador simples e genérico que permite a distinção única e uniforme de qualquer recurso.

Um exemplo prático do uso dessas ferramentas é o DBPedia (http://dbpedia.org/), um portal que disponibiliza dados da Wikipedia (https://www.wikipedia.org/) usando o paradigma de *Web* Semântica, permitindo que sistemas externos obtenham informações com valor semântico.

4.1. A Web Semântica aplicada à dados abertos e sensíveis

Para fomentar a disponibilidade dos dados relevantes que são coletados e armazenados em sistemas de informação, surge o conceito de Dados Conectados (do inglês, *Linked Data*). Trata-se de uma estratégia que visa conectar e relacionar dados heterogêneos, resultando numa Web de Dados (do inglês, *Web of Data*). A partir disso, com a utilização de ferramentas da Web Semântica, é possível fornecer acesso a determinados dados, vinculando-se diversas bases através da web.

Com o objetivo de promover a transparência, iniciativas como a *Open Government Partnership* (https://www.opengovpartnership.org) estimulam governos a disponibilizar dados relevantes relacionados às atividades de gestão pública e que sejam de interesse da população. Para tornar essas informações públicas, tecnologias como a web semântica são utilizadas, estendendo-se o conceito de *Linked Data* para *Linked Open Data* (LOD). No Brasil, o projeto SPUK (http://igovsp.net/spuk/), desenvolvido em parceria com o Governo do Reino Unido e o

Núcleo de Informação e Coordenação do Ponto BR (NIC.BR), foi criado para proporcionar melhorias no ambiente de negócios por meio da transparência no Estado de São Paulo. A partir desse projeto e do Portal do Governo Aberto de São Paulo (http://www.governoaberto.sp.gov.br/), documentos, informações e dados governamentais de domínio público estão disponíveis para a livre utilização pela sociedade, garantindo que esses dados possam ser reutilizados para produzir novas informações e aplicações digitais para a sociedade.

Porém, nem todos os dados podem ser públicos. A legislação brasileira trata como exceção a abertura de dados particulares, capazes de identificar indivíduos, ferir seu direito à privacidade ou sua honra, dados considerados sigilosos ou aqueles que possam comprometer a segurança nacional (BRASIL, 2015b). Ainda, a Lei Geral de Proteção de Dados (LGPD) determina as principais condições sobre a maneira como os dados pertencentes a cidadãos devem ser recolhidos e tratados e estabelece penalidades para aqueles que descumprirem as determinações contidas na lei (BRASIL, 2018).

Apesar de a regulamentação da LGPD para a área da saúde não ser clara por não haver uma definição explícita de como e quais dados podem ser trocados, é sabido que dados pessoais e sensíveis compõem boa parte dos conjuntos de dados que são coletados em práticas e serviços de saúde e que, portanto, devem ser protegidos. Os dados não devem ser coletados irrestritamente e sim com uma finalidade, a qual deve ser repassada ao titular, bem como os procedimentos e as práticas para executar o tratamento dos dados (FANTONELLI et al., 2020). Além disso, os projetos que envolvem dados em saúde também estão sujeitos a comitês de ética que avaliam questões relacionadas com a partilha de dados. Assim, para aplicar as ferramentas da Web Semântica aos dados de saúde e permitir a troca de informações entre SIS (interoperabilidade), aspectos de segurança devem ser considerados.

4.2. O uso da *Web* Semântica na área da saúde

Através da combinação das ferramentas da Web Semântica com técnicas tradicionais de interoperabilidade técnica, como o uso de *webservices*, é possível garantir a comunicação e troca de informações entre sistemas. No âmbito da saúde, ao passo em que mais bases de dados são integradas e os seus dados combinados, a capacidade de prover suporte ao processo de tomada de decisão, tanto em intervenções médicas quanto em questões administrativas, pode ser aprimorada.

A utilização da Web Semântica para assegurar a interoperabilidade semântica em SIS, em comparação a outras tecnologias existentes, justifica-se pela possibilidade de

implementação e configuração de ferramentas em sistemas novos e legados, o baixo custo, a fomentação do uso por diversas entidades internacionais para disseminar a importância da abertura de dados como incentivo a uma maior transparência, pelo crescimento da quantidade de SIS que oferecem suporte a essa tecnologia e pela importância da semântica na troca de informações em saúde. Além disso, a possibilidade de realizar consultas (*queries*) federadas, ou seja, pesquisas em bases de dados hospedadas em diferentes organizações, representa uma característica desejável para potencializar a integração de dados de saúde.

Yamada et. al. especificaram e desenvolveram uma ontologia para consumir, integrar, analisar e disponibilizar dados de uma rede de saúde mental regional no Brasil (YAMADA et al., 2018). A proposta dos autores visa auxiliar na gestão da rede de saúde mental no Brasil, a fim de permitir a integração e interoperabilidade entre bases de dados independentes e possibilitar a geração de indicadores relevantes à saúde mental, a fim de melhorar a tomada de decisão. Tal ontologia representa uma base de conhecimento comum no domínio da saúde mental, permitindo fazer inferências sobre tratamentos, sintomas, métodos de diagnóstico e prevenção, auxiliando os profissionais de saúde nas decisões clínicas (YAMADA et al., 2020).

Yoshiura et. al. propuseram um modelo conceitual baseado em Web Semântica para auxiliar nos processos de concepção, desenvolvimento e implementação de Observatórios de Saúde e/ou a adequação dos já existentes (YOSHIURA et al., 2018). Segundo os autores, o uso de tecnologias da Web Semântica pode aprimorar os observatórios em termos de padronização de dados, gestão do conhecimento e interoperabilidade com outras aplicações.

Hitzler e Janowicz enfatizam em (HITZLER; JANOWICZ, 2011) o crescente uso da *web* semântica. Tal crescimento é possível graças à aplicação do paradigma da *web* semântica não estar vinculado a um tipo específico de conhecimento ou área. Pelo contrário, a *web* semântica vem dar suporte a ambientes onde a interdisciplinaridade e a heterogeneidade estão implícitas na rotina. Neste sentido, é justificado o aumento no número de convenções e pessoas interessadas no assunto, sejam elas oriundas do meio acadêmico ou comercial.

Para Valle et al. em (VALLE et al., 2005), a adoção do paradigma da *web* semântica, aliado a outras estruturas e tecnologias amplamente utilizadas, como, por exemplo, a *Extensive Markup Language* (XML) e os *webservices*, visa promover a interoperabilidade entre os sistemas de informação em saúde. Esta junção permite a união necessária entre uma formalização sintática e a garantia de um significado (semântico) para o dado que está sendo intercambiado entre as aplicações. Tal junção abre um leque de exploração destes dados não somente por computadores, mas também por humanos, pois seus respectivos significados são mantidos independentemente de onde estes dados estejam sendo tratados ou exibidos.

Conhecendo-se ambas as formalizações (semânticas e sintáticas) a interoperabilidade entre os sistemas legados pode ser atingida adequando-os ou então se lançando mão de *webservices* para consumir, organizar e marcar tais dados. Trata-se de uma estratégia já consolidada, mas que permanece útil e eficiente até o momento.

Lopes e Oliveira (LOPES; OLIVEIRA, 2011) classificam os dados de saúde como dispersos através de sistemas independentes e fragmentados em silos de dados fechados. Além disso, os *softwares* disponíveis são primitivos, inconsistentes e estáticos. Para resolver esse problema, os autores propõem um *framework* para criar aplicações semânticas mais rapidamente e migrar sistemas legados para a *web* semântica, somado a um ponto de acesso SPARQL para publicar os dados desejados.

Jiang, Solbrig e Chute apresentam em (JIANG; SOLBRIG; CHUTE, 2013) uma ferramenta baseada em *web* semântica para contribuir com definições de termos da Classificação Internacional de Doenças, versão 11 (CID-11). Essa classificação permite que parâmetros de determinado termo sejam estabelecidos por um consenso de especialistas, através de sugestões e validações colaborativas e distribuídas. Um desses parâmetros é denominado "definição textual". Para sugerir definições textuais de forma automatizada com o uso da web semântica, buscas foram realizadas no DBPedia e comparadas com os valores encontrados na *Unified Medical Language System* (UMLS) e na *Systematized Nomenclature of Medicine – Clinical Terms* (SNOMED-CT). Os resultados são compilados e sugeridos, posteriormente, a especialistas para a devida validação.

Belleau et. al. apresentam em (BELLEAU et al., 2008) um projeto de código aberto, denominado Bio2RDF, que utiliza tecnologias da web semântica para agregar a compilar conhecimento biomédico através da integração de numerosas bases de dados, constituindo uma das maiores redes de dados conectados disponíveis.

Abhishek et al. apresentam em (ABHISHEK; M.P., 2016) uma abordagem para a criação de uma ontologia que descreve todos os pontos do Programa Nacional Indiano de Controle e Gestão da Tuberculose (RNTCP). Tal ontologia tem como objetivo auxiliar a tomada de decisão e gestão inerentes ao referido programa. Dada a complexidade da estratégia de controle da tuberculose na Índia, evidenciada pelos cinco níveis de atenção ao paciente, onde cada nível possui seus respectivos gestores e um determinado grau de liberdade para tomada de decisão, ficou clara a robustez da abordagem utilizando-se a web semântica. O alicerce utilizado para a construção do conjunto de ferramentas descrito no trabalho foi a *Basic Formal Ontology* (BFO, http://basic-formal-ontology.org/), onde todos os termos relacionados ao RNTCP foram mapeados numa estrutura hierárquica. Tal estrutura garante a consistência semântica necessária

para a aplicação de algoritmos, implementados utilizando-se a *Semantic Web Rule Language* (SWRL), para a extração de conhecimento e inferências a partir da base de dados.

Ogundele et al. desenvolveram uma ontologia para categorizar, esclarecer e consolidar os fatores de influência que afetam o comportamento de adesão ao tratamento dos pacientes com TB (OGUNDELE et al., 2015). A ontologia não apenas categoriza os fatores de influência na adesão ao tratamento, mas também fornece evidências na forma de *links* para estudos clínicos que validam a importância de determinado fator. A ontologia é composta por cinco classes principais, a saber: *Evidence* (evidência científica), *TABInfluencingFactor* (representa uma característica ou um grupo de características de um paciente com TB que foi identificado como influente na adesão e é informado por pesquisas em uma ou mais comunidades), *Work* (estudo científico), *Place* (onde o estudo foi realizado) e *Interdependency* (para representar uma descoberta de uma interdependência entre fatores). Os autores reforçam o potencial da ontologia para capturar múltiplos fatores influenciadores multifacetados e suas inter-relações. A ontologia tem a capacidade de levar a compreensão do comportamento de adesão ao tratamento além do que pode ser fundamentado por um único especialista humano, de forma a suportar a tomada de decisões baseada em evidências para o manejo de pacientes com TB.

Shaban-Nejad et al. apresentam em (SHABAN-NEJAD et al., 2017) uma plataforma, denominada PopHR, baseada em conhecimento para integrar, analisar e visualizar dados de saúde populacional. Segundo os autores, os sistemas de informação atuais são limitados para integrar e apresentar dados, oferecendo pouco suporte à tomada de decisão. É proposto, então, uma aplicação para automatizar o processo de extração e integração de grandes volumes de dados de fontes distribuídas e heterogêneas, através da web semântica. O sistema tem como objetivo facilitar a tomada de decisão baseada em evidências, sendo capaz de calcular indicadores de saúde e permitir aos usuários interpretá-los no contexto de determinada população e uma evidência relevante.

Ding, Wald e Wills propõem em (DING; WALD; WILLS, 2015) um sistema de apoio à decisão para viagens com foco em pessoas com deficiência seguindo o princípio de *Linked Data*. O sistema consiste de duas fases, sendo uma fase de integração de dados e outra de suporte à decisão. A primeira fase é responsável por extrair dados heterogêneos, obter uma correlação e publicá-los através dos princípios dos Dados Conectados. A segunda fase utiliza o conhecimento produzido pela primeira para encontrar informações otimizadas sobre viagens para usuários, baseado em suas capacidades físicas.

Vandervalk, McCarthy e Wilkinson apresentam em (VANDERVALK; MCCARTHY; WILKINSON, 2008) um *framework* denominado CardioSHARE. Trata-se de um *webservice*

que oferece um ponto de acesso SPARQL para disponibilização de conhecimento a partir de análises de dados clínicos sobre doenças cardíacas, embora, segundo os pesquisadores, seja possível aplicar o *framework* a qualquer tipo de dados.

Tilahun et al. apresentam em (TILAHUN et al., 2014) a criação de uma base de dados RDF a partir de um conjunto de dados relacionados ao HIV provenientes do Observatório Global da OMS. Essa base foi, então, interligada com outras, como o DBPedia e Bio2RDF, além de permitir consultas SPARQL e fornecer uma ferramenta gráfica para navegação pelos dados disponíveis.

Em (BURANARACH; SUPNITHI; CHALORTHAM, 2009), os autores propõem um *framework* baseado em web semântica para suportar o gerenciamento do tratamento de doenças crônicas a partir do fornecimento de suporte à decisão e de sistemas de informação clínica. Através da web semântica, uma plataforma de conhecimento é proposta para permitir a integração de bases de dados de pacientes, ontologias e repositórios de literatura médica para apoiar cuidados de saúde baseados em evidência. As ontologias são usadas para aquisição de conhecimento e modelagem de dados. Como trabalho futuro, é proposta a aplicação do *framework* em serviços de saúde e a condução da avaliação de uso.

Song et al. desenvolveram, em (SONG; PARK; JIN, 2014), um mecanismo de coleta e disponibilização de informações em saúde baseado em metadados e ontologias que garante a interoperabilidade semântica através da indexação de informações.

Sernadela, González-castro e Oliveira desenvolveram uma ferramenta de migração que pode ser acoplada aos sistemas tradicionais com a finalidade de trazer conhecimento, regras de inferência e consulta a dados federados (SERNADELA; GONZÁLEZ-CASTRO; OLIVEIRA, 2017). Destinada ao domínio biomédico, essa ferramenta oferece serviços de integração via web semântica que ajudam desenvolvedores e pesquisadores a criar sistemas de informação semanticamente aprimorados.

Zolhavarieh, Parry e Bai sugerem um modelo para uso da WS em Sistemas de Suporte à Decisão Clínica (ZOLHAVARIEH; PARRY; BAI, 2017). O modelo discorre sobre como o conhecimento pode ser descoberto e atualizado com as tecnologias da web semântica por meio de uma estrutura denominada "*knowledge broker*". O modelo é dividido em: descoberta de conhecimento, construção do conhecimento e avaliação de qualidade.

Machado et al. sugerem que a *web* semântica e seus recursos desempenham um papel fundamental na integração de dados, interoperabilidade de dados e descoberta de conhecimento e que todas essas características têm o potencial de impulsionar a saúde, auxiliando na

compreensão das doenças, tratamentos adequados e na melhoria da prática clínica diária (MACHADO et al., 2013).

Zaveri et al. descrevem em (ZAVERI et al., 2013) o processo de conversão e publicação dos dados do *Global Health Observatory* (GHO) em RDF. Ao fornecer os dados do GHO e interligá-los com outros conjuntos de dados, é possível obter informações sobre importantes tópicos relacionados à saúde em diferentes países, facilitando o trabalho de profissionais de saúde nas tarefas de análise de dados.

5. SEGURANÇA DA INFORMAÇÃO

Um problema crítico em relação à interoperabilidade em SIS é a segurança dos dados. Os dados dos pacientes, como informações pessoais e registros de tratamento, devem permanecer confidenciais, ou seja, apenas indivíduos e sistemas autorizados e que atendam a todos os requisitos de segurança devem ser capazes de acessá-los. Uma violação de segurança traz consequências financeiras e jurídicas e representa um forte desestímulo às iniciativas de compartilhamento de dados, pois muitas instituições temem que o acesso aos dados possa ser explorado indevidamente.

Mesmo que as questões de privacidade sejam resolvidas, não há acordo sobre a infraestrutura técnica específica necessária para dar suporte a essa tarefa (PETERSON et al., 2016). A troca de dados em SIS deve ser cuidadosamente planejada e protegida contra acessos não autorizados. Portanto, a definição de níveis de acesso e a execução de transferências seguras de dados são cruciais para aumentar a segurança no processo de intercâmbio de informação sigilosa entre sistemas.

Os mecanismos de segurança da informação visam proteger os dados de acessos não autorizados. Assim, ao implementá-los, é possível garantir que as propriedades de segurança chave, comumente referidas como Tríade CID - Confidencialidade, Integridade e Disponibilidade -, não sejam violadas na presença de ameaças e falhas deliberadas ou acidentais (TCHERNYKH et al., 2016). As propriedades CID são definidas da seguinte maneira:

- Confidencialidade: considera que apenas pessoas autorizadas têm acesso às bases de dados, sistemas e serviços através de um conjunto de regras e restrições (ZISSIS; LEKKAS, 2012);
- Integridade: as informações devem ser precisas, completas e consistentes, para que apenas indivíduos selecionados possam alterá-las. Os dados devem ser protegidos contra exclusão, modificação ou fabricação maliciosa (BODIN; GORDON; LOEB, 2005);
- Disponibilidade: um sistema deve estar disponível para uso por um usuário ou máquina sempre que necessário e sem interrupções (KARTHIK; ANANTHANARAYANA, 2018).

Outros princípios da segurança da informação complementam a tríade CID, como:

- Autenticidade: se refere à manutenção das condições iniciais dos dados de maneira autêntica, da mesma forma que foram produzidos ou armazenados. No caso da

verificação da identidade de usuários, como ela é única e imutável, sua autenticidade pode ser verificada através de um mecanismo de autenticação para garantir a correta identificação de usuários ou sistemas externos (KARTHIK; ANANTHANARAYANA, 2018);

- Autorização: processo que ocorre após uma autenticação válida para verificar os privilégios/níveis de acesso de determinada entidade externa (DI VIMERCATI; FORESTI; SAMARATI, 2007);
- Não-repúdio: refere-se a uma entidade não ser capaz de negar ações realizadas com sua identidade única. É de natureza semelhante a uma assinatura manual de um documento (CHIANG; HUANG, 2003; TRIPATHI; AGRAWAL, 2014).
- Privacidade: refere-se à não-intrusão no espaço pessoal ou nas informações pessoais de um indivíduo (GRODZINSKY; TAVANI, 2005). A privacidade pode ser preservada pela proteção dos identificadores e dos quasi-identificadores (aqueles que, quando combinados, podem permitir distinguir registros e possivelmente identificar um indivíduo se alguma informação complementar estiver disponível) (MOTWANI; XU, 2007).

Em relação à segurança da informação na Web Semântica, é importante notar que sua especificação não fornece algoritmos de segurança implementados, os quais devem ser desenvolvidos como recursos adicionais (algoritmos, padrões e ferramentas) e aplicados às suas camadas de segurança (THURAISINGHAM, 2005). A próxima subseção apresenta uma visão geral de técnicas de segurança da informação utilizadas no contexto da Web Semântica para permitir o manuseio e a interoperabilidade de dados de saúde.

5.1. Métodos de criptografia

A criptografia pode ser usada para limitar o acesso e para armazenar e compartilhar informações com segurança. Os dados em sua forma original são chamados de texto simples. Por outro lado, os dados transformados e ilegíveis são chamados de texto cifrado. Chave é uma entrada aleatória para o algoritmo de criptografia que é usado para transformar um texto simples em texto cifrado (criptografia/ encriptação) ou o processo oposto (descriptografia/desencriptação). Em geral, a criptografia simétrica e assimétrica são os dois métodos de criptografia amplamente utilizados, cada um com suas vantagens, desvantagens e algoritmos próprios (MAMUN et al., 2017).

Na criptografia simétrica, também conhecida como chave secreta ou criptografia de

chave compartilhada, a criptografia e a descriptografia são realizadas com a mesma chave. Assim, a chave deve ser distribuída de forma segura, o que é uma limitação. A criptografia com uma chave mais longa é mais difícil de quebrar do que a feita com uma chave menor (KUMAR; MUNJAL; SHARMA, 2011). Este método é usado principalmente para garantir a integridade e confidencialidade de dados.

A criptografia assimétrica, também conhecida como criptografia de chave pública, utiliza duas chaves: uma chave pública, que pode ser conhecida por todos, e uma chave privada, que deve ser conhecida apenas pelo seu dono, resolvendo o problema de distribuição das chaves. Uma chave é derivada de outra usando uma função unilateral. Este método fornece propriedades de confidencialidade, autenticidade, integridade e não repúdio. Os dados criptografados com uma chave só podem ser descriptografados com seu par. No entanto, a criptografia de chave pública é baseada em funções matemáticas e é computacionalmente intensiva, o que a torna pouco eficiente em termos de desempenho (KUMAR; MUNJAL; SHARMA, 2011). As técnicas de criptografia assimétrica são quase 1000 vezes mais lentas do que as técnicas simétricas porque requerem mais poder de processamento computacional (STALLINGS, 2005). Além disso, existem limitações relacionadas à quantidade de dados que podem ser criptografados usando criptografia assimétrica, o que pode ser um fator preocupante para seu uso em SIS e contextos de *big data* (MAMUN et al., 2017).

5.2. Abordagens de segurança para manipulação de dados eletrônicos de saúde por meio da Web Semântica: uma revisão de escopo[1]

Questões de segurança na Web Semântica sempre estiveram em discussão desde sua definição inicial. Embora inovador, o funcionamento de todas as camadas não é trivial, pois cada tecnologia e protocolo adicionam alguma complexidade à solução. Com relação à segurança dos dados, não é diferente.

Uma revisão de escopo da literatura foi conduzida para fornecer uma visão ampla dos mecanismos de segurança aplicados às tecnologias da Web Semântica que permitem seu uso em contextos que manuseiam dados de saúde, bem como identificar possíveis lacunas de pesquisa na literatura e características ou fatores-chave relacionados à segurança no Web Semântica.

As revisões de escopo são ideais para fornecer uma visão geral de um determinado

[1] A informação contida nesta subseção refere-se parcialmente ao conteúdo do artigo *Security approaches for electronic health data handling through the Semantic Web: a scoping review*, aceito para publicação no periódico *Semantic Web – Interoperability, Usability, Applicability*. ISSN 1570-0844. Disponível em http://www.semantic-web-journal.net/system/files/swj3088.pdf.

tópico, bem como para determinar sua cobertura e dar uma indicação clara do volume de literatura e estudos disponíveis (MUNN et al., 2018). A metodologia da revisão foi baseada em (ARKSEY; O'MALLEY, 2005) e em diretrizes definidas pelo *Joanna Briggs Institute* (PETERS et al., 2015), que recomendam uma estrutura de cinco estágios para revisões de escopo:

1. Identificação do objetivo e das questões de pesquisa
    - Quais são os principais mecanismos aplicados às tecnologias da Web Semântica para proteger os dados de saúde?
    - Quais propriedades de segurança estão sendo consideradas?
    - Quais são as lacunas de conhecimento em relação à segurança para o tratamento de dados de saúde por meio da Web Semântica?

2. Identificação dos estudos relevantes
    - Estratégia de busca
        i. Bases de dados: PubMed/MEDLINE, IEEE Xplore Digital Library, Scopus, Embase, Web of Science, ProQuest, and Cochrane Database of Systematic Reviews
        ii. *String* de busca: *health AND "semantic web" AND (security OR privacy OR "access control" OR integrity OR confi-dentiality OR cryptography)*
    - Critérios de elegibilidade
        i. Artigos completos, revisões, artigos de conferência, editoriais, cartas
        ii. Texto na língua inglesa
        iii. Materiais publicados entre maio de 2001, quando o termo Web Semântica foi definido (BERNERS-LEE; HENDLER; LASSILA, 2001), até julho de 2021.
        iv. Publicações que não fazem referência às palavras-chave "*health*", "semantic web" e pelo menos um relacionado à segurança no título ou no resumo foram excluídos, bem como trabalhos que apresentassem apenas propostas, modelos não implementados ou soluções não aplicadas à Web Semântica.
    - Busca recursiva
        i. Verificação das listas de referências dos artigos selecionados

3. Seleção dos estudos

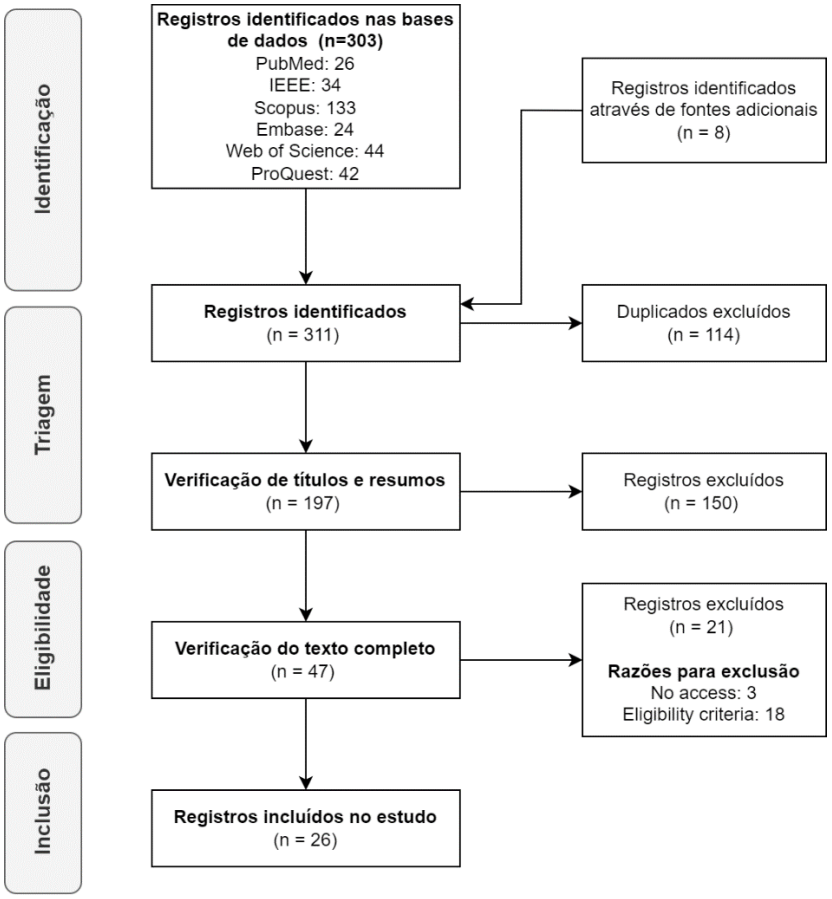
- Análise e seleção dos artigos com base no resumo, título e texto completo
- Uso do diagrama *Preferred Reporting Items for Systematic Review and Meta-Analysis* (PRISMA) (MOHER et al., 2009)

4. Mapeamento dos dados
    - Definição da estratégia de extração de dados dos artigos selecionados para responder adequadamente às perguntas do estudo.

5. Compilar, resumir e relatar os resultados

Inicialmente, foram encontrados 303 artigos nas bases de dados selecionadas e 8 foram identificados por outras fontes (total de 311 artigos), dos quais 197 foram selecionados após a retirada de 114 duplicatas. Após a triagem dos títulos e resumos, o número de artigos foi reduzido para 47. No entanto, uma avaliação de texto completo para elegibilidade excluiu 21 artigos adicionais porque 3 estudos estavam inacessíveis (sem acesso gratuito ou institucional) e 18 não atendiam aos critérios de elegibilidade. Finalmente, 26 (47-21) estudos foram incluídos na revisão. A Figura 3 apresenta o diagrama PRISMA.

Figura 3. Diagrama PRISMA



De acordo com a literatura explorada, as pesquisas sobre abordagens de segurança para lidar com dados eletrônicos de saúde por meio da Web Semântica têm aumentado gradativamente desde 2010. Os resultados mostraram que os artigos selecionados foram publicados entre 2005 e 2021, mas o volume foi maior entre 2014 e 2019. Dos 26 estudos incluídos, há 15 artigos de pesquisa, 10 artigos em conferências e 1 editorial. A Tabela 1 sumariza a análise dos artigos incluídos na revisão de escopo.

Os seguintes atributos e recursos de segurança foram abordados pelos artigos: autenticação, autorização, integridade, disponibilidade, confidencialidade, privacidade e procedência. Dependendo da finalidade da solução, cada estudo normalmente fornece mecanismos para implementar um ou mais de cada atributo ou recursos. Três categorias foram definidas para classificar os artigos e refletir seus objetivos, a saber: Controle de Acesso (15 artigos), Infraestrutura de Interoperabilidade (3 artigos) e Controle de Privacidade (8 artigos).

Nos artigos da categoria de Controle de Acesso, embora alguns trabalhos envolvem controle de privacidade, isso é feito no contexto de um mecanismo mais amplo que, por meio de políticas de acesso baseadas em regras e semântica, permite inferir se um determinado agente possui os direitos de acesso necessários, de modo a garantir a privacidade dos indivíduos. As soluções geralmente funcionam como uma ponte entre as abordagens clássicas para controle de acesso e tecnologias da Web Semântica, mas, geralmente, são complexas para implementar nos sistemas de informação de saúde existentes.

Por outro lado, os artigos da categoria de Controle de Privacidade propõem abordagens focadas em garantir a privacidade pelo mapeamento de requisitos (por exemplo, leis e regulamentos) por meio de modelos e ontologias, e não apresentam um conjunto completo de ferramentas para controle de acesso.

Por sua vez, os estudos classificados como Infraestrutura de Interoperabilidade são mais abrangentes, pois fornecem ferramentas que abrangem as duas categorias anteriores por meio de uma infraestrutura completa para interoperabilidade funcional e semântica de dados de saúde.

Na saúde, proteger a confidencialidade dos dados é crucial. Em (DAO et al., 2013), a criptografia assimétrica é utilizada, a qual lida com o problema de distribuição segura da chave de criptografia. Embora os autores não tenham definido mecanismos de controle de acesso, eles mostraram que o uso de criptografia em contextos sensíveis é uma maneira eficiente de garantir a confidencialidade e proteger os dados de leitores não autorizados, porque apenas aqueles em posse da chave são capazes de ler a mensagem. No entanto, não está claro como os autores lidam com a quantidade de dados que podem ser criptografados, devido às limitações na

criptografia assimétrica (MAMUN et al., 2017), o que pode ser um fator preocupante para lidar com grandes volumes de dados.

Vários estudos compartilham as mesmas tecnologias subjacentes para entregar suas soluções, mostrando a flexibilidade das ferramentas da Web Semântica para permitir a implementação de mecanismos de segurança e proteger dados sensíveis, e ainda permitir a interoperabilidade e integração desses dados. A Web Semântica pode atuar como uma ponte para o uso conjunto de tecnologias semânticas e clássicas.

Os resultados da revisão mostram que as principais lacunas da literatura se referem à ausência de uma arquitetura computacional completa capaz de cobrir todas as propriedades de segurança desejadas em ambientes que tratam de dados sensíveis através da Web Semântica, e à alta complexidade de implementação das soluções existentes que, na maioria das vezes, demandam mudanças não triviais nos SIS que não foram desenvolvidos inicialmente com a necessidade da semântica em consideração. Ainda, idealmente, os mecanismos de confidencialidade, privacidade e proveniência devem coexistir em favor de uma solução de privacidade mais abrangente para o tratamento de dados de saúde. No entanto, nenhum estudo apresentou uma solução que considerasse todos esses mecanismos em conjunto.

Embora os artigos classificados na categoria Infraestrutura de Interoperabilidade desta pesquisa possam tocar nessas questões, as soluções podem, ao mesmo tempo, representar um desestímulo ao uso de tecnologias da Web Semântica para promover a interoperabilidade e a integração semântica de dados de saúde, devido à complexidade envolvida. No entanto, o desenvolvimento de uma solução *plug-and-play* a ser oferecida como um serviço por provedores de computação em nuvem confiáveis e seguros, conforme sugerido por (ALRAJA et al., 2021), pode representar uma forma de simplificar a implementação de mecanismos de segurança na Web Semântica.

Embora o paradigma da Web Semântica tenha sido criado em 2001, os resultados sugerem que a aplicação e o uso de suas tecnologias continuam em crescimento, sendo a área da saúde particularmente interessante pela capacidade de inferência para vinculação de registros e derivação de políticas de acesso. No entanto, a complexidade de uma determinada solução tende a aumentar à medida que mais tecnologias e ferramentas são incorporadas, o que pode ser visto como uma desvantagem principalmente para soluções existentes não inicialmente projetadas para a Web Semântica.

Tabela 1. Sumarização dos resultados da revisão da literatura

Título do artigo	**Confidencialidade**	**Integridade**	**Disponibilidade**	**Privacidade**	**Provenance**	**Autenticação**	**Autorização**	**Categoria**	**Mecanismos e tecnologias**
A fine-grained context-aware access control model for health care and life science linked data (LIU; WANG, 2016)	-	-	-	-	-	x	x	Controle de Acesso	- XACML - OWL Ontology - SWRL rules - Inference/reasoning engine
A semantic authorization model for pervasive healthcare (LI; CHU; YAO, 2014)	-	-	-	-	-	-	x	Controle de Acesso	- Ontology - Inference/reasoning engine
A SWRL bridge to XACML for clouds privacy compliant policies (RAHMOUNI et al., 2014)	-	-	-	x	-	x	x	Controle de Acesso	- XACML - OWL Ontology - SWRL rules
Access control management for e-Healthcare in cloud environment (SUN; YONG; SOAR, 2014)	-	-	-	-	-	x	x	Controle de Acesso	- XACML - OWL Ontology - RBAC security model
An access control framework for pervasive mobile healthcare systems utilizing cloud services (POULYMENOPOULOU; MALAMATENIOU; VASSILACOPOULOS, 2012)	-	-	-	-	-	-	x	Controle de Acesso	- OWL Ontology - SWRL Rules - Inference/reasoning engine - RBAC security model
Behavior-Based Access Control for Distributed Healthcare Environment (YARMAND; SARTIPI; DOWN, 2008)	-	-	-	x	-	-	x	Controle de Acesso	- Ontology - Inference/reasoning engine
Context-aware Access Control Using Semantic	-	-	-	-	-	-	x	Controle de Acesso	- XACML - OWL Ontology

Policies (DERSINGH; LISCANO; JOST, 2008)									- SWRL rules - Inference/reasoning engine
Establishment of access levels for health sensitive data exchange through semantic web (LIMA et al., 2018b)	-	-	-	-	-	x	x	Controle de Acesso	- Ontology - Jena - JSON-LD - Semantic API
Flexible access to patient data through e-Consent (KONDYLAKIS et al., 2015)	-	-	-	x	x	-	x	Controle de Acesso	- Ontology - Abstract annotation model - RDF
Multi Authority Access Control in a Cloud EHR System with MA-ABE (DIXIT; JOSHI; GEOL CHOI, 2019)	x	-	-	x	-	x	x	Controle de Acesso	- Encryption (MA-ABE) - ABAC security model - Ontology - SWRL rules - Inference/reasoning engine
Privacy compliance and enforcement on European healthgrids: An approach through ontology (RAHMOUNI et al., 2010)	-	-	-	x	-	-	x	Controle de Acesso	- OWL Ontology - SWRL rules
Securing access to sensitive RDF data (PAPAKONSTANTINOU et al., 2014)	-	-	-	x	x	-	x	Controle de Acesso	- Ontology - Abstract annotation model - RDF
Semantic privacy-preserving framework for electronic health record linkage (LU; SINNOTT, 2018)	-	-	-	x	-	-	x	Controle de Acesso	- XACML - OWL Ontology - SWRL rules -

									Inference/reasoning engine
Towards a semantic medical internet of things (DRIDI; SASSI; FAIZ, 2018)	x	-	-	x	-	x	x	Controle de Acesso	- Ontology - RDF - IoT devices
Using OWL and SWRL to represent and reason with situation-based access control policies (BEIMEL; PELEG, 2011)	-	-	-	-	-	-	x	Controle de Acesso	- OWL Ontology - SWRL rules - Description Logic - Inference/reasoning engine
ARTEMIS: towards a secure interoperability infrastructure for healthcare information systems (BONIFACE; WILKEN, 2005)	x	x	-	x	-	x	x	Infraestrutura de Interoperabilidade	- Semantic Web Services - Ontology - Inference/reasoning engine - Encryption (Triple-DES)
Secure semantic smart healthcare (S3HC) (TIWARI et al., 2019)	x	x	x	x	-	x	x	Infraestrutura de Interoperabilidade	- Ontology - SWRL rules - Inference/reasoning engine - SPARQL - RDF/XML Security - Encryption - Hash functions - IoT devices
Security Framework for Tuberculosis Health Data Interoperability Through the Semantic Web (LIMA et al., 2021b)	x	x	x	-	-	x	x	Infraestrutura de Interoperabilidade	- Ontology - Hybrid cryptography - Hash functions - Jena - SPARQL

									- D2Rq Server - HyperGraphQL - Semantic APIs - JSON-LD - RDF
A Model-driven Privacy Compliance Decision Support for Medical Data Sharing in Europe (BOUSSI RAHMOUNI et al., 2011)	-	-	-	x	-	-	-	Controle de Privacidade	- Ontology - SWRL rules - Inference/reasoning engine
An integrated framework for privacy protection in IoT — Applied to smart healthcare (ALRAJA et al., 2021)	-	-	-	x	-	-	-	Controle de Privacidade	- OWL Ontology - Inference/reasoning engine - IoT devices
An Ontology for a HIPAA compliant cloud services (JOSHI; YESHA; FININ, 2016)	-	-	-	x	-	-	-	Controle de Privacidade	- OWL Ontology
COC: An ontology for capturing semantics of circle of care (DONG; SAMAVI; TOPALOGLOU, 2015)	-	-	-	x	x	-	x	Controle de Privacidade	- Ontology - RDF - SPARQL
Improving privacy in health care with an ontology-based provenance management system (CAN; YILMAZER, 2020)	-	-	-	x	x	-	-	Controle de Privacidade	- Ontology
Knowledge-based personalized search engine for the Web-based Human Musculoskeletal System Resources (HMSR) in biomechanics (DAO et al., 2013)	x	-	-	-	-	-	-	Controle de Privacidade	- OWL Ontology - Multi-agent semantic crawler for the Web - Asymmetric encryption

Ontology for Attack Detection: Semantic-Based Approach for Genomic Data Security (NOOR et al., 2017)	-	-	-	x	-	-	-	Controle de Privacidade	- OWL Ontology - SWRL rules - Jena - Inference/reasoning engine
Preserving patients' privacy in health scenarios through a multicontext-aware system (HUERTAS CELDRÁN et al., 2017)	-	-	-	x	-	-	-	Controle de Privacidade	- OWL Ontology - SWRL rules - Jena - Inference/reasoning engine

5.3. Blockchain

A arquitetura altamente resiliente e a natureza distribuída do Blockchain o tornam uma ferramenta interessante para fornecer uma maneira confiável de aprimorar o controle de identidade, segurança de dados e compartilhamento de conteúdo. Compreender os tipos e formas em que os dados são armazenados e gerenciados determina o grau de utilidade, flexibilidade, desempenho e custo de uma infraestrutura baseada em Blockchain, permitindo sua integração com outras tecnologias, como, por exemplo, a Web Semântica. Portanto, observar questões como análise de dados e governança, além do desenvolvimento de aplicativos, é fundamental para o fortalecimento das comunidades e indústrias de pesquisa.

Uma abordagem válida para lidar com requisitos de segurança e permitir uma infraestrutura descentralizada para compartilhamento de dados é o uso de uma tecnologia de livro-razão distribuída (do inglês, *Distributed Ledger Technology* - DLT). Refere-se aos protocolos tecnológicos que permitem o funcionamento seguro de uma base de dados digital descentralizada. Permite acesso simultâneo, validação, atualização de registros de maneira imutável em uma rede espalhada por várias entidades ou locais, e o armazenamento de todas as informações de maneira segura. Depois que as informações são armazenadas, elas se tornam um banco de dados imutável e são regidas pelas regras da rede (MAULL et al., 2017; SUNYAEV, 2020).

