



UNIVERSIDADE FEDERAL DE SANTA CATARINA
CENTRO TECNOLÓGICO
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA E GESTÃO DO
CONHECIMENTO

Maico Oliveira Buss

Transformação Digital na Gestão Descentralizada de Dados Digitais de Saúde Aplicada à
Pesquisa Clínica.

Florianópolis

2022

Maico Oliveira Buss

**Transformação Digital na Gestão Descentralizada de Dados Digitais de Saúde Aplicada
à Pesquisa Clínica.**

Tese submetida ao Programa de Pós-graduação em Engenharia e Gestão do Conhecimento da Universidade Federal de Santa Catarina para a obtenção do título de Doutor em Engenharia e Gestão do Conhecimento.

Orientador: Prof. Fernando Álvaro Ostuni Gauthier, Dr.

Coorientador: Prof. Marcelo Macedo, Dr.

Florianópolis

2022

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

Oliveira Buss, Maico
Transformação Digital na Gestão Descentralizada de Dados
Digitais de Saúde Aplicada à Pesquisa Clínica / Maico
Oliveira Buss ; orientador, Fernando Álvaro Ostuni
Gauthier, coorientador, Marcelo Macedo, 2022.
106 p.

Tese (doutorado) - Universidade Federal de Santa
Catarina, Centro Tecnológico, Programa de Pós-Graduação em
Engenharia e Gestão do Conhecimento, Florianópolis, 2022.

Inclui referências.

1. Engenharia e Gestão do Conhecimento. 2. Dados de
Saúde. 3. Pesquisas Clínicas. 4. Privacidade. I. Álvaro
Ostuni Gauthier, Fernando. II. Macedo, Marcelo. III.
Universidade Federal de Santa Catarina. Programa de Pós
Graduação em Engenharia e Gestão do Conhecimento. IV. Título.

Maico Oliveira Buss

Título: Transformação Digital na Gestão Descentralizada de Dados Digitais de Saúde Aplicada à Pesquisa Clínica.

O presente trabalho em nível de doutorado foi avaliado e aprovado por banca examinadora composta pelos seguintes membros:

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

Prof. Edevard José de Araújo, Dr.
Universidade Federal de Santa Catarina

Prof. Rogério Cid Bastos, Dr.
Universidade Federal de Santa Catarina

Prof. Vinicius Faria Culmant Ramos, Dr.
Universidade Federal de Santa Catarina

Certificamos que esta é a **versão original e final** do trabalho de conclusão que foi julgado adequado para obtenção do título de doutor em Engenharia e Gestão do Conhecimento.

Roberto Carlos dos Santos Pacheco, Dr.
Coordenador do Programa de Pós-Graduação

Prof. Fernando Álvaro Ostuni Gauthier, Dr.
Orientador

Florianópolis, 2022.

A Deus,
início e fim de tudo.
A minha filha Helena,
principal produção de minha vida.
A minha esposa Lucimara,
pelo apoio e compreensão.
A toda minha família,
base sólida de criação.

AGRADECIMENTOS

A realização deste trabalho não teria sido possível se não fosse todo o apoio que recebi. Deste modo, gostaria de agradecer a alguns daqueles que foram suporte desta caminhada.

Início agradecendo primeiramente a Deus, pelos dons que me concede diariamente. A minha família, Lucimara minha esposa e Helena nossa filha, companheiras em todas as horas. Ao meu pai Marcos, mãe Margarete, meu irmão Murilo, que mesmo longe, se fazem presentes sempre que preciso. Com eles a todo o restante da família, avó, avô, sogros, cunhados, compadres, todos sempre prontos para apoiar e festejar desde as menores conquistas.

Deixo aqui o meu registro do agradecimento a Universidade Federal de Santa Catarina e ao Departamento de Engenharia e Gestão do Conhecimento e toda sua equipe técnica, docente e discente. Meu especial agradecimento ao meu orientador Prof. Fernando Álvaro Ostuni Gauthier, pela confiança em tantos momentos de dúvidas, parceiro pela Universidade e amigo para a vida. Ao professor Marcelo Macedo, coorientador deste trabalho, por todo o auxílio e conversas que foram indispensáveis para a conclusão deste trabalho.

Agradecer a banca de qualificação que foi de suma importância para guiar os passos necessários para consolidar a estrutura deste trabalho, bem como agradecer previamente a banca de defesa que aceitou o desafio de contribuir e auxiliar no refinamento final da pesquisa realizada.

As minhas colegas de trabalho no Hospital Universitário, que me suportaram durante esse período de divagações revolucionárias e que, muitas vezes, desdobraram-se para permitir meu avanço nesta pesquisa.

Aos colegas de curso, de onde obtive novos amigos e inúmeras histórias para a vida.

E, por fim, agradeço a todos que contribuíram de forma direta ou indireta e que torceram por mim ao longo de mais esta caminhada.

Muito obrigado.

Que homem é um homem que não torna o mundo melhor?

(Jan Guillou, 1999)

RESUMO

As pesquisas clínicas, são processos de investigação científica envolvendo seres humanos. Nesses estudos as pessoas se voluntariam para testar novos tratamentos, intervenções ou testes como meio de prevenir, detectar, tratar ou controlar doenças ou condições médicas. Frente a sua relevância para a humanidade, as pesquisas clínicas precisam se valer da utilização dos melhores padrões científicos, éticos e de respeito aos seus participantes. O presente estudo possui como propósito, desenvolver um framework para apoiar a criação de conhecimento através das pesquisas clínicas, ao mesmo tempo em que se pretende colocar em evidência, a importância do conhecimento e consentimento dos participantes com o que acontece com os seus dados clínicos e pessoais, a fim de que estes se tornem agentes ativos das pesquisas das quais são sujeitos. Como metodologia, empregou-se a *Design Science Research Methodology (DSRM)*. A partir do reconhecimento e classificação em seis grupos (problemas de ordem financeira, tecnológica, social e política, científica, clínica e de privacidade) dos desafios encontrados na condução das pesquisas clínicas, foram definidos requisitos que auxiliaram na elaboração do framework. A concepção do framework foi guiada por preceitos que buscaram, além de cumprir com o objetivo de preservar a propriedade e a privacidade dos dados dos participantes de ensaios clínicos, também intervir nos demais problemas levantados na literatura. O framework baseia-se na descentralização dos dados pessoais dos participantes, sendo estruturado aplicando padrões adotados internacionalmente de modo a maximizar seu reuso e a sua interoperabilidade. Para tal aplicou-se os conceitos do Social Linked Data (Solid) que, ao mesmo tempo traz respostas para os problemas de propriedade, privacidade e rastreabilidade dos dados, e busca atrelar o produto da tese a uma comunidade já consolidada e engajada em ressignificar o cuidado com a privacidade digital. De modo a analisar a aplicabilidade do framework, um protótipo de aplicativo web foi implementado em um cenário representativo daquele encontrado no dia a dia da condução de ensaios clínicos. Efetuada a classificação dos atuais desafios na condução de pesquisas clínicas e estruturados os conceitos do framework, foi realizada a avaliação de sua aplicação por meio da instanciação de um protótipo como prova de conceito. Este procedimento permitiu concluir-se que a pesquisa desenvolvida contribuiu ao fornecer parâmetros para a concepção de modelos derivados do framework proposto, intervindo positivamente na frente aos desafios impostos aos campos da privacidade e confiabilidade na transformação digital por que passam as pesquisas clínicas.

Palavras-chave: dados de saúde; pesquisas clínicas; privacidade.

ABSTRACT

Clinical research is a process of scientific investigation involving human beings. In these studies, people volunteer to test new treatments, interventions, or tests as a means of preventing, detecting, treating, or controlling disease or medical conditions. Given its relevance to humanity, clinical research needs to use the best scientific and ethical standards and respect for its participants. The purpose of this study is to develop a framework to support the creation of knowledge through clinical research, at the same time as it intends to highlight the importance of knowledge and consent of the participants with what happens with their clinical data and individuals, so that they become active agents of the research to which they are subjects. As a methodology, the Design Science Research Methodology (DSRM) was used. Based on the recognition and classification into six groups (financial, technological, social and political, scientific, clinical and privacy problems) of the challenges encountered in conducting clinical research, requirements were defined that helped in the elaboration of the framework. The design of the framework was guided by precepts that sought, in addition to complying with the objective of preserving the ownership and privacy of data from clinical trial participants, also to intervene in other problems raised in the literature. The framework is based on the decentralization of the participants' personal data, being structured by applying internationally adopted standards in order to maximize its reuse and interoperability. To this end, the concepts of Social Linked Data (Solid) were applied, which, at the same time, brings answers to the problems of ownership, privacy and traceability of data, and seeks to link the product of the thesis to a community already consolidated and engaged in resignifying the beware of digital privacy. In order to analyze the applicability of the framework, a prototype of a web application was implemented in a scenario representative of that found in the day-to-day of conducting clinical trials. After classifying the current challenges in conducting clinical research and structuring the concepts of the framework, an evaluation of its application was carried out through the instantiation of a prototype as a proof of concept. This procedure allowed us to conclude that the research developed contributed by providing parameters for the design of models derived from the proposed framework, positively intervening in front of the challenges imposed on the fields of privacy and reliability in the digital transformation that clinical research undergoes.

Keywords: Clinical trial; Data decentralization; Privacy.

LISTA DE FIGURAS

Figura 1 - Fluxo dos dados nas pesquisas clínicas	18
Figura 2 - Abordagem em relação a centralização ou descentralização dos dados nas publicações	51
Figura 3 - Declaração RDF simples.	54
Figura 4 - Estrutura da SNOMED CT	58
Figura 5 - Arquitetura Solid	60
Figura 6 - Ilustração exemplificando o funcionamento do WAC	61
Figura 7 - Etapas da Metodologia Design Science Research	65
Figura 8 – Ilustração do fluxo de atividades Framework	71
Figura 9 - Visão ilustrativa do funcionamento do modelo	73
Figura 10 - Recorte ilustrando a estrutura dos pods utilizados no Framework	74
Figura 11 - Estatísticas BioPortal	76
Figura 12 - Tela do Protégé exibindo classes da ontologia	80
Figura 13 - Exemplo de visualização dos relacionamentos entre classes e	80
Figura 14 - Recorte de um arquivo RDF representando um recurso do tipo Paciente	81
Figura 15 - Criação do Solid Pod do Pesquisador	83
Figura 16 - Tela inicial da APCL	84
Figura 17 - Login no provedor	84
Figura 18 - APCL após a conexão com o Solid Pod do Pesquisador	85
Figura 19 - Exemplo de interação com dados privados e públicos	86
Figura 20 – Exemplo de página de autorização de acesso	87
Figura 21 - Permissões da APCL no Pod v2apcl.....	87
Figura 22 - Exibição da data de nascimento gravada no Pod v1apcl	88
Figura 23 - Tentativa de acesso a dado do voluntário sem permissão	89

LISTA DE QUADROS

Quadro 1 - Trabalhos do EGC que utilizaram a DSR para desenvolver frameworks	24
Quadro 2 - Categorização dos achados na bibliografia	33
Quadro 3 - Relação das categorias e o aparecimento de suas características na literatura.....	35

LISTA DE TABELAS

Tabela 1 - Estratégia de Busca	32
Tabela 2 - Documentos norteadores das pesquisas clínicas	36
Tabela 3 - Tópicos que devem constar em um protocolo de estudo clínico.....	42
Tabela 4 - Comparação entre (meta)paradigmas de pesquisa	66
Tabela 5 - Requisitos do Framework.....	68