Uma DLT é baseada nos princípios de rede ponto-a-ponto (P2P) e em primitivas criptográficas. Cada membro da rede possui uma credencial única (par de chaves público-privada) e armazena uma cópia idêntica do livro-razão e contribui para o processo coletivo de validação e certificação de transações digitais para a rede (WÜST; GERVAIS, 2017).

As informações são criptografadas e assinadas digitalmente para garantir sua autenticidade e precisão. Blockchain é um exemplo de DLT, que ficou famoso com a moeda digital Bitcoin em 2008 apresentada por Satoshi Nakamoto (provavelmente uma pessoa fictícia). Nakamoto publicou um artigo (NAKAMOTO, 2008) propondo um mecanismo de pagamento *online* entre duas partes sem depender de terceiros. Cada transação seria vinculada à anterior em um livro-razão público à prova de violação e, ao examinar o histórico de transações da moeda, a rede poderia evitar o problema de gastos duplicados (ISMAIL et al., 2019).

Para controlar as transações, várias transações são agrupadas em uma estrutura chamada de bloco, que é identificada exclusivamente por seu *hash* e carimbo de data / hora. Os blocos são então vinculados cronologicamente usando *hashes*, compondo uma cadeia de transações

(PUTHAL et al., 2018). As funções *hash* são algoritmos que transformam uma mensagem de comprimento aleatório e geram uma saída de comprimento fixo que atua como um identificador exclusivo de uma mensagem. Qualquer mínima alteração na mensagem original levará a um valor de *hash* totalmente diferente (TEH; TAN; ALAWIDA, 2019).

Blockchain é considerada uma tecnologia segura para automatizar o mecanismo de confiança sem uma autoridade central mitigando o risco devido às suas propriedades e mecanismos, tais como: i) o uso de criptografia para validar transações confiáveis; ii) imutabilidade, o que garante que os registros não podem ser alterados ou excluídos; iii) proveniência / rastreabilidade, que permite rastrear a origem de todas as transações do livro razão; iv) descentralização, i.e., todos os membros têm acesso ao banco de dados (livro-razão) com controle viabilizado por algoritmos de consenso; e v) transparência, ou seja, as transações são improváveis de serem corrompidas e são visíveis para todas as partes autorizadas (ØLNES; UBACHT; JANSSEN, 2017; HUGHES et al., 2019).

Um algoritmo de consenso é um mecanismo que permite que usuários ou máquinas se coordenem em uma configuração distribuída. Ele deve garantir que todos os membros de um sistema possam concordar com uma única fonte de verdade, mesmo que alguns dos membros falhem. É estruturado como um processo de tomada de decisão onde um grupo de entidades expressa suas opiniões individuais para construir a decisão que fornece a melhor estimativa de um processo ou sistema. É uma forma de resolução onde os indivíduos precisam apoiar a decisão da maioria, de forma consensual (DU et al., 2017; BHARDWAJ; DATTA, 2020).

Uma rede blockchain pode ser pública (sem permissão), privada (com permissão) ou federada (consórcio). Qualquer pessoa pode ingressar em uma rede pública e o acordo é obtido por meio de um mecanismo de consenso. Em uma rede privada, as transações estão disponíveis apenas para um grupo de participantes de uma ou mais organizações e membros/nós validadores especiais de uma organização são responsáveis pela obtenção de consenso. Finalmente, em uma rede federada, nenhuma organização é responsável pelo consenso e validação de blocos, mas sim um conjunto de nós predeterminados (PUTHAL et al., 2018).

Um ou outro componente importante presente em versões mais avançadas de redes Blockchain são os contratos inteligentes (do inglês, *smart contracts*). Contratos inteligentes são programas (código) armazenados em um Blockchain que são executados quando condições predeterminadas são atendidas. Eles normalmente são usados para automatizar a execução de um consenso, de modo que todos os participantes possam ter a certeza imediata do resultado, sem o envolvimento de qualquer intermediário ou perda de tempo. Também podem automatizar um fluxo de trabalho, disparando a próxima ação quando as condições forem atendidas

(KÕLVART; POOLA; RULL, 2016; ZHENG et al., 2020).

Com a evolução da tecnologia e o melhor entendimento da utilidade de uma DLT, surgiram diversas opções de plataformas, entre elas: Hyperledger Fabric, R3 Corda, Ethereum e IOTA. Como cada tipo de plataforma possui suas próprias características, o desenvolvedor de uma rede Blockchain deve considerar as especificidades do seu projeto para escolher a opção mais adequada. A Tabela 2 apresenta as características de cada plataforma.

Tabela 2 - Comparação entre as principais DLT

Característica	Ethereum	Hyperledger Fabric	R3 Corda	IOTA
Descrição da plataforma	Plataforma Blockchain genérica	Plataforma Blockchain modular	DLT especializado para o mercado financeiro	DLT especializado para Internet das Coisas (do inglês, *Internet of Things* – IoT)
Governança	Desenvolvedores da rede Ethereum	Linux Foundation	R3	IOTA Foundation
Modo de operação	Não permissionado, público ou privado	Permissionado, privado	Permissionado, privado	Não permissionado, público ou privado
Consenso	Prova de trabalho (do inglês, *Proof-of-Work* - PoW)	Múltiplas abordagens	Múltiplas abordagens	Consenso probabilístico rápido (do inglês, *Fast Probabilistic Consensus* – FPC)
Contratos inteligentes	Código em Solidity (linguagem específica para contratos inteligentes)	Código em Go ou Java	Código em Kotlin ou Java	Código em Solidity
Moeda	Ether	Nenhuma	Nenhuma	MIOTA

Fonte: Adaptado de (VALENTA; SANDNER, 2017)

DLTs oferecem oportunidades para pesquisas clínicas através da manutenção do histórico de transações e acessos, disponibilização em tempo real de dados individualizados, e a capacidade de definir permissões para acessar e auditar dados para garantir sua integridade e segurança (LIANG et al., 2018).

Cada vez mais são disponibilizados novos projetos baseados em Blockchain para a

promoção da ciência e da área da saúde, como, por exemplo, o projeto Enigma, desenvolvido pelo *Massachusetts Institute of Technology* (MIT). O Enigma é uma abordagem de Blockchain que permite o compartilhamento seguro de dados em grande escala e em um ambiente controlável pelo usuário. Com este tipo de implementação, os dados podem ser compartilhados entre quaisquer usuários ou grupos de usuários, sejam eles pesquisadores, profissionais de saúde ou pacientes. A ideia por trás da técnica é a privacidade diferencial, ou seja, o usuário pode ajustar a dose de equilíbrio entre dados públicos e privados compartilhados entre entidades aprovadas (ZYSKIND; NATHAN; PENTLAND, 2015).

Pesquisadores da área nutricional mostraram, usando a DLT *Hyperledger Framework* (THE HYPERLEDGER WHITE PAPER WORKING GROUP, 2018), como apoiar pesquisas científicas mapeando a eficiência do gerenciamento de pesquisas de alimentos funcionais (MACHADO; RICCIARDI; BEATRIZ P P OLIVEIRA, 2020). Ainda, composto por profissionais de diferentes áreas, o projeto *Open Blockchain for Science Community* promove o conceito de ciência aberta compartilhando suas experiências de uso da tecnologia Blockchain na ciência (BARTLING, 2018).

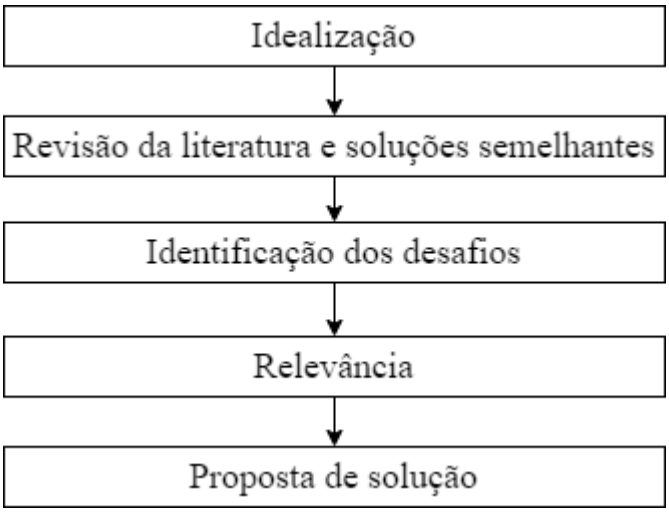
As investigações sobre soluções baseadas em Blockchain para o gerenciamento de pesquisas clínicas estão crescendo rapidamente. O pacote *BlocktTrial* fornece ferramentas úteis para controlar o acesso aos dados de pacientes e pesquisadores de acordo com os requisitos do protocolo de pesquisa (MASLOVE et al., 2018). Além disso, um aplicativo de gerenciamento de consentimento para ensaios clínicos simplificou o compartilhamento de dados para pesquisas científicas (ALBANESE et al., 2020). Finalmente, cientistas desenvolveram um método de gerenciamento de direitos digitais para promover a reutilização de dados de pesquisa, aproveitando o recurso de contratos inteligentes oferecido por algumas plataformas Blockchain (PÃNESCU; MANTA, 2018).

6. MATERIAIS E MÉTODOS

Este capítulo detalha os métodos utilizados no desenvolvimento deste trabalho. O capítulo foi dividido de acordo com os principais objetivos do projeto, nomeadamente, a integração de sistemas de informação em saúde, a disponibilização de um *framework* de segurança e, finalmente, o desenvolvimento de ferramentas para coleta, gerenciamento e compartilhamento de dados de pesquisas científicas.

A pesquisa baseou-se nas necessidades apontadas por profissionais de saúde e pesquisadores, com foco na facilidade de obtenção e visualização de dados de pacientes de TB provenientes de várias origens e nas dificuldades enfrentadas em pesquisas científicas conduzidas nos serviços de TB. O método científico aplicado neste trabalho de pesquisa compreende um conjunto de etapas para reconhecer adequadamente os desafios, questões em aberto e a relevância dos temas para, então, propor soluções capazes de atender às demandas identificadas. Assim, o desenvolvimento dos módulos/componentes de *software* (cada um com sua finalidade específica) baseou-se numa abordagem científica, conforme representado na Figura 4.

Figura 4. Representação do método científico



6.1. Integração de sistemas de informação em saúde de Tuberculose

Nesta seção, são apresentados o modelo de integração e interoperabilidade de dados, a metodologia de análise da necessidade de integração e o procedimento de solicitação de informações e acesso aos SIS e bases de dados governamentais.

6.1.1. Modelo de integração de dados

O modelo de integração considera 6 sistemas de informação e bases de dados em saúde governamentais, conforme descrito na Tabela 3:

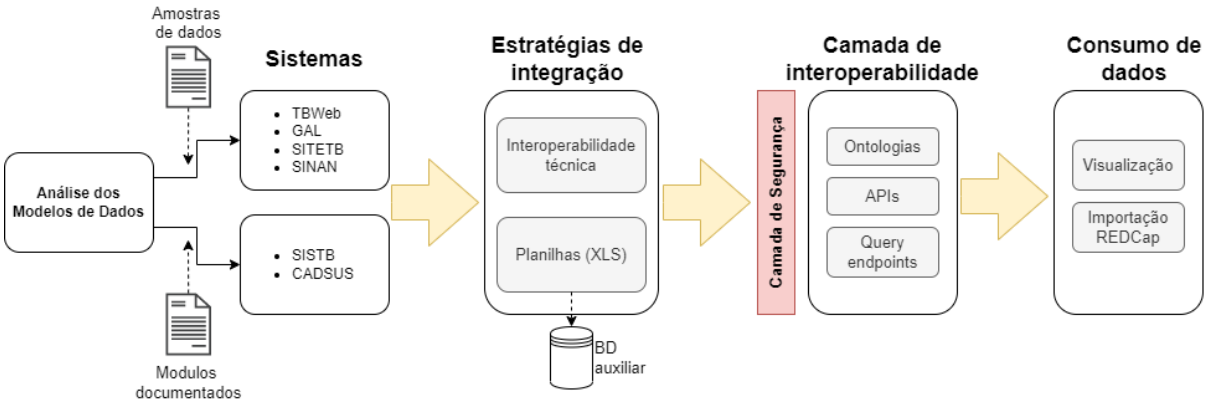
Tabela 3. Sistemas de informação e bases de dados de saúde governamentais.

Sistema	Objetivo	Abrangência	Tipo de arquitetura	APIs ou *Web Services* disponiveís?
SISTB	Informações sobre o paciente, tratamento, exames, internações, TDO, contatos	Regional (Ribeirão Preto - SP e região)	Web	Sim
TBWeb	Vigilância epidemiológica	Estadual (São Paulo)	Web	Sem informação
GAL	Controle de rotinas laboratoriais e requisição e realização de exames	Nacional (segmentado em instâncias estaduais)	Web	Sem informação
SITETB	Notificação e seguimento de paciente com TB resistente a medicamentos	Nacional	Web	Sim
SINAN	Notificação e investigação de casos de doenças e agravos de notificação compulsória	Nacional	Desktop/Web	Sem informação
CADSUS	Base de dados do Cadastro Nacional de Usuários do SUS	Nacional	Web	Sim

A Figura 5 apresenta o modelo de integração. Como há pouca informação disponível sobre as características das bases de dados, conjuntos de dados históricos dos sistemas TBWeb, GAL, SITETB e SINAN foram utilizados para mapear as variáveis disponíveis em cada sistema, o que permitiu a definição dos respectivos dicionários de dados[2]. No caso do SISTB e do CADSUS, os dicionários de dados já estavam documentados pelos administradores dos sistemas.

[2] Disponível em https://redbox.technology/dic.php

Figura 5. Modelo de integração de dados da TB



Os dicionários são essenciais para a integração e agregação de semântica, pois refletem a estrutura de dados de cada sistema. Com base neles, foi estabelecido o modelo de integração capaz de atender a duas formas de integração, a saber: i) através da interoperabilidade técnica entre os sistemas, o que permite a extração de dados programaticamente; ii) através do fornecimento manual de um conjunto de dados (por exemplo, dados exportados e estruturados em planilhas). Tal abordagem foi necessária para contornar os entraves burocráticos enfrentados para se obter informações e autorização para interoperar com os sistemas governamentais (seção 6.1.3).

Apesar de fornecidos em formatos estruturados, os dados provenientes dos sistemas que compõem o modelo de informação não apresentam semântica nativamente agregada, o que dificulta a interpretação e a geração de conhecimento. A anotação semântica, ou seja, o fornecimento explícito de significado aos dados, depende do uso de vocabulários bem definidos, como as ontologias (BAX, 2021), e pode ser realizada, por exemplo, pela inclusão de *tags* em páginas *web* ou pelo uso de *softwares* específicos para criação de bancos de dados em formato semântico (RDF *triplestores*), como o OpenLink Virtuoso (IDEHEN, 2013) e o D2R Server (BIZER; SEABORNE; CYGANIAK, 2004; BIZER; CYGANIAK, 2006; CYGANIAK; BIZER, 2006).

No contexto deste trabalho, foi utilizada uma camada de interoperabilidade semântica idealizada com foco no Programa Nacional de Controle da Tuberculose (PNCT) desenvolvida usando o paradigma da *Web* Semântica. Essa camada fornece os recursos e ferramentas necessários para extração de conhecimento e conteúdo dos sistemas em formatos semânticos, como o RDF, além de disponibilizar uma ontologia capaz de representar os diferentes tipos de dados armazenados sistemas apresentados na Tabela 3 (PELLISON et al., 2017, 2020; PELLISON, 2021).

O modelo de integração também prevê uma ferramenta de visualização, de modo a fornecer um ponto de acesso único e facilitar o acesso aos dados. Essa ferramenta permite a realização de buscas por indivíduos registrados nos vários sistemas através de um conjunto de filtros, como o número do Cadastro de Pessoa Física (CPF) ou do Cartão Nacional de Saúde (CNS), nome, nome da mãe e data de nascimento. O objetivo é proporcionar uma visão rápida (somente leitura) dos dados existentes em cada um dos sistemas. Além disso, como descrito nas seções 7.2.1 e 7.6, também permite realizar a importação de determinadas variáveis para um banco de dados de pesquisa científica.

6.1.2. Justificativa da necessidade de integração

No Brasil, há uma quantidade considerável de dados relacionados à admissão de pacientes com TB, ao acompanhamento do tratamento, à resistência da doença, à mortalidade, entre outros, os quais são continuamente gerados. A assistência a esses pacientes é realizada em diferentes níveis e envolve o uso de várias plataformas que registram dados em bancos de dados não compartilhados. Esse cenário contém barreiras críticas para a recuperação de informações devido à falta de padronização, resultando em dificuldades para produzir *insights* estratégicos, orientações clínicas e relatórios para apoiar os profissionais nos serviços de saúde e na criação de novas políticas públicas de combate à Tuberculose.

Rinner et. al. definiram uma metodologia para averiguar a necessidade de um ambiente integrado para a troca de informações em saúde entre diferentes serviços e especialidades, com base no grau de fragmentação informacional dos dados de saúde de um paciente (RINNER et al., 2016). Segundo os autores, a capacidade de um sistema de saúde proporcionar uma continuidade de cuidado eficaz é inversamente proporcional ao grau de fragmentação informacional, ou seja, quanto maior for a integração de dados de saúde, melhor será o seguimento do paciente.

Com base nesse método e na quantidade considerável de sistemas de informação apresentados na Tabela 3, além de outros sistemas transversais como o eSUS-AB (utilizado na atenção básica) e o Sistema de Informações sobre Mortalidade (SIM), verifica-se no contexto da TB um alto grau de fragmentação informacional, com dados relevantes espalhados e isolados em diversos sistemas.

6.1.3. Acesso aos sistemas de informação e bases de saúde governamentais

A gerência descentralizada dos SIS e bases de dados governamentais, espalhada por diversos grupos de trabalho, muitas vezes isolados, dificulta a integração e o acesso à

informação (CARDOSO; NETO; CHIORO, 2021).

Neste trabalho, alcançou-se sucesso parcial nas tentativas de contato para solicitação de informações e verificação da viabilidade de integração dos sistemas de informação de interesse para a área da TB. As solicitações foram encaminhadas a diferentes grupos do Ministério da Saúde sob a premissa de execução de um projeto de pesquisa em andamento (ver seção 7.6) com financiamento público, o qual prevê atividades de integração de dados no contexto da TB. Entretanto, até o momento da finalização do documento desta tese, os avanços resumiram-se à obtenção de conjuntos de dados históricos, os quais foram utilizados na definição do modelo de integração descrito anteriormente.

Em relação à base de dados do CADSUS, um termo de uso e confidencialidade foi celebrado entre o Departamento de Informática da SUS (DATASUS) e a Faculdade de Medicina de Ribeirão Preto (FMRP) da Universidade de São Paulo (USP), intermediado pelo Laboratório de Informática em Sáude (LIS[3], coordenado pelo Prof. Dr. Domingos Alves). A assinatura do referido termo permitiu o acesso à documentação e credenciais de acesso ao ambiente de homologação e produção (*web services*) desenvolvidos pelo DATASUS, sempre com a garantia de privacidade e confidencialidade. O único servidor/*software* autorizado a interoperar com o CADSUS pertence ao LIS, o qual foi desenvolvido no âmbito deste trabalho.

6.2. *Framework* de segurança

Nesta seção são apresentados os pilares e propriedades da segurança da informação abordados pelo conjunto de mecanismos de segurança e as aplicabilidades do *framework*.

6.2.1. Pilares e propriedades da segurança da informação

O *framework* de segurança proposto tem a intenção de contemplar os pilares e propriedades da segurança da informação descritos na seção 5: confidencialidade, integridade, disponibilidade, autenticidade, autorização, não-repúdio e privacidade.

A confidencialidade é garantida através de uma técnica de criptografia híbrida (combinação de métodos de criptografia simétrica e assimétrica) (MAMUN et al., 2017), enquanto funções *hash* são utilizadas para verificação de integridade. Mecanismos de autenticação por chaves, a definição de níveis de acesso e a assinatura de mensagens salvaguardam a autenticidade, autorização e o não-repúdio. A privacidade é preservada através da desidentificação de dados, ou seja, da remoção de campos sensíveis, previamente mapeados, de um determinado conjunto de dados.

[3] https://lis.fmrp.usp.br/

6.2.2. Aplicabilidades

O *framework* de segurança pode ser aplicado em dois cenários, nomeadamente na:

i) Camada de interoperabilidade semântica: fornecer uma solução abrangente e robusta a ser aplicada na camada de interoperabilidade semântica para o PNCT da TB, desenvolvida por Pelisson et al. (PELLISON et al., 2017, 2020; PELLISON, 2021). A solução foi concebida para ser de fácil implementação por sistemas externos que façam uso da camada de interoperabilidade. O manual de uso do *framework* de segurança está disponível no APÊNDICE A – MANUAL DE UTILIZAÇÃO DO *FRAMEWORK* SEGURANÇA.

ii) Desidentificação de dados de pesquisas científicas: oferecer uma ferramenta para remover dados marcados como sensíveis em um conjunto de dados coletados a partir de sistemas de captura eletrônica de dados. Os dados sensíveis devem ser mapeados, impreterivelmente, na etapa de definição dos instrumentos de coleta.

6.3. Ferramentas para coleta, gerenciamento e compartilhamento de dados de pesquisas científicas de Tuberculose

Nesta seção são apresentados os sistemas de captura eletrônica de dados (EDCs) e a plataforma Blockchain utilizados no projeto e a estratégia de anotação semântica dos dados obtidos através de instrumentos de coleta.

6.3.1. Sistemas de captura eletrônica de dados

Embora apresentem recursos de compartilhamento de dados limitados (METKE-JIMENEZ; HANSEN, 2019), o uso de sistemas eletrônicos para coleta e análise de dados vem crescendo na última década. Os EDCs eliminam os riscos associados aos instrumentos em papel e aprimoram a coleta de dados de alta qualidade necessários para a realização de pesquisas em saúde (PASALIC et al., 2018), e também podem reduzir o custo e o tempo dispendido em atividades de monitoria e gerenciamento de dados (DILLON et al., 2014).

Nesta pesquisa, os EDCs REDCap e KoBoToolbox foram utilizados. Eles foram escolhidos por serem *softwares* livres, estáveis, amplamente utilizados e bem documentados, com APIs disponíveis para integração com outros sistemas.

O REDCap é um *software web* baseado em metadados criado em 2004 por uma equipe da Vanderbilt University para a condução de pesquisas clínicas clássicas e longitudinais, fornecendo aos pesquisadores uma ferramenta para o *design* e desenvolvimento de ferramentas

eletrônicas de captura de dados (HARRIS et al., 2009; WRIGHT, 2016). O REDCap é um *software* gratuito, apesar de não ser considerado de código aberto, o qual pode ser instalado e gerenciado por uma pequena equipe de TI após a obtenção de uma licença de uso (KLIPIN et al., 2014).

Desenvolvido pela *Harvard Humanitarian Initiative*, o KoBoToolbox representa um conjunto de ferramentas gratuito e de código aberto para coleta de dados e análise básica. Foi inicialmente construído para uso em ambientes desafiadores em países em desenvolvimento (HARVARD, 2018). O KoBoToolbox o *software* Enketo como base (SUSTAINABLE ENGINEERING LABORATORY, 2012) e oferece um ambiente para construção e uso de formulários, *online* e *offline*, para uso em qualquer navegador moderno, graças aos recursos do HTML5. O *software* utiliza o padrão XLSForm, que simplifica a autoria de formulários em planilhas em formato legível por humanos (SUSTAINABLE ENGINEERING LABORATORY; ODK TEAM,). Como ferramenta de auxílio, um construtor de formulários intuitivo está disponível ou, ainda, os formulários podem ser criados em arquivos Excel XLS e posteriormente importados.

O sistema REDCap é amplamente utilizado pela comunidade científica para coletar e gerenciar dados de pesquisa, permitindo que pesquisadores conduzam seus estudos por conta própria. No entanto, o *software* pode apresentar alguns problemas de usabilidade durante a coleta de dados, como *interface* gráfica poluída, degradação gradual do desempenho e falta de operação *offline* sem depender de um aplicativo móvel. Por outro lado, apesar de apresentar funcionalidades mais básicas, o KoBoToolbox oferece estilos modernos e permite que os usuários trabalhem *offline* diretamente do navegador *web*, sem a necessidade de instalação de aplicativos auxiliares.

A escolha do REDCap e do KoBoToolbox como sistemas de coleta eletrônica de dados justifica-se pela ausência de custos financeiros para aquisição de licenças de uso, pela ampla aceitação pela comunidade científica (no caso do REDCap, o *software* é produto de um projeto de pesquisa, com publicações relacionadas) e pela existência de recursos computacionais que permitem a interoperabilidade desses sistemas com outras plataformas (característica essencial no contexto deste trabalho).

6.3.2. Plataforma Blockchain

A plataforma de Blockchain Kaleido foi utilizada neste trabalho de pesquisa. Kaleido foi desenvolvido como uma plataforma Blockchain como Serviço (do inglês, *Blockchain as a*

Service - BaaS) que faz uso de recursos de computação em nuvem para a implementação de redes Blockchain privadas (CONSENSYS, 2018). A plataforma tem uma operação simples para permitir que qualquer entidade use a tecnologia Blockchain para melhorar o desempenho, a segurança, a globalização e a escalabilidade de seu modelo de negócios e o alcance de seus dados (REBELO, 2019).

Kaleido utiliza o protocolo Ethereum como base, um *software* distribuído de código aberto para operacionalizar redes Blockchain. Ethereum é utilizado em todo o mundo para implantar rapidamente redes blockchain privadas flexíveis com alto rendimento de transações e baixo custo (WOOD, 2014).

6.3.3. Anotação semântica

Em pesquisas em saúde, a anotação semântica pode ajudar a descrever os dados que estão sendo coletados. Normalmente, cada estudo conta com diversos instrumentos de coleta, totalizando centenas de campos a serem preenchidos durante o andamento da pesquisa e, posteriormente, a anotação pode ser útil para extrair e vincular diferentes conjuntos de dados de pesquisa descritos por um vocabulário (SINGHAL; SRIVASTAVA, 2014).

Para reforçar a representação dos dados coletados, os campos em formulários de pesquisa podem ser anotados com vocabulários semânticos. O REDCap oferece a possibilidade de incluir anotações para cada campo, as quais não são exibidas nos formulários, mas ficam disponíveis para o *designer* e nas exportações de dados para aumentar a interpretabilidade dos dados (WRIGHT, 2016). Essa anotação pode ser realizada através de uma ontologia ou de uma *tag* da *Humanitarian Exchange Language* (HXL), dependendo da preferência do usuário. HXL é um projeto do Escritório das Nações Unidas para a Coordenação de Assuntos Humanitários para a coordenação da resposta a desastres, com tecnologias da *web* semântica. Utiliza marcação simples por meio de *hashtags* e tem como objetivo contribuir para a automatização de processos para melhorar o fluxo de informações para os tomadores de decisão (KESSLER; HENDRIX, 2015). O KoBoToolbox suporta nativamente o uso de HXL. Ao criar um XLSForm, o usuário deve simplesmente inserir uma coluna extra na planilha e preenchê-la com *hashtags* HXL identificando o tipo de informação em cada coluna.

Para tornar a anotação semântica mais flexível, uma forma adicional foi implementada, representando uma alternativa para a anotação direta nos EDCs. Além disso, o KoBoToolbox não oferece suporte a anotações que não sejam as *tags* HXL. Nesse caso, a anotação pode ser realizada em forma de metadados, os quais são utilizados para mapear os campos de instrumentos a determinadas terminologias/ontologias. Assim, os dados podem ser anotados

sempre que necessário e utilizados pelo usuário da maneira desejada.

É importante ressaltar que o mapeamento, feito diretamente nos formulários ou por meio de metadados, deve ser realizado durante a fase de definição dos instrumentos de coleta da pesquisa, em paralelo com a seleção dos campos sensíveis (identificadores e quasi-identificadores).

7. RESULTADOS

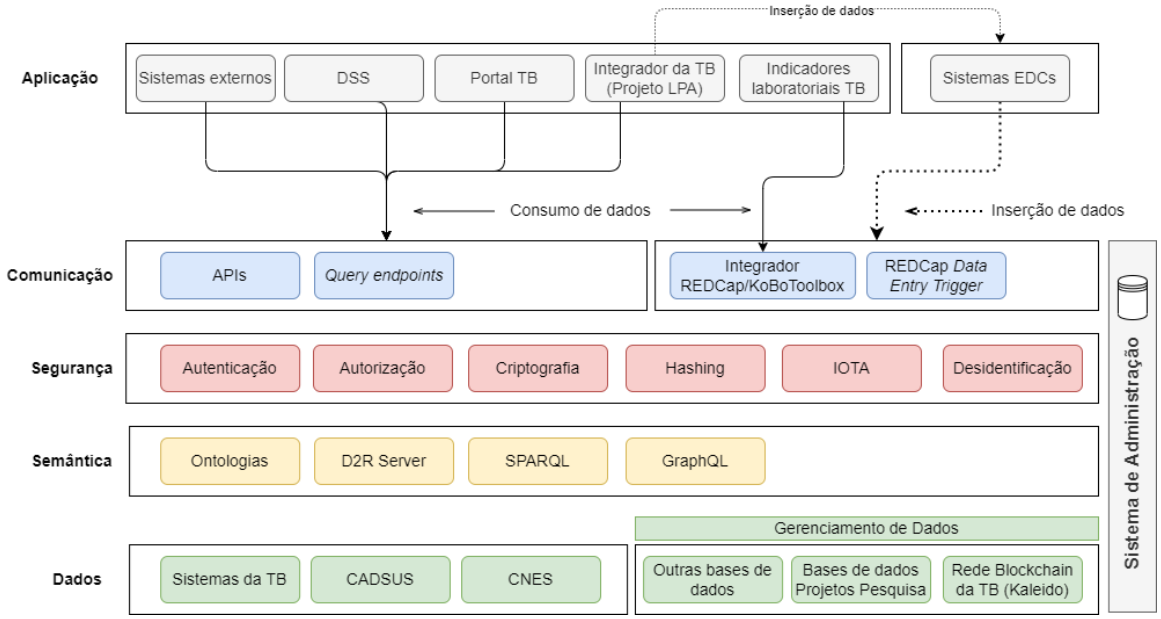
7.1. Arquitetura de *software*

A arquitetura de software abrange a organização de um sistema e quais componentes são responsáveis por realizar um conjunto específico de funções. (GARLAN, 1999; CLEMENTS et al., 2003). Uma arquitetura definida em camadas permite distribuir a complexidade de um sistema em componentes menores, facilitando sua compreensão, desenvolvimento e manutenção (SAVOLAINEN; MYLLARNIEMI, 2009).

No contexto deste trabalho, considerando as abordagens de integração de dados da TB e a coleta e gerenciamento de dados de pesquisas, definiu-se uma arquitetura interconectada composta por camadas (Figura 6), nomeadamente:

- Camada de Dados: representação e armazenamento dos dados;
- Camada Semântica: recursos para agregação de valor semântico e recuperação de dados;
- Camada de Segurança: mecanismos de segurança para proteção de dados sensíveis;
- Camada de Comunicação: pontos de acesso variados para interoperabilidade com sistemas externos; *middlewares* para os sistemas EDCs e interação com os dados de pesquisa;
- Camada de Aplicação: representação dos sistemas externos aptos ao consumo de dados.

Figura 6. Arquitetura de software

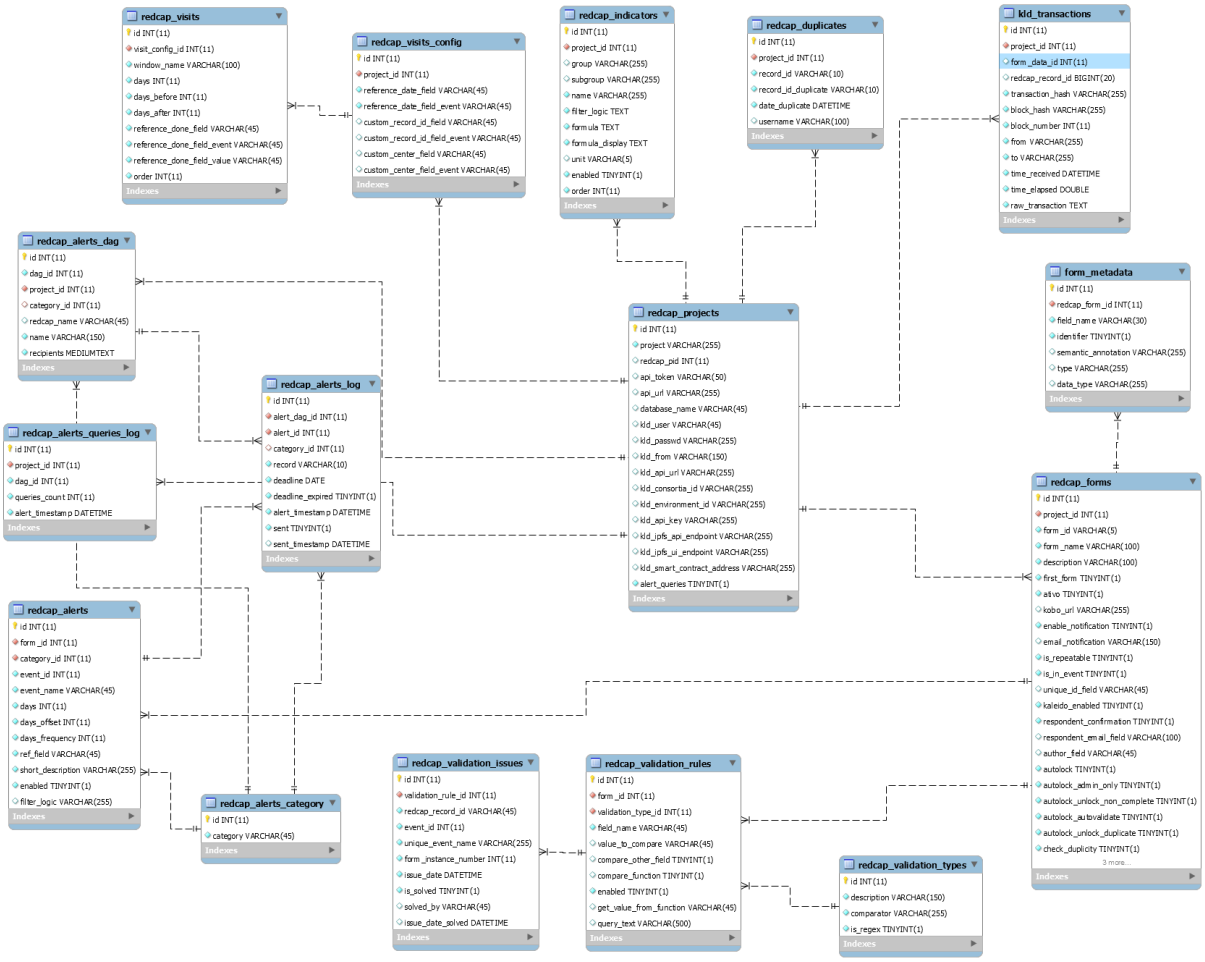


Sistema de Administração da Arquitetura

Os parâmetros de configuração da arquitetura podem ser gerenciados através de um

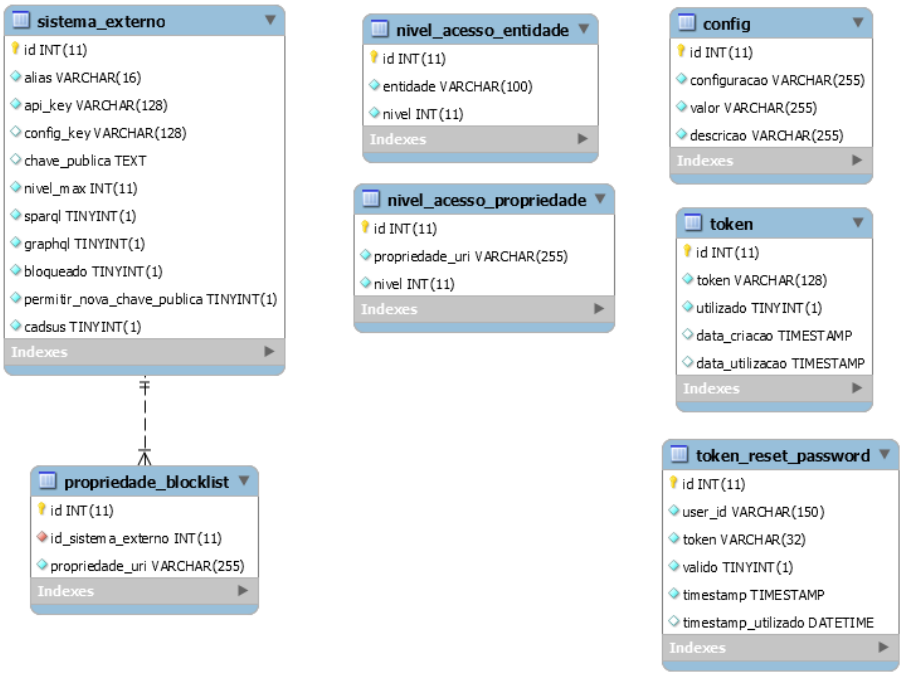
sistema *web*[4] que centraliza todos os metadados necessários para a operação dos módulos de Interoperabilidade e Integração (seção 7.2), Segurança (seção 7.3) e Gerenciamento de Dados de Pesquisas Científicas (seção 7.4), além das credenciais e pontos de acesso da Rede Blockchain da TB (seção 7.5). Todos os metadados são armazenados em um banco de dados relacional (MySQL), conforme os modelos entidade-relacionamento apresentados na Figura 7 (tabelas relacionadas ao Módulo de Gerenciamento de Dados, à integração REDCap/KoBoToolbox e à Rede Blockchain da TB) e na Figura 8 (tabelas relacionadas ao Módulo de Segurança).

Figura 7. Modelo Entidade-Relacionamento – Módulo de Gerenciamento de Dados, integração REDCap/KoBoToolbox e Rede Blockchain da TB



[4] Disponível em https://admin.redbox.technology:3007/redbox-admin

Figura 8. Modelo Entidade-Relacionamento – Módulo de Segurança



A Figura 9 exibe uma das *interfaces* do sistema de administração, a qual permite a gestão de formulários e diversos outros parâmetros relacionados às atividades de coleta e gerenciamento de dados de pesquisa.

Figura 9. *Interface* de configuração dos metadados de um projeto do REDCap.

Dados do Projeto - LPA

Detalhes Formulários Validação de campos Calendário de eventos/visitas Alertas Indicadores Uploads Dados submetidos Blockchain Explorer

Id	Nome	Nome interno REDCap	Nº Formulário	Primeiro formulário?	Ativo?
0	Dados Demográficos	dados_pessoais	0	false	true
0	1 Recrutamento	recrutamento	0	false	true
0	2A Confirmação de Elegibilidade - Dad...	a_confirmao_de_elegibilidade_dados_p...	0	false	true
0	2B Confirmação de Elegibilidade - Dad...	b_confirmao_de_elegibilidade_dados_s...	0	false	true
0	FALHA DE RECRUTAMENTO	falha_de_recrutamento	0	false	true
0	FALHA DE RECRUTAMENTO - LAB	falha_de_recrutamento_lab	0	false	true
0	3 Inclusão	incluso	0	false	true
0	4A Inclusão - Revisão de Prontuário	a_incluso_reviso_de_pronturio	0	false	true
0	4B Inclusão - Revisão de Banco de Dad...	b_incluso_reviso_de_banco_de_dados	0	false	true
0	5 Resultados Laboratório	resultados_laboratrio	0	false	true
0	6 Revisão de Prontuário	reviso_de_pronturio	0	false	true
0	7 Visita Final	visita_final	0	false	true
0	LPA1 e LPA2 - Cultura	lpa1_e_lpa2_cultura	0	false	true

7.2. Módulo de Interoperabilidade e Integração

Este módulo é constituído por dois componentes independentes, os quais foram desenvolvidos com dois focos distintos: os sistemas da TB e os sistemas EDC.