LISTA DE ABREVIATURAS E SIGLAS

OMS Organização Mundial da Saúde

ANVISA Agência Nacional da Vigilância Sanitária

FDA US Food and Drug Administration

EMA European Medicines Agency

EHR Electronic Health Record

ECR Ensaio Clínico Randomizado

ISO International Organization for Standardization

SNOMED CT *Systematized Nomenclature of Medicine–Clinical Terms*

ICD *International Classification of Diseases*

CID Classificação Internacional de Doenças

NCBO National Center for Biomedical Ontology

LGPD Lei Geral de Proteção de Dados

GPDR *General Data Protection Regulation*

SUMÁRIO

1	INTRODUÇÃO	15
1.1	CONSIDERAÇÕES INICIAIS	15
1.2	IDENTIFICAÇÃO DO PROBLEMA DA PESQUISA	17
1.3	PERGUNTA DA PESQUISA	20
1.4	DEFINIÇÃO DOS OBJETIVOS	21
1.4.1	Objetivo Geral.....	21
1.4.2	Objetivos Específicos	21
1.5	JUSTIFICATIVA DA PROPOSTA.....	21
1.6	ADERÊNCIA AO PPEGC	23
1.7	PROCEDIMENTOS METODOLÓGICOS	26
1.7.1	CARACTERIZAÇÃO DA PESQUISA.....	26
1.7.2	PESQUISA TECNOLÓGICA.....	27
1.7.3	METODOLOGIA DA PESQUISA.....	29
1.8	DELIMITAÇÕES.....	30
1.9	ORGANIZAÇÃO DO TRABALHO	30
2	REFERENCIAL TEÓRICO	32
2.1	TRABALHOS RELACIONADOS	32
2.2	PESQUISAS CLÍNICAS	35
2.2.1	CONTEXTUALIZANDO A PESQUISA CLÍNICA.....	35
2.2.2	O PROTOCOLO DOS ENSAIOS CLÍNICOS.....	41
2.2.3	DESVIOS, VIOLAÇÕES E ALTERAÇÕES DE PROTOCOLO.....	45
2.2.4	PROBLEMAS DE INTEGRIDADE NOS PROTOCOLOS	45
2.2.5	O TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO.....	47
2.2.6	O DESAFIO DO CONSENTIMENTO NAS PESQUISAS CLÍNICAS.....	48
2.3	FERRAMENTAS	50

2.4	LINKED DATA	53
2.4.1	RESOURCE DESCRIPTION FRAMEWORK (RDF).....	53
2.5	ONTOLOGIAS.....	55
2.5.1	FOAF	56
2.6	SNOMED-CT	56
2.7	SOLID.....	58
2.7.1	A ARQUITETURA SOLID.....	60
2.8	CONSIDERAÇÕES FINAIS DO CAPÍTULO.....	63
3	DESENVOLVIMENTO DO FRAMEWORK.....	64
3.1	CONSIDERAÇÕES INICIAIS	64
3.1.1	IDENTIFICAÇÃO DO PROBLEMA E MOTIVAÇÃO.....	66
3.1.2	DEFINIÇÃO DOS OBJETIVOS PARA UMA SOLUÇÃO.....	66
3.1.3	DESIGN/PROJETO E DESENVOLVIMENTO.....	67
3.2	LEVANTAMENTO DOS REQUISITOS.....	68
3.3	ARQUITETURA DO FRAMEWORK	70
3.4	CONSIDERAÇÕES FINAIS DO CAPÍTULO.....	77
4	AMBIENTE EXPERIMENTAL.....	78
4.1	ONTOLOGIA DO DOMÍNIO ENSAIOS CLÍNICOS.....	78
4.2	INSTANCIAMENTO DO MODELO	82
4.3	AVALIAÇÃO.....	89
5	CONSIDERAÇÕES FINAIS.....	92
5.1	CONCLUSÃO	92
5.2	TRABALHOS FUTUROS	94
	REFERÊNCIAS	96
	APÊNDICE 1	103
	APÊNDICE 2	104

1 INTRODUÇÃO

1.1 CONSIDERAÇÕES INICIAIS

Pesquisas Clínicas, ensaios clínicos ou estudos clínicos são os termos utilizados para denominar um processo de investigação científica envolvendo seres humanos. São estudos em que as pessoas se voluntariam para testar novos tratamentos, intervenções ou testes como meio de prevenir, detectar, tratar ou controlar doenças ou condições médicas. Algumas investigações examinam como as pessoas respondem a uma nova intervenção, ou novas maneiras e combinações de aplicação de intervenções já existentes e, ainda, quais efeitos colaterais e resultados podem ocorrer. Isso ajuda a determinar se uma nova intervenção funciona, se é segura e se é mais eficaz do que as intervenções já disponíveis (KARBWANG; CRWLEY, 2002; MÉDICAS, 2022; TRIALS, 2021).

Dentre os vários tipos de pesquisa clínica existentes, esta tese se concentrará naqueles estudos conduzidos em seres humanos com objetivo de comprovar que um medicamento, vacina, produto para saúde, cosmético ou alimento, sujeitos ao controle sanitário, é seguro e funciona, denominando-os Ensaios Clínicos conforme o conceito da Agência Nacional da Vigilância Sanitária – ANVISA:

“pesquisa conduzida em seres humanos com o objetivo de descobrir ou confirmar os efeitos clínicos e/ou farmacológicos e/ou qualquer outro efeito farmacodinâmico do medicamento experimental e/ou identificar qualquer reação adversa ao medicamento experimental e/ou estudar a absorção, distribuição, metabolismo e excreção do medicamento experimental para verificar sua segurança e/ou eficácia;” (ANVISA, 2015)

A definição da Organização Mundial da Saúde (OMS) para um ensaio clínico é *“Qualquer estudo de pesquisa que atribua prospectivamente participantes humanos ou grupos de humanos a uma ou mais intervenções relacionadas à saúde para avaliar os efeitos sobre os resultados de saúde”*(HULLEY et al., 2013; TRIALS, 2021).

As intervenções dos ensaios clínicos incluem, mas não estão restritas a:

- drogas experimentais
- células e outros produtos biológicos
- vacinas
- dispositivos médicos
- procedimentos e tratamentos médicos e cirúrgicos
- terapias psicoterapêuticas e comportamentais
- mudanças no serviço de saúde

- estratégias de cuidados preventivos e
- intervenções educacionais.

Os pesquisadores também podem realizar pesquisas clínicas para avaliar testes de diagnóstico ou de triagem e novas maneiras de detectar e tratar doenças, ajudando a melhorar os serviços de saúde ao elevar os padrões de tratamento disponíveis. A maioria das intervenções médicas modernas são resultado direto de pesquisas clínicas (TRIALs, 2021).

A quantidade de estudos clínicos vem crescendo exponencialmente no Brasil nos últimos anos. Segundo dados obtidos do *Clinical Trials*¹, base de dados internacional que registra os estudos clínicos no mundo, existiam cerca de 2 mil estudos clínicos sendo conduzidos no Brasil em 2018, passando para mais de 3100 em 2021. Porém ainda hoje, existe um grande preconceito e desinformação envolvendo ações relacionadas as pesquisas clínicas (SBPPC, 2022).

Cedro e Duque (2019) ressaltam que não há dúvidas quanto aos inúmeros avanços nas ciências biomédicas obtidos por meio dos ensaios clínicos. Também apontam que, ao longo do tempo, esse ramo de estudos tem sofrido novas regulações e aperfeiçoamento de seus protocolos, mas ainda há grandes desafios a serem encarados. Apesar das exigências já existentes de regulação quanto ao registro e publicidade, é possível a constatação de ocorrências de pesquisas onde os dados não são devidamente registrados, e quando o são, esse procedimento ocorre de forma inconsistente ou incompleta, colocando em risco a segurança, a reprodutibilidade e a qualidade desses estudos.

A falta de reprodutibilidade, em específico, foi amplamente estudada e as publicações científicas médicas foram consideradas, em geral, não reproduzíveis. Ioannidis (2018) estimou uma taxa de cerca de 80% de estudos que não permitem, da forma como foram publicados, sua reprodução.

Frente a estes apontamentos, atualmente existe uma grande expectativa quanto à uma transformação digital na área com a incorporação de novas tecnologias que possam assegurar a confiança, rastreabilidade e irreversibilidade dos dados de pesquisa de ensaios clínicos. (CEDRO; DUQUE, 2019).

¹ <https://www.clinicaltrials.gov/>

1.2 IDENTIFICAÇÃO DO PROBLEMA DA PESQUISA

O desenvolvimento de medicamentos é um processo desafiador. Apenas um em cada 10 medicamentos em desenvolvimento chega ao mercado (HAY et al., 2014). Essa baixa taxa de entrada no mercado é um fator que contribui para os altos custos de desenvolvimento de medicamentos. Dimasi; Grabowski; Hansen (2016) indicam em seu estudo que o desenvolvimento de um medicamento da bancada para o mercado custa cerca de US\$ 2,6 bilhões. A necessidade de acesso a um grupo adequado de pacientes para a execução de ensaios clínicos é outra característica deste tipo de pesquisa. Alguns desses estudos exigem milhares de participantes, cada um dos quais deve atender a critérios precisos para participar. Atendendo aos rigorosos critérios de qualificação impostos pelos investigadores, apenas cerca de 5% dos candidatos acabam por constituir o grupo que realmente viabilizará os ensaios clínicos. Portanto, não é de surpreender que 80% desses estudos estejam atrasados devido a problemas de recrutamento e acompanhamento dos participantes. Longas fases de recrutamento prolongam a execução dos ensaios clínicos, aumentando assim o tempo necessário para que novos medicamentos sejam estudados e aprovados, deixando os pacientes esperando anos por novas opções de tratamento (ANGELETTI; CHATZIGIANNAKIS; VITALETTI, 2018).

Conforme apontado, a indústria farmacêutica global vem enfrentando desafios relacionados ao aumento de custo e necessidade de aceleração do desenvolvimento de medicamentos. Este duplo desafio - economia de tempo e custo – obrigou a globalização simultânea dos ensaios clínicos para os países em desenvolvimento. No entanto, as diferenças em processos regulatórios, questões éticas, perícia médica, prática clínica e infraestrutura de saúde, entre países desenvolvidos e os em desenvolvimento, tornam a manutenção da conformidade e qualidade em âmbito global, difícil e exigente. Para gerenciar esses ensaios clínicos complexos em vários países, a estratégia da indústria farmacêutica tem sido terceirizar o processo de execução dos ensaios clínicos (HANG et al., 2021).

No capítulo 2 deste trabalho faz-se um aprofundamento detalhado sobre as características dos ensaios clínicos. No momento é imperativo demonstrar que o atual padrão mais comumente aplicado de fluxo de dados nos ensaios clínicos consiste em diferentes etapas realizadas independentemente umas das outras (WENG et al., 2013), conforme pode ser observado na Figura 1. Neste modelo, os dados são criados a partir de distintas fontes (dispositivos inteligentes, locais de ensaios clínicos), processados e analisado por diferentes organizações em sua forma preferida e formato (HANG et al., 2021).

Durante a execução dos ensaios clínicos é gerada uma quantidade significativa de dados no intuito de aprovar novos medicamentos, instrumentos e procedimentos médicos ou cirúrgicos em participantes humanos. Os investigadores coletam dados dos participantes durante um período pré-determinado, incluindo, mas não somente, dados demográficos, clínicos, laboratoriais, de imagens, efeitos colaterais ou complicações causadas pela intervenção em análise.



Fonte: adaptado de (HANG et al., 2021).

Geralmente, o ensaio clínico requer colaboração entre diversas partes interessadas, incluindo órgãos reguladores, empresas farmacêuticas, centros clínicos e, o mais importante, as pessoas com as características necessárias para participarem do estudo. Para cada parte interessada existe um diferente conjunto de ferramentas que funciona como infraestrutura para a devida execução do ensaio clínico (BHATT, 2011; GARDNER; LIDZ; HARTWIG, 2005; RANG, 2013).