O primeiro refere-se à disponibilidade de um conjunto de recursos computacionais para a integração dos diferentes sistemas utilizados no contexto da TB para, posteriormente, permitir a visualização dos dados através de um ponto de acesso unificado e centrado no paciente. O

segundo funciona como uma *interface* para os *softwares* REDCap e KoBoToolbox, permitindo a comunicação entre os dois sistemas, a integração com os outros módulos da arquitetura e ferramentas externas, e o fornecimento de funcionalidades adicionais capazes de aprimorar a experiência de coleta e gerenciamento de dados de pesquisa.

7.2.1. Integração dos sistemas da TB

Este componente visa suportar a troca de informações e a integração dos sistemas apresentados na Tabela 3. A seguir, são apresentados os recursos disponíveis e implementados pelo Módulo de Interoperabilidade de Integração.

Integração com o Cadastro Nacional de Usuários do SUS (CADSUS) e com o Cadastro Nacional de Estabelecimentos de Saúde (CNES)

O Ministério da Saúde, por meio do DATASUS, disponibiliza diversos *web services*[5] de uso gratuito para interoperabilidade com sistemas externos. O CADSUS e o CNES podem ser utilizados para consulta dos dados demográficos dos usuários do SUS e das informações dos estabelecimentos de saúde do Brasil, respectivamente. Outros serviços só estão disponíveis para órgãos do governo.

Foi desenvolvida, então, uma *interface* capaz de consumir dados desses serviços e disponibilizá-los para uso pelo SISTB e sistemas correlacionados. O acesso ao CADSUS requer uma autorização específica, a qual foi solicitada e concedida mediante termo de uso e confidencialidade, conforme descrito na seção 6.1.3. O DATASUS disponibiliza manuais e exemplos de uso do CADSUS na sua página *web*[6]. Em relação ao CNES, até o momento, não é necessária autorização de acesso, visto que são dados públicos. O LIS da FMRP-USP mantém o servidor autorizado a comunicar com os serviços do DATASUS.

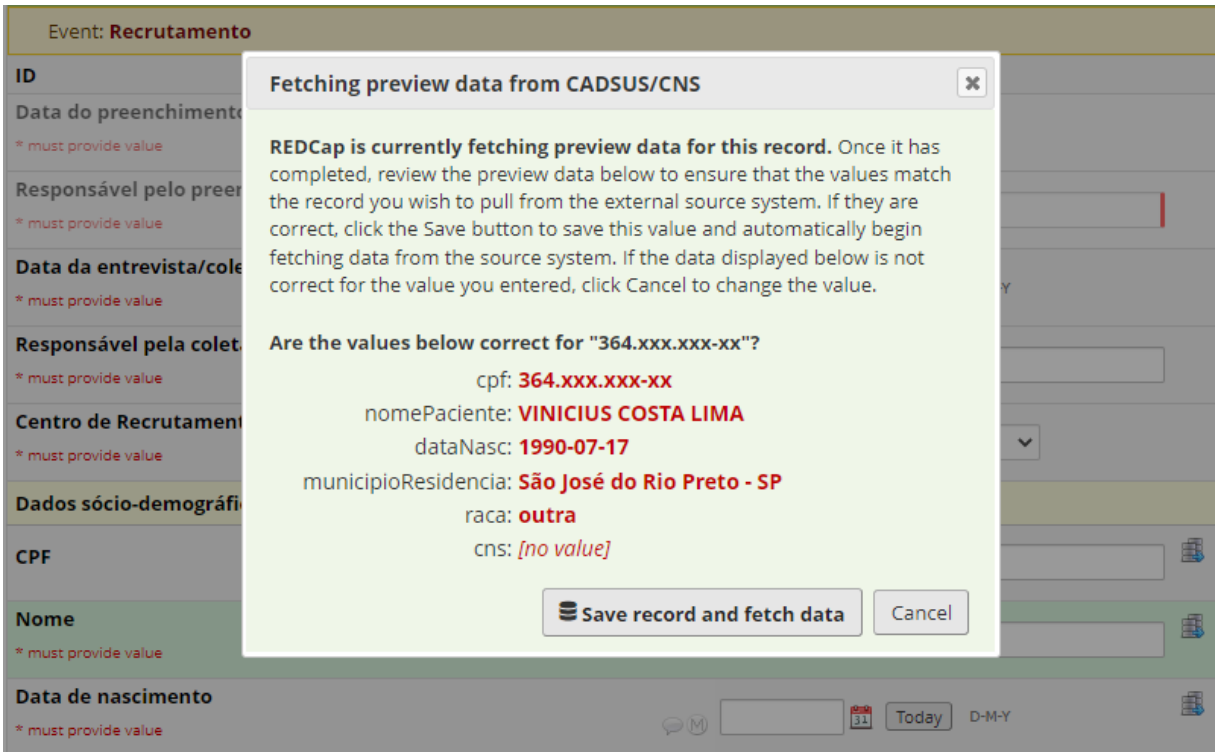
Para consumir dados do CADSUS e CNES, a *interface* recebe os dados de entrada, como o CPF e CNS para o CADSUS e o CNPJ para o CNES (filtros), e envia requisições para os servidores do DATASUS. Os dados recebidos são formatados em JSON e disponibilizados para uso pelo SISTB.

O REDCap também foi integrado ao CADSUS através do módulo *Dynamic Data Pull* (DDP) do *software* e a *interface* disponibilizada pelo LIS. Esse módulo permite o consumo de dados de sistemas externos e o seu uso nos instrumentos de coleta de uma determinada pesquisa, o que permite a obtenção de informações atualizadas e reduz a carga de trabalho do coletador

[5] Catálogo disponível em https://servicos-datasus.saude.gov.br/
[6] Disponível em https://servicos-datasus.saude.gov.br/detalhe/tgKoKpju8s

de dados. A Figura 10 apresenta a integração do CADSUS com o REDCap.

Figura 10. Consumo de dados do CADSUS através do REDCap



Anotação semântica dos dados

Os dados da TB podem receber marcação semântica de forma manual ou automatizada. No caso do SISTB, as páginas *web* do sistema foram marcadas manualmente para uma maior precisão, utilizando a ontologia proposta por (PELLISON et al., 2020). Os demais sistemas da TB são anotados de forma automatizada através do D2R Server, que aplica as ontologias e disponibiliza os dados no formato RDF, além de processar consultas SPARQL. A Figura 11 apresenta um exemplo de marcação manual do SISTB.

Figura 11. Marcação manual do SISTB

```
<div itemscope itemtype="http://sistb-dev.ddns.net/ontology/Paciente">
    Nome paciente: <span itemprop="nomePaciente">PACIENTE NOVE</span>
    CPF: <span itemprop="cpf"></span>
    Cartão Nacional de Saúde: <span itemprop="cartaoSus">80000270</span>
    Número do Prontuário: <span itemprop="nroProntuario">10304561</span>
    Número Hygia: 270
    Data de Nascimento: <span itemprop="dataNascimento">06/04/1990</span>
    Idade: <span itemprop="idade">27</span>
    Sexo: <span itemprop="genero">Feminino</span>
    Gestante: <span itemprop="gestante">Não</span>
    Nome da Mãe: <span itemprop="nomeMae">MÃE NOVE</span>
    Etnia: <span itemprop="etnia">Branco</span>
    Naturalidade: <span itemprop="naturalidade"></span>
    Escolaridade: <span itemprop="escolaridade">0 ano(s) concluído(s)</span>
    Tipo de Ocupação: <span itemprop="ocupacao">Outra</span>
    Outro tipo de ocupação: <span itemprop="ocupacao"></span>
    Observações: <span itemprop="observacoesPaciente"></span>
    Número SINAN: <span itemprop="nroSinan">9</span>
</div>
```

O D2R requer acesso a um banco de dados relacional para conversão dos dados para o formato RDF. Obviamente, por questões de segurança, tais bases de dados não têm o acesso permitido. Portanto, é necessário armazenar os dados obtidos dos sistemas da TB (via APIs ou planilhas de dados) em um banco intermediário, mesmo que temporariamente. A Figura 12 apresenta a marcação realizada através do D2R Server.

Figura 12. Marcação automatizada através do D2R Server

paciente #100

Resource URI: http://sistb-dev.ddns.net/d2rq-dev/resource/paciente/100

Home | All paciente

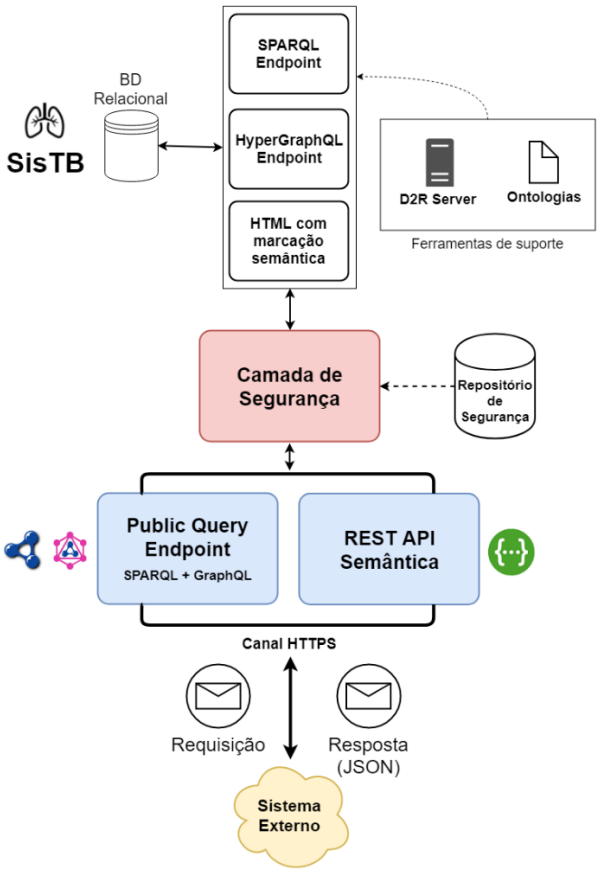
Property	Value
vocabPaciente:cartao_nacional_saude	80000100
vocabPaciente:cep	
vocabPaciente:cidade	RIBEIRAO PRETO (pt-br)
vocabPaciente:cod_paciente	100 (xsd:integer)
vocabPaciente:cpf	
vocabPaciente:data_nascimento	1997-07-26 (xsd:date)
vocabPaciente:endereco	Rua 100 (pt-br)
vocabPaciente:escolaridade	0 (xsd:integer)
vocabPaciente:estado	SP (pt-br)
vocabPaciente:etnia	Pardo (pt-br)
vocabPaciente:gestante	N
vocabPaciente:idade	20 (xsd:integer)
rdfs:label	paciente #100
vocabPaciente:naturalidade	(pt-br)
vocabPaciente:nome	PACIENTE TREZE (pt-br)
vocabPaciente:nome_mae	NULL (pt-br)
vocabPaciente:nro_hygia	100 (xsd:integer)
vocabPaciente:nro_prontuario	8.98001E+14
vocabPaciente:nro_sinan	6032378 (xsd:int)
vocabPaciente:ocupacao	(pt-br)
vocabPaciente:paciente_observacoes	(pt-br)

Disponibilização de APIs e query endpoints para o SISTB

O SISTB conta com uma API e um *query endpoint* capazes de produzir respostas em formatos semânticos, como RDF e JSON-LD e permitem o consumo dos dados com anotação semântica, os quais são extraídos das páginas *web* do SISTB ou do *endpoint* SPARQL do D2R Server, incluindo os dados dos outros sistemas da TB, quando disponíveis. O *public query endpoint* permite o processamento de consultas SPARQL e GraphQL (projeto HyperGraphQL - https://www.hypergraphql.org). Esses recursos são protegidos pelo Módulo de Segurança (seção 7.3.1). A Figura 13 apresenta a arquitetura de interoperabilidade do SISTB.

Uma documentação da API está disponível[7] para que os usuários avaliem os modelos de requisição e resposta e testem os *endpoints*. O modelo de requisição inclui uma chave de API, uma assinatura, o formato de dados desejado e outros parâmetros que variam entre os *endpoints*. O modelo de resposta inclui um código HTTP, uma mensagem informativa, a chave simétrica criptografada e os dados criptografados. Nesta *interface* de testes, todos os resultados são apresentados criptografados devido à confidencialidade dos dados.

Figura 13. Visão macro da arquitetura de interoperabilidade do SISTB



[7] Disponível em: https://api.tuberculosis.cloud/tb-api-dev/apiDocs

Integração com o SINAN, SITETB, GAL e TBWeb

Conforme descrito nas seções 6.1.2 e 6.1.3, os sistemas governamentais da TB são gerenciados por diferentes grupos ligados, em geral, ao Ministério da Saúde, com baixa cooperação, o que gera uma fragmentação a nível informacional. Soma-se, ainda, as dificuldades de contato e a obtenção de acesso às *interfaces* de programação dos sistemas, o que limita as operações de interoperabilidade desejadas.

Por esse motivo, uma abordagem alternativa foi adotada para permitir o uso e a integração dos dados de TB armazenados nos sistemas SINAN, SITETB, GAL e TBWeb. Em determinadas situações e com as devidas justificativas, como é o caso do projeto apresentado na seção 7.6, é possível solicitar acesso à informação desses sistemas para uso específico. Tratam-se de dados estáticos estruturados em planilhas, mas que ainda possibilitam a inferência do modelo de dados e a extração e o processamento das informações.

Uma *interface web* foi desenvolvida para permitir o carregamento e o processamento dessas planilhas. Uma vez enviadas ao servidor, os dados são extraídos e armazenados em uma base de dados temporária, com base nos dicionários de dados. A partir desse ponto, tais dados podem receber marcação semântica, ser integrados e visualizados.

Em paralelo, como já se conhece os modelos de dados, componentes de *software* foram preparados para uma eventual conexão, via *web services* ou APIs, com os sistemas da TB de interesse. Isso significa que o avanço para uma ferramenta capaz de obter dados atualizados em tempo real depende da autorização das entidades que gerenciam esses sistemas e da disponibilização das credenciais de acesso e manuais de utilização dos recursos de interoperabilidade.

Visualização dos dados

O processo de integração dos dados requer, preferencialmente, a disponibilização de uma camada de visualização para os dados obtidos nos diversos sistemas, de forma a fornecer um ponto de acesso unificado e centrado no paciente. Dessa forma, profissionais de saúde, gestores e pesquisadores podem consultar a informação de um determinado paciente de forma centralizada, sem a necessidade de aceder a simultaneamente vários sistemas com diferentes credenciais de acesso.

Uma ferramenta de visualização foi disponibilizada no contexto de um projeto de pesquisa que prevê a integração dos sistemas da TB e o uso dos dados em suas análises de desfechos. Denominado “Integrador da TB”, tal contribuição é descrita na seção 7.6, a qual está

disponível para uso pelos pesquisadores para consulta e importação de dados para a base de dados do projeto no REDCap.

7.2.2. Coleta e gerenciamento de dados de pesquisa de TB

Este componente tem como objetivos permitir a transmissão de dados do *software* KoBoToolbox para o REDCap, habilitar a integração do REDCap com ferramentas externas e oferecer funcionalidades úteis para agregar valor semântico aos dados, de forma a aprimorar a experiência de coleta e gerenciamento de dados em um projeto de pesquisa.

Integração do REDCap com o KoBoToolbox

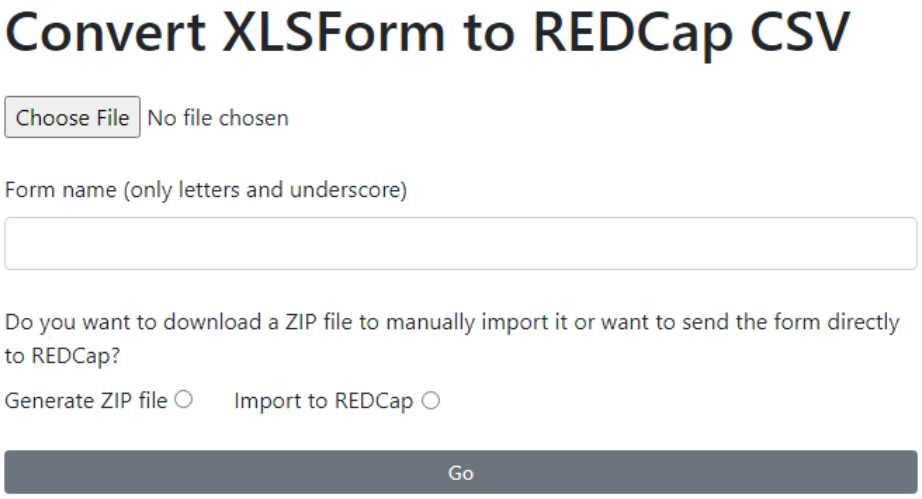
A utilização do REDCap e do KoBoToolbox, em conjunto, ajuda a mitigar os problemas de usabilidade e/ou a ausência de determinadas características de cada ferramenta. Entretanto, faz-se necessária a comunicação entre os dois *softwares*, de modo que seja transparente para o usuário final.

A integração das ferramentas parte do pressuposto que o gerenciamento de dados e o acompanhamento da evolução do banco de dados é sempre realizado pelo REDCap, podendo a coleta ser feita pelo KoBoToolbox, devido à sua capacidade de coleta *offline* (sem a necessidade de instalação de aplicativos adicionais) e melhor usabilidade. Portanto, para efetivar a integração, é preciso garantir que um mesmo instrumento de coleta exista, simultaneamente, nos dois *softwares* e que o formato de dados utilizado pelo KoBoToolbox seja aceito pelo REDCap. Nesse sentido, foram desenvolvidas duas ferramentas: i) um conversor de formulários do padrão XLSForm do KoBoToolbox para o formato *Comma-Separated Value* (CSV) do REDCap; e ii) um integrador (*middleware*) para coordenar o envio de dados do KoBoToolbox para o REDCap.

Conversor de formulários[8]. Considerando que os instrumentos são construídos com padrões distintos em cada *software*, é desejável um conversor para que o usuário não precise criar o mesmo formulário duas vezes. Portanto, os formulários do REDCap podem ser criados automaticamente através da conversão dos formulários desenvolvidos no padrão XLSForm. A Figura 14 apresenta a *interface web* do conversor.

[8] Disponível em https://redbox.technology/form_converter/ (acesso público)

Figura 14. Conversor de formulários



O conversor suporta todos os tipos de campos comuns, como: texto, data, data e hora, hora, inteiro, decimal, cálculo, seleção única, seleção múltipla, arquivos e notas. O nome da variável e os valores atribuídos às opções em seleções simples e múltiplas, por exemplo, são mantidos, de modo que os instrumentos tenham estrutura correspondente em ambos os sistemas. A lógica de salto (do inglês, *skip logic*) definida no KoBoToolbox é traduzida para a lógica de ramificação (do inglês, *branching logic*) do REDCap, bem como as regras de validação (limiares mínimos e máximos).

No processo de construção do formulário no padrão XLSForm, há uma particularidade relacionada às questões de seleção múltipla (*checkboxes*). Esse tipo de pergunta precisa ter o nome do campo começando com 'checkbox_'. Isso é necessário para garantir uma identificação correta de uma estrutura de perguntas de seleção múltipla durante a transferência de dados do KoBoToolbox para o REDCap. Antes de iniciar o processo de conversão, a convenção de nomenclatura será pré-verificada pelo módulo conversor. Se alguma inconsistência for detectada, a conversão falhará e o usuário será informado com o erro detectado.

Integrador. Para permitir o envio de dados, os serviços REST nativos do KoBoToolbox devem ser habilitados nas configurações do formulário para enviar instantaneamente os dados coletados para o integrador por meio de uma solicitação POST. A URL[9] do integrador e as credenciais básicas de autenticação HTTP devem ser fornecidas.

O integrador recebe os dados coletados no KoBoToolbox como um objeto JSON, que é analisado para remover elementos desnecessários não relacionados aos dados de interesse (em geral, metadados não utilizados). É realizada uma verificação para identificar se os dados se referem a um novo registro ou a um já existente no REDCap, de modo a preparar os dados

[9] https://redbox.technology/kobo_redcap.php

adequadamente para transmissão e armazenamento. Em seguida, uma solicitação é enviada à API do REDCap para importar os dados.

O integrador também pode ser configurado para armazenar os dados coletados no KoBoToolbox em outras bases de dados diferentes do REDCap. Como exemplo, dados referentes à pandemia de COVID-19, como o número de casos e óbitos, foram coletados através de formulários eletrônicos do KoBoToolbox por pesquisadores e voluntários do Portal COVID-19 Brasil[10].

REDCap Data Entry Trigger

O REDCap oferece um recurso para monitorar os dados armazenados no sistema. Trata-se de um *trigger* (gatilho) que é disparado sempre que ocorre a inserção ou atualização de dados, o que permite executar operações específicas e complementares que podem aprimorar o gerenciamento de dados da pesquisa.

Assim, vários processamentos podem ocorrer dependendo das configurações definidas para um determinado instrumento, como: envio de notificações por *e-mail* (tanto para o respondente quanto para a equipe de pesquisa), verificação de duplicidade de registros, bloqueio instantâneo do registro (para evitar alterações nos dados), autovalidação, entre outros. Na prática, o *trigger* comunica-se com o Módulo de Gerenciamento de Dados (seção 7.4) para a execução de etapas adicionais.

Além disso, a comunicação com a rede Blockchain de compartilhamento de dados da TB (seção 7.5) ocorre através deste recurso. Nesse caso, os dados recebem tratamentos específicos, como a desidentificação e a agregação de semântica (criação de um objeto JSON-LD complementar).

Serviço de Ontologias

O Serviço de Ontologias fornece ferramentas práticas para potencializar o uso das ontologias e permitir a integração contínua de diferentes fontes de dados, capaz de se adaptar à evolução das ontologias, garantir a disponibilidade e evitar a perda de dados.

Uma ontologia pode ser convertida em um formulário de coleta, onde cada propriedade é mapeada como um campo de formulário. O nome e o tipo de um campo são obtidos a partir do nome da propriedade e do tipo associado (texto é o tipo padrão). Os valores mínimo e máximo definidos como restrições nas propriedades da ontologia também são convertidos.

[10] https://ciis.fmrp.usp.br/covid19/

De forma semelhante, este serviço possibilita a criação de uma ontologia baseada em um formulário (processo inverso). Esta funcionalidade conta com uma aplicação externa, nomeadamente o D2R Server, para converter conteúdos relacionais em semânticos, permitindo uma rápida conversão entre estes formatos através da criação automática de ontologias baseadas no *schema* relacional. Assim, é possível aproveitar esse recurso para definir uma ontologia a partir de um instrumento de coleta de dados. Para isso, é criada uma tabela temporária em um banco de dados relacional, onde cada coluna representa um campo do instrumento. Em seguida, o software D2R gera uma ontologia utilizando a estrutura da tabela, ou seja, convertendo colunas em propriedades, que podem ser posteriormente customizadas.

Essas funcionalidades suportam formulários no formato CSV (REDCap) ou XLSForm (KoBoToolbox). O carregamento dos ficheiros necessários (formulário ou ontologia) é realizado através de uma interface semelhante àquela apresentada na Figura 14. A Tabela 4 apresenta um exemplo de uma ontologia gerada a partir de um instrumento contendo os dados de tratamento do paciente de TB.

Tabela 4. Exemplo de correspondência entre instrumentos de coleta e ontologias

Instrument		**Ontology**	
		TB treatment	
Field	**Type**	**Property**	**Range**
Start date	textbox with date validation	http://vocab.redbox.technology/vocab /treatment/start_date	Literal (date)
TB clinical form	Multiple choice with single answer	http://vocab.redbox.technology /vocab/treatment/clinical_form	Literal
Discharge date	textbox with date validation	http://vocab.redbox.technology /vocab/treatment/discharg_date	Literal (date)

7.3. Módulo de Segurança

7.3.1. Mecanismos de segurança aplicado à arquitetura de interoperabilidade semântica do SISTB[11]

Uma questão fundamental relativa à interoperabilidade em sistemas de informação de saúde é a segurança de dados. Nesse sentido, motivado pela arquitetura de interoperabilidade

[11] A informação contida nesta subseção refere-se parcialmente ao conteúdo do artigo *Security Framework for Tuberculosis Health Data Interoperability Through the Semantic Web*, publicado no periódico *International Journal of Web Portals (IJWP)*. 2021. ISSN 1938-0194. Disponível em http://doi.org/10.4018/IJWP.2021070103.

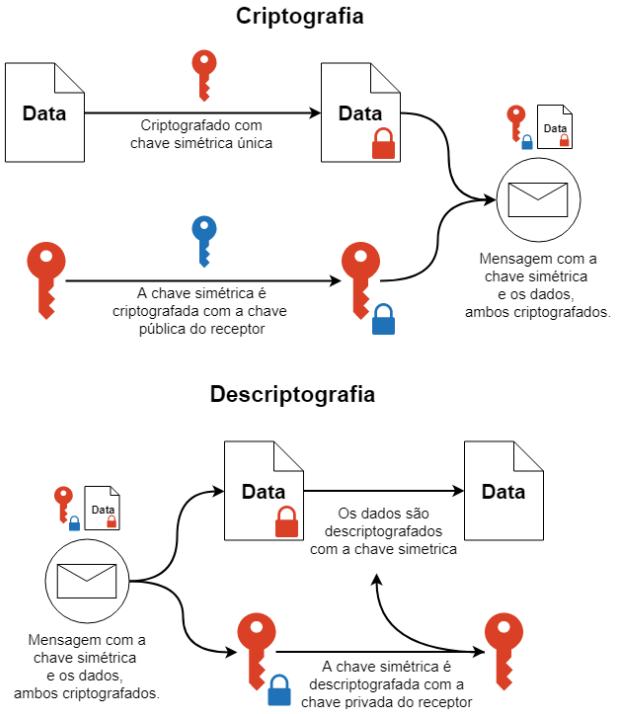
semântica apresentada por Pellison et. al. (PELLISON et al., 2017, 2020), desenvolveu-se uma camada de segurança robusta para proteger dados sensíveis compartilhados pela Web Semântica.

Padrões para uma Web Semântica segura foram descritos, como a integração segura de e o conceito de confiança (do inglês, *trust*) (THURAISINGHAM, 2005). Algoritmos de segurança podem ser desenvolvidos para integrar bancos de dados heterogêneos espalhados pela *web* com segurança. Apesar dessas abordagens, afirma-se que os mecanismos de segurança devem ser aplicados a todas as camadas na pilha da Web Semântica e não apenas nas camadas relacionadas à segurança. Com isso, várias barreiras de segurança coexistiriam e se complementariam para manter os dados seguros.

Criptografia Híbrida

A criptografia simétrica protege a informação de maneira eficiente, mas apresenta o problema de distribuição de chaves. Embora a criptografia assimétrica lide com esse problema, ela só é capaz de cifrar uma quantidade limitada de dados. Assim, uma abordagem híbrida, i.e., uma combinação de criptografia simétrica e assimétrica (MAMUN et al., 2017), foi projetada e aplicada à camada de segurança. Assim, é possível resolver, ao mesmo tempo, o problema de distribuição de chaves e a criptografia de grandes quantidades de dados. A Figura 15 mostra como a abordagem híbrida pode ser usada para criptografar e descriptografar dados.

Figura 15. Processo de criptografia e descriptografia, utilizando Criptografia Híbrida



Para iniciar o processo de criptografia, uma chave simétrica aleatória é gerada. Em seguida, os dados são criptografados com essa chave usando o algoritmo *Advanced Encryption Standard* (AES). Depois disso, a chave simétrica também é criptografada, mas usando a chave pública do receptor por meio do algoritmo *Rivest–Shamir–Adleman* (RSA). Os dados e a chave criptografados são empacotados na mesma mensagem de resposta e enviados pelo canal de comunicação. Somente o receptor poderá obter a chave secreta e, depois, os dados solicitados. No processo de descriptografia, o receptor começará descriptografando a chave secreta com sua chave privada (do mesmo par). Em seguida, a chave secreta deve ser usada para descriptografar os dados.

Ferramentas de desenvolvimento

As ferramentas foram escolhidas considerando principalmente sua maturidade e compatibilidade com os recursos da Web Semântica. Todas as ferramentas selecionadas são de código aberto e/ou gratuitas.

O *framework* de segurança, como a camada de interoperabilidade (API semântica), foi desenvolvido utilizando Java EE, versão 8, devido à existência de ferramentas que auxiliam na criação de aplicações semânticas, como as bibliotecas Apache Jena (https://jena.apache.org/) e Apache Any23 (https://any23.apache.org/). O primeiro fornece todas as interfaces, classes e métodos necessários para trabalhar com dados em RDF e outros formatos semânticos, como o JSON-LD. O segundo é usado para extrair dados de páginas HTML semanticamente marcadas e transformá-las em formato RDF. Finalmente, para armazenar metadados de segurança, o banco de dados relacional MySQL foi selecionado. Este repositório armazena todos os metadados para suportar mecanismos de autenticação e autorização, como, por exemplo, as informações sobre as chaves e os níveis de acesso atribuídos aos dados e aos sistemas externos.

Camada de Segurança para a Arquitetura de Interoperabilidade da TB

A arquitetura de interoperabilidade funcional e semântica permite o compartilhamento de dados de saúde de TB. Vários *endpoints* estão disponíveis para recuperar informações relevantes. Há outros recursos incluídos, como um *middleware* para permitir acesso semântico aos bancos de dados relacionais (D2R Server), um *endpoint* SPARQL, um *endpoint* GraphQL para dados semânticos, um conjunto de ontologias e páginas HTML com marcação semântica, conforme descrito na seção 7.2.1.

As contribuições deste trabalho para a arquitetura de interoperabilidade incluem uma camada de segurança e um *endpoint* de consultas (*Public Query Endpoint*, responsável pelo processamento de *queries* SPARQL e GraphQL) (LIMA et al., 2021b).

Gerenciamento de chaves e autenticação. Como mencionado anteriormente, a criptografia híbrida é uma combinação de algoritmos simétrico (AES) e assimétrico (RSA). Para o primeiro, uma chave secreta de 128 bits será gerada automaticamente pela API. Cada resposta será criptografada usando uma chave simétrica distinta. Para o segundo método, é necessário estabelecer um repositório de chaves para armazenar as chaves públicas dos sistemas externos para o processo de autenticação. O par de chaves deve ter pelo menos 2048 bits e a chave pública precisa ser instalada no repositório de chaves pelo administrador do sistema externo por meio de uma URL que é fornecida. Assim, a chave simétrica será criptografada usando a chave pública armazenada, habilitando a descriptografia somente pela entidade em posse da respectiva chave privada.

A chave da API é exclusiva e é gerada por um administrador por meio do software de administração da arquitetura. Esse sistema foi desenvolvido para permitir a configuração rápida dos metadados necessários para a integração com sistemas externos e para controle dos níveis de acesso.

Posteriormente, a chave de API é enviada ao administrador do sistema externo criptografada com sua chave pública. Esse processo ocorre somente após a instalação da chave pública do sistema externo no repositório para garantir que a chave da API seja transmitida com segurança. Além disso, é importante observar que, como o protocolo HTTPS está sendo usado, a chave da API e a assinatura permanecerão seguras durante as requisições, e, no caso de uma violação improvável, os dados permanecerão criptografados e ilegíveis para usuários não autorizados. A chave da API pode ser invalidada a qualquer momento.

Autorização. Como parte da estratégia para aumentar o controle sobre os dados, um mecanismo de autorização foi estabelecido através da atribuição de níveis de acesso aos sistemas externos e à propriedades de ontologias (LIMA et al., 2018b). Quatro níveis de acesso foram definidos e são representados por cores, como verde, amarelo, vermelho e preto. O verde é o nível inferior e é o valor padrão para todas as propriedades. Amarelo é o nível intermediário, enquanto vermelho representa o nível mais alto. As propriedades com o nível preto ou listadas em uma *blocklist* serão sempre censuradas e não serão transferidas para o solicitante. A Figura 16 apresenta os sistemas externos registrados no software de administração, com seus respectivos níveis de acesso (representados por cores).

Figura 16 - Software de Administração da Arquitetura de Interoperabilidade

SisTB API - Configurações de Segurança

Sistemas Externos | Nivel Acesso - Entidades | Nivel Acesso - Propriedades | Blocklist - Propriedades | Tokens

Alias	Nível de Acesso Máximo	SPARQL	HyperGraphQL
external-01	2	true	true
external-02	3	true	true
external-03	1	false	false

Esse mecanismo pode ser facilmente aplicado a todos os dados que são semanticamente marcados com propriedades específicas de uma determinada ontologia (escopo inferior) ou a uma entidade inteira no sistema (escopo mais alto). Cada propriedade de ontologia é verificada individualmente. Se o sistema não atender ao nível de acesso necessário, a propriedade ainda é incluída na resposta, mas com seu valor censurado. Por exemplo, se a propriedade *birth_date* da entidade *Patient* tiver um nível amarelo atribuído, ela estará disponível para alguns usuários e não para outros. Nesse caso, somente os sistemas com nível de acesso definido como amarelo ou vermelho poderão recuperar as informações, excluindo os sistemas com nível verde. A Figura 17 ilustra o cenário mencionado, mostrando uma resposta enviada para (a) sistemas com nível amarelo e vermelho e (b) para sistemas com nível verde.

Figura 17 - Resposta enviada para (a) sistemas com nível amarelo e vermelho e (b) para sistemas com nível verde.

```
{
    "@graph" : [{
        "@id" : "_:b0",
        "@type" : "http://sistb-dev.ddns.net/d2rq/vocab/Paciente",
        "nome" : "PACIENTE TESTE 10",
        "data_nascimento" : "1990-04-30",
        "idade" : "29",
        "genero" : "Masculino",
        "endereco" : "Rua teste, 105",
        "telefone" : "555555-555"
    }],
    "@context" : {
        //conteúdo omitido
    }
}

{
    "@graph" : [{
        "@id" : "_:b0",
        "@type" : "http://sistb-dev.ddns.net/d2rq/vocab/Paciente",
        "nome" : "PACIENTE TESTE 10",
        "data_nascimento" : "NÍVEL DE ACESSO INSUFICIENTE",
        "idade" : "29",
        "genero" : "Masculino",
        "endereco" : "Rua teste, 105",
        "telefone" : "555555-555"
    }],
    "@context" : {
        //conteúdo omitido
    }
}
```

Outro mecanismo de autorização que foi implementado está relacionado ao *endpoint* de

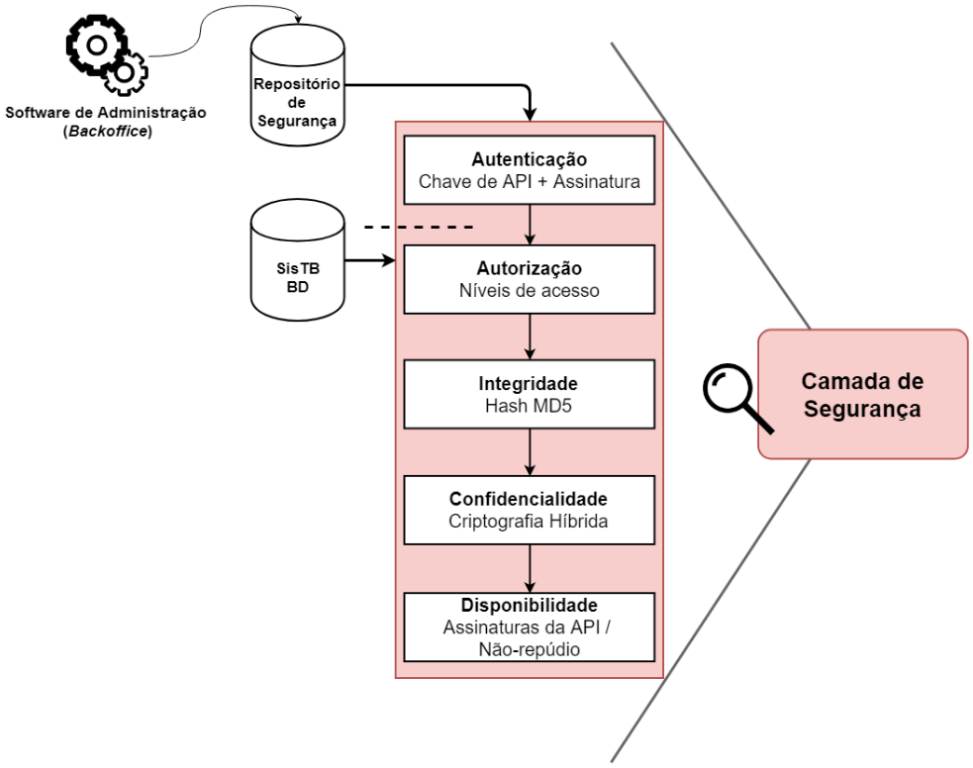
consulta (*Public Query Endpoint*). A arquitetura tem dois *endpoints* de consulta protegidos: o SPARQL, fornecido pelo D2R Server; e o GraphQL fornecido por um HyperGraphQL Server. Como se destinam a publicar dados abertos, nenhum recurso de segurança está disponível para implementar autenticação e autorização de forma nativa. Por isso, esses *endpoints* foram protegidos com *firewalls* e isolados em uma rede privada que somente a API tem acesso.

No entanto, quando os dados estão sendo publicados por meio das tecnologias da Web Semântica, normalmente espera-se que tais *endpoints* estejam disponíveis para permitir consultas semânticas sobre o conjunto de dados. Dessa forma, optou-se por desenvolver um *endpoint* de consulta disponível publicamente, para atuar como uma "ponte" para os que estão protegidos. Só são aceitas consultas de usuários autorizados, usando a mesma chave de API (e sua assinatura) emitida no momento do registro do sistema externo. No software de administração da arquitetura, é possível configurar no repositório de segurança se uma determinada chave de API tem direitos de acesso aos serviços SPARQL e/ou HyperGraphQL. O resultado da consulta também passa pelo processo de criptografia híbrida e o modelo de resposta é o mesmo implementado pela API, conforme descrito anteriormente.

Integridade. A verificação de integridade é outro mecanismo fundamental que deve ser reforçado na interoperabilidade dos dados de saúde. Para isso, a camada de segurança inclui na resposta metadados adicionais, a saber: um *hash* calculado sobre os dados brutos, uma assinatura do *hash*; e uma assinatura dos dados brutos. O *hash* é calculado pelo algoritmo Message-Digest 5 (MD5) e ambas as assinaturas são geradas pelo algoritmo SHA256+RSA usando a chave privada da API para garantir a autenticidade dos dados e garantir a propriedade de não-repúdio. Posteriormente, quando o sistema externo receber a resposta da API e decifrar a mensagem, os *hash* podem ser recalculados e confrontados com o fornecido pela API para examinar se os dados não foram modificados e se, de fato, foram enviados a partir da origem esperada.

Fluxo de integração. O fluxo de integração é simples, mas implementa mecanismos eficientes e complementares que são capazes de garantir a troca segura de dados ponto-a-ponto. O sistema externo inicia o fluxo através de uma requisição enviada para a API ou pelo *endpoint* público de consulta. A Figura 18 sumariza a estrutura da camada de segurança e os mecanismos implementados.