Tradicionalmente, a qualidade de um ensaio clínico é avaliada por auditorias conduzidas pelo patrocinador nos locais de realização das atividades além de inspeções regulatórias pelos órgãos de controle aplicáveis. Bhatt (2011) apontava que das inspeções de rotina do *Food and Drug Administration* (FDA) nos ensaios clínicos avaliados, 59% tiveram indicações de ações de correções voluntárias, 1% teve indicações de correções oficiais e 40% não possuíam nenhuma ação de correção necessária. Já em inspeções “motivadas”, aquelas decorrentes de denúncias ou amostragem de fiscalização por exemplo, os estudos com indicações para correções oficiais chegam a 23% da amostra. Como comparação, na Índia, de 23 inspeções conduzidas, 12 (52%) tiveram indicações de ações de correções voluntárias e 11 (48%) tiveram indicações para correções oficiais.

Outro desafio para garantir a qualidade dos ensaios clínicos é a sua reprodutibilidade. Ioannidis (2018) traz a estimativa de que cerca de 80% dos estudos clínicos não são

reproduzíveis. A qualidade dos dados dos ensaios clínicos pode ser diretamente afetada por erros na aquisição e transcrição dos dados dos participantes. Em sua publicação Ioannidis (2018), confirma a necessidade da existência de requisitos mais estritos e de maior supervisão nesta fase de coleta.

Em geral, garantir a qualidade dos dados não é um trabalho trivial. É necessário conhecimento tanto da atividade quanto da modelagem do domínio em que se atua, para que se tenha condições de manter a qualidade dos dados em um nível cientificamente aceitável (BENCHOUFI; PORCHER; RAVAUD, 2017). Estima-se que o monitoramento de ensaios clínicos por meio de visitas locais pode consumir um quarto de todo o orçamento do patrocinador para um ensaio clínico. Como a complexidade e a distribuição geográfica destes estudos cresceram, tornou-se cada vez mais caro manter este tipo de abordagem presencial (HIRANO et al., 2020; OLSEN et al., 2016).

Para as agências reguladoras como o US Food and Drug Administration (FDA) nos Estados Unidos, European Medicines Agency (EMA) na União Europeia, Agência Nacional de Vigilância Sanitária (ANVISA) no Brasil e a Pharmaceuticals and Medical Devices Agency (PMDA) no Japão, a auditoria de dados é um grande desafio, pois não há uma maneira fácil e segura de acessar ou visualizar dados complexos, que garanta a rastreabilidade desde a etapa de anuência para uso até seu processamento. Esta é uma demanda tão relevante que a melhoria do processo é uma área extremamente ativa de pesquisas (BHATT, 2011; HIRANO et al., 2020).

Também há de se considerar o ponto de vista do participante dos estudos clínicos. Durante a execução dos ensaios, a coleta de dados de alta qualidade é absolutamente vital. Por esta razão, os centros de pesquisas exigem a realização de testes e observações regulares em suas instalações para garantir a precisão da coleta dos dados. Aproximadamente 70% dos potenciais participantes vivem a mais de duas horas do centro de pesquisa mais próximo (DIMASI; GRABOWSKI; HANSEN, 2016). É, portanto, comum que os pacientes se desloquem a esses centros para exames e observações regulares, às vezes várias vezes por semana durante a duração do estudo. Essas dificuldades, por vezes, superam os benefícios percebidos de participar de um estudo, aumentando inevitavelmente a taxa de abandono dos participantes (ANGELETTI; CHATZIGIANNAKIS; VITALETTI, 2018).

De acordo com uma pesquisa conduzida por (PICKARD; SWAN, 2014), 85% dos entrevistados apontam a preocupação com sua privacidade como sendo uma barreira para compartilhar informações de saúde. Os participantes relatam que receiam que os dados

coletados podem ser usados para obter ou inferir informações confidenciais sobre a vida privada, hábitos, atividades e relações dos usuários, todas diretamente relacionadas com a privacidade dos indivíduos. Cerca de metade dos entrevistados demonstraram preocupação com uma possível re-identificação de suas informações médicas e de saúde fornecidas de forma anônima. Já se os dados fossem irreversivelmente anonimizados, 71% dos entrevistados estariam dispostos a compartilhá-los com os pesquisadores. Os autores complementam afirmando que com a crescente demanda por compartilhamento dos dados dos estudos clínicos pode aumentar o risco de invasões de privacidade ou violações de confidencialidade das informações dos participantes, possibilitando a ocorrência de danos sociais/morais ou econômicos (ANGELETTI; CHATZIGIANNAKIS; VITALETTI, 2018).

Crisigiovanni (2017) expõe que apenas reconhecer que a propriedade dos dados deve permanecer sendo dos participantes não é o suficiente. Atualmente a proteção de dados pessoais não é garantida. É necessária a busca por uma economia baseada em dados mais sustentável e responsável, com foco em soluções inovadoras e serviços personalizados que melhor atendam às necessidades reais dos participantes, contribuindo para melhorar suas vidas e a sociedade como um todo, protegendo sempre os seus dados pessoais.

Compreender esses problemas e abordá-los adequadamente é fundamental para o desenvolvimento de soluções de saúde digital bem-sucedidas. À medida que a tecnologia se torna mais acessível e economicamente viável à maior parte da população, o papel dos dados digitais de saúde se tornará cada vez mais vital em ensaios clínicos (HUTCHINGS et al., 2020). Angeletti et al. (2018) defendem a importância de se disponibilizar novos métodos que facilitem o recrutamento de participantes para os estudos, de maneira tal que se garanta o respeito, a privacidade, a confidencialidade e a rastreabilidade dos dados, a fim de, além dos benefícios diretamente provenientes dos ensaios clínicos, maximizar a motivação em se tornarem participantes desses estudos.