Figura 18 - Estrutura da camada de segurança e os mecanismos implementados



É essencial que todas as solicitações enviadas à API utilizem o protocolo HTTPS, o que garante mais uma camada de proteção através da criptografia do canal de comunicação. Cada sistema externo é identificado por uma chave de API que deve ser assinada com sua chave privada (usando o algoritmo SHA256+RSA). Tanto a chave da API quanto a assinatura devem ser incluídas como parâmetros em todas as solicitações enviadas para a API.

A estrutura de segurança apresentada neste projeto visa proteger dados sensíveis para fins de interoperabilidade. Todas as técnicas computacionais que suportam a solução, como criptografia e *hashing*, são tecnologias bem conhecidas que, quando combinadas, podem oferecer recursos de segurança robustos. Desta forma, cada sistema externo que deseje interoperar poderá facilmente atender todos os requisitos técnicos necessários.

Todas as respostas são produzidas em formatos legíveis por máquina e humanos. A mensagem de resposta é enviada no formato JSON e, após o processo de descriptografia, os dados solicitados ficam disponíveis nos formatos JSON-LD ou RDF (com semântica). A linguagem de programação usada pelo sistema externo não deve ser uma barreira, considerando que qualquer linguagem de programação moderna implementa bibliotecas para enviar solicitações HTTP, criptografar e descriptografar dados, calcular *hashes*, assinar dados e decodificar mensagens Base64.

A solução baseada em criptografia híbrida permite uma transmissão ponto-a-ponto

segura de dados de saúde, solucionando, de uma maneira elegante, o problema de distribuição de chaves e a quantidade de dados que podem ser criptografados. Além disso, o uso de chaves e assinaturas de API, a definição de níveis de acesso a dados e o cálculo de códigos *hashes* atendem aos requisitos de autenticação, autorização, confidencialidade, integridade e disponibilidade. O protocolo HTTPS também é essencial para proteger o canal de comunicação, o que representa uma camada de segurança adicional.

Ao aplicar esses mecanismos às tecnologias da Web Semântica, as necessidades de segurança podem ser satisfeitas (THURAISINGHAM, 2005), contribuindo com a confiabilidade dos dados e promovendo seu uso em contextos que não foram inicialmente construídos para disseminação de dados abertos. Apesar da existência de outras soluções que contribuem com questões de segurança na Web Semântica, ressalta-se algumas desvantagens relacionadas ao aumento da complexidade da estrutura de dados, como o rápido crescimento de grafos RDF, devido à geração de novas tuplas com os metadados necessários (GERBRACHT, 2008), e a necessidade de definição e uso de ontologias não-triviais para controle de privacidade (DENKER; KAGAL; FININ, 2005).

Finalmente, de modo a facilitar o uso dos *endpoints* da arquitetura de interoperabilidade, os requisitos técnicos foram compilados em um manual, o qual está disponível no APÊNDICE A – MANUAL DE UTILIZAÇÃO DO *FRAMEWORK* SEGURANÇA.

7.3.2. Verificação adicional de integridade dos dados do SISTB[12]

Devido à sensibilidade dos dados, um dos principais desafios é fornecer garantias de que um determinado registro de saúde não foi corrompido. Uma funcionalidade implementada no SISTB (sistema sob gerência do LIS) permite que sistemas externos/usuários verifiquem a integridade e imutabilidade de um registro usando funções *hash* e uma DLT (LIMA et al., 2021a).

IOTA

Uma alternativa às arquiteturas tradicionais de Blockchain são os Grafos Acíclicos Dirigidos - do inglês, *Directed Acyclic Grap*h (DAG). Os *ledgers* DAG fornecem muitos ganhos em relação ao Blockchain tradicional, incluindo desempenho, escalabilidade e custos

[12] A informação contida nesta subseção refere-se parcialmente ao conteúdo do artigo *A mechanism for verifying the integrity and immutability of Tuberculosis data using IOTA distributed ledger technology*, apresentado na conferência *dHealth - Health Informatics meets Digital Health* (Áustria, 2021) e publicado no periódico *Studies in Health Technology and Informatics*. 2021. ISBN 978-1-64368-181-8. Disponível em https://doi.org/10.3233/SHTI210099.

de transação. Neste trabalho, adotamos a DLT IOTA, desenvolvida em 2015, sem taxas de transação, blocos ou mineração. Com base no princípio *Machine-to-Machine (*M2M), a IOTA foi projetada especificamente para o setor de Internet das Coisas (IoT) com uma arquitetura interconectada por meio de uma rede emaranhada, a rede Tangle, um DAG na qual vértices representam transações e arestas representam validações. (BHANDARY; PARMAR; AMBAWADE, 2020a).

O gerenciamento do fluxo de dados pode acontecer em canais protegidos com a troca de Mensagens Autenticadas Mascaradas – do inglês, *Masked Authenticated Messages* (MAM). O protocolo MAM é uma das características mais notáveis da IOTA. Ele permite que os dados sejam compartilhados com segurança por meio de um canal criptografado. Dessa forma, é possível garantir que o destinatário de uma transação receba os dados íntegros e de uma fonte confiável (BHANDARY; PARMAR; AMBAWADE, 2020b). Para descriptografar os dados armazenados nas transações, a chave deve ser conhecida, para que apenas as partes autorizadas possam acessar os dados.

A comunidade IOTA mantém duas redes públicas, cada uma tem seu próprio Tangle: a Mainnet e a Devnet. Um nó privado pode ser instalado e conectado à rede.

Funcionamento da solução

Neste trabalho, foi realizada a instalação de um nó privado na rede pública IOTA Devnet, seguida do desenvolvimento de um mecanismo para realizar verificações de integridade dos dados armazenados no SISTB.

O nó privado foi implantado em um servidor virtual. O *software*, denominado Hornet[13], é fornecido pela comunidade IOTA. Apesar de estar conectado a uma rede pública, possuir um nó privado permite que uma organização interaja com segurança com o Tangle, com permissões específicas e sem depender da disponibilidade de nós de terceiros.

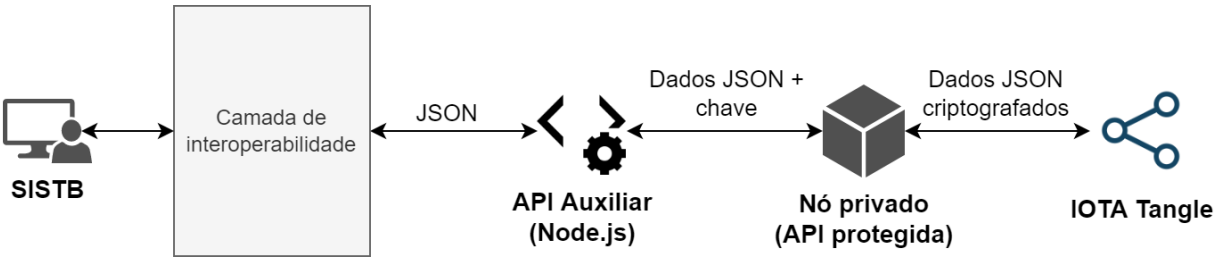
O nó IOTA expõe automaticamente uma API para permitir a interoperabilidade com aplicativos externos. A API permite operações de escrita e leitura, ou seja, envio e recuperação de transações, bem como a comunicação por meio de canais confidenciais com dados criptografados (MAM).

Para se comunicar com o nó, uma API auxiliar foi desenvolvida na linguagem Node.js, a qual se conecta diretamente à API do nó usando bibliotecas de programação mantidas pela comunidade IOTA. Essa API é capaz de enviar/recuperar transações de/para o Tangle e

[13] Disponível em https://github.com/gohornet/hornet

funcionar como uma ponte para a rede. Por fim, por meio de *endpoints* específicos da camada de interoperabilidade semântica, os sistemas autorizados conseguiram se conectar à API auxiliar para interagir com o Tangle. A Figura 19 resume o fluxo de comunicação e a integração entre os componentes.

Figura 19. Fluxo de comunicação



O SISTB necessita da API auxiliar para interagir com o Tangle. Quando um registro de saúde é criado ou atualizado no SISTB, um gatilho de entrada de dados calcula um *hash* para os dados e chama a API auxiliar, fornecendo o ID do registro, o *hash* e um carimbo de data/hora, no formato JSON. Em seguida, é estabelecido um canal de comunicação com o nó IOTA, que efetivamente submete a transação (com os dados) ao Tangle. A Figura 20 apresenta um exemplo de dados armazenados no Tangle.

Figura 20. Dados no formato JSON armazenados em uma transação no Tangle.

MESSAGE JSON


```
{
   "cod": "11245875",
   "hash": "E9E16315EAE1308D458BCBE079000C0632F5F2C35100DAF6B14330646E241536",
   "timestamp": "2020-12-07 12:34:05"
}
```


Considerando que esta solução utiliza uma rede compartilhada, é importante manter a confidencialidade dos dados no Tangle. Para tanto, foi utilizado o protocolo MAM, que requer a mensagem e uma chave simétrica conhecida apenas pela API auxiliar para criptografar e descriptografar os dados. Os sistemas não têm acesso à chave, pois a API auxiliar é o único ponto que grava e lê os dados do Tangle.

Por fim, um sistema externo que queira verificar a integridade de um dado específico pode calcular o *hash* dos dados por conta própria e compará-lo com o *hash* armazenado no Tangle. Se os *hashes* corresponderem, significa que os dados não foram alterados. No caso dos usuários finais do SISTB, as interfaces/telas do sistema apresentam o *hash* calculado em tempo real (quando o usuário navega até a página que contém os dados), o qual é comparado com o

hash armazenado no Tangle. Se houver uma incompatibilidade, o usuário é alertado sobre uma possível violação de integridade dos dados.

7.4. Módulo de Gerenciamento de Dados de Pesquisas Científicas

O gerenciamento de dados é um processo contínuo e representa uma atividade crítica na pesquisa, devido à sua importância para a qualidade e confiabilidade dos dados, que devem atender aos requisitos e parâmetros especificados no protocolo de pesquisa (KRISHNANKUTTY et al., 2012). É fundamental que esse processo ocorra em paralelo com a coleta de dados. O gerente de dados é responsável pela validação dos dados, que inclui a verificação da consistência, integridade e precisão.

O módulo de gerenciamento de dados é composto de várias funcionalidades complementares e independentes, as quais podem ser configuradas, ativadas ou desativadas pelo *software* de administração da arquitetura, dependendo da necessidade de cada projeto. Esses recursos foram desenvolvidos para uso integrado ao *software* REDCap, sendo:

- Verificação de participantes duplicados em tempo real;
- Validação automática e semi-automática de registros;
- Sistema de alerta de tempos de preenchimento;
- Agenda dinâmica de visitas do estudo;
- Modelo de Plano de Gerenciamento de Dados padronizado.

Verificação de participantes duplicados em tempo real

O processo de verificação de duplicidade de registros é automático e ocorre em tempo real para identificar registros duplicados em um mesmo centro de pesquisa ou entre centros distintos, considerando que um participante pode vir a ser acompanhado em mais de um local de atendimento. Este mecanismo é particularmente importante para ajudar os centros a controlarem os seus registros e, principalmente, evitar vieses na fase de análise.

Sempre que um participante é incluído no sistema REDCap, é realizada uma verificação em todo o banco de dados, com base em um identificador pré-definido para o estudo (nº prontuário ou CPF, por exemplo). Se uma duplicidade for detectada, uma notificação é imediatamente disparada para o *e-mail* do usuário que gravou os dados, o qual deverá verificar se realmente trata-se de uma duplicidade e informar o gerente de dados. Caso se trate de uma duplicidade confirmada, o usuário deverá indicar qual registro deseja manter. O Quadro 1 representa um exemplo de mensagem que é enviada ao usuário.

Quadro 1. Exemplo de notificação de duplicidade

Assunto: Identificação - Duplicidade de registro

Prezado(a),

Foi identificado um registro duplicado no seu centro - nº XX-YY.
Registros duplicados: XX-WW, XX-ZZ

Solicita-se a verificação da duplicidade e a indicação de qual registro deseja manter.

Atenciosamente,
O Gerente de Dados do Projeto XPTO

Validação automática de preenchimentos de registros

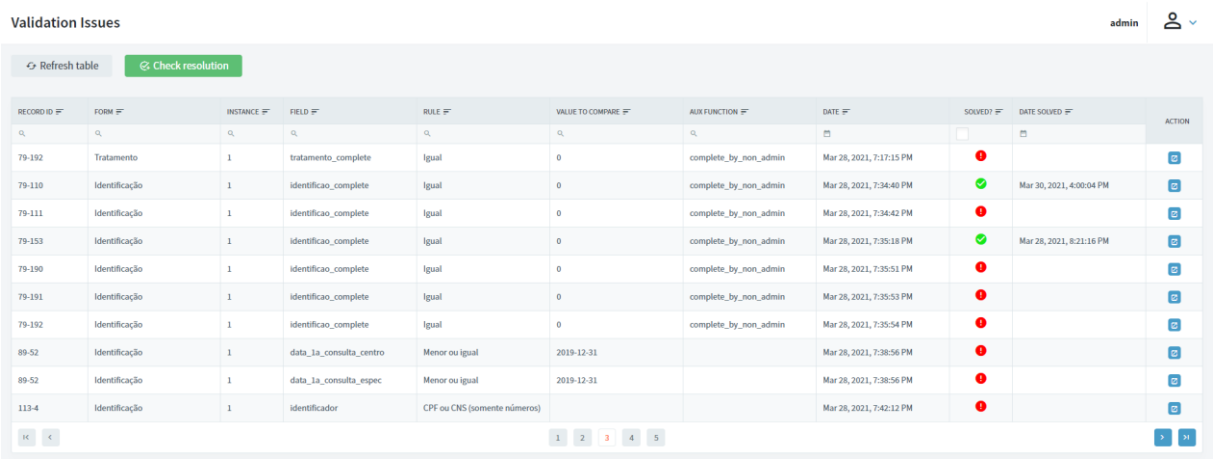
Um procedimento automático de validação baseado em regras percorre cada campo de todos os instrumentos de coleta em busca de qualquer inconsistência. As regras devem ser pré-definidas como metadados através do sistema de administração e representam o formato ou intervalo de valores esperados para um determinado campo. O procedimento é executado várias vezes ao dia para verificar, ao mesmo tempo, novos problemas e verificar a resolução dos previamente identificados.

Quando um problema é detectado, uma *query* é aberta no REDCap e o coletor de dados é alertado por e-mail. Uma *query* trata-se de um pedido formal de verificação de um determinado campo, efetuado pelo gerente de dados e direcionado ao usuário que editou o campo. Se nenhuma inconsistência for detectada em um formulário, o mesmo é marcado como validado e bloqueado para edição.

Ressalta-se que todo esse processo ocorre de forma automática, sem necessidade de intervenção por parte do gerente de dados. Somente a checagem da resolução das *queries* abertas devem ser realizadas manualmente pelo gerente de dados, pois envolve uma resposta textual obrigatória por parte do coletador de dados que deve ser cuidadosamente verificada.

Um painel de fácil visualização está disponível para acompanhar os problemas de validação detectados e as respectivas resoluções, conforme apresentado na Figura 21. Nesse painel, é possível identificar o número do registro, o instrumento de coleta, o campo problemático, a regra aplicada ao campo e as datas de detecção e resolução do problema.

Figura 21. Painel de validação de campos de instrumentos de coleta do REDCap



Validation Issues

admin

Refresh table | Check resolution

RECORD ID	FORM	INSTANCE	FIELD	RULE	VALUE TO COMPARE	AUX FUNCTION	DATE	SOLVED?	DATE SOLVED	ACTION
79-192	Tratamento	1	tratamento_complete	Igual	0	complete_by_non_admin	Mar 28, 2021, 7:17:15 PM			
79-110	Identificação	1	identificao_complete	Igual	0	complete_by_non_admin	Mar 28, 2021, 7:34:40 PM		Mar 30, 2021, 4:00:04 PM	
79-111	Identificação	1	identificao_complete	Igual	0	complete_by_non_admin	Mar 28, 2021, 7:34:42 PM			
79-153	Identificação	1	identificao_complete	Igual	0	complete_by_non_admin	Mar 28, 2021, 7:35:18 PM		Mar 28, 2021, 8:21:16 PM	
79-190	Identificação	1	identificao_complete	Igual	0	complete_by_non_admin	Mar 28, 2021, 7:35:51 PM			
79-191	Identificação	1	identificao_complete	Igual	0	complete_by_non_admin	Mar 28, 2021, 7:35:53 PM			
79-192	Identificação	1	identificao_complete	Igual	0	complete_by_non_admin	Mar 28, 2021, 7:35:54 PM			
89-52	Identificação	1	data_1a_consulta_centro	Menor ou igual	2019-12-31		Mar 28, 2021, 7:38:56 PM			
89-52	Identificação	1	data_1a_consulta_espec	Menor ou igual	2019-12-31		Mar 28, 2021, 7:38:56 PM			
113-4	Identificação	1	identificador	CPF ou CNS (somente números)			Mar 28, 2021, 7:42:12 PM			

1 2 3 4 5

Sistema de alerta de tempos de preenchimento

Foi desenvolvido um sistema de alertas para enviar periodicamente notificações aos centros de pesquisa sobre coleta de dados pendentes com base nos eventos programados de cada estudo. Por meio de lembretes, o sistema auxilia os pesquisadores a manter os dados dos participantes atualizados de acordo com o protocolo formal e, assim, evitar violações críticas.

As notificações são enviadas por e-mail para uma lista de destinatários previamente definida para cada centro, contendo os formulários e os prazos para cada registro, conforme apresentado no Quadro 2.

Quadro 2. Exemplo de notificação de alerta de tempos de preenchimentos

Assunto: ALERTAS - PREENCHIMENTOS PENDENTES - REDCap

Centro: HOSPITAL 01

PID 26-1 - CRF[14] Condições de Transporte - Dia 1 - até 09/01/2022 - PRAZO EXPIRADO
PID 26-1 - CRF Rastreamento da Amostra 1 - Dia 1 - até 09/01/2022 - PRAZO EXPIRADO
PID 26-1 - CRF Rastreamento da Amostra 2 - Dia 1 - até 09/01/2022 - PRAZO EXPIRADO

Atenciosamente,
O Gerente de Dados do Projeto XPTO

Agenda dinâmica de visitas dos participantes estudo

Na pesquisa em saúde, boa parte dos dados são coletados durante visitas dos participantes nos centros de pesquisa. Por isso, acompanhar o cronograma de visitas e seu estado (realizado, não realizado, pendente) é essencial para não falhar nenhum marco do projeto.

[14] Do inglês, *Case Report Form*

Pensando nisso, um painel foi desenvolvido para fornecer uma visualização rápida de todas as visitas dos participantes. Cada linha no painel representa um participante e cada coluna uma visita. A cor da célula representa o estado da visita (verde: realizada; vermelha: não realizada; amarela: pendente/aguardando o participante). As datas são calculadas com base em um campo de data de referência (por exemplo, o dia de uma intervenção ou inclusão no estudo) e na variação de dias para cada evento. O painel é criado em tempo real com dados *online* extraídos do banco de dados do REDCap, economizando tempo dos pesquisadores que costumam criar seu próprio painel usando planilhas. A Figura 22 mostra o painel de um estudo com 21 visitas.

Figura 22. Agenda de visitas

Calendário

UERJ | Enviar

CSV | Excel | Imprimir

Pesquisar

Participante	Data Referência	M1	M3	M4	M5	M6	M7	M8	M9	M10	M11	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11
33-10	27-11-2020	27-12-2020	25-02-2021	27-03-2021	26-04-2021	26-05-2021	25-06-2021	25-07-2021	24-08-2021	23-09-2021	23-10-2021	04-12-2020	11-12-2020	18-12-2020	25-12-2020	01-01-2021	08-01-2021	15-01-2021	22-01-2021	29-01-2021	05-02-2021	12-02-2021
33-100	12-01-2021	11-02-2021	12-04-2021	12-05-2021	11-06-2021	11-07-2021	10-08-2021	09-09-2021	09-10-2021	08-11-2021	08-12-2021	19-01-2021	26-01-2021	02-02-2021	09-02-2021	16-02-2021	23-02-2021	02-03-2021	09-03-2021	16-03-2021	23-03-2021	30-03-2021
33-101	12-01-2021	11-02-2021	12-04-2021	12-05-2021	11-06-2021	11-07-2021	10-08-2021	09-09-2021	09-10-2021	08-11-2021	08-12-2021	19-01-2021	26-01-2021	02-02-2021	09-02-2021	16-02-2021	23-02-2021	02-03-2021	09-03-2021	16-03-2021	23-03-2021	30-03-2021
33-102	12-01-2021	11-02-2021	12-04-2021	12-05-2021	11-06-2021	11-07-2021	10-08-2021	09-09-2021	09-10-2021	08-11-2021	08-12-2021	19-01-2021	26-01-2021	02-02-2021	09-02-2021	16-02-2021	23-02-2021	02-03-2021	09-03-2021	16-03-2021	23-03-2021	30-03-2021
33-103	12-01-2021	11-02-2021	12-04-2021	12-05-2021	11-06-2021	11-07-2021	10-08-2021	09-09-2021	09-10-2021	08-11-2021	08-12-2021	19-01-2021	26-01-2021	02-02-2021	09-02-2021	16-02-2021	23-02-2021	02-03-2021	09-03-2021	16-03-2021	23-03-2021	30-03-2021
33-104	13-01-2021	12-02-2021	13-04-2021	13-05-2021	12-06-2021	12-07-2021	11-08-2021	10-09-2021	10-10-2021	09-11-2021	09-12-2021	20-01-2021	27-01-2021	03-02-2021	10-02-2021	17-02-2021	24-02-2021	03-03-2021	10-03-2021	17-03-2021	24-03-2021	31-03-2021
33-105	13-01-2021	12-02-2021	13-04-2021	13-05-2021	12-06-2021	12-07-2021	11-08-2021	10-09-2021	10-10-2021	09-11-2021	09-12-2021	20-01-2021	27-01-2021	03-02-2021	10-02-2021	17-02-2021	24-02-2021	03-03-2021	10-03-2021	17-03-2021	24-03-2021	31-03-2021
33-107	13-01-2021	12-02-2021	13-04-2021	13-05-2021	12-06-2021	12-07-2021	11-08-2021	10-09-2021	10-10-2021	09-11-2021	09-12-2021	20-01-2021	27-01-2021	03-02-2021	10-02-2021	17-02-2021	24-02-2021	03-03-2021	10-03-2021	17-03-2021	24-03-2021	31-03-2021
33-11	27-10-2020	26-11-2020	25-01-2021	24-02-2021	26-03-2021	25-04-2021	25-05-2021	24-06-2021	24-07-2021	23-08-2021	22-09-2021	03-11-2020	10-11-2020	17-11-2020	24-11-2020	01-12-2020	08-12-2020	15-12-2020	22-12-2020	29-12-2020	05-01-2021	12-01-2021
33-110	13-01-2021	12-02-2021	13-04-2021	13-05-2021	12-06-2021	12-07-2021	11-08-2021	10-09-2021	10-10-2021	09-11-2021	09-12-2021	20-01-2021	27-01-2021	03-02-2021	10-02-2021	17-02-2021	24-02-2021	03-03-2021	10-03-2021	17-03-2021	24-03-2021	31-03-2021

Mostrando de 1 até 10 de 43 registros

Anterior 1 2 3 4 5 Próximo

Modelo de Plano de Gerenciamento de Dados padronizado

Os Planos de Gerenciamento de Dados estão se tornando um componente essencial em pesquisas científicas. Inclusive, agências de financiamento nacionais e internacionais estão a tornar obrigatória a sua disponibilização como documento complementar dos projetos. No Brasil, por exemplo, a Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) já considera a apresentação do PGD como obrigatória na maioria de seus editais.

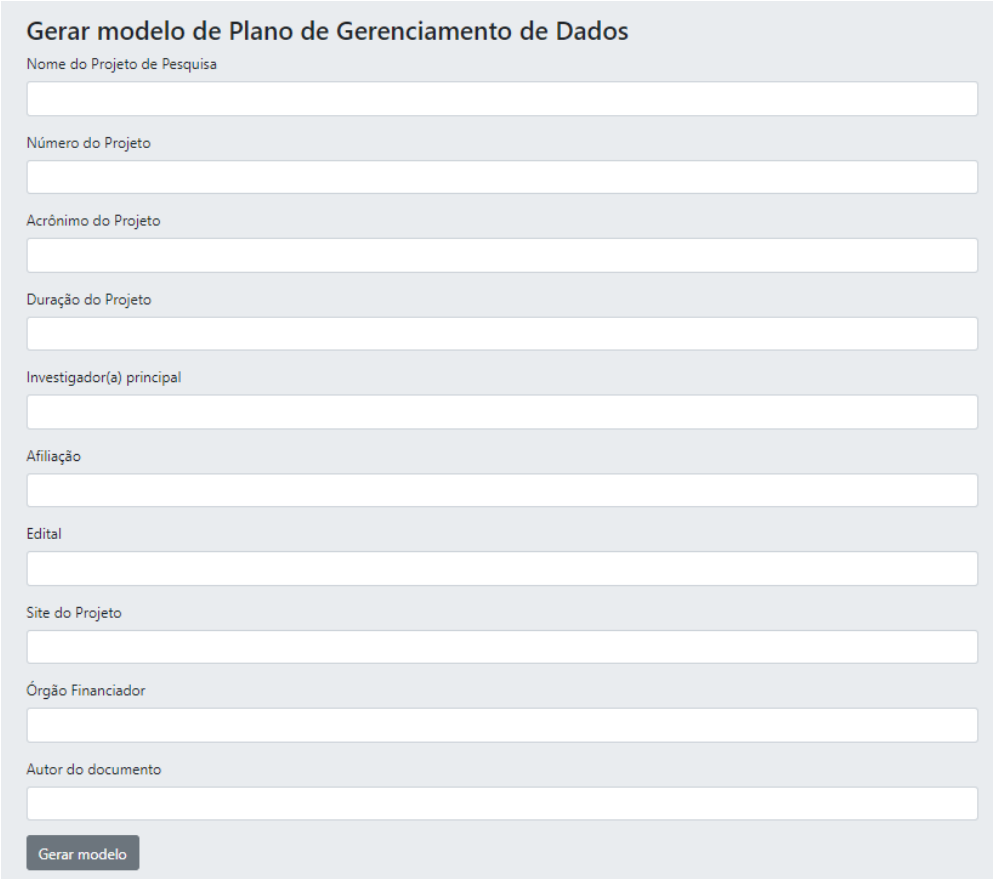
Os PGDs geralmente são planos descritivos de alto nível que prescrevem os dados a serem gerados por um projeto de pesquisa, como esses dados serão armazenados (com segurança, conforme necessário), quem terá acesso, qual documentação e metadados serão criados e as intenções de preservação a longo prazo (BURNETTE; WILLIAMS; IMKER, 2016).

Considerando essa tendência, um modelo de PGD foi proposto, o qual está disponível

para consulta no APÊNDICE B - MODELO DE PLANO DE GERENCIAMENTO DE DADOS. Os tópicos considerados fundamentais para a construção de um plano completo e abrangente foram abordados, como: a apresentação de informações básicas, dos dados e recursos do estudo, questões de privacidade e segurança dos dados (incluindo um protocolo de segurança e violação de dados), aspectos éticos e regulatórios, propriedade intelectual, requisitos para acesso aos dados, procedimentos de arquivamento e uso futuro dos dados, e, finalmente, estratégias de gerenciamento de dados e monitoria do estudo. Além do mais, o PGD proposto se propõe a padronizar o processo de gerenciamento de dados e monitoria do estudo, pois trata-se de um conjunto de atividades fundamental para a garantia de consistência e qualidade dos dados.

Uma ferramenta foi desenvolvida para automatizar a geração da versão inicial do PGD, com base no modelo proposto. Para criar uma cópia do documento, já com as informações básicas do projeto pré-preenchidas, basta preencher um formulário eletrônico com tais informações, conforme apresentado na Figura 23. Após o preenchimento, um *link* de acesso é exibido para acessar o documento.

Figura 23. Gerador de Plano de Gerenciamento de Dados



O referido modelo considera o uso do REDCap como sistema EDC e se trata de uma proposta de PGD não exaustiva, podendo não atender a determinadas especificidades de um projeto de pesquisa e, portanto, deve ser utilizado e revisado com cautela. Ainda, algumas seções devem ser completadas de acordo com o cenário do projeto.

7.5. Rede Blockchain para compartilhamento de dados de pesquisa de Tuberculose[15]

Considerando que os dados de pesquisa devem permitir a replicação e validação dos estudos, eles devem ser confiáveis, rastreáveis, reutilizáveis e interpretáveis (PARK; GREENE, 2018). Uma preocupação comum refere-se à como compartilhar dados de forma eficaz e superar os desafios de segurança ao trabalhar com dados confidenciais (BENCHOUFI; RAVAUD, 2017; ZHUANG et al., 2018; BERNARDI et al., 2019). Os responsáveis pelos dados também devem atender aos requisitos de integridade e privacidade (RUOTSALAINEN, 2010).

Uma abordagem válida para atender aos requisitos de segurança e disponibilizar uma infraestrutura descentralizada para compartilhamento de dados é através do uso de uma DLT, como o Blockchain, combinado com outros mecanismos como, por exemplo, para a desidentificação de dados para garantir privacidade e o uso de *hashes* para verificação de integridade.

Nesse sentido, uma rede Blockchain foi desenvolvida como uma solução para disponibilizar conjuntos de dados desidentificados de TB com foco em pesquisas (LIMA et al., 2021c). Esses conjuntos destinam-se principalmente a apoiar a replicação de experimentos e e novas pesquisas, como estudos orientados que fazem o uso de dados para desenvolver algoritmos inovadores de inteligência artificial/aprendizado de máquina para suporte à decisão clínica e aprimoramento do diagnóstico de TB.

O banco de dados é construído em tempo real à medida que os dados são coletados e o pré-processamento é realizado logo antes de gravar os dados na rede. Metadados semânticos também são disponibilizados juntamente com dados. A plataforma garante redundância, disponibilidade, integridade e autenticidade dos dados. A rede é escalável e projetada para oferecer alto desempenho com baixo consumo de armazenamento.

Os principais objetivos da rede são:

1. construir continuamente um banco de dados desidentificado composto de dados de pesquisa de TB;

[15] A informação contida nesta subseção refere-se parcialmente ao conteúdo do artigo *A permissioned blockchain network for security and sharing of de-identified Tuberculosis research data in Brazil*, publicado no periódico *Methods of Information in Medicine*. 2021. ISSN 0026-1270. Disponível em https://doi.org/10.1055/s-0041-1727194.

2. aumentar a transparência, disponibilidade e integridade dos dados brutos coletados nos sistemas EDC;
3. possibilitar o compartilhamento de dados com anotação semântica;
4. atender às necessidades das partes interessadas, como pacientes, pesquisadores e organizações de TB.

O repositório pode beneficiar participantes dos estudos, pesquisadores, organizações de saúde e centros de pesquisa. Visando as necessidades deles, a solução entrega diversos conjuntos de dados relacionados à TB, incluindo variados tipos de dados, como socioeconômicos, clínicos, genéticos, medicamentosos, resultados de exames, entre outros. No entanto, é importante notar que a disponibilidade dessas categorias de dados depende do desenho dos estudos. Ou seja, o conjunto de dados será composto pelo tipo de dados que estão sendo coletados em um determinado estudo. Pesquisadores e organizações de TB poderão aproveitar esses dados para realizar estudos baseados em dados. Os pacientes podem ser beneficiários indiretos dos resultados de novos estudos e sua privacidade será preservada.

Os participantes da rede podem ser organizações ou membros de uma organização. Os membros podem possuir e controlar um nó na rede. Quanto mais organizações houver, mais descentralizada será a rede. Um pesquisador individual pode se credenciar como consumidor de dados, o qual terá acesso à funcionalidades de leitura e exportação dos conjuntos de dados.

Ferramentas de desenvolvimento

Para permitir a interoperabilidade de dados entre os sistemas EDC e a plataforma Blockchain Kaleido, foi desenvolvido um Motor de Integração (MI) usando a linguagem PHP v7.2. O MI funciona como um integrador e é responsável por integrar APIs de ambos os lados para permitir efetivamente a comunicação entre esses sistemas e preparar os dados antes de enviá-los à rede Blockchain. Um banco de dados relacional MySQL (https://www.mysql.com/) está disponível para armazenar metadados, *hashes* de transações da rede Blockchain e identificadores de registros.

Os metadados são necessários para permitir um bom funcionamento do MI. Inclui configurações do REDCap e KoBoToolbox, como credenciais de API e identificadores de instrumentos de coleta de dados. Além disso, o mapeamento da semântica e dos identificadores pessoais para os campos dos formulários pode ser feito através de metadados.

Por fim, para gerenciar os metadados, entregar uma ferramenta de exploração de cadeias e fornecer um recurso de exportação dos conjuntos de dados, foi desenvolvido um sistema de gerenciamento nas linguagens de programação C# e JavaScript. O sistema conta com três níveis

de acesso: acesso total (para administradores), gerenciamento de metadados do projeto (para pesquisadores e organizações) e acesso somente leitura para exportação e visualização de dados (para pesquisadores).

Anotação semântica

O processo de "semantização" é feito através do mapeamento dos campos do instrumento de coleta com propriedades das ontologias. Os dados brutos são armazenados em uma transação na rede Blockchain no formato JSON. De maneira semelhante, os dados marcados semanticamente são armazenados no formato JSON for Linking Data (JSON-LD) (W3C, 2010). Esta é a maneira mais simples de compartilhar dados com significado anexado. Quando os dados são obtidos em outro contexto, como um aplicativo externo, o desenvolvedor pode escolher qual tipo de dado é mais adequado (RUTA et al., 2017; UGARTE, 2017).

O objeto JSON-LD é gerado dinamicamente pelo MI após receber dados de um sistema EDC. Por isso, o mapeamento de campos para propriedades de ontologias deve ser pré-definido como metadados. No caso do REDCap, como já referido (6.3.1), o mapeamento pode ser feito diretamente no *software* (recurso de anotação de campo). Este recurso exclui a necessidade de configuração do banco de metadados, pois o MI extrairá a anotação dos dados enviados pelo REDCap. Mas, se uma anotação não for encontrada, o MI verificará se há correspondência nos metadados.

Desidentificação

De forma similar ao processo de anotação semântica, metadados podem ser usados para marcar um campo como um identificador. Considerando os requisitos de privacidade e o recurso de imutabilidade de dados do blockchain, o valor de um identificador não será armazenado de forma alguma, ou seja, ocorrerá um processo de desidentificação antes de gravar os dados na transação. Identificadores e quasi-identificadores, como nome do paciente, número de telefone e endereço, são automaticamente removidos dos dados estruturados pelo MI.

Os identificadores e quasi-identificadores são mapeados na etapa de elaboração dos instrumentos de coleta de dados de uma pesquisa clínica. Normalmente, os investigadores principais do estudo realizam várias rodadas de validação desses instrumentos para garantir que todos os campos necessários estejam disponíveis para alcançar um resultado desejável e selecionar corretamente os identificadores.

Então, este processo deve ser refletido nos sistemas EDC (construção dos instrumentos

de coleta e mapeamento dos identificadores). Isso garante que os identificadores não sejam exportados para o repositório Blockchain. O REDCap remove todos os identificadores marcados antes de exportar os dados, porém o KoBoToolbox não possui esse recurso. Assim, ao utilizar o REDCap, o mapeamento não precisa ser replicado nos metadados da solução, o que é obrigatório para o sistema KoBoToolbox.

O MI executa etapas adicionais para mitigar o risco de reidentificação dos dados adicionando ruído aos dados coletados, que incluem: armazenar apenas um hash do identificador do participante; mudança de datas; definição de faixas etárias; e definição de grupos de localização. As datas são deslocadas para o futuro de forma consistente para preservar os intervalos. Os campos de idade são convertidos em intervalos. Os locais, quando disponíveis, são expandidos para uma macrorregião para diminuir a precisão.

É importante notar que, embora existam métodos de anonimização mais fortes, como privacidade diferencial e k-anonimato, nesta solução é necessário realizar a desidentificação em tempo real após a coleta de dados e antes de armazená-los no repositório Blockchain. Esses algoritmos geralmente dependem de um conjunto de dados de tamanho mínimo e, portanto, não são adequados para este caso.

Retração de dados

Considerando que o *ledger* de uma rede blockchain é imutável, o que não permite a exclusão de informações, é necessário um mecanismo para sinalizar os dados como indisponíveis. O objetivo é evitar a exportação de dados armazenados em transações sinalizadas, devido à detecção de alguma inconsistência ou porque os dados foram removidos do banco de dados do sistema EDC. Embora a transação ainda seja visível no explorador da plataforma Kaleido, os dados são codificados e ilegíveis nesta interface.

Um histórico de transações é mantido fora da cadeia - porém criptografado e com redundância - para permitir uma rastreabilidade rápida e fácil. O histórico é necessário para o bom funcionamento da solução e não é acessível aos membros gerais da rede.

Adicionalmente, para promover a transparência na rede, um sistema de *log* de eventos foi estabelecido e disponibilizado aos membros da rede (pesquisadores e organizações) como um repositório inviolável. Isso significa que todas as alterações realizadas nos dados, como operações de adição/edição e retrações, serão registradas. Essas alterações podem ser facilmente rastreadas por qualquer pessoa.

Funcionamento da rede Blockchain

Na plataforma blockchain Kaleido, uma rede permissionada baseada em Ethereum foi criada usando o *ledger* distribuído Quorum, um protocolo especialmente projetado para redes Blockchain privadas, onde há apenas um único membro possuindo todos os nós, ou, uma rede Blockchain de consórcio, onde vários membros cada um possui uma parte da rede (JP MORGAN CHASE, 2016). Este protocolo aprimora a implementação clássica do Ethereum com suporte para transações privadas, contratos inteligentes e mecanismos de consenso alternativos.

Mecanismo de consenso

O algoritmo Raft, desenvolvido por pesquisadores da Universidade de Stanford (ONGARO; OUSTERHOUT, 2014), é utilizado como um sistema de consenso. O Raft elege um líder (minerador) no conjunto de nós, que é responsável por aceitar as solicitações dos clientes e gerenciar a replicação do *ledger* para outros nós. O líder minera e replica novos blocos e o consenso é alcançado quando todos os nós escrevem as transações em seus *ledgers*. Se o líder ficar indisponível, o restante da rede elegerá automaticamente um novo líder após um determinado período de tempo limite, sem indisponibilidade da rede (ZHANG, 2018). Todos os nós estão conectados uns aos outros para trocar constantemente os dados mais recentes para permanecer atualizados.

Raft não minera blocos, a menos que haja transações pendentes. Isso resulta em economias de armazenamento significativas, especialmente quando a carga da transação é baixa, porque nenhum bloco vazio contendo zero transações está sendo gerado.

Fluxo de dados

O protocolo Ethereum oferece um campo especial na estrutura das transações onde os dados podem ser codificados em formato binário. A API da rede recebe dados do MI que incluem: os dados coletados em sistemas EDC, tanto nos formatos JSON (data_to_store) quanto JSON-LD (data_semantic_annotation); o projeto (project_id); o registro (record_id); o nome do instrumento (form_name); o nome do autor (author); a operação (operation) relacionada aos dados da pesquisa, por exemplo, adicionar ou editar; e um *hash* dos dados. Para não expor dados, a API não permite leitura de dados, somente gravação.

Nesse caso, o MI é o sistema externo que armazenará e recuperará os dados na/da cadeia. O processo de integração começa com o REDCap e o KoBoToolbox enviando os dados

coletados ao MI por meio de seus módulos de gatilho. Antes de escrever os dados na cadeia, é necessário um processamento para remover alguns campos, por exemplo, metadados internos dos sistemas EDC, desidentificar os dados e agregar semântica neles criando-se um objeto JSON-LD. O repositório de metadados é consultado para verificar os campos de cada instrumento.