1.3 PERGUNTA DA PESQUISA

Como apoiar a criação de conhecimento por meio dos ensaios clínicos, garantindo a confiabilidade, integridade, a rastreabilidade e a privacidade dos dados dos participantes?

1.4 DEFINIÇÃO DOS OBJETIVOS

1.4.1 Objetivo Geral

Construir um framework para orientar a transformação digital na gestão descentralizada de dados de saúde, buscando a preservação da propriedade proporcionando confiabilidade e rastreabilidade, aplicado a criação de conhecimento por meio de pesquisas clínicas.

1.4.2 Objetivos Específicos

1. Classificar os desafios encontrados na utilização dos dados pessoais em ensaios clínicos;
2. Levantar as técnicas atualmente utilizadas para tratamento dos dados clínicos e pessoais em ensaios clínicos e suas vantagens e desvantagens;
3. Estruturar num framework os critérios e padrões de modo a garantir confiabilidade, rastreabilidade e a privacidade dos dados dos participantes em ensaios clínicos;
4. Analisar a aplicabilidade do framework proposto em um cenário representativo ao real.

1.5 JUSTIFICATIVA DA PROPOSTA

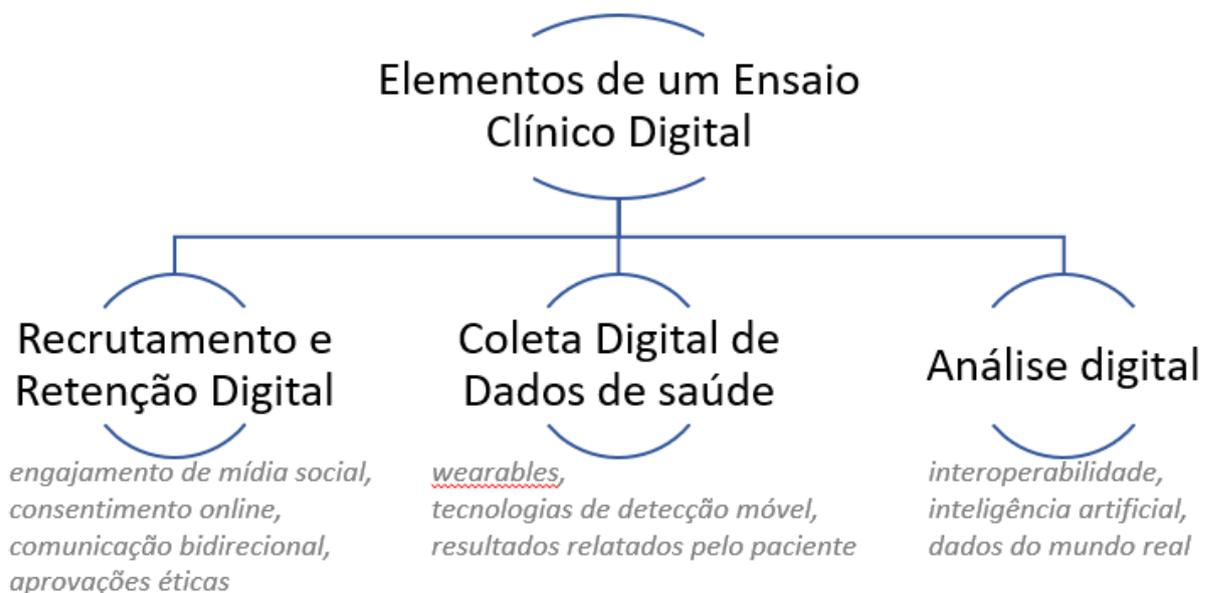
Cada novo medicamento e tratamento começou com voluntários participando de ensaios clínicos. Devemos os altos padrões atuais de assistência médica a estudos que foram conduzidos no passado sob a orientação das normas internacionais que guiam este tipo de pesquisa. Além de testar novos medicamentos e dispositivos, os ensaios clínicos fornecem uma base científica para aconselhar e tratar os pacientes. Mesmo quando os pesquisadores não obtêm os resultados que previram, os resultados dos testes podem ajudar a apontar os cientistas na direção correta (CECCHINI et al., 2019; NOVITZKE, 2008).

Ensaio clínicos bem desenhados e executados, oferecem uma oportunidade para que os participantes desempenhem um papel ativo na contribuição para o avanço do conhecimento nas ciências biomédicas além dos seus próprios cuidados de saúde, obtendo acesso a novos tratamentos na fase de pesquisa antes que eles estejam amplamente disponíveis, aumentando suas opções quando não há um tratamento disponível ou mesmo quando a opção padrão não é mais indicada. Porém, mesmo com os esforços regulatórios para a proteção desses participantes

e para garantir que as pessoas tenham informações confiáveis ao decidirem participar de um ensaio clínico, alguns riscos podem ser inevitáveis devido à incerteza inerente aos estudos de pesquisa médica (ANGELETTI; CHATZIGIANNAKIS; VITALETTI, 2018; CECCHINI et al., 2019; NOVITZKE, 2008).

Outros desafios concernentes a execução de ensaios clínicos tradicionais, podem dificultar a condução eficiente das pesquisas. A literatura aponta para ineficiências operacionais relacionadas à identificação, recrutamento, aquisição de dados e acompanhamento dos participantes que, inflacionam os custos, aumentam a carga social dos participantes e estendem os já longos prazos de ensaios clínicos, influenciando para uma baixa participação de voluntários nestes estudos (CECCHINI et al., 2019; NOVITZKE, 2008).

Como forma de se beneficiar dos avanços tecnológicos recentes, vem sendo apresentado o conceito de ensaio clínico digital, que promete ser uma forma de superar alguns dos desafios tradicionais dos ensaios clínicos. Um ensaio clínico digital é aquele que usa a tecnologia para melhorar o recrutamento, a retenção, a coleta e a análise de dados (INAN et al., 2020).



Fonte: baseado em (INAN et al., 2020)

A oportunidade de reduzir os custos e esforços por meio da utilização de tecnologias digitais, ao mesmo tempo em que se caminha para uma experiência de pesquisa centrada no usuário, tem capturado a atenção de especialistas em ensaios clínicos, tecnologias e análise de

dados para discutir estratégias e desafios para a implementação do uso dessas tecnologias com responsabilidade para preservar o valor e a confiança dos ensaios clínicos (CECCHINI et al., 2019; INAN et al., 2020).

Com a modernização do processo de pesquisa, surge a necessidade de esforços para modernizar os padrões de privacidade, segurança, ética e regulamentação. A digitalização e o monitoramento remoto aumentam a necessidade de medidas de segurança para proteção contra violações durante a coleta, transmissão e/ou armazenamento de dados. Essas violações que podem colocar os participantes em risco e causar ações legais e perdas econômicas para os envolvidos. Agindo em relação a isso, organizações como a FDA adotaram a segurança cibernética como item de relevante preocupação. Os riscos e desafios relacionados ao gerenciamento dos ensaios clínicos digitais vem sendo uma área de estudo bastante incentivada e várias estratégias diferentes vem sendo aplicadas na busca pela manutenção da privacidade dos participantes sem que se tenha perdas em relação a validade e qualidade dos ensaios clínicos. Para esta tese foi efetuada uma revisão da literatura no sentido de se classificar os desafios encontrados na utilização dos dados pessoais em ensaios clínicos e levantar as técnicas atualmente propostas para o gerenciamento desses dados (CECCHINI et al., 2019; INAN et al., 2020).

O levantamento realizado, detalhado no capítulo 2, evidencia a necessidade da continuidade de estudos relacionados a questões ligadas a privacidade, utilização de novas tecnologias e avanços na área de transparência e confiabilidade dos dados dos ensaios clínicos. Este trabalho se fundamenta na criação de um framework abrangente que considere aspectos relacionados à preservação da privacidade dos participantes dos ensaios clínicos ao mesmo tempo que pretende conferir maior confiabilidade a estes estudos por meio da rastreabilidade dos dados utilizados.

1.6 ADERÊNCIA AO PPEGC

Esta tese aborda a transformação digital na gestão de dados digitais de saúde pessoal. Sua aplicação se dá por meio do desenvolvimento de um framework para suporte à criação de conhecimento nas pesquisas clínicas com a aplicação de recursos que visam garantir a privacidade dos participantes destas pesquisas, por meio da provisão de confiabilidade, rastreabilidade e proteção de seus dados pessoais. Este trabalho está inserido na área de concentração da Engenharia do Conhecimento, pois considera em seu arcabouço de definições

o conhecimento como um processo e produto, tangível ou intangível, efetivado na relação entre pessoas e agentes não humanos para a geração de valor, visando a pesquisa e o desenvolvimento de métodos, técnicas e ferramentas para a construção de modelos e sistemas de conhecimento em atividades intensivas em conhecimento (PPEGC, 2022).

Em relação a linhas de pesquisa, esta tese se enquadra naquela da Engenharia do Conhecimento Aplicada às Organizações, que aborda a concepção, desenvolvimento e implantação de soluções da EC em organizações públicas e privadas e, naquela da Modelagem e Representação do Conhecimento (PPEGC, 2022).

Neste contexto, disciplinas como Teoria Geral dos Sistemas, Fundamentos de Engenharia do Conhecimento, Métodos e Técnicas de Extração de Conhecimento e Ambientes Computacionais Distribuídos para Engenharia do Conhecimento oferecem o arcabouço teórico capaz de fundamentar iniciativas como a proposta deste trabalho, ao capacitarem os alunos com os conceitos básicos ligados a estratificação e classificação dos tipos de conhecimento envolvidos em cada um dos processos e atividades organizacionais, modelagem desse conhecimento e características e desenvolvimento de ontologias, conceitos amplamente aplicados nesta iniciativa.

A relação deste trabalho com os trabalhos já publicados do PPGEGC, é perceptível com base na análise de que até março de 2022 foram publicados mais de 78 trabalhos, entre teses e dissertações, contendo como mote o desenvolvimento de framework ou modelo de conhecimento, tanto na Engenharia do Conhecimento (EC) quanto na Gestão do conhecimento (GC). Entre estes trabalhos, foi efetuada uma análise com base nos seus títulos e resumos com vista a buscar referências aplicáveis para o alcance dos objetivos centrais propostos nesta tese. Apesar de não terem sido encontrados trabalhos que abrangessem a preocupação com dados pessoais sensíveis na mesma área de aplicação desta tese, foi possível identificar correlações na abordagem metodológica utilizada, tanto em trabalhos de dissertação como em teses de doutorado, todas utilizando a Design Science Research (DSR), conforme exposto no Quadro 1.

Quadro 1 - Trabalhos do EGC que utilizaram a DSR para desenvolver frameworks

Ano	Título	Autor	Orientador	Tipo
2021	<i>Framework</i> de indicadores para gestão da inovação em empresas de base tecnológica	BERNARD, André Peressoni	Prof. ^a Dr. ^a Gertrudes Aparecida Dandolini	Dissertação
2021	<i>Framework</i> Conceitual Baseado em	KASTER, Gerson Bovi	Prof. Dr. Aires José Rover	Dissertação