Se houver um mapeamento indicando que o campo pode ser um identificador, ele será removido. Conforme mencionado anteriormente, ao utilizar o REDCap, os campos contendo dados pessoais são removidos automaticamente antes da exportação, ou seja, o MI não precisa realizar a etapa do processo de desidentificação. No entanto, o mesmo não ocorre ao usar o KoBoToolbox. Finalmente, se um mapeamento para uma anotação semântica for encontrado, um objeto JSON-LD é construído para representar o contexto de dados. A Figura 24 mostra um exemplo de dados representados em JSON-LD usando ontologias definidas pela comunidade *schema.org* (SCHEMA.ORG, 2012).

Figura 24. Dados desidentificados no formato JSON-LD.

```
{
  "@context": {
    "gender": "https://schema.org/gender",
    "job_title": "https://schema.org/jobTitle",
    "hiv": {
      "@id": "https://schema.org/InfectiousDisease",
      "@type": "xsd:boolean"
    },
    "tuberculosis": {
      "@id": "https://schema.org/InfectiousDisease",
      "@type": "xsd:boolean"
    },
    "geo": "http://schema.org/geo",
    "latitude": {
      "@id": "http://schema.org/latitude",
      "@type": "xsd:float"
    },
    "longitude": {
      "@id": "http://schema.org/longitude",
      "@type": "xsd:float"
    },
    "xsd": "http://www.w3.org/2001/XMLSchema#"
  },
"gender": "male",
  "job_title": "construction worker",
  "hiv": "true",
  "tuberculosis": "true",
  "geo": {
    "latitude": "40.75",
    "longitude": "73.98"
  }
}
```

* The JSON-LD object is built after deidentification

Após a fase de processamento, os dados são armazenados na Blockchain por meio da API da rede. É necessário incluir credenciais de autenticação na solicitação e usar o protocolo HTTPS por motivos de segurança. A solicitação é síncrona e um objeto JSON contendo a transação Blockchain é retornado como resposta para confirmar que os dados foram armazenados corretamente. Uma vez recebida a resposta, o *hash* da transação, juntamente com o ID do registro do sistema EDC, são inseridos no banco de dados de metadados, de modo a

manter um histórico externo para permitir a rastreabilidade e facilitar a consulta das transações.

Acesso aos dados

Os dados armazenados nas transações são serializados e codificados em um formato binário chamado *Recursive Length Prefix* (RLP), que é um esquema de serialização de objetos eficiente usado no Ethereum (WOOD, 2014). O processo de serialização e codificação para escrita é tratado pela própria rede Blockchain. No entanto, o processo inverso deve ser realizado pelo usuário para se obter os dados em um formato legível por humanos. Para isso, é necessária uma biblioteca de programação que possa decodificar RLP, como a biblioteca web3.js (ETHEREUM, 2015).

A plataforma Kaleido oferece um recurso de explorador de blocos para navegar na rede Blockchain. No entanto, os dados armazenados não podem ser visualizados devido a limitações no formato RLP. Para decodificá-lo, o esquema (parâmetros) descrito na API da rede deve ser conhecido. Isso é definido na Interface Binária da Aplicação - do inglês, *Application Binary Interface* (ABI), uma maneira padrão de interagir com o ecossistema Ethereum (ETHEREUM, 2014).

Considerando que o explorador de blocos do Kaleido não conhece a ABI, ele não consegue decodificar os dados. Para resolver esse problema, aplicativos externos podem ser desenvolvidos para interagir com os dados da rede. Assim, foi desenvolvida uma ferramenta para permitir a visualização dos dados. Esta ferramenta faz parte do sistema utilizado na gestão de metadados, onde os usuários podem navegar por uma lista de transações que estão relacionadas a registros enviados em sistemas EDC, permitindo rastrear dados na rede. Além disso, o campo de dados é decodificado de acordo com a ABI usando a biblioteca web3.js. Após isso, todas as informações granulares, como os dados brutos em JSON e/ou a versão semântica em JSON-LD, o nome do formulário, o tipo de operação e o autor podem ser consultados. A Figura 25 mostra os detalhes de uma transação.

Figura 25. Detalhes da transação Ethereum

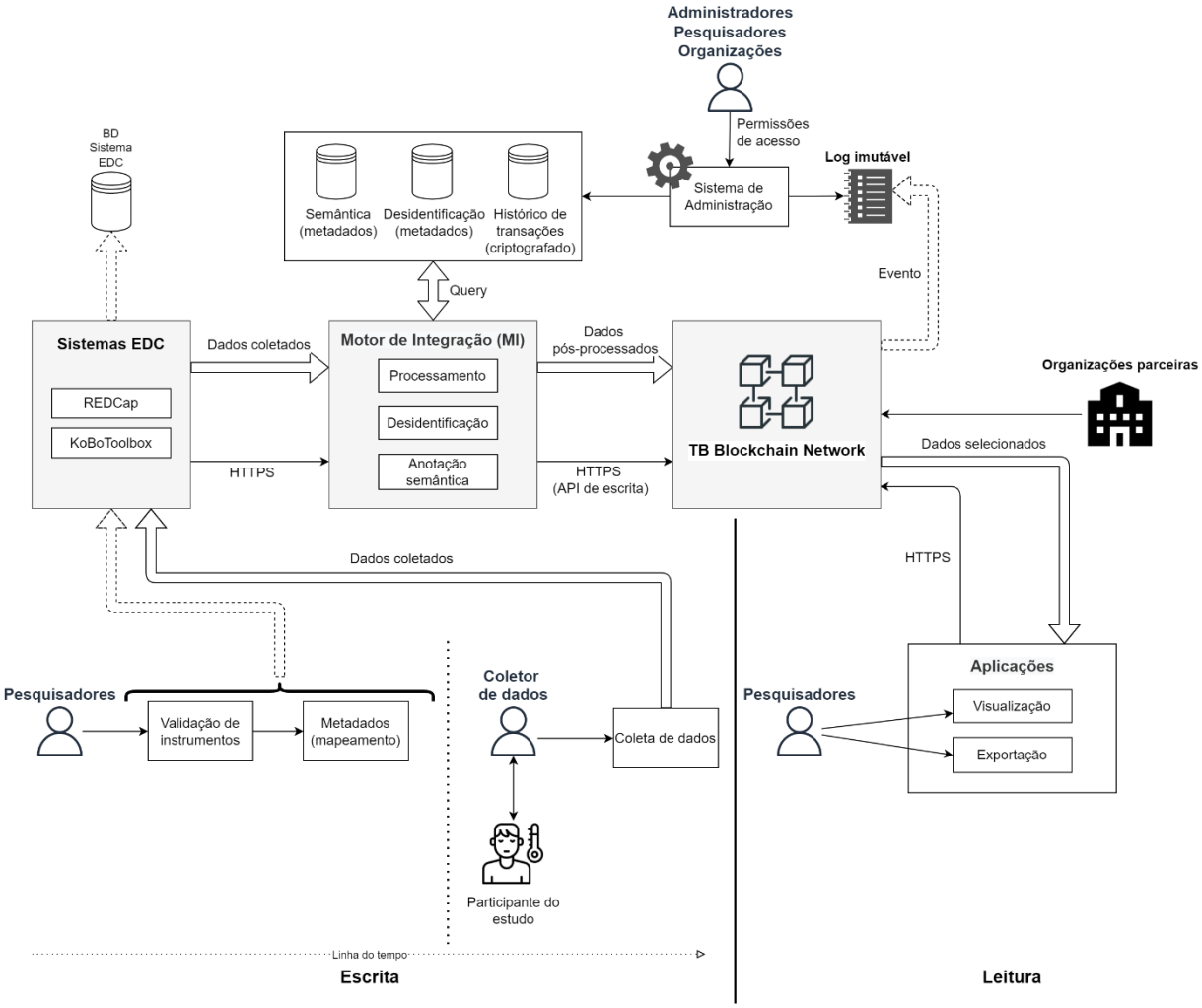
```
TRANSACTION DETAILS
from: 0x356d7897e1c145d7715ba8429cc9e826ec4f0d37
to: 0x219613b8aa1adb6fcb41027f5e93c4b68b1748eb
timestamp: 2020-11-04 16:59:03 -3:00
blockNumber: 667
blockHash: 0x244d1c1d07146dd70743f3335b00a85fef00fae9d2e35f15d129d2a46c622424
previousHash: 0xe859c77727b22dde7ddc0ef037ed6c5573b75a5ee5d82728d6f7ba1b17dedf1b
hash: 0x95feacf1dd87d9e57b68ca4c1fd679ac2f58bed8ebb45fa9a5444c526345e15c

DATA INFORMATION
Record ID: 24
Projeto ID: 5
Form: elegibility
Operação: add
Raw data hash: 1f342be2712af1b426cded1567c0c5276b34c3eb1f58a5f4f8bc3cb0640751ec

RAW DATA: { "gender": "male", "job_title": "construction worker", "hiv": "true",
"tuberculosis": "true", "geo": { "latitude": "40.75", "longitude": "73.98" } }
SEMANTIC DATA: //JSON-LD OMMITED - SAME AS FIG. 2
```

Outra opção para explorar dados é exportar registros. Isso é útil para análise estatística ou aprendizado de máquina, por exemplo. Ao selecionar o conjunto de dados desidentificado desejado, os usuários podem acessar os dados nos formatos JSON ou CSV, que podem ser facilmente interpretados. Além disso, é possível escolher entre os dados brutos ou com anotações semânticas, quando disponíveis. Os usuários também podem verificar a integridade dos dados comparando o *hash* na Blockchain com um *hash* autocalculado.

O processo de exportação de dados é baseado no histórico de transações. Esse histórico *off-chain* é criptografado e acessível apenas pelos módulos do sistema de gerenciamento (não há interação do usuário com esses dados), permitindo a rastreabilidade para verificar facilmente todas as modificações dos dados. Os usuários podem optar por exportar todas as versões dos dados ou apenas a última versão (a última coletada pelo EDC usando um determinado instrumento - se os dados salvos forem editados, uma nova versão será enviada à rede). No entanto, transações retiradas devido a erros humanos ou indisponibilidade de dados não serão incluídas nos resultados da exportação. Ao exportar os dados, a solução vincula os registros de um mesmo participante.

A Figura 26 resume o funcionamento da rede Blockchain.

Figura 26. Fluxo de informação e funcionamento da rede Blockchain



Considerações finais

A solução proposta satisfaz os requisitos das partes interessadas, nomeadamente doentes/participantes de estudos de investigação, investigadores e organizações de TB. Esforços foram feitos para aprimorar o processo de anonimização para respeitar e proteger a privacidade pessoal. Para pesquisadores e organizações, a solução é capaz de entregar qualquer tipo de dado (clínico, socioeconômico, genético, etc), desde que estejam disponíveis no conjunto de dados. Em relação aos requisitos gerais, a abordagem baseada em Blockchain é capaz de abordar questões de segurança de integridade, disponibilidade e privacidade de dados, devido às características inerentes à tecnologia.

O risco de reidentificação é reduzido se os identificadores forem mapeados corretamente. Além disso, o mecanismo de retração de dados ajuda a mitigar erros humanos. No entanto, não se pode negar que há uma pequena possibilidade de reidentificação se, por exemplo, um subconjunto de dados for combinado com algum conjunto de dados publicamente

disponível.

É importante observar que o conjunto de dados é construído em tempo real à medida que os dados são coletados em sistemas EDC. O conjunto de dados crescerá com o tempo e, no início da pesquisa, pode ser pequeno e impreciso. Além disso, a disponibilidade de dados com anotação semântica depende do mapeamento dos campos dos instrumentos de coleta.

7.6. Caso de Uso – Projeto Avaliação do LPA

Atualmente, a OMS recomenda o uso dos testes moleculares em linha, ou *Line Probe Assay* (LPA), em duas versões - Genotype MTBDRplus (LPA-1) e o MTBDRsl (LPA-2) - para a detecção de resistência aos principais fármacos de primeira e segunda linha, respectivamente, na investigação de MDR-TD e XDR-TD. A adoção destes testes moleculares aumentou o número de pacientes com DR/MDR/XDR-TB detectada e reduziu significativamente o tempo entre o aparecimento dos sintomas e o início do tratamento. Entretanto, são escassos os estudos de validação do uso do LPA na rotina diagnóstica da DR-TB em países de alta carga da doença.

Assim, considerando que os métodos LPA-1 e LPA-2 apresentam uma ótima escolha para avaliação de resistência em amostras clínicas obtidas de pacientes com suspeita de TB, foi estabelecido um protocolo clínico, por meio do projeto *Validação da performance do Line Probe Assay como método de diagnóstico rápido para tuberculose resistente em centros de referência no Brasil*, ou, resumidamente, projeto *Avaliação do LPA*, para analisar a acurácia dos testes moleculares LPA-1 e LPA-2, em condições de rotina, nos laboratórios participantes localizados nos estados do Amazonas, Bahia, Rio de Janeiro, e São Paulo. O projeto é financiado pelo PNCT da Secretaria de Vigilância em Saúde do Ministério da Saúde.

Uma particularidade do estudo está relacionada ao tipo de dados que determinados centros de pesquisa irão coletar, definidos como dados primários e secundários. Primários referem-se aos dados sociodemográficos, radiológicos e laboratoriais coletados por meio de fichas clínicas (CRFs) após a assinatura do Termo de Consentimento Livre Esclarecido (TCLE) pelo participante. Secundários são os dados coletados em SIS, nomeadamente, no TBWeb (para os centros do Estado de São Paulo), SINAN (para os centros do Estado do Rio de Janeiro), SITE-TB e GAL, e posteriormente transcritos para as fichas clínicas. Toda a coleta de dados é realizada através do sistema de coleta eletrônica REDCap.

O preenchimento das fichas clínicas, principalmente àquelas relacionadas aos dados secundários, revela-se uma tarefa repetitiva e dispendiosa devido à necessidade de acesso, pesquisa e transcrição de dados provenientes dos vários sistemas. Observa-se, então, que a disponibilidade de um sistema integrado, capaz de centralizar o acesso à informação, é

potencialmente vantajoso para agilizar o processo de busca e coleta de dados dos participantes do estudo. Ainda, no caso de dados secundários, a possibilidade de selecionar e importar dados para o REDCap pode contribuir com os objetivos de recrutamento dos centros de pesquisa e com a redução de atrasos na transcrição e disponibilização dos dados para a pesquisa.

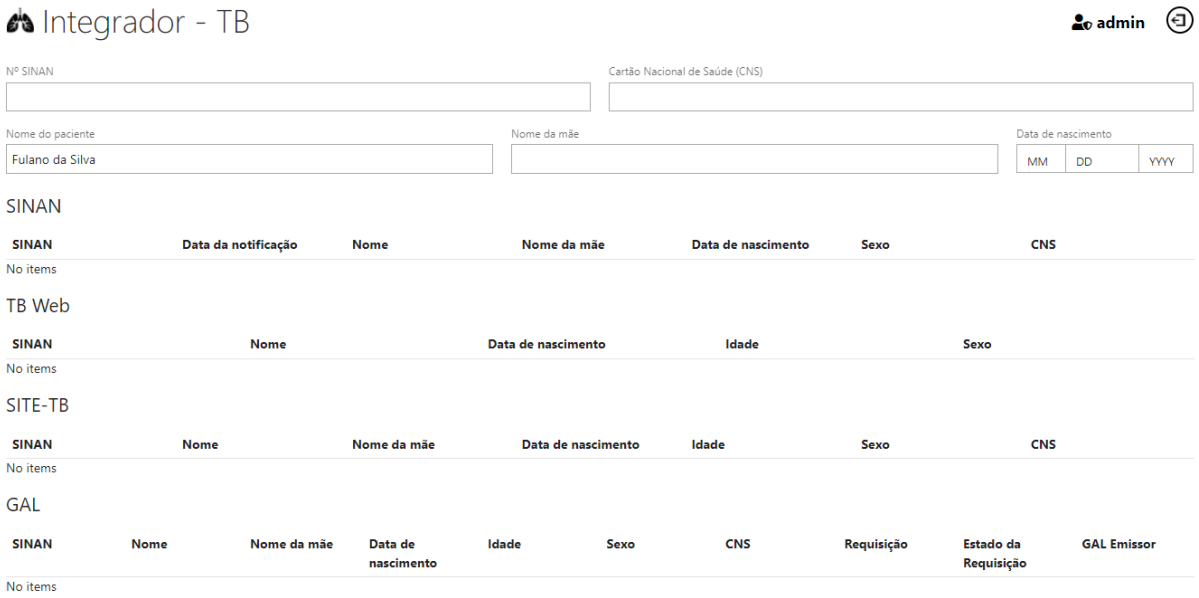
Assim, de forma a auxiliar na obtenção de dados nos SIS e apoiar as atividades de gerenciamento de dados, parte dos recursos desenvolvidos no contexto desta tese foram aplicados no Projeto Avaliação do LPA.

A necessidade de coleta de dados secundários, conforme referido anteriormente, faz com que a integração dos SIS relacionados à TB seja de grande interesse. Assim, uma ferramenta para consulta dos dados provenientes desses sistemas e importação para a base de dados do REDCap foi desenvolvida exclusivamente para o projeto[16]. Conforme descrito na seção 7.2.1, o Módulo de Interoperabilidade e Integração permite a integração de dados armazenados em planilhas e, posteriormente, possibilitará a integração em tempo real. As planilhas são, justamente, a maneira utilizada para disponibilizar dados para o projeto (além da consulta manual nos sistemas), as quais são preparadas e enviadas pelos responsáveis por cada sistema, mediante o envio de solicitações de acesso à informação no escopo do projeto de pesquisa.

O *software* de consulta/visualização permite buscar um paciente através de filtros pré-definidos. A Figura 27 apresenta a interface principal do sistema com as bases de dados do SINAN, TBWeb, SITETB e GAL. Quando um paciente é encontrado, uma linha é adicionada sob cada sistema (somente naqueles que o paciente está disponível). Ao selecionar uma linha, pode-se explorar o restante dos dados, de acordo com o dicionário de dados de cada sistema. Ainda, determinados dados podem ser importados para a base de dados do REDCap. O sistema permite a importação de campos previamente mapeados com as variáveis dos formulários de coleta disponíveis no REDCap. A importação ocorre mediante ação do usuário, indicando-se o número do participante registrado no REDCap. Todos os campos mapeados são importados, os quais devem ser posteriormente revisados.

[16] Disponível em https://admin.redbox.technology:3007/integrador-tb

Figura 27. Integrador da TB do Projeto Avaliação do LPA



Os outros recursos disponibilizados para o projeto incluem: a verificação de duplicidade de registros em tempo real, especialmente útil para os centros que coletam dados secundários de modo a evitar repetição equivocada de registros; o envio de alertas diários de prazos/janelas de inserção de dados, de modo a evitar atrasos e desvios de protocolo; a validação automática do preenchimento de alguns formulários, exceto daqueles que demandam revisão humana para garantia de precisão e veracidade da informação; e, finalmente, a transmissão dos dados desidentificados da base de dados de pesquisa para a Rede Blockchain TB, o que compõe gradualmente um repositório de dados compartilhado.

7.7. Resultados complementares

7.7.1. Cálculo de Indicadores Laboratoriais de Tuberculose[17]

Os indicadores para monitoramento e controle da TB refletem a atuação dos serviços de saúde na qualidade da assistência aos pacientes. No Brasil, em geral, estão relacionados à detecção, diagnóstico, coinfecção TB-HIV e desfecho dos casos da doença (FARIAS et al., 2020). Além disso, indicadores como a taxa de positividade de cultura, taxa de contaminação, taxa de positividade da baciloscopia e taxa de resistência a medicamentos são usados para avaliar a qualidade dos processos laboratoriais relacionados à TB. Os indicadores estão definidos em diversos documentos oficiais, como, por exemplo, no Manual de Recomendações

[17] A informação contida nesta subseção refere-se ao conteúdo do trabalho *Monitoring and evaluation indicators for tuberculosis based on laboratory routine data managed through REDCap*, apresentado na conferência HCist - *International Conference on Health and Social Care Information Systems and Technologies* (2021).

para o Controle da Tuberculose no Brasil (MINISTÉRIO DA SAÚDE, 2019a), e em outros manuais operacionais, dependendo dos objetivos do conjunto específico de indicadores.

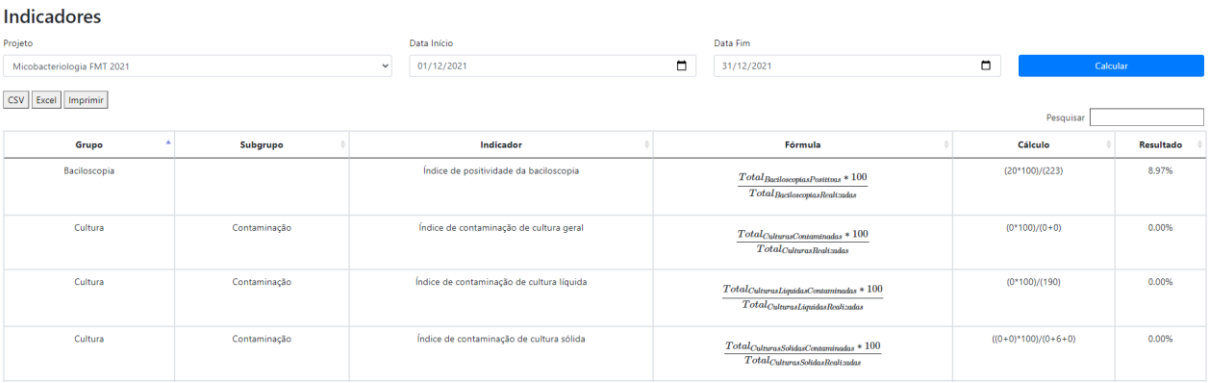
No Laboratório de Micobacteriologia da Fundação de Medicina Tropical Doutor Heitor Vieira Dourado, localizado em Manaus, Brasil, os dados de rotina laboratorial eram coletados em diversas planilhas independentes, dificultando a gestão e o uso dos dados. Devido ao grande volume de dados, o cálculo manual de indicadores periódicos obrigatórios é trabalhoso e propenso a erros.

Após uma série de interações com a equipe do laboratório, foi definida uma estratégia para alterar a forma de coleta de dados para o REDCap. As planilhas foram analisadas e mapeadas para definir instrumentos capazes de coletar os mesmos tipos de dados, porém com maior flexibilidade, praticidade e dinamismo. São coletados dados para o cadastro de pacientes, rastreamento de amostras e resultados de exames (baciloscopia, cultura, suscetibilidade antimicrobiana e TRM-TB). Em 2021, foram registrados dados de 2507 pacientes.

Para facilitar a tarefa de cálculo dos indicadores, foi desenvolvida uma ferramenta web, apresentada na Figura 28, na qual basta o usuário selecionar o conjunto de dados e o período de tempo desejado. Com base nas fórmulas pré-definidas, o sistema extrai apenas os dados necessários do REDCap e apresenta os indicadores calculados em uma tabela. O usuário pode exportar ou imprimir os resultados.

O objetivo principal é abordar indicadores de desempenho que ajudem a avaliar a qualidade dos processos laboratoriais relacionados à TB. A ferramenta desenvolvida se baseia em dados reais, não identificados, para entregar uma ferramenta amigável e adaptável a qualquer fórmula de cálculo, permitindo reduzir o tempo gasto em uma tarefa dispendiosa e repetitiva.

Figura 28. Interface do software para cálculo de indicadores laboratoriais da TB



Grupo	Subgrupo	Indicador	Fórmula	Cálculo	Resultado
Baciloscopia		Índice de positividade da baciloscopia	$\frac{Total_{BaciloscopiasPositivas} * 100}{Total_{BaciloscopiasRealizadas}}$	(20*100)/(223)	8.97%
Cultura	Contaminação	Índice de contaminação de cultura geral	$\frac{Total_{CulturasContaminadas} * 100}{Total_{CulturasRealizadas}}$	(0*100)/(0+0)	0.00%
Cultura	Contaminação	Índice de contaminação de cultura líquida	$\frac{Total_{CulturasLiquidasContaminadas} * 100}{Total_{CulturasLiquidasRealizadas}}$	(0*100)/(190)	0.00%
Cultura	Contaminação	Índice de contaminação de cultura sólida	$\frac{Total_{CulturasSolidasContaminadas} * 100}{Total_{CulturasSolidasRealizadas}}$	((0+0)*100)/(0+6+0)	0.00%

Cada indicador e as respectivas fórmulas são armazenados em um banco de dados relacional. O sistema exige um formato exclusivo para as fórmulas, para que as variáveis

possam ser substituídas por valores apropriados extraídos do REDCap através da sua API. O Quadro 3 apresenta a fórmula para cálculo da taxa de positividade da baciloscopia no formato humanamente legível e no formato esperado pelo software.

Quadro 3. Exemplo de fórmula - Taxa de positividade da baciloscopia

$$\frac{Total_{BaciloscopiasPositivas} * 100}{Total_{BaciloscopiasRealizadas}}$$

A fórmula acima é traduzida para o formato abaixo para interpretação pelo software.
Considera palavras-chave específicas e o nome dos campos (variáveis) dos dicionários de dados do REDCap.

(#COUNT[resultado_baciloscopia=positivo]#*100)/(#COUNT[realizacao_baciloscopia=sim]#)

7.7.2. Divulgação da relevância das estratégias digitais de saúde para apoiar os serviços de TB no contexto da pandemia de COVID-19[18]

A pandemia de COVID-19 é um dos surtos de doenças infecciosas mais devastadores registrados na história, que segue contabilizando óbitos diariamente. Ademais, a transmissibilidade observada em pacientes assintomáticos dificulta o controle da doença, desafiando a compreensão da dinâmica do vírus.

Para enfrentar o desafio de combater a TB durante a pandemia de COVID-19, a pesquisa e a inovação podem fornecer acesso equitativo a ferramentas e abordagens inovadoras. As tecnologias digitais podem ajudar pacientes e cuidadores a melhorar a experiência nas unidades de saúde e processos de tratamento, além de automatizar o gerenciamento das informações de saúde (WORLD HEALTH ORGANIZATION, 2015b, 2015c).

O breve texto apresentado em (LIMA et al., 2021d), em formato de Carta ao Editor, buscou apresentar as ações dos pesquisadores do LIS para oferecer suporte ao manejo e atenção aos pacientes de TB nos serviços públicos do Estado de São Paulo, como a plataforma para TDO remoto (ALBANO DOS SANTOS et al., 2019; CREPALDI et al., 2019), modelos de predição de abandono de tratamento (HOKINO YAMAGUTI et al., 2020) e um sistema capaz de monitorar e predizer a evolução dos casos de COVID-19 em uma região, de forma a emitir alertas antecipados aos serviços de TB.

[18] A informação contida nesta subseção refere-se ao conteúdo do texto *The relevance of digital health strategies to support tuberculosis services in the COVID-19 pandemic context*, publicado como um Carta ao Editor no periódico *Journal of Multiprofessional Health Research* (2021). ISSN 2675-8849. Disponível em: https://journalmhr.com/index.php/jmhr/article/view/16.

7.7.3. Sistemas de suporte à decisão para TB

Ferramentas de apoio à decisão são imprescindíveis para orientar a prática de cuidados em saúde e sustentar as decisões de gestores que irão influenciar diretamente na qualidade do atendimento prestado à população. A disponibilização de um conjunto de dados de referência como um Sistema de Suporte à Decisão (do inglês, DSS - *Decision Support System*) para integração, análise, comparação e visualização de dados saúde através da integração de fontes heterogêneas e dispersas é o principal aspecto a ser considerado.

DSS é um termo geral utilizado para qualquer sistema de informação capaz de aprimorar a capacidade de uma pessoa ou grupo de identificar problemas e tomar decisões que possam ajudar a resolver tais problemas (POWER; POWER, 2008). Esses sistemas podem ser categorizados como: orientados a dados, orientados a documentos, orientados a conhecimento e orientados a modelos. De acordo com sua categoria, os sistemas poderão tomar decisões através da coleta e processamento de dados, documentos não estruturados, regras de inferência e análise de entradas e parâmetros dos usuários, respectivamente.

Um DSS deve oferecer funcionalidades de análise inteligente, compilação, processamento e visualização de um grande volume de dados obtidos a partir de diversas fontes, automatizando a integração de bases de dados heterogêneas, como bases de informações administrativas, de dados clínicos, de sistemas epidemiológicos, da literatura, entre outros.

7.7.3.1. Revisão de escopo da literatura[19]

Considerando a relevância dos DSSs e sua potencial aplicabilidade na TB, foi proposto um protocolo de revisão de escopo para verificar e formalizar, pela primeira vez, a literatura disponível relacionada ao desenvolvimento e/ou implementação de sistemas de apoio à decisão para a Tuberculose. A proposta de revisão visa determinar os temas que os DSSs abordam na linha de atenção à TB, definir categorias, apontar os principais resultados e possíveis lacunas de pesquisa e identificar onde há falta de ferramentas para subsidiar os processos decisórios. Detalhes do protocolo podem ser consultados em (LIMA et al., 2019a).

A revisão permitiu identificar diversos trabalhos relevantes. Em geral, abrangem questões gerenciais/epidemiológicas e clínicas. Além disso, alguns estudos abordam a coinfecção TB/HIV, TB latente e TB resistente a medicamentos. No entanto, não foram encontradas referências a tecnologias da Web Semântica e ao compartilhamento seguro de

[19] A informação contida nesta subseção refere-se ao conteúdo do artigo *Decision support systems for Tuberculosis: protocol for a scoping review* publicado no periódico *Health Informatics Vision: From Data via Information to Knowledge* (2019). ISBN 978-1-61499-987-4. Disponível em https://doi.org/10.3233/SHTI190027.

dados.

Houben et al. definem um modelo para TB amigável e intuitivo que permite capacitar gestores locais e reforçar as discussões sobre políticas públicas no âmbito dos Programas Nacionais de Controle de Tuberculose, de modo a auxiliar gestores a compreender suas próprias epidemias de TB, planejar respostas, elaborar propostas para financiamentos e avaliar o impacto das mudanças implementadas (HOUBEN et al., 2016). O modelo é implementado no software *TIME Impact*, o qual permite pode ser personalizado para diferentes cenários epidemiológicos para explorar as atividades de tratamento e prevenção, incluindo casos de TB resistente e TB/HIV. O sistema inclui um conjunto de dados demográficos e estimativas de incidência, prevalência, mortalidade e notificações para vários países, obtidos a partir do *Global Tuberculosis Programme* (GTB). Por fim, os autores reforçam que o modelo *TIME Impact* serve como um ponto focal para discussões e deve ser considerado como parte do ciclo de planejamento dos programas nacionais de TB, fortalecendo os processos políticos e de tomada de decisão a nível nacional, além de contribuir com o aumento da transparência.

Zwerling et al. estabeleceram um modelo de análise de custo-efetividade de regimes de tratamento de TB de primeira linha (UFM) com uma estrutura altamente simplificada e capaz de fornecer resultados semelhantes em valor e interpretação a um modelo complexo (ZWERLING et al., 2016), direcionado a tomadores de decisão que demandam dados personalizados para sua realidade específica. O UFM está disponível através de um website interativo para permitir que os usuários visualizem os diferentes parâmetros disponíveis. Os parâmetros podem ser modificados para se obter rápidas análises de custo-efetividade. Os UFMs são limitados pela qualidade e precisão dos dados de entrada. Assim, os autores apontam que, quando projeções precisas são necessárias, essa abordagem pode não ser ideal e deve ser usada para informar decisões preliminares antes que modelos mais detalhados sejam criados. Portanto, modelos simplificados permitem uma maior flexibilidade nos parâmetros de entrada, facilitando a compreensão de fatores importantes em análises de custo-efetividade adaptados a realidades específicas, onde a construção de modelos mais detalhados pode não ser viável técnica e economicamente.

Nishikiori e Van Weezenbeek propuseram uma metodologia para modelar diferentes estratégias de busca ativa de casos (ACF, do inglês, *Active Case Finding*) para planejamento de projetos e intervenções a nível nacional, baseado na quantidade estimada de casos de TB pulmonar e seus custos associados (NISHIKIORI; VAN WEEZENBEEK, 2013). Um modelo foi projetado para estimar os parâmetros de custo-efetividade do ACF para diferentes populações-alvo em vários cenários epidemiológicos, representado por ferramenta para web

disponível para apoiar programas nacionais de tuberculose e parceiros na formulação de atividades de busca ativa de casos.

Dowdy et al. criaram uma ferramenta de modelagem simples e flexível que permite que usuários não especialistas definam seu cenário local de acordo com três parâmetros epidemiológicos (incidência de TB, proporção de novos casos de TB multirresistente e incidência de HIV em adultos) e custos unitários locais de diagnóstico e tratamento da TB (DOWDY et al., 2014). Tais dados são usados como estimativas em um *framework* de análise para gerar projeções de 5 anos de incidência de TB, mortalidade e custos de controle para nove estratégias diagnósticas. Uma versão está disponível gratuitamente para download e os usuários podem ajustar parâmetros para fornecer um ajuste mais preciso à sua própria região, modificando o software (um manual é fornecido).

No campo clínico, através de uma plataforma eletrônica, o *TB Consilium*, criado pela OMS e pela Sociedade Respiratória Europeia (ERS, do inglês *European Respiratory Society*) fornece conselhos baseados em evidências globalmente para comitês locais do programa e clínicos independentes sobre como lidar com casos de tuberculose resistente e outros casos de TB de difícil tratamento (como a coinfecção com o vírus HIV), contribuindo com a prevenção do aumento da resistência à medicamentos (D'AMBROSIO et al., 2017). O TB Consilium também permite monitorar e avaliar a qualidade das práticas clínicas em um determinado país ou cenário, em termos de prevenção, diagnóstico e tratamento com base em diretrizes baseadas em evidências. O *TB Consilium* opera com rapidez, fornecendo uma resposta às questões clínicas dentro de 48 horas. Após a submissão de uma solicitação de avaliação de caso pelo clínico, o qual especifica os detalhes e dúvidas através de um formulário eletrônico, o programa ativa um dos coordenadores clínicos, que escolhe dois especialistas para avaliar o caso. Outras funcionalidades incluem: apoio às questões transfronteiriças de migrantes; aconselhamento específico sobre TB por um médico, enfermeiro ou psicólogo mediante solicitação de pacientes, além de possibilitar pedidos de apoio motivacional a outros pacientes que sobreviveram à doença; e a extensão do suporte clínico relacionado a micobactérias não-tuberculosas.

Shabut et al. apresentam uma técnica para detectar anticorpos específicos de TB através da análise de imagens digitais de ensaios de imunoabsorção enzimática (ELISA, do inglês, *Enzyme-Linked Immunosorbent Assay*) com sinais de cores produzidos por biossensores (SHABUT et al., 2018). Os autores ressaltam que a portabilidade e a viabilidade técnica e financeira do método de diagnóstico automático de TB apresentado no sistema pode beneficiar milhões de pessoas, especialmente em locais remotos, onde poucos especialistas estão disponíveis.

Cazella, Feyh e Ben apresentam um aplicativo para dispositivos móveis, chamado *Tuber*, que disponibiliza o conteúdo de diretrizes e orientações clínicas para Tuberculose baseadas em papel (CAZELLA; FEYH; BEN, 2014). Como fonte de informação, foi utilizado o conteúdo do *Manual de Recomendações para o Controle da Tuberculose no Brasil*. A aplicação foi dividida em dois perfis definidos: 1) Crianças menores de dez anos e 2) Adultos ou adolescentes com idade maior ou igual a dez anos. Cada perfil segue um fluxo específico para calcular uma pontuação para rastrear pacientes com potencial TB. Os autores afirmam que o aplicativo abrange desde o diagnóstico da Tuberculose até o seu tratamento, além de auxiliar de forma eficiente no processo de tomada de decisão dos profissionais de saúde frente à demanda de tratamento nos centros de saúde, permitindo acesso rápido às informações necessárias para diagnosticar e tratar doentes com TB.

Osamor, Azeta e Ajulo apresentam um sistema web para pré-selecionar e aconselhar indivíduos com potencial TB, com base em perguntas respondidas relacionadas a sintomas comuns da doença (OSAMOR; AZETA; AJULO, 2014). O sistema utiliza um algoritmo baseado em regras para dois casos: onde há suspeita de TB e onde não há suspeita de TB. O objetivo principal é ajudar os médicos a reduzir a proporção paciente-médico nos hospitais e, finalmente, o tempo de consulta a longo prazo. O resultado da avaliação de usabilidade revela que o protótipo do aplicativo desenvolvido possui uma classificação de “Boa Usabilidade”.

Alguns trabalhos buscam aplicar a lógica difusa (Fuzzy) para construir sistemas de suporte à decisão para auxílio ao diagnóstico da TB. De maneira similar, os trabalhos de (SHARMA; KOCHHER, 2018), (SEMOGAN et al., 2011) e (SOUNDARARAJAN; SURESHKUMAR; ANUSUYA, 2012) propõem o desenvolvimento de um DSS para TB baseado em regras e lógica Fuzzy para determinar o estágio atual da doença com base na intensidade de sintomas (*scores* definidos por médicos especialistas).

O sistema desenvolvido por Hossain et al. permite a compreensão da relação entre sinais e sintomas e o grau de suspeita de TB de um paciente de forma explícita e transparente (HOSSAIN et al., 2017). O software é definido como um *Belief Rule Based Expert System* (BRBES), o qual pode fornecer uma plataforma de apoio à decisão para médicos para um diagnóstico com tempo reduzido e de baixo custo. O sistema tem a capacidade de lidar com vários tipos de incertezas para diagnosticar a TB, permitindo a medição do nível de suspeita da doença, levando em consideração seus vários sinais e sintomas. Os experimentos, realizados com base nos dados de 100 pacientes, mostram que os resultados gerados pelo BRBES são mais confiáveis do que os produzidos por um especialista humano, bem como por um sistema especialista baseado em regras difusas (Fuzzy).

Maidantchik et al. desenvolveram o sistema web NeuralTB, que visa apoiar o diagnóstico de TB em unidades de saúde em áreas de recursos limitados (MAIDANTCHIK et al., 2007). O sistema compreende redes neurais artificiais como modelo para diagnosticar a infecção e oferece interfaces amigáveis e intuitivas para registro de sintomas, monitoramento de pacientes e recuperação de resultados. Os dados coletados durante todo o processo podem ser usados para identificar possíveis novos fatores ou sintomas, uma vez que a transmissão da infecção pode evoluir. Essas informações também podem ser usadas por entidades governamentais para controle da TB.

7.7.3.2. Proposta de DSS integrado para TB[20]

Um modelo conceitual para a concepção de DSS integrado para TB (DSS-TB) foi proposto em (LIMA et al., 2019b), a ser alcançado através da transformação do Ecossistema da Tuberculose (CREPALDI et al., 2019), dos sistemas de informação governamentais e de *softwares* específicos de apoio à decisão identificados na literatura num conjunto integrado de ferramentas que permitirão a integração de dados de TB. Espera-se como principais contribuições do DSS-TB a redução da fragmentação da informação de saúde dos pacientes e a promoção da qualidade e disponibilidade dos dados.

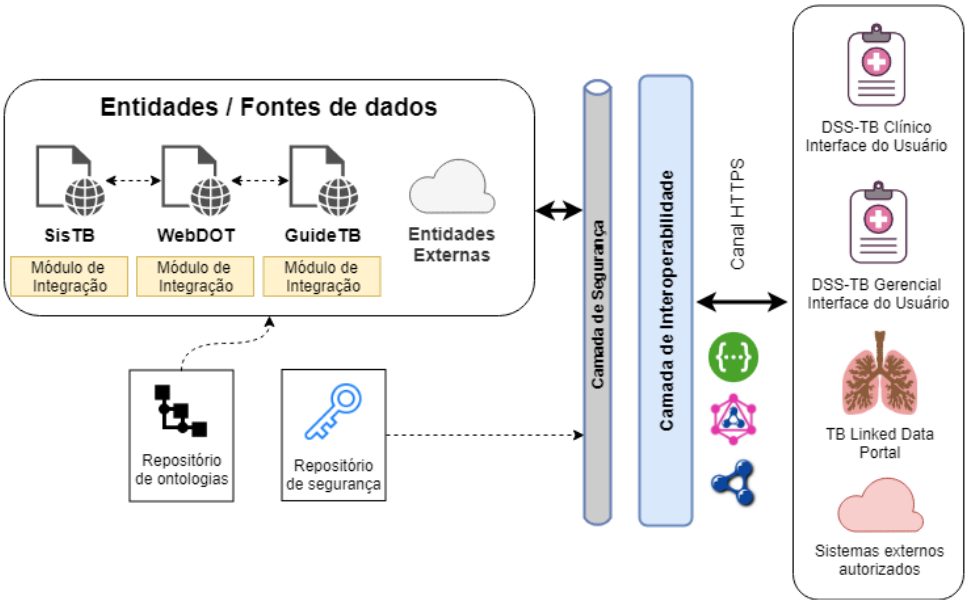
Propõe-se um sistema será multi-orientado, categorizado como orientado a dados (dados coletados de várias fontes), orientado a documentos (dados não estruturados de diretrizes, manuais, etc.), orientado a conhecimento (regras, procedimentos, etc.) e orientado a modelos (entradas do usuário).

O DSS-TB será composto por módulos integrados e categorizados de acordo com o seu objetivo, i.e., apoio clínico e gerencial/epidemiológico. Cada segmento fornecerá informações específicas para profissionais de saúde, gerentes e formuladores de políticas públicas. Todos os módulos contribuirão com informações relacionadas à Tuberculose. No entanto, também permanecerão disponíveis para serem usados separadamente em intervenções específicas. Além disso, para transparência, um Portal de Dados Vinculados TB estará disponível para disseminação pública de informações consolidadas de TB (LIMA et al., 2018a), bem como duas interfaces de usuário (*softwares* clientes) para interação visual com dados integrados, que conterão recursos relevantes de acordo com cada público-alvo. A Figura 29 ilustra o modelo conceitual do DSS-TB proposto.

---

[20] A informação contida nesta subseção refere-se ao conteúdo do artigo *Proposal of an integrated decision support system for Tuberculosis based on Semantic Web* publicado no periódico *Procedia Computer Science* (2019). ISSN 1877-0509. Disponível em https://doi.org/10.1016/j.procs.2019.12.219.

Figura 29. Modelo conceitual de DSS para TB



7.7.4. Guia rápido para profissionais de saúde – aplicativo para apoio ao cuidado do paciente de TB[21]

O uso de guias e manuais de procedimentos e recomendações como ferramentas de apoio à decisão é muito comum e também uma simples maneira de estruturar, organizar e obter alguma informação relevante que poderá impactar no atendimento e no cuidado ao paciente. Em geral, o conteúdo é disponibilizado como texto estático e com a informação apresentada de forma passiva e declarativa, em papel ou como um documento eletrônico no formato PDF.

Nesse sentido, nota-se que o acesso à informação contida nesses documentos não é simples, o que demanda um tempo considerável para o profissional de saúde selecionar a informação desejada devido, por exemplo, à dificuldade em encontrar uma seção específica. O uso de ferramentas digitais, como aplicativos móveis, pode simplificar o processo e promover o acesso à informação organizada.

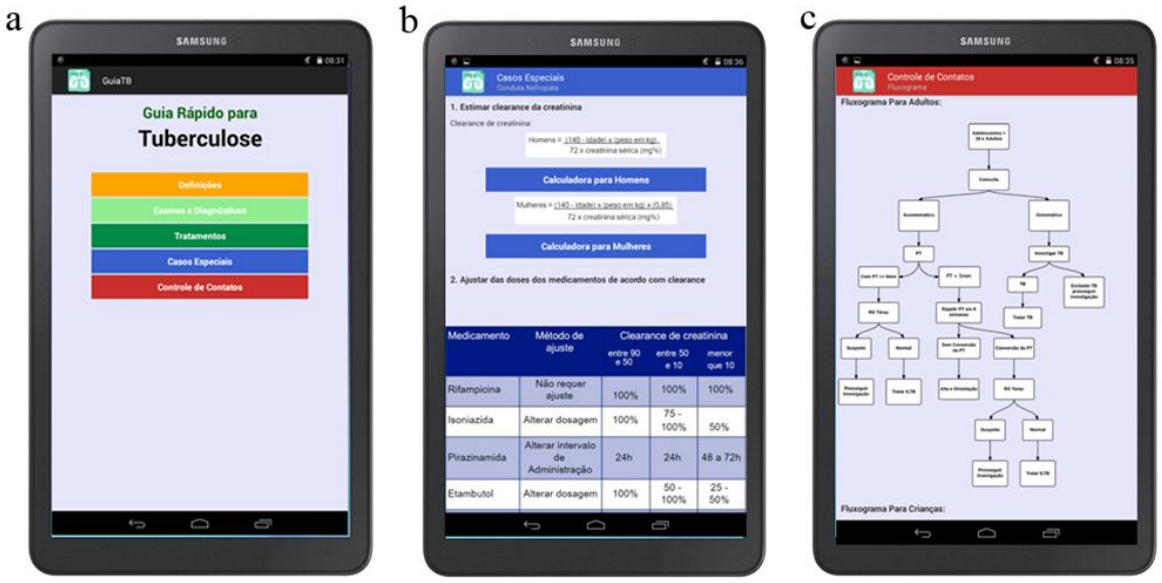
No contexto da TB, um protótipo funcional de aplicativo móvel foi desenvolvido com o conteúdo do documento "Recomendações para o Controle da Tuberculose - Guia Rápido para Profissionais de Saúde" (BRASIL, 2011c), elaborado pelo Ministério da Saúde e fornecido como um arquivo PDF, a qual foi disponibilizada para *download* através da loja de aplicativos *Google Play Store*, identificada como "Guia TB" (LIMA et al., 2017).

---

[21] A informação contida nesta subseção refere-se ao conteúdo do artigo *From guidelines to decision-making: using mobile applications and semantic web in the practical case of guides to support patients* publicado no periódico *Procedia Computer Science* (2017). ISSN 1877-0509. Disponível em https://doi.org/10.1016/j.procs.2017.11.104.

A Figura 30 apresenta a tela principal da aplicação (a), um exemplo de funcionalidade de calculadora para estimar a depuração de creatinina (b) e o fluxograma de rastreio da Tuberculose (c).

Figura 30. Protótipo Funcional para Android - (a) tela principal; (b) calculadora de depuração de creatinina



Através dos botões disponíveis na tela principal, é possível navegar pelo conteúdo do guia. Cada seção foi codificada com uma cor específica para facilitar a comunicação visual e identificação do assunto a ser abordado, da mesma maneira que é feito no guia em papel. As informações foram coletadas através da cópia e reprodução do conteúdo do documento em papel. Ressalta-se que os documentos são de domínio público e, basicamente, os dados estão disponíveis em forma de texto livre e imagens, os quais devem ser extraídos de forma manual.

O aplicativo permite a realização de buscas básicas para ajudar na obtenção da informação desejada. Além do conteúdo disponibilizado de forma estruturada, inclui calculadoras específicas que fornecem respostas rápidas para o usuário de acordo com a necessidade (ex: calculadora de depuração de creatinina). Ainda, a navegação baseada em fluxograma aprimora o processo de rastreamento da Tuberculose, permitindo que o usuário insira notas em cada estágio e realize buscas na web em bases de dados públicas sobre algum termo específico presente no guia.

Destaca-se, também, a possibilidade de se utilizar aplicativos como o GuiaTB como ferramentas educativas, devido a sua capacidade de disseminar informações relevantes sobre determinado domínio de saúde.

8. CONSIDERAÇÕES FINAIS

Este projeto de pesquisa teve como objetivo propor uma arquitetura de software, operacionalizada como uma plataforma computacional, capaz de disponibilizar recursos tecnológicos para a área da TB. Espera-se que os resultados desta tese contribuam com melhorias no acesso à informação de pacientes com TB, através do modelo de interoperabilidade e integração segura de diferentes SIS utilizados no contexto da doença (para, por exemplo, notificação, diagnóstico e tratamento), e na condução de pesquisas científicas por meio de uma experiência de coleta, gerenciamento e compartilhamento de dados aperfeiçoada. Procurou-se, ainda, alinhar as atividades desenvolvidas neste projeto com a *End TB Strategy* da OMS, com base, principalmente, no pilar de intensificação da investigação científica e inovação tecnológica.

Foi mostrado que a complexidade e heterogeneidade dos dados em saúde deve ser abordada em prol de um uso mais eficiente dos dados. Nesse sentido, a Web Semântica apresenta-se como uma ferramenta essencial para enfrentar esse cenário, dada a importância da semântica nos processos de troca e interpretação de informações entre SIS.

A interoperabilidade, integração e disponibilização de dados de TB pode aprimorar a qualidade dos dados e reduzir a carga de trabalho de profissionais de saúde e gestores e, consequentemente, garantir uma maior oferta e continuidade de cuidados de saúde aos indivíduos acometidos com a doença. É importante lembrar da necessidade da implementação de mecanismos de segurança que permitam o manuseio de dados de saúde sensíveis, que, no contexto deste projeto, foram disponibilizados à arquitetura de interoperabilidade semântica desenvolvida para o PNCT.

No que toca à investigação científica, ressalta-se sua importância na busca por novas formas de diagnóstico e tratamento da TB. O suporte às atividades de coleta e gerenciamento de dados tende a contribuir com a geração de resultados de maior qualidade. Portanto, a disponibilização de ferramentas tecnológicas com foco nessas atividades visa facilitar o processo como um todo e permitir ao pesquisador direcionar a maior parte do seu tempo na execução adequada do protocolo de pesquisa.

O compartilhamento de dados de pesquisas, por sua vez, permite o reuso e o embasamento de pesquisas futuras na área da TB. Nesse contexto, a Web Semântica, novamente, é essencial para agregar significado aos dados para aumentar sua interpretabilidade. Ainda, o uso de uma rede Blockchain como plataforma de compartilhamento de dados permite oferecer à comunidade científica um mecanismo robusto e seguro, capaz de armazenar e

distribuir dados em grande escala.

Finalmente, destaca-se que, embora todo o trabalho desenvolvido tenha sido motivado pelo cenário da TB no Brasil, os resultados apresentados são aplicáveis em outras áreas da saúde.

8.1. Limitações do trabalho realizado

A principal limitação deste trabalho refere-se à falta de acesso aos sistemas governamentais (seção 6.1.3) que manuseiam dados relacionados à TB. Apesar da possibilidade de acesso intermediada por alguns pesquisadores (por meio de credenciais de acesso individuais), essa abordagem não permite a interoperabilidade e integração dos sistemas, a qual demanda credenciais e documentação adequadas para se efetivar comunicação entre sistemas. As dificuldades de comunicação, os vários grupos de trabalho responsáveis por cada sistema e a inexistência de um procedimento para solicitação formal de acesso foram os principais entraves encontrados.

Procurou-se, então, encontrar alternativas que, embora limitadas, pudessem permitir uma integração parcial dos dados. Entretanto, o uso de planilhas é uma abordagem restritiva, pois a falta de relações entre determinados campos/variáveis e as dificuldades para se compreender os tipos e valores esperados resultam numa eventual perda de informações.

Por fim, os potenciais custos associados à plataforma Blockchain (utilizada no trabalho apresentado na seção 7.5) foram responsáveis pela não disponibilização da rede de compartilhamento de dados da TB para uso real. Inicialmente, a oferta de recursos gratuitos mostrava-se suficiente para uma versão inicial. Porém, esse cenário foi alterado, o que tornou a escalabilidade da rede muito limitada. Ressalta-se, no entanto, que o modelo proposto pode ser replicado para outra plataforma Blockchain.

8.2. Desenvolvimentos subsequentes e proposta de trabalhos futuros

Alguns pontos específicos podem ser abordados em trabalhos futuros. Em primeiro lugar, apesar de terem sido obtidas opiniões e *feedbacks* de pesquisadores, profissionais de saúde e usuários finais, é desejável proceder com uma validação formal das ferramentas propostas, de forma a conhecer a percepção de utilidade e usabilidade de cada recurso computacional proposto nesta tese.

A arquitetura de *software* proposta neste trabalho pode servir de base para a criação de sistemas de suporte à decisão na área da TB. Assim, novas ferramentas podem ser desenvolvidas para apoio clínico, operacional e administrativo.

No caso do *software* de integração e visualização de dados (“Integrador da TB”, ver seção 7.6) pode ser melhorado para oferecer uma navegação mais abrangente e precisa nos dados obtidos a partir dos diferentes sistemas de informação como, por exemplo, através da possibilidade de se pesquisar um conjunto de registros em cada sistema, através de filtros específicos, em contraste à forma de pesquisa atual centrada no indivíduo.

Por fim, no cenáro da TB, são desejáveis soluções para controle da qualidade dos dados armazenados em sistemas de informação em saúde e aqueles coletados em pesquisas científicas, além da condução de estudos analíticos para compreender o impacto operacional e financeiro gerado pela falta de interoperabilidade e integração dos sistemas de informação governamentais.

REFERÊNCIAS

ABHISHEK, K.; M.P., S. An Ontology based Decision support for Tuberculosis Management and Control in India. **International Journal of Engineering and Technology**, v. 8, n. 6, p. 2860–2877, 2016. Disponível em: <http://www.enggjournals.com/ijet/abstract.html?file=16-08-06-247>.

ABOUZAHR, C.; BOERMA, T. Health information systems: the foundations of public health. **Bulletin of the World Health Organizatio**, v. 83, p. 578–583, 2005.

ALBANESE, G. et al. Dynamic consent management for clinical trials via private blockchain technology. **Journal of Ambient Intelligence and Humanized Computing**, n. i, 2020. Disponível em: <https://doi.org/10.1007/s12652-020-01761-1>.

ALBANO DOS SANTOS, L. R. et al. The perception of health providers about an artificial intelligence applied to Tuberculosis video-based treatment in Brazil: a protocol proposal. **Procedia Computer Science**, v. 164, p. 595–601, 2019. Disponível em: <https://doi.org/10.1016/j.procs.2019.12.225>.

ALICE, M.; MENEZES, A.; TRAJMAN, A. Bottlenecks and recommendations for the incorporation of new technologies in the tuberculosis laboratory network in Brazil. **Jornal Brasileiro de Pneumologia**, v. 38, n. July, p. 766–770, 2012.

ALRAJA, M. N. et al. An integrated framework for privacy protection in IoT — Applied to smart healthcare. **Computers and Electrical Engineering**, v. 91, n. February 2020, p. 107060, 2021. Disponível em: <https://doi.org/10.1016/j.compeleceng.2021.107060>.

ANDRADE, A. Q. de. **A Tomada de Decisão e os Sistemas de Informação em Saúde**. 2008. Universidade Federal de Minas Gerais, 2008.

ANDREAS, F. Data Life Cycle Management Pilot Projects and Implications for Research Data Management at Universities of Applied Sciences. 2021.

ARKSEY, H.; O'MALLEY, L. Scoping studies: Towards a methodological framework. **International Journal of Social Research Methodology: Theory and Practice**, v. 8, n. 1, p. 19–32, 2005.

BART, T. Comparison of electronic data capture with paper data collection: Is there really an advantage? **Business Briefing**, v. 30, n. March 2002, p. 1–4, 2003. Disponível em: <http://www.dreamslab.it/dlab/media/docs/eclinica.pdf>.

BARTLING, S. **Blockchain for Science**. Disponível em: <https://www.blockchainforscience.com/>. Acesso em: 27 abr. 2020.

BAX, M. P. Uso de Dicionário Semântico de Dados na anotação de modelos de dados dimensionais para geração de indicadores de desempenho. p. 128–141, 2021.

BEIMEL, D.; PELEG, M. Using OWL and SWRL to represent and reason with situation-based access control policies. **Data and Knowledge Engineering**, v. 70, n. 6, p. 596–615, 2011. Disponível em: <http://dx.doi.org/10.1016/j.datak.2011.03.006>.

BELLEAU, F. et al. Bio2RDF: Towards a mashup to build bioinformatics knowledge

systems. **Journal of Biomedical Informatics**, v. 41, n. 5, p. 706–716, 2008.

BENCHOUFI, M.; RAVAUD, P. Blockchain technology for improving clinical research quality. **Trials**, v. 18, n. 1, p. 1–5, 2017.

BERNARDI, F. et al. Blockchain Based Network for Tuberculosis: A Data Sharing Initiative in Brazil. In: Studies in Health Technology and Informatics, **Anais**...IOS Press, 2019. Disponível em: <http://ebooks.iospress.nl/ISBN/978-1-61499-987-4>.

BERNERS-LEE, T.; HENDLER, J.; LASSILA, O. The Semantic Web. A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities. **Scientific American**, v. 284, n. 5, p. 34–43, 2001.

BHANDARY, M.; PARMAR, M.; AMBAWADE, D. A Blockchain Solution based on Directed Acyclic Graph for IoT Data Security using IoTA Tangle. **Proceedings of the Fifth International Conference on Communication and Electronics Systems (ICCES 2020)**, n. ICCES, p. 827–832, 2020a.

BHANDARY, M.; PARMAR, M.; AMBAWADE, D. Securing Logs of a System - An IoTA Tangle Use Case. **Proceedings of the International Conference on Electronics and Sustainable Communication Systems, ICESC 2020**, n. ICESC, p. 697–702, 2020b.

BHARDWAJ, R.; DATTA, D. Consensus Algorithm. In: KHAN, M. A. et al. (Ed.). **Decentralised Internet of Things: A Blockchain Perspective**. Cham: Springer International Publishing, 2020. p. 91–107.

BIZER, C.; CYGANIAK, R. D2R Server - Publishing Relational Databases on the Semantic Web. **5th International Semantic Web Conference**, n. June, p. 26, 2006. Disponível em: <http://www4.wiwiss.fu-berlin.de/bizer/pub/Bizer-Cyganiak-D2R-Server-ISWC2006.pdf>.

BIZER, C.; SEABORNE, A.; CYGANIAK, R. **D2RQ - Accessing Relational Databases as Virtual RDF Graphs**. Disponível em: <http://d2rq.org/>. Acesso em: 8 abr. 2020.

BODIN, L. D.; GORDON, L. A.; LOEB, M. P. Evaluating information security investments using the analytic hierarchy process. **Communications of the ACM**, v. 48, n. 2, p. 78–83, 2005. Disponível em: <http://portal.acm.org/citation.cfm?doid=1042091.1042094>.

BONIFACE, M.; WILKEN, P. ARTEMIS: Towards a secure interoperability infrastructure for healthcare information systems. **Studies in Health Technology and Informatics**, v. 112, p. 181–189, 2005.

BOUSSI RAHMOUNI, H. et al. A model-driven privacy compliance decision support for medical data sharing in Europe. **Methods of Information in Medicine**, v. 50, n. 4, p. 326–336, 2011.

BRASIL. MINISTÉRIO DA SAÚDE. **A experiência brasileira em sistemas de informação em saúde**. [s.l: s.n.]v. 2

BRASIL. Lei nº 8.080. . 1990.

BRASIL. Lei Geral de Proteção de Dados Pessoais. . 2018.

BRASIL, G. do E. de S. P. **Guia de Web Semântica**. [s.l: s.n.]

BRASIL, G. do E. de S. P. **Guia de Dados Abertos**. [s.l: s.n.]

BRASIL, M. da S.-S. de V. em S. **Manual de recomendações para o controle da tuberculose no Brasil**. [s.l: s.n.]

BRASIL, M. da S. **A experiência brasileira em sistemas de informação em saúde**. [s.l: s.n.]v. 1

BRASIL, M. da S. Doenças infecciosas e parasitárias: guia de bolso. **Cadernos de Saúde Pública**, v. 22, n. 11, p. 2498–2498, 2010.

BRASIL, M. da S. Portaria nº 2073. . 2011 b.

BRASIL, M. da S. **Recomendações para o Controle da Tuberculose - Guia Rápido Para Profissionais da Súde**, 2011c. .

BRASIL, M. da S. Portaria nº 204. . 2016.

BURANARACH, M.; SUPNITHI, T.; CHALORTHAM, N. A Semantic Web Framework to Support Knowledge Management in Chronic Disease Healthcare. **Metadata and Semantic Research**, p. 164–170, 2009. Disponível em: <http://dx.doi.org/10.1007/978-3-642-04590-5_15>.

BURNETTE, M.; WILLIAMS, S.; IMKER, H. From Plan to Action: Successful Data Management Plan Implementation in a Multidisciplinary Project. **Journal of eScience Librarianship**, v. 5, n. 1, p. e1101, 2016.

CAN, O.; YILMAZER, D. Improving privacy in health care with an ontology-based provenance management system. **Expert Systems**, v. 37, n. 1, p. 1–18, 2020.

CARDOSO, G.; NETO, C.; CHIORO, A. Afinal, quantos Sistemas de Informação em Saúde de base nacional existem no Brasil? **Cadernos de Saúde Pública**, v. 37, n. 7, 2021.

CAZELLA, S. C.; FEYH, R.; BEN, Â. J. A Decision Support System for Medical Mobile Devices Based on Clinical Guidelines for Tuberculosis. **Advances in Intelligent Systems and Computing**, 2014.

CHIANG, T.; HUANG, Y. Group Key and the Multicast Security in Ad Hoc Networks. **Proceedings of the 2003 International Conference on Parallel Processing Workshops (ICPPW03) IEEE**, 2003.

CLEMENTS, P. et al. Documenting software architectures: Views and beyond. **Proceedings - International Conference on Software Engineering**, n. June, p. 740–741, 2003.

CONSENSYS. **Kaleido Blockchain Platform web site**. Disponível em: <https://kaleido.io/>. Acesso em: 18 abr. 2020.

CREPALDI, N. Y. et al. SISTB: an ecosystem for monitoring TB. **Procedia Computer Science**, v. 164, p. 587–594, 2019. Disponível em: <https://doi.org/10.1016/j.procs.2019.12.224>.

CUNHA, M.; JÚNIOR, M. S.; ALMEIDA, H. Dificuldades com Integração e Interoperabilidade de Sistemas de Informação nas Instituições Públicas de Ensino: um Estudo de Caso no CEFET-AL. **XII Simpósio de Engenharia de …**, 2005.

CYGANIAK, R.; BIZER, C. D2R Server : A Semantic Web Front-end to Existing Relational Databases. **XML Tage**, p. 2–4, 2006.

D'AMBROSIO, L. et al. A new free-cost e-service supporting clinicians to manage their difficult-to treat tuberculosis cases: The European Respiratory Society-World Health Organization tuberculosis Consilium. 2017.

DAO, T. T. et al. Knowledge-based personalized search engine for the Web-based Human Musculoskeletal System Resources (HMSR) in biomechanics. **Journal of Biomedical Informatics**, v. 46, n. 1, p. 160–173, 2013. Disponível em: <http://dx.doi.org/10.1016/j.jbi.2012.11.001>.

DDI WORKING GROUP, D. D. I. Data Documentation Initiative (DDI) Technical Specification Part I : Overview. n. October, 2009.

DENKER, G.; KAGAL, L.; FININ, T. Security in the Semantic Web using OWL. **Information Security Technical Report**, v. 10, n. 1, p. 51–58, 2005.

DERSINGH, A.; LISCANO, R.; JOST, A. Context-aware Access Control Using Semantic Policies. **Special issue on Autonomic Computing Systems and Applications ,Ubiquitous Computing and Communication Journal, UBICC**, v. ACSA-Spe, p. 19–32, 2008.

DI VIMERCATI, S.; FORESTI, S.; SAMARATI, P. Authorization and Access Control. In: PETKOVIĆ, M.; JONKER, W. (Ed.). **Security, Privacy, and Trust in Modern Data Management**. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007. p. 39–53.

DILLON, D. G. et al. Open-source electronic data capture system offered increased accuracy and cost-effectiveness compared with paper methods in Africa. **Journal of Clinical Epidemiology**, v. 67, n. 12, p. 1358–1363, 2014. Disponível em: <http://dx.doi.org/10.1016/j.jclinepi.2014.06.012>.

DING, C.; WALD, M.; WILLS, G. Linked Data-driven decision support for accessible travelling. **W4A 2015 - 12th Web for All Conference**, p. 4–5, 2015. Disponível em: <http://www.scopus.com/inward/record.url?eid=2-s2.0-84953391506&partnerID=40&md5=dff6628a6c1cc47ba4dff9638d88ef17>.

DIXIT, S.; JOSHI, K. P.; GEOL CHOI, S. Multi Authority Access Control in a Cloud EHR System with MA-ABE. **Proceedings - 2019 IEEE International Conference on Edge Computing, EDGE 2019 - Part of the 2019 IEEE World Congress on Services**, p. 107–109, 2019.

DONG, X.; SAMAVI, R.; TOPALOGLOU, T. COC: An ontology for capturing semantics of circle of care. **Procedia Computer Science**, v. 63, p. 589–594, 2015. Disponível em: <http://dx.doi.org/10.1016/j.procs.2015.08.389>.

DOWDY, D. W. et al. A user-friendly, open-source tool to project impact and cost of diagnostic tests for tuberculosis. **eLife**, v. 3, p. 1–24, 2014.

DRIDI, A.; SASSI, S.; FAIZ, S. Towards a semantic medical internet of things. **Proceedings of IEEE/ACS International Conference on Computer Systems and Applications, AICCSA**, v. 2017- Octob, p. 1421–1428, 2018.

DU, M. et al. A review on consensus algorithm of blockchain. **2017 IEEE International Conference on Systems, Man, and Cybernetics, SMC 2017**, v. 2017- Janua, p. 2567–2572, 2017.

DURMUŞ, V.; UYDACI, M. A Legal Framework for Healthcare: Personal Data Protection for Health Law in Turkey. In: **Handbook of Research on Intrusion Detection Systems**. [s.l.] IGI Global, 2020. p. 219–236.

ESCOLA NACIONAL DE ADMINISTRAÇÃO PÚBLICA. Introdução à Interoperabilidade. p. 1–15, 2015.

ETHEREUM. **Contract ABI Specification**. Disponível em: <https://solidity.readthedocs.io/en/v0.6.6/abi-spec.html>. Acesso em: 22 abr. 2020.

ETHEREUM. **web3.js - Ethereum JavaScript API**. Disponível em: <https://github.com/ethereum/web3.js>. Acesso em: 22 abr. 2020.

EUROPEAN COMISSION. **COMMISSION RECOMMENDATION of 2 July 2008 on cross-border interoperability of electronic health record systemsOfficial Journal of the European Union**. [s.l: s.n.].

FANTONELLI, M. et al. Lei geral de proteção de dados e a interoperabilidade na saúde pública. . 2020, p. 166–171.

FARIAS, R. C. et al. Indicadores Operacionais do Controle da Tuberculose no Município de Belém-Pará. **Cogitare Enfermagem**, v. 25, 2020.

FERREIRA, M. I.; FELIX, N. Interoperabilidade numa perspectiva Hospitalar. p. 1–17, 2009.

FITZGERALD, D. W.; STERLING, T. R.; HAAS, D. W. Mycobacterial diseases: mycobacterium tuberculosis. **Principles and practice of infectious diseases**, p. 2852–2886, 2005.

GARLAN, D. Software Architecture: a Roadmap David Garlan. **Design**, 1999.

GERACI, A. et al. **IEEE Standard Computer Dictionary. A Compilation of IEEE Standard Computer Glossaries**. [s.l: s.n.]

GERBRACHT, S. Possibilities to encrypt an RDF-graph. **2008 3rd International Conference on Information and Communication Technologies: From Theory to Applications, ICTTA**, 2008.

GEZER, V.; BERGWEILER, S. Service and Workflow Engineering based on Semantic Web Technologies. **The Tenth International Conference on Mobile Ubiquitous Computing, Systems, Services and Technologies. International Conference on Mobile Ubiquitous Computing, Systems, Services and Technologies (UBICOMM-16), located at Tenth International Conference on M**, n. October, 2016.

GRODZINSKY, F. S.; TAVANI, H. T. P2P networks and the Verizon v. RIAA case: Implications for personal privacy and intellectual property. **Ethics and Information Technology**, v. 7, n. 4, p. 243–250, 2005.

HARRIS, P. A. et al. Research electronic data capture (REDCap)-A metadata-driven methodology and workflow process for providing translational research informatics support. **Journal of Biomedical Informatics**, v. 42, n. 2, p. 377–381, 2009. Disponível em: <http://dx.doi.org/10.1016/j.jbi.2008.08.010>.

HARVARD, H. I. **KoBoToolbox Website**. Disponível em: <https://www.kobotoolbox.org/>. Acesso em: 11 abr. 2020.

HAUX, R. Health information systems - Past, present, future. **International Journal of Medical Informatics**, v. 75, n. 3- 4 SPEC. ISS., p. 268–281, 2006.

HITZLER, P.; JANOWICZ, K. Semantic Web – Interoperability , Usability , Applicability. v. 0, n. 0, p. 1–2, 2011.

HOKINO YAMAGUTI, V. et al. Development of CART model for prediction of tuberculosis treatment loss to follow up in the state of São Paulo, Brazil: A case–control study. **International Journal of Medical Informatics**, v. 141, n. December 2019, 2020.

HOSSAIN, M. S. et al. A Belief Rule Based Expert System to Assess Tuberculosis under Uncertainty. **Journal of Medical Systems**, v. 41, n. 3, 2017.

HOUBEN, R. M. G. J. et al. TIME Impact - a new user-friendly tuberculosis (TB) model to inform TB policy decisions. **BMC Medicine**, v. 14, n. 1, p. 1–10, 2016. Disponível em: <http://dx.doi.org/10.1186/s12916-016-0608-4>.

HUERTAS CELDRÁN, A. et al. Preserving patients' privacy in health scenarios through a multicontext-aware system. **Annales des Telecommunications/Annals of Telecommunications**, v. 72, n. 9–10, p. 577–587, 2017.

HUGHES, L. et al. Blockchain research, practice and policy: Applications, benefits, limitations, emerging research themes and research agenda. **International Journal of Information Management**, v. 49, n. February, p. 114–129, 2019. Disponível em: <https://doi.org/10.1016/j.ijinfomgt.2019.02.005>.

HUMPHREY, C. e-Science and the Life Cycle of Research. p. 1–5, 2005.

IDEHEN, K. Introducing OpenLink Virtuoso: Universal Data Access without Boundaries. 2013.

IROJU, O. et al. Interoperability in Healthcare: Benefits, Challenges and Resolutions. **International Journal of Innovation and Applied Studies**, v. 3, n. 1, p. 262–270, 2013. Disponível em: <http://www.ijias.issr-journals.org/abstract.php?article=IJIAS-13-090-01>.

ISABEL, R. et al. The performance of health services in providing Directly Observed Treatment (DOT) in households to control. **Revista da Escola de Enfermagem da USP**, v. 42, n. 4, p. 627–633, 2008a.

ISABEL, R. et al. Desempenho de serviços de saúde no Tratamento Diretamente Observado

no domicílio para controle da tuberculose. **Revista da Escola de Enfermagem da USP**, v. 42, n. 4, p. 628–634, 2008b.

ISMAIL, L. et al. Towards a blockchain deployment at UAE University: Performance evaluation and blockchain taxonomy. **ACM International Conference Proceeding Series**, v. Part F1481, p. 30–38, 2019.

JACCOUD, L.; VIEIRA, F. S. **FEDERALISMO, INTEGRALIDADE E AUTONOMIA NO SUS: DESVINCULAÇÃO DA APLICAÇÃO DE RECURSOS FEDERAIS E OS DESAFIOS DA COORDENAÇÃO**. [s.l: s.n.].

JIANG, G.; SOLBRIG, H. R.; CHUTE, C. G. Using Semantic Web technology to support icd-11 textual definitions authoring. **Journal of Biomedical Semantics**, v. 4, n. 1, p. 11, 2013. Disponível em: <http://jbiomedsem.biomedcentral.com/articles/10.1186/2041-1480-4-11>.

JOSHI, K. P.; YESHA, Y.; FININ, T. An Ontology for a HIPAA compliant cloud services. In: 4th International IBM Cloud Academy Conference ICACON 2016, **Anais**...2016.

JP MORGAN CHASE. Quorum Whitepaper. p. 1–8, 2016. Disponível em: <https://www.blocksg.com/single-post/2017/12/27/Quorum-Whitepaper%0Ahttps://github.com/jpmorganchase/quorum-docs/blob/master/Quorum Whitepaper v0.1.pdf>.

KARTHIK, N.; ANANTHANARAYANA, V. S. An Ontology Based Trust Framework for Sensor-Driven Pervasive Environment. **AMS 2017 - Asia Modelling Symposium 2017 and 11th International Conference on Mathematical Modelling and Computer Simulation**, p. 147–152, 2018.

KESSLER, C.; HENDRIX, C. The Humanitarian eXchange Language: Coordinating disaster response with semantic web technologies. **Semantic Web**, v. 6, n. 1, p. 5–21, 2015.

KITSOS, P.; PAPPA, P. Mobile Communications Privacy. In: **Encyclopedia of Information Science and Technology, Third Edition**. [s.l.] IGI Global, 2015.

KLIPIN, M. et al. The process of installing REDCap, a web based database supporting biomedical research: the first year. **Applied clinical informatics**, v. 5, n. 4, p. 916–929, 2014.

KÕLVART, M.; POOLA, M.; RULL, A. Smart Contracts. In: KERIKMÄE, T.; RULL, A. (Ed.). **The Future of Law and eTechnologies**. Cham: Springer International Publishing, 2016. p. 133–147.

KONDYLAKIS, H. et al. Flexible access to patient data through e-consent. **MOBIHEALTH 2015 - 5th EAI International Conference on Wireless Mobile Communication and Healthcare - Transforming Healthcare through Innovations in Mobile and Wireless Technologies**, n. September, 2015.

KONOPKA, B. M. Biomedical ontologies - A review. **Biocybernetics and Biomedical Engineering**, v. 35, n. 2, p. 75–86, 2015. Disponível em: <http://dx.doi.org/10.1016/j.bbe.2014.06.002>.

KOWALCZYK, S. T. Modelling the Digital Research Data Lifecycle. v. 12, p. 331–361,

[s.d.]

KRISHNANKUTTY, B. et al. Data management in clinical research : An overview. **Indian Journal of Pharmacology**, v. 44, n. 2, p. 168–173, 2012.

KUMAR, Y.; MUNJAL, R.; SHARMA, H. Comparison of Symmetric and Asymmetric Cryptography with Existing Vulnerabilities and Countermeasures. **IJCSMS International Journal of Computer Science and Management Studies**, v. 11, n. 03, p. 2231–5268, 2011. Disponível em: <www.ijcsms.com>.

LAW, E. L. C. et al. Understanding, scoping and defining user experience: A survey approach. In: Conference on Human Factors in Computing Systems - Proceedings, **Anais**...2009.

LI, Z.; CHU, C. H.; YAO, W. A semantic authorization model for pervasive healthcare. **Journal of Network and Computer Applications**, v. 38, n. 1, p. 76–87, 2014. Disponível em: <http://dx.doi.org/10.1016/j.jnca.2013.06.006>.

LIANG, X. et al. Integrating blockchain for data sharing and collaboration in mobile healthcare applications. **IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, PIMRC**, v. 2017- Octob, p. 1–5, 2018.

LIMA, R. R. de et al. Development of a Knowledge Portal for Tuberculosis based on Semantic Web. **Book of industry papers, poster papers and abstracts of the HCist 2018 - International Conference on Health and Social Care Information Systems and Technologies**, p. 287–290, 2018a.

LIMA, J. C. D.; CARVALHO, C. L. D. Ontologias - OWL (Web Ontology Language). p. 22, 2005.

LIMA, V. et al. Decision Support Systems for Tuberculosis: Protocol for a Scoping Review. In: Studies in Health Technology and Informatics, **Anais**...IOS Press, 2019a. Disponível em: <http://ebooks.iospress.nl/ISBN/978-1-61499-987-4>.

LIMA, V. et al. Proposal of an integrated decision support system for Tuberculosis based on Semantic Web. **Procedia Computer Science**, v. 164, p. 552–558, 2019b. Disponível em: <https://doi.org/10.1016/j.procs.2019.12.219>.

LIMA, V. et al. A Mechanism for Verifying the Integrity and Immutability of Tuberculosis Data sing IOTA Distributed Ledger Technology. **Studies in Health Technology and Informatics**, v. 279, p. 130–135, 2021a.

LIMA, V. C. et al. From guidelines to decision-making: Using mobile applications and semantic web in the practical case of guides to support patients. **Procedia Computer Science**, v. 121, p. 803–808, 2017. Disponível em: <https://doi.org/10.1016/j.procs.2017.11.104>.

LIMA, V. C. et al. Establishment of access levels for health sensitive data exchange through semantic web. **Procedia Computer Science**, v. 138, p. 191–196, 2018b. Disponível em: <https://doi.org/10.1016/j.procs.2018.10.027>.

LIMA, V. C. et al. Security Framework for Tuberculosis Health Data Interoperability Through the Semantic Web. **International Journal of Web Portals**, v. 13, n. 2, p. 36–57,

2021b.

LIMA, V. C. et al. A Permissioned Blockchain Network for Security and Sharing of De-identified Tuberculosis Research Data in Brazil. **Methods of Information in Medicine**, 2021c.

LIMA, V. C. et al. The relevance of digital health strategies to support tuberculosis services in the COVID-19 pandemic context. **Journal of Multiprofessional Health Research**, p. 72–74, 2021d.

LIU, Z.; WANG, J. A fine-grained context-aware access control model for health care and life science linked data. **Multimedia Tools and Applications**, v. 75, n. 22, p. 14263–14280, 2016. Disponível em: <http://dx.doi.org/10.1007/s11042-016-3269-6>.

LIYANAGE, H.; KRAUSE, P.; LUSIGNAN, S. de. Using ontologies to improve semantic interoperability in health data. **Journal of Innovation in Health Informatics**, v. 22, n. 2, p. 309–315, 2015. Disponível em: <http://hijournal.bcs.org/index.php/jhi/article/view/159/838>.

LOPES, P.; OLIVEIRA, J. L. A semantic web application framework for health systems interoperability. **Proceedings of the first international workshop on Managing interoperability and complexity in health systems - MIXHS '11**, p. 87, 2011. Disponível em: <http://dl.acm.org/citation.cfm?id=2064747.2064768>.

LU, Y.; SINNOTT, R. O. Semantic privacy-preserving framework for electronic health record linkage. **Telematics and Informatics**, v. 35, n. 4, p. 737–752, 2018. Disponível em: <https://doi.org/10.1016/j.tele.2017.06.007>.

LYNCH, C. Big data: How do your data grow? **Nature**, v. 455, n. 7209, p. 28–29, 2008.

MACHADO, C. M. et al. The semantic web in translational medicine: Current applications and future directions. **Briefings in Bioinformatics**, v. 16, n. 1, p. 89–103, 2013.

MACHADO, T. B.; RICCIARDI, L.; BEATRIZ P P OLIVEIRA, M. Blockchain technology for the management of food sciences researches. **Trends in Food Science and Technology**, n. November 2019, p. 1–10, 2020. Disponível em: <https://doi.org/10.1016/j.tifs.2020.03.043>.

MADURO-ABREU, A. et al. Transparência da informação pública no Brasil: uma análise da acessibilidade de Big Data para o estudo das interfaces entre mudanças climáticas, mudanças produtivas e saúde. **Revista Eletrônica de Comunicação, Informação e Inovação em Saúde**, v. 14, n. 1, 2020.

MAIDANTCHIK, C. et al. NEURALTB WEB SYSTEM - Support to the Smear Negative Pulmonary Tuberculosis Diagnosis. **Proceedings of the Ninth International Conference on Enterprise Information Systems**, p. 198–203, 2007. Disponível em: <http://www.scitepress.org/DigitalLibrary/Link.aspx?doi=10.5220/0002366401980203>.

MAMUN, A. Al et al. BigCrypt for Big Data Encryption. p. 93–99, 2017.

MASLOVE, D. M. et al. Using Blockchain Technology to Manage Clinical Trials Data: A Proof-of-Concept Study. **JMIR Medical Informatics**, v. 6, n. 4, p. e11949, 2018.

MAULL, R. et al. Distributed ledger technology: Applications and implications. **Strategic Change**, v. 26, n. 5, p. 481–489, 2017.

METKE-JIMENEZ, A.; HANSEN, D. FHIRCap: Transforming REDCap forms into FHIR resources. In: AMIA Joint Summits on Translational Science proceedings. AMIA Joint Summits on Translational Science, **Anais**...2019.

MICHENER, W. K. Ten Simple Rules for Creating a Good Data Management Plan. p. 1–9, 2015.

MINISTÉRIO DA SAÚDE. **Manual de Recomendações para o Controle da Tuberculose**. [s.l: s.n.]

MINISTÉRIO DA SAÚDE, B. **MANUAL DE RECOMENDAÇÕES PARA O CONTROLE DA TUBERCULOSE NO BRASIL - 2ª edição**. [s.l: s.n.]

MINISTÉRIO DA SAÚDE, B. Portaria nº 1434. . 2020, p. 231–234.

MINISTÉRIO DA SAÚDE BRASIL. **Brasil Livre da Tuberculose: Plano Nacional pelo Fim da Tuberculose como Problema de Saúde Pública**. [s.l: s.n.]

MOHER, D. et al. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement (Reprinted from Annals of Internal Medicine). **Annals of Internal Medicine**, v. 151, n. 4, p. 264–269, 2009.

MOTWANI, R.; XU, Y. Efficient Algorithms for Masking and Finding. In: VLDB Conference, Vienna. **Anais**... Vienna: ACM, 2007.

MUNN, Z. et al. Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. **BMC Medical Research Methodology**, v. 18, n. 1, p. 1–7, 2018.

NAKAJIMA, H. **Tuberculosis : a global emergencyWorld health 1993 ; 46(4) : 3-31**, 1993. . Disponível em: <https://www.undp-capacitydevelopment-health.org/en/capacities/focus/health-information-systems/>.

NAKAMOTO, S. Bitcoin: A Peer-to-Peer Electronic Cash System. 2008. Disponível em: <https://bitcoin.org/bitcoin.pdf>.

NISHIKIORI, N.; VAN WEEZENBEEK, C. Target prioritization and strategy selection for active case-finding of pulmonary tuberculosis: A tool to support country-level project planning. **BMC Public Health**, v. 13, n. 1, 2013.

NOOR, S. et al. Ontology for attack detection: Semantic-based approach for genomic data security. **Journal of Medical Imaging and Health Informatics**, v. 7, n. 6, p. 1309–1323, 2017.

OGUNDELE, O. A. et al. An Ontology for Tuberculosis Treatment Adherence Behaviour. p. 1–10, 2015.

OLIVEIRA, T.; PASSOS, L.; TOLFO, D. Implantação de sistemas informatizados na saúde: uma revisão sistemática. **Revista Eletrônica de Comunicação, Informação e Inovação em**

Sáude, v. 11, n. 3, p. 1–11, 2017.

ØLNES, S.; UBACHT, J.; JANSSEN, M. Blockchain in government: Benefits and implications of distributed ledger technology for information sharing. **Government Information Quarterly**, v. 34, n. 3, p. 355–364, 2017.

ONGARO, D.; OUSTERHOUT, J. In search of an understandable consensus algorithm. **Proceedings of the 2014 USENIX Annual Technical Conference, USENIX ATC 2014**, p. 305–319, 2014.

OSAMOR, V. C.; AZETA, A. A.; AJULO, O. O. Tuberculosis-Diagnostic Expert System: An architecture for translating patients information from the web for use in tuberculosis diagnosis. **Health Informatics Journal**, v. 20, n. 4, p. 275–287, 2014.

PÃNESCU, A. T.; MANTA, V. Smart Contracts for Research Data Rights Management over the Ethereum Blockchain Network. **Science and Technology Libraries**, v. 37, n. 3, p. 235–245, 2018. Disponível em: <https://doi.org/10.1080/0194262X.2018.1474838>.

PAPAKONSTANTINOU, V. et al. Securing access to sensitive RDF data. **Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)**, v. 8798, p. 455–460, 2014.

PARK, Y.; GREENE, C. S. A parasite's perspective on data sharing. **GigaScience**, v. 7, n. 11, p. 1–3, 2018.

PASALIC, D. et al. Implementing an Electronic Data Capture System to Improve Clinical Workflow in a Large Academic Radiation Oncology Practice. **JCO Clinical Cancer Informatics**, n. 2, p. 1–12, 2018.

PELLISON, F. C. et al. Development and evaluation of an interoperable system based on the semantic web to enhance the management of patients' tuberculosis data. **Procedia Computer Science**, v. 121, p. 791–796, 2017. Disponível em: <https://doi.org/10.1016/j.procs.2017.11.102>.

PELLISON, F. C. et al. Data Integration in the Brazilian Public Health System for Tuberculosis: Use of the Semantic Web to Establish Interoperability. **JMIR Medical Informatics**, v. 8, n. 7, 2020.

PELLISON, F. C. **Desenvolvimento e Avaliação de um Sistema Interoperável para a Gestão da Informação de Pacientes com Tuberculose**. 2021. 2021.

PEREIRA, S. M.; MARINHO, J. M. Characteristics of healthcare services associated with adherence to tuberculosis treatment. **Revista de Saúde Pública**, v. 43, n. 6, p. 1–8, 2009.

PERES RONDELLI, J. P. **Desenvolvimento de uma ferramenta de geração de relatórios para auxiliar o acompanhamento do tratamento de tuberculose**. 2017. 2017.

PETERS, M. D. J. et al. Guidance for conducting systematic scoping reviews. **International Journal of Evidence-Based Healthcare**, v. 13, n. 3, p. 141–146, 2015.

PETERSON, K. et al. A Blockchain-Based Approach to Health Information Exchange Networks. **Mayo Clinic**, n. 1, p. 10, 2016. Disponível em:

<http://www.colleaga.org/sites/default/files/12-55-blockchain-based-approach-final.pdf>.

POULYMENOPOULOU, M.; MALAMATENIOU, F.; VASSILACOPOULOS, G. An Access Control Framework for Pervasive Mobile Healthcare Systems Utilizing Cloud Services. **International ICST Conference on Wireless Mobile Communication and Healthcare**, v. 83, n. January, p. 380–385, 2012. Disponível em: <http://www.scopus.com/inward/record.url?eid=2-s2.0-84869598486&partnerID=tZOtx3y1>.

POWER, D. J.; POWER, D. J. Understanding Data-Driven Decision Support Systems. v. 0530, 2008.

PUTHAL, D. et al. Everything You Wanted to Know about the Blockchain: Its Promise, Components, Processes, and Problems. **IEEE Consumer Electronics Magazine**, v. 7, n. 4, p. 6–14, 2018.

RAHMOUNI, H. B. et al. Privacy compliance and enforcement on European healthgrids: An approach through ontology. **Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences**, v. 368, n. 1926, p. 4057–4072, 2010.

RAHMOUNI, H. B. et al. A SWRL bridge to XACML for clouds privacy compliant policies. **CLOSER 2014 - Proceedings of the 4th International Conference on Cloud Computing and Services Science**, p. 27–37, 2014.

REBELO, J. **Blockchain Technology Impact on Supply Chain Management**. 2019. 2019.

REDE INTERAGENCIAL DE INFORMAÇÃO PARA A SAÚDE. Indicadores básicos para a saúde no Brasil: conceitos e aplicações. **Organização Pan-Americana da Saúde**, p. 349, 2008. Disponível em: <http://bases.bireme.br/cgi-bin/wxislind.exe/iah/online/?IsisScript=iah/iah.xis&src=google&base=REPIDISCA&lang=p&nextAction=lnk&exprSearch=7469&indexSearch=ID>.

REZENDE, S. et al. Determinantes sociais da saúde e custos catastróficos associados ao diagnóstico e tratamento da tuberculose. **Jornal Brasileiro de Pneumologia**, v. 46, n. 5, p. 1–5, 2020.

RINNER, C. et al. Improving the informational continuity of care in diabetes mellitus treatment with a nationwide Shared EHR system: Estimates from Austrian claims data. **International Journal of Medical Informatics**, v. 92, p. 44–53, 2016. Disponível em: <http://dx.doi.org/10.1016/j.ijmedinf.2016.05.001>.

ROBU, I.; ROBU, V.; THIRION, B. An introduction to the Semantic Web for health sciences librarians. **Journal of the Medical Library Association : JMLA**, v. 94, n. 2, p. 198–205, 2006.

RUOTSALAINEN, P. Privacy and security in teleradiology. **European Journal of Radiology**, v. 73, n. 1, p. 31–35, 2010.

RUSSELL, D. G. Mycobacterium tuberculosis: Here today, and here tomorrow. **Nature Reviews Molecular Cell Biology**, v. 2, n. 8, p. 569–577, 2001.

RUTA, M. et al. Semantic Blockchain to Improve Scalability in the Internet of Things. **Journal of Internet of Things (OJIOT)**, v. 3, n. 1, 2017. Disponível em:

<http://www.ronpub.com/ojiot>.

SÁ, L. D. de et al. Implantação da estratégia DOTS no controle da Tuberculose na Paraíba: entre o compromisso político e o envolvimento das equipes do programa saúde da família (1999-2004). **Ciência & Saúde Coletiva**, v. 16, n. 9, p. 3917–3924, 2011. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1413-81232011001000028&lng=pt&tlng=pt>.

SAVOLAINEN, J.; MYLLARNIEMI, V. Layered architecture revisited — Comparison of research and practice. In: 2009 Joint Working IEEE/IFIP Conference on Software Architecture European Conference on Software Architecture, **Anais**...2009.

SCHEMA.ORG. **Schema.org vocabulary**. Disponível em: <https://schema.org/>. Acesso em: 15 abr. 2020.

SEMOGAN, A. R. C. et al. A rule-based fuzzy diagnostics decision support system for tuberculosis. **Proceedings - 2011 9th International Conference on Software Engineering Research, Management and Applications, SERA 2011**, p. 60–63, 2011.

SERNADELA, P.; GONZÁLEZ-CASTRO, L.; OLIVEIRA, J. L. SCALEUS: Semantic Web Services Integration for Biomedical Applications. **Journal of Medical Systems**, v. 41, n. 4, 2017.

SHABAN-NEJAD, A. et al. PopHR: a knowledge-based platform to support integration, analysis, and visualization of population health data. **Annals of the New York Academy of Sciences**, v. 1387, n. 1, p. 44–53, 2017.

SHABUT, A. M. et al. An intelligent mobile-enabled expert system for tuberculosis disease diagnosis in real time. **Expert Systems with Applications**, v. 114, p. 65–77, 2018. Disponível em: <https://doi.org/10.1016/j.eswa.2018.07.014>.

SHARMA, R.; KOCHHER, R. Fuzzy decision support system for tuberculosis detection. **Proceedings of the 2017 IEEE International Conference on Communication and Signal Processing, ICCSP 2017**, v. 2018- Janua, p. 2001–2005, 2018.

SHIFERAW, A. M. et al. Routine health information system utilization and factors associated thereof among health workers at government health institutions in East Gojjam Zone , Northwest Ethiopia. p. 1–9, 2017.

SINGHAL, A.; SRIVASTAVA, J. Generating semantic annotations for research datasets. **ACM International Conference Proceeding Series**, 2014.

SONG, T. M.; PARK, H. A.; JIN, D. L. Development of health information search engine based on metadata and ontology. **Healthcare Informatics Research**, v. 20, n. 2, p. 88–98, 2014.

SOUNDARARAJAN, K.; SURESHKUMAR, S.; ANUSUYA, C. Diagnostics Decision Support System for Tuberculosis using Fuzzy Logic. **IRACST - International Journal of Computer Science and Information Technology & Security (IJCSITS)**, v. 2, n. 3, p. 684–689, 2012.

STALLINGS, W. **Cryptography and Network Security: Principles and Practices**. [s.l:

s.n.]

STRASSER, C. **Research Data Management: A Primer**. [s.l: s.n.]. Disponível em: <http://www.niso.org/publications/primer-research-data-management>.

SUN, L.; YONG, J.; SOAR, J. Access control management for e-Healthcare in cloud environment. **ICST Transactions on Scalable Information Systems**, v. 1, n. 2, p. e3, 2014.

SUNYAEV, A. Distributed Ledger Technology. In: **Internet Computing: Principles of Distributed Systems and Emerging Internet-Based Technologies**. Cham: Springer International Publishing, 2020. p. 265–299.

SUSTAINABLE ENGINEERING LABORATORY, C. U. **Enketo - Smart paper**. Disponível em: <https://enketo.org>. Acesso em: 12 abr. 2020.

SUSTAINABLE ENGINEERING LABORATORY, C. U.; ODK TEAM, U. of W. **XLSForm documentation**. Disponível em: <https://xlsform.org/en/>. Acesso em: 11 abr. 2020.

TCHERNYKH, A. et al. Towards understanding uncertainty in cloud computing with risks of confidentiality, integrity, and availability. **Journal of Computational Science**, 2016. Disponível em: <http://dx.doi.org/10.1016/j.jocs.2016.11.011>.

TEH, J. Sen; TAN, K.; ALAWIDA, M. A chaos-based keyed hash function based on fixed point representation. **Cluster Computing**, v. 22, n. 2, p. 649–660, 2019. Disponível em: <https://doi.org/10.1007/s10586-018-2870-z>.

THE HYPERLEDGER WHITE PAPER WORKING GROUP. **An Introduction to HyperledgerJuly 2018**. [s.l: s.n.]. Disponível em: <https://www.hyperledger.org/wp-content/uploads/2018/07/HL_Whitepaper_IntroductiontoHyperledger.pdf>.

THURAISINGHAM, B. Security standards for the semantic web. **Computer Standards and Interfaces**, v. 27, n. 3, p. 257–268, 2005.

TILAHUN, B. et al. Design and development of a linked open data-based health information representation and visualization system: Potentials and preliminary evaluation. **Journal of Medical Internet Research**, v. 16, n. 10, p. e31, 2014.

TIWARI, S. M. et al. Secure semantic smart healthcare (S3HC). **Journal of Web Engineering**, v. 17, n. 8, p. 617–646, 2019.

TRIPATHI, R.; AGRAWAL, S. Comparative Study of Symmetric and Asymmetric Cryptography. **International Journal of Advance Foundation and Research in Computer**, v. 1, n. 6, p. 2348–4853, 2014. Disponível em: <https://pdfs.semanticscholar.org/e0e4/810c5276f9c05cc82425fcf911f206c52bef.pdf>.

UGARTE, H. A more pragmatic Web 3.0: Linked Blockchain Data. **Personal Communication**, n. June, 2017.

VALENTA, M.; SANDNER, P. Comparison of Ethereum, Hyperledger Fabric and Corda. n. June, p. 1–8, 2017.

VALLE, E. Della et al. The Need for Semantic Web Service in the eHealth. **First International Workshop on Mediation in Semantic Web Services (MEDIATE 2005)**, p. 35–50, 2005.

VANDERVALK, B.; MCCARTHY, L.; WILKINSON, M. CardioSHARE: Web Services for the Semantic Web. **Web**, n. August 2016, p. 1–6, 2008. Disponível em: <papers2://publication/uuid/A2C2A586-056E-45BF-9DEA-B96B19FA4224>.

VERNADAT, F. B. Enterprise modeling and integration : principles and applications. In: **Anais**...1996.

W3C, J.-L. W. G. **JSON for Linking Data**. Disponível em: <https://json-ld.org/>. Acesso em: 8 abr. 2020.

W3C, W. W. W. C.-. **Resource Description Framework**. Disponível em: <https://www.w3.org/RDF/>. Acesso em: 8 abr. 2020.

W3C, W. W. W. C.-. **HTML Microdata Working Draft**. Disponível em: <https://www.w3.org/TR/microdata/>.

WOOD, G. Ethereum: a secure decentralised generalised transaction ledger. **Ethereum Project Yellow Paper**, p. 1–32, 2014.

WORLD HEALTH ORGANISATION. **Xpert MTB/RIF assay for the diagnosis of pulmonary and extrapulmonary TB in adults and childrenWorld Health Organisation**. [s.l: s.n.]. Disponível em: <http://www.ncbi.nlm.nih.gov/pubmed/25473701>.

WORLD HEALTH ORGANIZATION. What is DOTS ? A Guide to Understanding the WHO-recommended TB Control Strategy Known as DOTS. **Prevention and Control**, p. 1–39, 1999.

WORLD HEALTH ORGANIZATION. The End TB Strategy- WHO. 2014. Disponível em: <http://www.who.int/tb/strategy/End_TB_Strategy.pdf?ua=1>.

WORLD HEALTH ORGANIZATION. Unite To End TB. 2015a.

WORLD HEALTH ORGANIZATION. Digital health for the End TB strategy : Progress since 2015 and future perspectives. n. February, p. 1–33, 2015b.

WORLD HEALTH ORGANIZATION. Digital Health for the End Tb Strategy: an Agenda for Action. **Who**, p. 24, 2015c. Disponível em: <https://www.who.int/cancer/en/>.

WORLD HEALTH ORGANIZATION. **Global Tuberculosis Report 2021**. [s.l: s.n.].

WRIGHT, A. REDCap: A Tool for the Electronic Capture of Research Data. **Journal of Electronic Resources in Medical Libraries**, v. 13, n. 4, p. 197–201, 2016. Disponível em: <http://dx.doi.org/10.1080/15424065.2016.1259026>.

WÜST, K.; GERVAIS, A. Do you need a Blockchain? **IACR Cryptology ePrint Archive**, n. i, p. 375, 2017. Disponível em: <https://eprint.iacr.org/2017/375.pdf>.

YAMADA, D. B. et al. Proposal of an ontology for mental health management in Brazil.

Procedia Computer Science, v. 138, p. 137–142, 2018. Disponível em: <https://doi.org/10.1016/j.procs.2018.10.020>.

YAMADA, D. B. et al. Ontology-based inference for supporting clinical decisions in mental health. **Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)**, v. 12140 LNCS, p. 363–375, 2020.

YARMAND, M. H.; SARTIPI, K.; DOWN, D. G. Behavior-based access control for distributed healthcare environment. **Proceedings - IEEE Symposium on Computer-Based Medical Systems**, n. July, p. 126–131, 2008.

YOSHIURA, V. T. et al. Towards a health observatory conceptual model based on the semantic web. **Procedia Computer Science**, v. 138, p. 131–136, 2018. Disponível em: <https://doi.org/10.1016/j.procs.2018.10.019>.

ZAVERI, A. et al. Publishing and interlinking the global health observatory dataset. **Semantic Web**, v. 4, n. 3, p. 315–322, 2013.

ZHANG, J. **Consensus Algorithms: PoA, IBFT or Raft?** Disponível em: <https://kaleido.io/consensus-algorithms-poa-ibft-or-raft>. Acesso em: 21 abr. 2020.

ZHENG, Z. et al. An overview on smart contracts : Challenges, advances and platforms. **Future Generation Computer Systems**, v. 105, p. 475–491, 2020. Disponível em: <https://doi.org/10.1016/j.future.2019.12.019>.

ZHUANG, Y. et al. Applying Blockchain Technology for Health Information Exchange and Persistent Monitoring for Clinical Trials. **AMIA ... Annual Symposium proceedings. AMIA Symposium**, v. 2018, p. 1167–1175, 2018.

ZISSIS, D.; LEKKAS, D. Addressing cloud computing security issues. **Future Generation Computer Systems**, v. 28, n. 3, p. 583–592, 2012. Disponível em: <http://dx.doi.org/10.1016/j.future.2010.12.006>.

ZOLHAVARIEH, S.; PARRY, D.; BAI, Q. Issues Associated With the Use of Semantic Web Technology in Knowledge Acquisition for Clinical Decision Support Systems: Systematic Review of the Literature. **JMIR Medical Informatics**, v. 5, n. 3, p. e18, 2017.

ZWERLING, A. et al. A simplified cost-effectiveness model to guide decision-making for shortened anti-tuberculosis treatment regimens. **The International Journal of Tuberculosis and Lung Disease**, v. 20, n. 2, p. 257–260, 2016. Disponível em: <http://openurl.ingenta.com/content/xref?genre=article&issn=1027-3719&volume=20&issue=2&spage=257>.

ZYSKIND, G.; NATHAN, O.; PENTLAND, A. Decentralizing Privacy: Using Blockchain to Protect Personal Data. In: IEEE Security and Privacy Workshops (SPW), **Anais**...2015.

APÊNDICE A – MANUAL DE UTILIZAÇÃO DO *FRAMEWORK* SEGURANÇA

Este apêndice descreve, de um ponto de vista técnico, a forma de utilização da API e dos *query endpoints* do Sistema da Tuberculose (SisTB) para integração com sistemas externos (SE).

O documento está organizado da seguinte maneira:

- Seção 1: apresenta a terminologia utilizada no manual.
- Seção 2: apresenta os requisitos essenciais que devem ser satisfeitos pelos sistemas externos, o procedimento para obter a chave de API e apresenta como instalar a chave pública do sistema externo no repositório de segurança.
- Seção 3: apresenta como gerar uma assinatura da chave de API para realizar o envio de requisições.
- Seção 4: apresenta as URLs dos ambientes de desenvolvimento e produção para envio de requisições.
- Seção 5: apresenta os modelos de respostas e o processo de descriptografia dos dados (leitura).

1. Terminologia

SisTB	Sistema da Tuberculose
SE	Sistema externo que irá comunicar com o SisTB.
Chave de API	Chave intransferível que identifica um SE
Par de chaves	Conjunto de chaves pública e privada geradas pelo SE.
AES	Algoritmo Advanced Encryption Standard utilizado para o processo de criptografia simétrica (com uma única chave).
RSA	Algoritmo *Rivest-Shamir-Adleman* utilizado para o processo de criptografia assimétrica (com par de chaves público-privada)
SHA256withRSA	Algoritmo de assinatura que combina algoritmo de síntese (hash) SHA-256 e o algoritmo de criptografia RSA.
Repositório de segurança	Conjunto de dados utilizado para garantir a segurança dos dados. Armazena, por exemplo, as chaves públicas dos SEs.
Base64	Método para codificação de dados para transferência de conteúdio na Internet.
SPARQL	Linguagem de consulta para conjunto de dados semânticos armazenados ou visualizados como RDF
GraphQL	Linguagem de consulta intuitiva criada pelo Facebook. O projeto HyperGraphQL permite o uso desta linguagem em conjuntos de dados semânticos.

2. Requisitos essenciais

2.1. Registro dos SEs no repositório de segurança

Os sistemas externos, primeiramente, serão registrados no repositório de segurança manualmente por um administrador. Em seguida, uma URL será fornecida ao SE para divulgação da chave de API e para instalação da chave pública (vide item 2c).

2.2.Requisitos técnicos

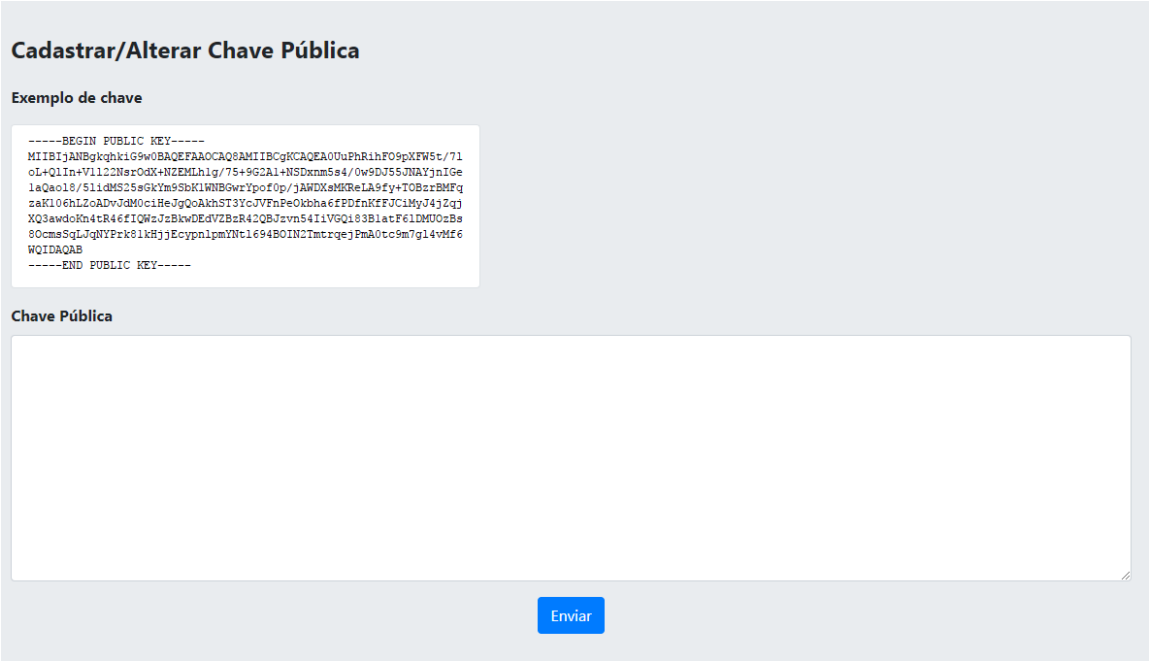
Os seguintes requisitos técnicos devem ser satisfeitos:

- Todas as requisições deverão ser enviadas via protocolo HTTPS.
- Todas as requisições deverão conter a chave de API e sua assinatura.
- O responsável pelo sistema externo deve gerar um par de chaves (pública/privada) RSA de 2048 bits.
- O sistema externo deverá ser capaz de reverter texto codificado em Base64.
- O sistema externo deverá ser capaz de descriptografar os dados recebidos utilizando sua chave privada. Não há limitação quanto à técnica usada ou à linguagem de programação.

2.3.Instalação da chave pública e obtenção da chave de API do SE

Ao acessar a URL fornecida pelo administrador ao SE, será possível proceder à instalação da chave pública e obter da chave de API. A Figura 1 apresenta a interface que permite executar tais ações.

Figura 1 - Interface para instalação da chave pública e obtenção da chave de API do SE



O usuário deverá gerar uma chave do tipo RSA com 2048 bits e colar o texto no campo denominado “Chave Pública”. Se a chave for válida, a página será atualizada e a chave de API será exibida codificada em Base64 e criptografada com a chave pública informada, conforme mostrado na Figura 2.

Figura 2 - Chave de API e chave pública do SE

Suas chaves

Chave de API criptografada:

zCsUI3ThrxTC41QHKxnBthc5W6yQli3fFiGIpEypv5EIj8k9t5TUnuDkmpviEIPJyJQLzs16tj/zt86pkqL2RZoNsq14vl1LZzkauxepR87QCSlLLIgm54jAdA79F5jTe6QAsJVNCdM
4Ywuz/apEzM3ix/jIsjBLP0xcugxhgCakuBhOUVF+p6BrgqUeX/VPaxcAqkk15kHzqoebH+iNBIyHPNbFonQ09E2YJK8D8UKt0SUuhvhEekQxPP14pwYz01HD2eNmjaNDG
1Qfx9PfYh4x8SNnjpSpIHKrSTE09kE9o2AP0lw7Bl3ZHSIvcNVAbpDlusEX48Kq/w1gBxN5vg==

Chave Pública:

-----BEGIN PUBLIC KEY-----
MIIBIjANBgkqhkiG9w0BAQEFAAOCAQ8AMIIBCgKCAQEA0UuPhRihFO9pXFW5t/7l
oL+QIIn+V1I22NsrOdX+NZEMLh1g/75+9G2A1+NSDxnm5s4/0w9DJ55JNAYjnIGe
1aQaol8/5lidMS25sGkYm9SbK1WNBGwrYpof0p/jAWDXsMKReLA9fy+TOBzrBMFq
zaK106hLZoADvJdM0ciHeJgQoAkhST3YcJVFnPeOkbha6fPDfnKfFJCiMyJ4jZqj
XQ3awdoKn4tR46fIQWzJzBkwDEdVZBzR42QBJzvn54IiVGQi83B1atF6lDMUOzBs
8OcmsSqLJqNYPrk81kHjjEcypnlpmYNtl694BOIN2TmtrqejPmA0tc9m7gl4vMf6
WQIDAQAB
-----END PUBLIC KEY-----

Finalmente, deverá copiar o texto do campo “Chave de API criptografada”, reverter a codificação Base64 e descriptografá-la com a chave privada do SE, utilizando o algoritmo **RSA**. Esse procedimento pode ser realizado através de qualquer técnica e linguagem de programação à escolha do usuário.

A chave de API é representada por uma string de 128 caracteres, como, por exemplo:

Quadro 1 - Exemplo de chave de API descriptografada

Hf1EMLdQVGR0HVNSFp1BXnUguNGKiYDNUKDARILrY1cQamZBcKXGZL5Bplyf6YvqyNDoIF2qrH4OcZ9w1O5QFfXQVbRqzXKYXRNuUjcUCubJXn2cyoEVvuUmqBvmTZLf

3. Assinatura da chave de API

A chave de API deve ser assinada pelo SE utilizando o algoritmo **SHA256-RSA**. A Figura 3 exemplifica como gerar a assinatura utilizando a linguagem de programação Java.

Figura 3 - Exemplo de método para gerar assinatura em Java

```
import java.security.PrivateKey;
import java.security.Signature;
import java.io.UnsupportedEncodingException;
import java.security.InvalidKeyException;
import java.security.SignatureException;
import java.security.NoSuchAlgorithmException;
import org.apache.commons.codec.binary.Base64;

    public static String sign(String text, PrivateKey privateKey) throws
InvalidKeyException, SignatureException, NoSuchAlgorithmException,
UnsupportedEncodingException {

        Signature sign = Signature.getInstance("SHA256withRSA");
        sign.initSign(privateKey);
        sign.update(text.getBytes("UTF-8"));
        return new String(Base64.encodeBase64(sign.sign()), "UTF-8");
    }
```

4. Requisições

4.1. Interfaces gráficas para testes

As interfaces de testes permitem o envio de requisições para os diversos *endpoints* disponíveis. Entretanto, por questões de segurança, as respostas são enviadas criptografadas com a chave pública do SE, mesmo na versão de desenvolvimento. Para aceder aos dados, o usuário deverá descriptografar os dados através de mecanismos próprios.

- Desenvolvimento
  - Swagger - API Semântica (documentação e testes)
    - https://api.tuberculosis.cloud/sistb-api-dev/apiDocs
  - SPARQL/GraphQL endpoint
    - https:/api.tuberculosis.cloud/sistb-api-dev/query
- Produção
  - Swagger - API Semântica (documentação e testes)
    - https://api.tuberculosis.cloud/sistb-api/apiDocs
  - SPARQL/GraphQL endpoint

        - https:/api.tuberculosis.cloud/sistb-api/query

4.2. URLs

As URLs estão listadas de acordo com os *endpoints* disponíveis e com o ambiente desejado (desenvolvimento ou produção), sendo:

- Desenvolvimento
    - API Semântica:
        - https://api.tuberculosis.cloud/sistb-api-dev/api/
        - Consultar a documentação em https://api.tuberculosis.cloud/sistb-api-dev/apiDocs

    - SPARQL/GraphQL endpoint
        - https:/api.tuberculosis.cloud/sistb-api-dev/executeQuery
        - Método HTTP: POST
        - Parâmetros:
            - query - String
            - apiKey - String
            - signature - String
            - endpoint – String (valores: sparql | graphql)
        - Retorno: JSON - consulte o formato no item 5-a

- Produção
    - API Semântica:
        - https://api.tuberculosis.cloud/sistb-api/api/
        - Consultar a documentação em https://api.tuberculosis.cloud/sistb-api/apiDocs

    - SPARQL/GraphQL endpoint
        - https://api.tuberculosis.cloud/sistb-api/executeQuery
        - Método HTTP: POST
        - Parâmetros:
            - query - String
            - apiKey - String

- signature - String
- endpoint – String (valores: sparql | graphql)

▪ Retorno: JSON - consulte o formato no item 5-a

5. Respostas

5.1.Modelo

O modelo utilizado em todas as respostas da API e dos *endpoints* segue o padrão apresentado na Figura 4, o qual é codificado no formato JSON, conforme a Figura 5.

Figura 4 - Modelo de resposta

```
ResponseApi ⌄ {
    code                 integer($int32)
                         Status HTTP
    hashMD5              string
                         Hash MD5 do resultado descriptografado para verificação de integridade

    hashMD5Signature     string
                         Assinatura da API para o Hash MD5 (codificada em base64)

    initVector           string
                         Vetor de inicialização para o algoritmo de criptografia simétrica

    key                  string
                         Chave simétrica criptografada (codificada em base64)

    message              string
                         Mensagem informativa
    obj                   ⌄ {
                            description:        Resultado criptografado (codificado em base64)

                         }
    objSignature         string
                         Assinatura da API para os dados descriptografados (codificada em base64)

}
```

Figura 5 - Resposta no formato JSON com dados criptografados


```
{
  "code": 200,
  "message": "SUCCESS",
  "key":
"avxpebJR0FIrIGJZh+K9u4wWmgDpaN8HoyCcs+UrRIB4QJctK6QOpDtF0xQmARGnUmfO5TxJpUX7VLvzq/7HR03S/
OjqtSp21GHaVFHdn4vMdjB6QpEOwup2j7Rz4fyD/lLERtu0rQEZnG/IK4GocDbbvAv2HQDMHiHyCmfQCkHxjLiR71z
k6jvpijfupfLGmDK5804fr83eOY5Uv/VzUD8+B25x+s7V/uA5EqB9mJMnrDvnr1oAW+bdOYtYYz6O/KC/tkB15aqWI
CGh8guGpRlii8Ql3Tx8A8gA4zE9vviaRK/0N4nSFCdx8DsYr2qw+WifdRi2BF2tlfhhYT0OfQ==",
  "initVector": "uAQUWV5kgZHUt2kX",
  "obj":
"+gC2z2skXhsDTYaDy2j+kYmSINp0a0pHaNm5d1TOjpgnDDYWCC3EHpL37J1XG4mSiWKUHv2eomOl5wZAAqzNzDdEa
eBdrvC8+CIzJdY3dS1J4JW/Fw1v6Dg900+dpulSwECwEmhbDMiflfHuoqvcm+JjO81HkAWpgwwqDAkOwea3QhIOHQE
0+MtpxP1DJYvhqNuzwg5/szcg82wklUW04np54nwQX+vPuVYbkw6F4vj3Hhk3tHaRvtk2OQ+mEOK3tpOkNcMd2qBKm
mLmKbKZVpwFp/DxNmGcaLbLbQyq7zkknM6SgArBX499Bl5gROAXuTS3tCYbe6WfNqQ0VeUW8fXBnlKotW4CSD2xjcF
ZBNiquCB0/9wYTjAHOrO3DpS4GXGxZgt+VuiJF9QEbNixff0TzQ1D2OdiZ67wDi39LGR0zDlluyFNzTeps+HhGG/b/
HhWPbmun2RPjpktx3KA1vMrD0WYvW/slic2FSJrRRIcSNkRkoPCM/F7l42euTpTc423/IAZ4ynkhtWhjHRp4s5XD9r
fdDN2ZXt0pVUTmku349j7kELiHxTt14EjpERRSJAVlOXW1PmPk6lArmTee+GeRZPhyRqJlpgdUnnIjUqxIOMSXdYow
/fcXALt/8JTXVwgOYsdIurVmAK0aIFXEqmBO6riKKni8CJHN0y/xJg5MYgC3jdN5p4Kq2Ti4rJjHS0wI+NDWbPJ8lZ
PMYEtqyJiA3AysbRFRCIyqqVSq3qXAPvzgZw2sh+slMrFKUaiNphImBDmVKgKQwAEEDkdb/S8wg0VeH9S5vnJvw5d/
u//rlwH5h6Y6pec8pzrjfRZz5n5A4uO7f/IZLnu/Eac7BVkvrkV+126ztxXHlSDYbyoS2XrzeZaCPghJ10TPz72z1p
asJeWflK4fKWc1SRrQciyXdr6OvObcVT7ULxfQZtfY0DVGCeu47tcPFX8J4tnYkX1LYWfGLMVVJcy1gk5otYW1v01H
mG3+AwyaXDKobFhXsgalX+A9RqwCpHCrn7uAO35fK4b3tZ2CLNKLf3X3ZibLwUab0xjzMpJ2vCvc053Q6CgMghflbM
dMt/cFxUgfA1Z/mMtU+EyYik1CC8NOHjbGJJ4oEJe+9VcbM9zn9V9AlFazlcppKT75AV9izoPx4LIbQ6vNwGToXn+5
AIV7Ph/ZomzNVE3IB6RtJxCUWr1W48Xoco5pYj8QnzqfKlTHi3W/WzdyEAqU1LQyRHMJt/x/t3THjsG+gWUd8C/fxP
ZqJNCdiysFjnE4NoJSZFSefWcgmQKWHI3t0ZJ0+xa8pIHsnz+k9v9GLJFELqOP7LbBe5y3wTL14cDTZ8KGbcz1erMS
5ekdZrhr36k6OXid/U2XU8xDzz86qVk8pLfvTkuaqtYd8Z6gRRxdnWtVn1eQf2gORgK4SFwAl7rESfBgygKmrTklyl
+R0QauAXd3D73DrkE/meMQhnH9XnZra/sUoc/wlLt+2INOu4aGH5PKcjySL/Xhqng4BlKFtqyp+LQ1PWFbD661FuVP
tEO8e+ve33RAt1IuapNkCI7az9QurHkESt/sdyZrO5Twynl67iHh/RBjDCn2r/1pDverkYp0vKFSrtfHHzu1Lo7BoB
zpMUSX1idc0TM3UW6a1pg+cgO0wgORfLM0HJCRPa/f6zCRYhD0rz+qoKIJdzrN8eW1NqzCSC+0RbFEVG3nepl7cA/4
1Ot4ycPHu2DfxYvo1zVKcswVbn6X5ATTRXUJRx0i+3Me4V7k1yiZ3SMXdIzR5tCpNMbQ8H4DUB1YSh8M7VSTG2Hj/O
XaFopDp2O+coIWL2iZvzUE3VZg+OZiHqw5mCFtQVLtTGCegYXbcTH51VCENFj1mBhsYtUQh8xrXk+9d7zoEtEJ8Myp
uThXyUziA00vIybgTKKRtlf0bXXXSC9lvDhdXOMbf6HOCpB0tHRnEvtUg5MIdt57ouo+7fPzQj/pjL7B9ZUnJ+Mw7l
gGpbEPxI5O4PBOWO6oPu6ao=",
  "hashMD5": "e41cd1401d0fd8672a80fcc537073076",
  "hashMD5Signature":
"O/dWtkjpPPIdEVjd7JoKoPrAg2hzd+mpAki8WU1ZXSGWxpWrrXb3DpRw7BjEc5r4G5yMUeRxNLF3tTTENvsxMrmtU
awFE9Uh/QX/WbItigz0LL8gxEFAr4v3LzSjFVLspSyOc8Z5qwcVuHHEFJs7IFaN10+W/jBoOnrI2iFoqYpKq8XYUaE
j8iAtR32Fm0MCBUQVDtdR4wMZbG4CpgOuryMPqwM5pV+Zmuv9PYFoyObvHHFQyd6wjLZ1wSwcAFVaywAwOX2BbcMlp
QJhYwPQiq+DzCSa7Ix1bw/Ho7LPdCbsECB1RMxvw/O7xaCd2XHkeJROWlpmLSl6CwDTV8YT2w==",
  "objSignature":
"eGA7Jkns+nAxNZYASIgKVNSLIKbbWhVj4zhdMnih3YCAIeN8+8GhFkLA7gUMcaL41Sj3pzJe6oMuPQTxUEX3XRIQa
tGAc6wTuxmnMbQ8NNDoS0VJtHjvy54a34EcrU60xVMlkH95FNqwCIcN1sUZXEHfawdsRR6dweHMIB59vnjcP8nSs0D
OX0+qnG+yWLwATTVQGdBZynfcQ5v1NTgQc74yJGSG9Oe1MCsMQrNIOSEG5jIkMYHjSpQfFAXuTuX5CwwgvTIY85VA+
ZJR9/ZwjRkjBqR9V4jdXUF5NELcmfbnbl6UPdLJ3gDK3+H5mN2iBYnrFv1gh+MJOxuejF+YAQ=="
}
```


5.2.Leitura dos dados

O método de **criptografia híbrida** é utilizado para proteger os dados. Neste caso, os dados brutos são criptografados pelo algoritmo **AES** com uma chave simétrica. Por sua vez, a chave simétrica é também criptografada com a chave pública do SE pelo algoritmo **RSA**. Dessa forma, é possível criptografar grandes volumes de dados e distribuir de forma segura a chave simétrica.

Passos:

1. Obter a chave simétrica criptografada no objeto JSON da resposta, identificada por "key";
2. Descriptografar a chave simétrica utilizando a chave privada do SE e o algoritmo RSA (Figura 6);
3. Obter o resultado (dados) no objeto JSON da resposta, identificado por "obj";
4. Descriptografar os dados utilizando a chave simétrica e o algoritmo AES. (Figura 7).

Figura 6 - Exemplo de método para descriptografar a chave simétrica utilizando o algoritmo RSA em Java

```
import javax.crypto.Cipher;
import javax.crypto.NoSuchPaddingException;
import javax.crypto.IllegalBlockSizeException;
import javax.crypto.BadPaddingException;
import java.security.NoSuchAlgorithmException;
import java.security.InvalidKeyException;

    public static byte[] decrypt(Key key, byte[] ciphertext) throws
NoSuchAlgorithmException, NoSuchPaddingException, InvalidKeyException,
IllegalBlockSizeException, BadPaddingException {

        Cipher cipher = Cipher.getInstance("RSA");
        cipher.init(Cipher.DECRYPT_MODE, key);
        return cipher.doFinal(ciphertext);
    }
```

Figura 7 - Exemplo de método para descriptografar dados utilizando o algoritmo AES em Java

```
import javax.crypto.Cipher;
import javax.crypto.spec.IvParameterSpec;
import javax.crypto.spec.SecretKeySpec;

    public static String decrypt(byte[] conteudo, byte[] chave, String initVector) throws
Exception{

        Cipher decripta = Cipher.getInstance("AES/CBC/PKCS5Padding");
        SecretKeySpec key = new SecretKeySpec(chave, "AES");
        decripta.init(Cipher.DECRYPT_MODE, key,
                      new IvParameterSpec(initVector.getBytes("UTF-8")));
        return new String(decripta.doFinal(conteudo),"UTF-8");
    }
```

Observação: o vetor de inicialização do algoritmo AES deve ser obtido no objeto JSON da resposta, identificado por "initVector"

Opcional:

- Verificar a assinatura do objeto (dados brutos descriptografados), identificado por “objSignature”, de forma a validar a autenticidade do valor. A assinatura é gerada com a chave privada da API e pode ser verificada com a chave pública da API utilizando o algoritmo SHA256withRSA;
- Verificar a assinatura do hash MD5, identificada por “hashMD5Signature”, de forma a validar a autenticidade do valor. A assinatura é gerada com a chave privada da API e pode ser verificada com a chave pública da API utilizando o algoritmo SHA256withRSA;
- Recalcular o hash MD5 dos dados brutos (descriptografados) e comparar com o valor enviado na resposta, identificado por “hashMD5”, de forma a verificar a integridade dos dados.

A Figura 8 apresenta um exemplo de como verificar uma assinatura em Java utilizando o algoritmo SHA256withRSA.

Figura 8 - Exemplo de método para verificar uma assinatura em Java

```
import java.security.NoSuchAlgorithmException;
import java.security.InvalidKeyException;
import java.security.SignatureException;
import java.io.UnsupportedEncodingException;
import java.security.Signature;
import java.security.PublicKey;
import org.apache.commons.codec.binary.Base64;

    public static boolean verifySignature(String text, byte[] signature, PublicKey
publicKey) throws NoSuchAlgorithmException, InvalidKeyException, SignatureException,
UnsupportedEncodingException {

        Signature sign = Signature.getInstance("SHA256withRSA");
        sign.initVerify(publicKey);
        sign.update(text.getBytes("UTF-8"));
        return sign.verify(Base64.decodeBase64(signature));
    }
```

APÊNDICE B - MODELO DE PLANO DE GERENCIAMENTO DE DADOS

Este apêndice apresenta a estrutura sugerida para um Plano de Gerenciamento de Dados (PGD).

O modelo completo pode ser consultado através do endereço:

- https://redbox.technology/pgd_pdf.php

O sistema para gerar um PGD pré-preenchido pode ser acessado através do endereço:

- https://redbox.technology/pgd.php

Plano de Gerenciamento de Dados

Informações do documento e do projeto

Nº do projeto	{{NUMERO_PROJETO}}
Acrônimo	{{ACRONIMO_PROJETO}}
Título do projeto	{{NOME_PROJETO}}
Duração do projeto	{{DURACAO_PROJETO}}
Edital	{{EDITAL}}
Investigador(a) responsável	{{INVESTIGADOR_PRINCIPAL}} - {{AFILIACAO}}
URL Site	{{SITE_PROJETO}}
Financiador	{{FINANCIADOR}}

Título do documento	Plano de Gerenciamento de Dados
Autores	{{AUTOR}}
Partes interessadas	Equipe de TI e Gerenciamento de Dados, equipe coordenadora, equipes dos centros participantes, comitês externos, participantes, patrocinador

Conteúdo sugerido para o Plano de Gerenciamento de Dados

1 Introdução

2 Informações do projeto

3 Informações sobre os dados e recursos do estudo
- 3.1 Formas de coleta de dados
  - 3.1.1 Coleta online
  - 3.1.2 Coleta offline
- 3.2 Termo de Consentimento Livre e Esclarecido
- 3.3 Conjunto mínimo de dados
- 3.4 Coleta de dados de identificação pessoal
- 3.5 Obrigatoriedade de preenchimento dos dados
- 3.6 Manuais e treinamentos
- 3.7 Canais de comunicação

4 Privacidade e segurança dos dados
- 4.1 Infraestrutura computacional
- 4.2 Armazenamento, cópias de segurança e recuperação de dados
- 4.3 Medidas de segurança adotadas
- 4.4 Visão geral dos papéis e restrições de acessos
- 4.5 Gerenciamento de segurança e violação de dados
- 4.6 Envio e recebimento de dados identificados para processamento em lote
- 4.7 Desidentificação do participante da pesquisa
  - 4.7.1 Procedimento de decodificação para projetos com mascaramento

5 Aspectos éticos e regulatórios, propriedade intelectual e citação
- 5.1 Ética
- 5.2 Propriedade intelectual
  - 5.2.1 Registo de software
  - 5.2.2 Transferência de tecnologia
- 5.3 Citação
- 5.4 Conflitos de interesse

6 Acesso aos dados
- 6.1 Disponibilização e compartilhamento de dados
- 6.2 Disseminação, comunicação e documentação

7 Arquivamento e prova futura de informações
- 7.1 Melhores práticas para formatos de arquivo
- 7.2 Reuso e redistribuição de dados

8 Gerenciamento de dados e monitoria centralizada
- 8.1 Objetivos

8.2 Métricas e padrões
8.3 Processos
8.3.1 Boas práticas para inserção, modificação, exclusão ou exportação de dados
8.3.2 Monitoria centralizada - processo de validação
8.3.2.1 Processo de validação em simultâneo com monitoria clínica
8.3.2.2 Falha de recrutamento
8.3.3 Pós-processamento e curadoria
8.3.4 Importação de dados
8.3.5 Identificação de duplicidade
8.3.6 Monitoramento e indicadores de qualidade de projeto
8.3.7 Monitoramento de eventos adversos
8.3.8 Monitoramento de prazos de inserção de dados
8.3.9 Alterações no dicionário de dados
8.3.10 Randomização
8.4 Relatórios
8.4.1 Relatórios em tempo real
8.4.2 Relatórios específicos sob demanda
8.4.3 Relatórios de suporte para comitês externos
8.4.4 Análises

Referências

Apêndice I - Dicionário de dados

Apêndice II - eCRFs

Apêndice III - Protocolo de gerenciamento de segurança e violação de dados

APÊNDICE C - PUBLICAÇÕES E APRESENTAÇÕES DE TRABALHOS COMO PRIMEIRO AUTOR

Acesso aos artigos e trabalhos

Todos os artigos e trabalhos mencionados neste apêndice podem ser acessados através do DOI (quando disponível), pelo site do periódico, ou através do repositório disponível no endereço:

https://drive.google.com/drive/folders/193Z-ZAS77ORX7xuHtEdpOYeoy3daQBJa?usp=sharing

Artigos publicados

Lima VC, Bernardi FA, Alves D, Mello R, Rui G, Charters P, et al. A Permissioned Blockchain Network for Security and Sharing of Deidentified Tuberculosis Research Data in Brazil. Methods Inf Med. 2021.

Disponível em: https://doi.org/10.1055/s-0041-1727194

Background: Tuberculosis (TB) is an infectious disease and is among the top 10 causes of death in the world and Brazil is part of the top 30 high TB burden countries. Data collection is an essential practice in health studies and the adoption of electronic data capture systems (EDC) can positively increase the experience of data acquiring and analysis. Also, data sharing capabilities are crucial to the construction of efficient and effective evidence-based decision-making tools for managerial and operational actions in TB services. Data must be held secure and traceable, as well as available and understandable for authorized parties.

Objectives: In this sense, this work aims to propose a blockchain-based approach to build a reusable, decentralized, and de-identified dataset of TB research data, while increasing transparency, accountability, availability, and integrity of raw data collected in EDC systems.

Methods: After identifying challenges and gaps, a solution was proposed to tackle them, considering its relevance for TB studies. Data security issues are being addressed by a blockchain network and a lightweight and practical governance model. Research Electronic Data Capture (REDCap) and KoBoToolbox are used as EDC systems in TB research. Mechanisms to de-identify data and aggregate semantics to data are also available.

Results: A permissioned blockchain network was built using Kaleido Platform. An Integration Engine integrates the EDC systems with the blockchain network, performing de-identification and aggregating meaning to data. A governance model addresses operational and legal issues for the proper use of data. Finally, a management system facilitates the handling of necessary metadata and additional applications are available to explore the blockchain and export data.

Conclusions: Research data is an important asset not only for the research where it was generated, but also to underpin studies replication and support further investigations. The proposed solution allows the delivery of de-identified databases built in real-time by storing

data in transactions of a permissioned network, including semantic annotations, as data is being collected in TB research. The governance model promotes the correct use of the solution.

Lima VC, Pellison FC, Bernardi FA, Alves D, Rijo RPCL. Security Framework for Tuberculosis Health Data Interoperability Through the Semantic Web. Int J Web Portals. 2021;13(2):36–57.

Disponível em: https://doi.org/10.4018/ijwp.2021070103

One of the critical challenges in health information systems is interoperability. The clinical, strategic, and operational decision depends on quality data and the ability to exchange data inter-systems and inter-health organizations. The Semantic Web plays a cornerstone role as a technology to enable functional and semantic interoperability without substantial changes in existing systems and to improve data sharing capabilities and, consequently, its quality and completeness. However, a common concern is data security. In this context, this research proposes a framework for securing health data, with a real case scenario focused on tuberculosis data exchange over the Semantic Web. The needs of security for the Semantic Web were satisfied, helping to build trust in the data and promoting its use in contexts that were not initially created to openly disseminate data without a significant technological apparatus.

Lima VC, Rijo RPCL, Bernardi FA, Carvalho I, Neiva MB, Alves D. The relevance of digital health strategies to support tuberculosis services in the COVID-19 pandemic context. J Multiprof Heal Res. 2021;72–4.

Disponível em: https://journalmhr.com/index.php/jmhr/article/view/16

Artigos em revisão

Lima VC, Bernardi FA, Pellison FC, Júnior FB, Filho MEC, Alves D, Kritski AL, Galliez RM, Rijo RPCL. REDbox: a comprehensive semantic framework for data collection, management and sharing in tuberculosis research. 2021;1–17. Preprint.

Revista: Nature Scientifc Reports

Disponível em: https://doi.org/10.21203/rs.3.rs-1070973/v1

The outcomes of a clinical research directly depend on the correct definition of the research protocol, the data collection strategy and the data management plan. Furthermore, researchers often need to work within challenging contexts, such as in Tuberculosis services, where human and technological resources for research may be rare. The use of Electronic Data Capture systems, such as REDCap and KoBotoolbox, can help to mitigate such risks and to enable a reliable environment to conduct health research and promote results dissemination and data reusability. The proposed solution was based on needs pinpointed by researchers, considering the lack of an embracing solution to conduct research in low resources environments. The REDbox framework was built to enhance data collection, management and sharing in tuberculosis research, while providing a better user experience. The relevance of this article lies in the innovative approach to support TB research by combining existing technologies and developing support features. When focusing on positive aspects of each tool, it is possible to underpin tuberculosis research by improving data collection, management capability and security. Furthermore, the aggregation of meaning in raw data helps to promote the quality and the availability of research data.

Lima VC, Bernardi FA, Alves D, Rijo RPCL. Security approaches for electronic health data handling through the Semantic Web: a scoping review. 2022. Preprint.

Revista: Semantic Web – Interoperability, Usability, Applicability

Disponível em: http://www.semantic-web-journal.net/system/files/swj3088.pdf

Integration of health information systems are crucial to advance the effective delivery of healthcare for individuals and communities across organizational boundaries. Semantic Web

technologies may be used to connect, correlate, and integrate heterogeneous datasets spread over the internet. However, when working with sensitive data, such as health data, security mechanisms are needed. A scoping review of the literature was undertaken to provide a broad view of security mechanisms applied to, or along with, Semantic Web technologies that could allow its use with health data. Searches were conducted in the most relevant databases for the scope of this work. The findings were classified according to the main objective and features presented by each solution. Twenty-six studies were included in the review. They introduced mechanisms that addressed several security attributes, such as authentication, authorization, integrity, availability, confidentiality, privacy, and provenance. These mechanisms support access control frameworks, semantic and functional interoperability infrastructures, and privacy compliance solutions. The findings suggest that the application and use of Semantic Web technologies is still growing, with the healthcare area being particularly interested. The main security mechanisms for Semantic Web technologies, the key security attributes and properties, and the main gaps in the literature were identified, helping to understand the technical needs to mitigate the risks of handling personal health information over the Semantic Web. Also, this research has shown that complex and robust solutions are available to successfully address several security properties and features, depending on the context that the electronic health data is being managed.

Artigos publicados em *proceedings* de conferências

Lima VC, Lopes Rijo RPC, Pellison FC, De Lima RR, Cruz Correia RJ, Giuriati HT, et al. From guidelines to decision-making: Using mobile applications and semantic web in the practical case of guides to support patients. Procedia Comput Sci [Internet]. 2017;121:803–8.

Disponível em: https://doi.org/10.1016/j.procs.2017.11.104.

The use of guidelines, procedures and recommendations is very common and also a simple way to structure, organize and retrieve relevant information that, in turn, may impact on patient care. However, it is necessary to expedite the retrieval of stored information. By simplifying the access to the desired data, it is possible to make the process practical, fast and intuitive for the health professional. The main goal of this article is to present a development approach for a project that proposes decision-making tools, based on guidelines, to assist health professionals in the care process of patients with specific diseases (or a combination of them). These tools will be developed for use on mobile devices and as a web-based system, which will transform a passive and declarative manual into an interactive tool that will provide not only data and information, but also knowledge through semantic web paradigm. Further, we will show an example of usability and feasibility of the proposed application through a functional prototype developed for Android platform.

Lima VC, Alves D, Pellison FC, Yoshiura VT, Crepaldi NY, Rijo RPCL. Establishment of access levels for health sensitive data exchange through semantic web. Procedia Comput Sci [Internet]. 2018;138:191–6.

Disponível em: https://doi.org/10.1016/j.procs.2018.10.027

Data exchange in health information systems must be carefully planned and needs to be protected from unauthorized access due to sensibility of stored content. Security aspects like authentication, authorization and encryption must be considered in this context. The main goal of this article is to present the implementation of security mechanisms to a semantic API that allows data extraction from a regional health information system designed to create notifications and to follow patients diagnosed with Tuberculosis. Data semantically tagged will be mapped individually to several access levels. It will be showed how external systems can

connect, authenticate and retrieve only authorized data that are classified in the scope of its maximum access level.

Lima V, Bernardi F, Pellison F, Rijo R, Kritski A, Galliez R, et al. Decision Support Systems for Tuberculosis: Protocol for a Scoping Review. In: Studies in Health Technology and Informatics [Internet]. IOS Press; 2019. p. 3–6.

Disponível em: http://ebooks.iospress.nl/ISBN/978-1-61499-987-4

Tuberculosis (TB) represents a global challenge in terms of prevention, care and control. Decision support systems (DSS) can supply the necessary knowledge basis to underpin investigators, policy makers and health personnel actions and to provide crucial elements that can help reducing TB burden. Thus, the objectives of this work are to present the protocol to be followed for carrying out a scoping review to identify topics where DSSs are used, to define appropriate categories and to clarify main outcomes and research gaps. As part of the protocol, five electronic bibliographic databases will be searched for articles from 2006 to 2019 and two investigators will independently screen each work using the study inclusion criteria. Data extraction will be performed, and findings will be reported. The results will be used to provide a broad understanding of how DSSs for TB are being used.

Lima V, Pellison F, Bernardi F, Carvalho I, Rijo R, Alves D. Proposal of an integrated decision support system for Tuberculosis based on Semantic Web. Procedia Comput Sci [Internet]. 2019;164:552–8.

Disponível em: https://doi.org/10.1016/j.procs.2019.12.219

Epidemiological surveillance of Tuberculosis (TB) requires a strong integration of different health services, programs and levels of care. The deepening and broadening of data management techniques must be constantly carried out to increase the integrality of healthcare. Otherwise, knowledge extraction and clinical and administrative decision-making processes are significantly hampered, directly affecting the management and quality of health services. Thus, this work aims to establish a computerized decision support system capable of collecting, integrating and sharing TB health data in Brazilian Unified Public Health System. Also, it will allow the monitoring of infected patients and the visualization of consolidated information of regular TB and its resistant variants for health professionals and managers. The data will be

made available from heterogeneous, disconnected and unstructured sources by combining traditional web services, Semantic Web resources and security algorithms. A solid knowledge base applied to epidemiological surveillance, health information governance and clinical support will be enabled to integrate the multiple areas of TB patients care, as well as to support the creation of more accurate operational and diagnostics models

---

Lima V, Bernardi F, Rijo R, Ueyama J, Alves D. A Mechanism for Verifying the Integrity and Immutability of Tuberculosis Data sing IOTA Distributed Ledger Technology. Stud Health Technol Inform. 2021;279:130–5.

Disponível em: https://doi.org/10.3233/SHTI210099

Background: Intensified research and innovation and rapid uptake of new tools, interventions, and strategies are crucial to fight Tuberculosis, the world's deadliest infectious disease. The sharing of health data remains a significant challenge. Data consumers must be able to verify the consistency and integrity of data. Solutions based on distributed ledger technologies may be adequate, where each member in a network holds a unique credential and stores an identical copy of the ledger and contributes to the collective process of validating and certifying digital transactions. **Objectives**: This work proposes a mechanism and presents a use case in Digital Health to allow the verification of integrity and immutability of TB electronic health records **Methods**: IOTA was selected as a supporting tool due to its data immutability, traceability and tamper-proof characteristics. **Results**: A mechanism to verify the integrity of data through hash functions and the IOTA network is proposed. Then, a set of TB related information systems was integrated with the network. **Conclusion**: IOTA technology offers performance and flexibility to enable a reliable environment for electronic health records.

Lima VC, Bernardi FA, Domingues M, Kritski AL, Lopes Rijo RPC, Alves D. A computational infrastructure for semantic data integration towards a patient-centered database for Tuberculosis care. Procedia Comput Sci [Internet]. 2022;196:434–8.

Disponível em: https://doi.org/10.1016/j.procs.2021.12.033

Tuberculosis is an infectious disease that is among the top 10 causes of death in the world and Brazil ranks in the top 30 high TB Abstract burden countries. In this scenario, data integration and sharing are crucial to the construction of efficient and effective evidence- based decision-

making tools and to enable data-driven research. Through a socio-technical approach, this work proposes a Tuberculosis is an infectious disease that is among the top 10 causes of death in the world and Brazil ranks in the top 30 high TB computational infrastructure composed of a functional and semantic interoperability layer and security mechanisms to integrate burden countries. In this scenario, data integration and sharing are crucial to the construction of efficient and effective evidence- national level health information systems towards a patient-centered unified database to provide a broad view of a patient across based decision-making tools and to enable data-driven research. Through a socio-technical approach, this work proposes a several isolated databases. The CMIID, a medical information identifier for data harmonization, developed by the University of computational infrastructure composed of a functional and semantic interoperability layer and security mechanisms to integrate Porto, was used for the linkage of patients across these health information systems and to perform records anonymization for national level health information systems towards a patient-centered unified database to provide a broad view of a patient across privacy of personal health information. Through the integration of such systems, it is possible to gather, summarize and visualize several isolated databases. The CMIID, a medical information identifier for data harmonization, developed by the University of TB data in a single system, which can be useful for health professionals and managers. Therefore, this work sought to promote the Porto, was used for the linkage of patients across these health information systems and to perform records anonymization for integration of disparate systems and the availability of data to support decision-making and research, which are fundamental for privacy of personal health information. Through the integration of such systems, it is possible to gather, summarize and visualize improving the quality of TB services in Brazil.

Outros trabalhos apresentados em conferências

Sentiment analysis on Brazilian tweets regarding social isolation and quarantine during COVID-19 pandemic. HCist - International Conference on Health and Social Care Information Systems and Technologies. Vilamoura, Portugal. 2020.

Data Integration as a basis for decision-making and research in TB services in Brazil. IV Congresso Médico-Científico dos Açores (uMED). Ponta Delgada, Portugal. 2021.

REDbox - A comprehensive framework for data collection in tuberculosis research. 2º Congresso Latino Americano e Brasileiro do REDCap – REDCapCon. 2021.

Monitoring and evaluation indicators for tuberculosis based on laboratory routine data managed through REDCap. HCist - International Conference on Health and Social Care Information Systems and Technologies. Braga, Portugal. 2021

APÊNDICE D - PUBLICAÇÕES E APRESENTAÇÕES DE TRABALHOS COMO COAUTOR

Acesso aos artigos e trabalhos

Todos os artigos e trabalhos mencionados neste apêndice podem ser acessados através do DOI (quando disponível), pelo site do periódico, ou através do repositório disponível no endereço:

https://drive.google.com/drive/folders/193Z-ZAS77ORX7xuHtEdpOYeoy3daQBJa?usp=sharing

Artigos publicados

Pellison FC, Rijo RPCL, **Lima VC**, Crepaldi NY, Bernardi FA, Galliez RM, et al. Data Integration in the Brazilian Public Health System for Tuberculosis: Use of the Semantic Web to Establish Interoperability. JMIR Med Informatics. 2020;8(7).

Disponível em: https://doi.org/10.2196/17176

Background: Interoperability of health information systems is a challenge due to the heterogeneity of existing systems at both the technological and semantic levels of their data. The lack of existing data about interoperability disrupts intra-unit and inter-unit medical operations as well as creates challenges in conducting studies on existing data. The goal is to exchange data while providing the same meaning for data from different sources. **Objective**: To find ways to solve this challenge, this research paper proposes an interoperability solution for the tuberculosis treatment and follow-up scenario in Brazil using Semantic Web technology supported by an ontology. **Methods**: The entities of the ontology were allocated under the definitions of Basic Formal Ontology. Brazilian tuberculosis applications were tagged with entities from the resulting ontology. **Results**: An interoperability layer was developed to retrieve data with the same meaning and in a structured way enabling semantic and functional interoperability. **Conclusions**: Health professionals could use the data gathered from several data sources to enhance the effectiveness of their actions and decisions, as shown in a practical use case to integrate tuberculosis data in the State of São Paulo.

Neiva MB, Carvalho I, Filho EDSC, Barbosa-Junior F, Bernardi FA, Sanches TLM, de Oliveira LL, **Lima VC**, Miyoshi NSB, Alves D. Brazil: The emerging epicenter of COVID-19 pandemic. Rev Soc Bras Med Trop. 2020;53:1–8.

Disponível em: https://doi.org/10.1590/0037-8682-0550-2020

Introduction: Five months after the first confirmed case of COVID-19 in Brazil, the country has the second highest number of cases in the world. Without any scientifically proven drug or vaccine available combined with COVID-19's high transmissivity, slowing down the spread of the infection is a challenge. In an attempt to save the economy, the Brazilian government is slowly beginning to allow non-essential services to reopen for in-person customers.

Methods: In this study, we analyze, based on data analysis and statistics, how other countries evolve and under which conditions they decided to resume normal activity. In addition, due to the heterogeneity of Brazil, we explore Brazilian data of COVID-19 from the State Health Secretaries to evaluate the situation of the pandemic within the states.

Results: Results show that while other countries have flattened their curves and present low numbers of active cases, Brazil continues to see an increase in COVID-19 patients. Furthermore, a number of important states are easing restrictions despite a high percentage of confirmed cases.

Conclusions: All analyses show that Brazil is not ready for reopening, and the premature easing of restrictions may increase the number of COVID-19-related deaths and cause the collapse of the public health system.

Alves D, Oliveira LL de, Sanches TLM, Miyoshi NSB, Carvalho I, Neiva MB, **Lima VC**, Bernardi FA. PORTAL COVID-19 BRASIL: MONITORAMENTO E ANÁLISES DA SITUAÇÃO DO CORONAVÍRUS COVID-19. Rev Fontes Doc [Internet]. 2020;3:32–9.

Disponível em: https://aplicacoes.ifs.edu.br/periodicos/fontesdocumentais/article/view/621

No Brasil, há comportamentos distintos e regionalizados da disseminação da COVID 19 que caracterizam geograficamente e temporalmente a pandemia. Para lidar com essa complexidade foi criado um portal, denominado COVID-19 Brasil, que visa monitorar e analisar, os dados oriundos de fontes diversas. A estratégia adotada inclui uma auditoria detalhada da evolução da doença, desagregados no tempo e espaço. Permitiu-se assim, realizarmos inferências e acompanharmos o cenário real de várias cidades brasileiras e do Brasil de modo geral. Recentemente, em parceria com a Universidade do Porto, começamos a expandir nossa estratégia e monitorar também, a evolução da COVID-19 em Portugal. Palavras-Chave: Covid-19; auditoria de dados; análise e visualização de dados; portal de conteúdo.

Alves D, Yamada DB, Bernardi FA, Carvalho I, Eloi M, Neiva MB, **Lima VC**, Félix TM. Mapping, Infrastructure, and Data Analysis for the Brazilian Network of Rare Diseases: Protocol for the RARASnet Observational Cohort Study. JMIR Res Protoc. 2020;10(October 2020).

Disponível em: https://doi.org/10.2196/24826

Background: A rare disease is a medical condition with low prevalence in the general population, but these can collectively affect up to 10% of the population. Thus, rare diseases have a significant impact on the health care system, and health professionals must be familiar with their diagnosis, management, and treatment.

Objective: This paper aims to provide health indicators regarding the rare diseases in Brazil and to create a network of reference centers with health professionals from different regions of the country. RARASnet proposes to map, analyze, and communicate all the data regarding the infrastructure of the centers and the patients' progress or needs. The focus of the proposed study is to provide all the technical infrastructure and analysis, following the World Health Organization and the Brazilian Ministry of Health guidelines.

Methods: To build this digitized system, we will provide a security framework to assure the privacy and protection of each patient when collecting data. Systems development life cycle methodologies will also be applied to align software development, infrastructure operation, and quality assurance. After data collection of all information designed by the specialists, the computational analysis, modeling, and results will be communicated in scientific research papers and a digital health observatory.

Results: The project has several activities, and it is in an initial stage. Initially, a survey was given to all health care centers to understand the technical aspects of each network member, such as the existence of computers, technical support staff, and digitized systems. In this survey, we detected that 59% (23/39) of participating health units have electronic medical records, while 41% (16/39) have paper records. Therefore, we will have different strategies to access the data from each center in the data collection phase. Later, we will standardize and analyze the clinical and epidemiological data and use these data to develop a national network for monitoring rare diseases and a digital health observatory to make the information available. The project had its financing approved in December 2019. Retrospective data collection started in October 2020, and we expect to finish in January 2021. During the third quarter of 2020, we enrolled 40 health institutions from all regions of Brazil.

Conclusions: The nature of rare disease diagnosis is complex and diverse, and many problems will be faced in the evolution of the project. However, decisions based on data analysis are the best option for the improvement of the rare disease network in Brazil. The creation of RARASnet, along with all the digitized infrastructure, can improve the accessibility of information and standardization of rare diseases in the country.

Bernardi FA, **Lima VC**, Rijo RPCL, Alves D. Mais do que palavras: uma análise das emoções brasileiras durante a COVID-19. Rev Iber Sist e Tecnol Informação. 2021;(August).

Disponível em:

https://www.researchgate.net/publication/353919486_Mais_do_que_palavras_uma_analise_das_emocoes_brasileiras_durante_a_COVID-19

Este estudo teve como objetivo investigar e analisar as emoções do público frente aos eventos relacionados ao coronavírus no Brasil. A ciência de dados pode ajudar a explicar os fatos e reduzir o ruído que interfere na compreensão do real impacto de surtos virais em nossas vidas. A rede social Twitter foi analisada para verificar a reação da população ao isolamento social e à quarentena. Para explorar as preocupações do público em relação à evolução da COVID-19 e as posturas governamentais adotadas, uma amostra diária de tweets em língua portuguesa foi coletada para realização de análises de sentimentos e emoções. Ressalta-se a importância da disponibilidade e confiabilidade da informação, da decorrente necessidade da transparência de sistemas de informação e do empoderamento de profissionais de saúde e gestores locais para a tomada das melhores decisões de vigilância e/ou clínicas de forma transversal aos diferentes setores da saúde.

Artigos publicados em *proceedings* de conferências

Pellison FC, Lopes Rijo RPC, **Lima VC**, De Lima RR, Martinho R, Cruz Correia RJ, et al. Development and evaluation of an interoperable system based on the semantic web to enhance the management of patients' tuberculosis data. Procedia Comput Sci [Internet]. 2017;121:791–

Disponível em: https://doi.org/10.1016/j.procs.2017.11.102

Tuberculosis (TB), despite all the efforts and progress made in the last decade, continues to be very relevant in Brazil as well as in other Portuguese-speaking countries, so there is the need to join efforts for increasing the effectiveness of the fight against TB. Therefore, the epidemiological surveillance system of TB requires, not only the implementation of basic prevention and care actions, but also the strengthening of the integration between the different health services, programs and levels of care, whose resolvability varies according to financial, technical and human resources, as well as service infrastructure that comprise the network of attention. The deepening and broadening of data management techniques must constantly be carried out to achieve, at a higher level, integration and integrity of the ideal and desirable health care. Then, the goal of this article is to describe the research methods used to develop an information system using systems interoperability techniques based on the Semantic Web paradigm that will contribute to the activities of epidemiological surveillance and the follow-up of TB patients.

Bernardi F, **Lima V**, Pellison F, Mazzoncini de Azevedo Marques P, Rijo R, Galliez RM, et al. Blockchain Based Network for Tuberculosis: A Data Sharing Initiative in Brazil. In: Studies in Health Technology and Informatics [Internet]. IOS Press; 2019. p. 264–7.

Disponível em: http://ebooks.iospress.nl/ISBN/978-1-61499-987-4

Data sharing, information exchange, knowledge acquisition and health intelligence are the basis of an efficient and effective evidence-based decision- making tool. A decentralized blockchain architecture is a flexible solution that can be adapted to institutional and managerial culture of organizations and services. Blockchain can play a fundamental role in enabling data sharing within a network and, to achieve that, this work defines the high-level resources necessary to apply this technology to Tuberculosis related issues. Thus, relying in open-source

tools and in a collaborative development approach, we present a proposal of a blockchain based network, the TB Network, to underpin an initiative of sharing of Tuberculosis scientific, operational and epidemiologic data between several stakeholders across Brazilian cities.

Pellison FC, **Lima VC**, Rijo RPCL, Alves D. Integrating Tuberculosis data in State of São Paulo over Semantic Web: a proof of concept. Procedia Comput Sci. 2019;164:686–91.

Disponível em: https://doi.org/10.1016/j.procs.2019.12.236

Although tuberculosis is a curable disease and, in most cases, with low cost drugs, its mortality still is a global concern. This fact turns our attention to management issues and the difficulties related to retrieving data of interest that are powdered on many applications. This work presents a semantic web approach to achieve functional and semantic interoperability between two applications in State of São Paulo that contain tuberculosis data. By combining a theoretical-practical development, the geolocalization tool created is a proof of concept that could help managers to take strategic decisions and develop better health policies by showing the distribution of tuberculosis cases across the state. This work stands out the importance of working in solutions that could improve the quality of data in health field and daily activities of health professionals.

Crepaldi NY, **Costa Lima V**, Andrade Bernardi F, Albano dos Santos LR, Yamaguti VH, Pellison FC, et al. SISTB: an ecosystem for monitoring TB. Procedia Comput Sci [Internet]. 2019;164:587–94.

Disponível em: https://doi.org/10.1016/j.procs.2019.12.224

Tuberculosis (TB) is a curable disease, although it still kills thousands of people every year. There are several information systems and applications developed to assist the TB treatment, but many of these systems are isolated and are not able to exchange relevant information. In this way, a set of interoperable applications and decision support modules have been developed. A socio-technical approach was applied to incorporate key stakeholders in all development stages so that final tools fit the health service reality. A set of systems and applications, called the SISTB Ecosystem, has been designed to assist in the treatment of TB and to improve the routine of health professionals.

Albano dos Santos LR, Andrade Bernardi F, Prado GCS, **Costa Lima V**, Crepaldi NY, Marçal MA, et al. The perception of health providers about an artificial intelligence applied to Tuberculosis video-based treatment in Brazil: a protocol proposal. Procedia Comput Sci [Internet]. 2019;164:595–601.

Disponível em: https://doi.org/10.1016/j.procs.2019.12.225

Tuberculosis (TB) is an infectious-contagious disease that affects mainly the lungs, ranking in the 10 main causes of death in the world. Non-adherence or non-treatment of TB may prolong transmissibility, increase the risk of drug resistance and lead to patient death. One step forward is the use of smartphones for monitoring the medication intake via video (VDOT). VDOT is a more acceptable and cost-effective option than the traditional DOT (in person). However, the VDOT system requires a professional (verifying agent) to check all medication intake daily. Based on an initiative that aims to replace the verifying agent with an artificial intelligence tool capable of validating it automatically through computer vision techniques (AI-based VDOT), the main goal of this work is to measure the acceptance and perception of Brazilian health professionals on the AI-Based VDOT before implementing this technology by using a quantitative questionnaire as instrument. To achieve this, we have developed a proposal of study protocol that describes the steps to create, apply and validate this research.

Bernardi FA, **Lima VC**, Sampaio DM, dos Santos MC, Lopes Rijo RPC, Alves D. TBI Score - use of a mobile score system to aid the diagnosis of tuberculosis in children in Brazil. Procedia Comput Sci [Internet]. 2022;196:493–500.

Disponível em: https://doi.org/10.1016/j.procs.2021.12.041

Tuberculosis (TB) is a bacterial infectious disease that mainly affects the lungs and remains as one of the biggest public health problems in the world. The treatment methods currently available can cure almost all cases. Due to the difficulty of bacteriological confirmation of TB in children, the Brazilian Ministry of Health recommended the use of a scoring system for the diagnosis of pulmonary TB in childhood, covering aspects of clinical, radiological and epidemiological data. The general objective of this work is the development and availability of a mobile application based on the score described in the Manual of Recommendations for TB Control in Brazil. The application was organized to make the questionnaire flow linear, while

maintaining the accordance with the structure presented in the manual. The score adapted to the Brazilian context allows health professionals to underpin their decisions with reliable information.

Carvalho I, Bernardi FA, Neiva MB, **Lima VC**, de Oliveira LL, Brandão Miyoshi NS, et al. COVID-19 BR: A web portal for COVID-19 information in Brazil. Procedia Comput Sci [Internet]. 2022;196(2021):525–32.

Disponível em: https://doi.org/10.1016/j.procs.2021.12.045

Brazil is a large developing country that requires attention to regionalized behaviors regarding the dissemination of COVID-19. To deal with this complexity, the COVID-19 Brazil observatory was developed. The Portal aims to monitor and analyze data from different sources. Therefore, with a detailed audit, we centralized this information on the evolution of the disease, allowing for territorial and temporal monitoring. The daily publication of numbers about COVID-19 allowed anyone to follow the current scenario in several Brazilian cities. With about 1,7 million accesses, the Portal offers clarity and an easy understanding of the pandemic data in the country.

Bernardi FA, Yamada DB, de Oliveira BM, **Lima VC**, Félix TM, Alves D. The minimum dataset for rare diseases in Brazil: a systematic review protocol. Procedia Comput Sci [Internet]. 2022;196(2021):439–44.

Disponível em: https://doi.org/10.1016/j.procs.2021.12.034

The Minimum Data Set (MDS) can be used for subsidiarity decision-making and health planning. Besides, this strategy allows to identify obligatory points that must be adjusted to achieve sustainable management in the planning and development of relevant Health Information Systems for public health. Specifically, in the context of rare diseases, the MDS strategy can be very valuable. This systematic review will focus on research using MDS for rare diseases in several databases. We seek to answer the question: "What is the minimum data set used in registries for rare diseases?" Some outcomes of interests specific for MDS will involve information about epidemiology, clinical procedures, and therapeutic resources among other features. We hope that by standardizing data through a careful analysis of evidence from different sources of a common format, with shared specifications and structures, we can help

in the methodological transparency and reproducibility of results in the context of rare disease research.

Trabalhos/pôsteres apresentados em conferências

Development Of An Integrated Semantic Api For Reproductive Health: An Intrapartum Case. HCist - International Conference on Health and Social Care Information Systems and Technologies. Lisboa, Portugal. 2018.

Development Of A Knowledge Portal For Tuberculosis Based On Semantic Web. HCist - International Conference on Health and Social Care Information Systems and Technologies. Lisboa, Portugal. 2018.

From social sensibility to interdisciplinary collaboration: The Brazilian COVID-19 epidemiology reference portal. University Social Responsibility: Priorities for the Next decade. Pretória, África do Sul. 2021.