	Aprendizagem de Máquina Supervisionada para Concepção de Sistemas Judiciais de Agentes Inteligentes			
2021	k-SCAS: <i>Framework</i> do sistema de agronegócios de cafés especiais orientado ao conhecimento	TRAUER, Eduardo	Prof. Dr. Eduardo Moreira da Costa	Tese
2021	Indicadores socioeconômicos produzidos a partir de Big Data: Um <i>framework</i> para avaliação da qualidade estatística aplicado ao turismo	CAUMO, Rafael Bassegio	Prof. Dr. João Artur de Souza	Tese
2020	<i>Framework</i> conceitual do potencial de coprodução de inovação em ecossistemas de inovação	MARQUES, Maria Angelica Jung	Prof. ^a Dr. ^a Andrea Valéria Steil	Tese
2020	Gestão do conhecimento do cliente e destinos turísticos inteligentes: um <i>framework</i> para a gestão inteligente da experiência turística - SMARTUR	MUNIZ, Emerson Cleister Lima	Prof. ^a Dr. ^a Gertrudes Aparecida Dandolini	Tese
2019	Framework de análise de conhecimentos críticos às capacidades de resiliência organizacional	FRAGA, Bruna Devens	Prof. Dr. Gregório Varvakis	Tese
2018	<i>Metaframework</i> de coprodução em ambientes complexos para a geração de insumos estratégicos	NASCIMENTO, Everton Ricardo do	Prof. Dr. Paulo Mauricio Selig	Tese
2015	<i>Framework</i> baseado em conhecimento para análise de rede de colaboração científica	BORDIN, Andréa Sabedra	Prof. Dr. Alexandre Leopoldo Gonçalves	Tese

Fonte: Elaborado pelo autor

1.7 PROCEDIMENTOS METODOLÓGICOS

Esta seção visa descrever a metodologia utilizada no trabalho a fim de classificá-lo dentre as diferentes visões de pesquisas com validade científica. Para Silva; Menezes (2005), “[...] pesquisa é um conjunto de ações propostas para encontrar a solução para um problema, que tem por base procedimentos racionais e sistemáticos”. Assim, quanto ao ponto de vista de sua natureza, este trabalho pode ser associado ao ramo das pesquisas aplicadas, mais precisamente daquelas descritas como tecnológicas. Caracterizadas por ter como objetivo gerar conhecimentos para aplicação prática seja em forma de produtos ou processos, com vistas a resolver problemas específicos, envolvendo verdades e interesses locais. Para tal, ela faz uso dos conhecimentos obtidos através da pesquisa básica, em conjunto com a utilização da tecnologia existente.

1.7.1 CARACTERIZAÇÃO DA PESQUISA

O ato de pesquisar pode ser tomado de maneira simplista, como a busca por respostas à um problema proposto. Para (SILVA; MENEZES, 2001), “pesquisa é um conjunto de ações, propostas para encontrar a solução para um problema, que têm por base procedimentos racionais e sistemáticos. A pesquisa é realizada quando se tem um problema e não se tem informações para solucioná-lo.”

Uma pesquisa pode ser classificada de diferentes formas, tomando como parâmetros para essa classificação os seguintes itens:

- Natureza (SILVA; MENEZES, 2001): pesquisa básica e pesquisa aplicada;
- Abordagem do problema (SILVA; MENEZES, 2001): pesquisa quantitativa e pesquisa qualitativa;
- Objetivos (GIL, 2009): pesquisa exploratória, descritiva e explicativa;
- Procedimentos técnicos (GIL, 2009): pesquisa bibliográfica, documental, de levantamento, estudo de caso, ex-post-facto, pesquisa-ação e pesquisa participante;

Como mencionado anteriormente a pesquisa busca a criação de novo conhecimento, porém, para que este conhecimento gerado tenha valor científico a pesquisa deve obedecer a critérios pré-definidos “a) a existência de uma pergunta que se deseja responder; b) a elaboração de um conjunto de passos que permitam chegar à resposta; c) a indicação do grau de

confiabilidade na resposta obtida” (GOLDENBERG, 2004). Com isto exposto, pode-se concluir que para que uma pesquisa possa ser considerada como de caráter científico é necessária que seja constituída por uma investigação planejada, e que esta, através do ferramental da metodologia escolhida, seja capaz de contemplar os critérios citados.

Partindo desse arcabouço e em linha com a visão de Dresch, Lacerda e Antunes Jr. (2015), este trabalho possui como razão para a realização da pesquisa a busca por uma solução para um problema prático, com objetivo de propor um artefato e prescrever soluções, utilizando-se de método científico de cunho dedutivo, indutivo e abduutivo, orientado pelo método de pesquisa Design Science Research (DSR) e executando, para tanto, técnicas de investigação bibliográfica e documental, buscando atingir o propósito da pesquisa caracterizada como tecnológica.

1.7.2 PESQUISA TECNOLÓGICA

A chamada pesquisa tecnológica/de cunho tecnológico, vem ganhando cada vez mais espaço na academia, especialmente em áreas como engenharia e computação, campos do saber humano que se ocupam principalmente do desenvolvimento de novos artefatos nem sempre baseados no conhecimento científico clássico (FREITAS JUNIOR ET AL., 2014).

A pesquisa tecnológica está classificada quanto à sua natureza dentro do ramo das pesquisas aplicadas, o que a caracteriza como objetivando gerar conhecimentos para aplicação prática, seja como produtos ou processos, em vista de resolver problemas específicos. Para tal, ela faz uso dos conhecimentos obtidos através da pesquisa básica, em conjunto com a utilização da tecnologia existente.

Cupani (2006) afirma que a tecnologia não pode ser considerada, mera e simplesmente, a aplicação do conhecimento científico exatamente pelo fato de que, muitos dos seus resultados não provém, de forma linear, da ciência clássica. O conhecimento científico, em primeira análise, diferencia-se do conhecimento tecnológico a partir da constatação de que o primeiro propõe teorias de aplicação ampla, enquanto que o segundo é responsável pela geração de teorias bastante restritas, cujo foco é a solução de problemas específicos. Dessa forma, se entende a partir de Bunge (2003), que a tecnologia pode ser entendida como “o campo do conhecimento relativo ao projeto de artefatos e ao planejamento de sua realização, operação, ajuste, manutenção e monitoramento, à luz do conhecimento científico”. O artefato, para o autor, não necessariamente é algo claramente definido, mas pode ser considerado uma

modificação do estado de um sistema natural, por meio de uma mudança artificial (como o desvio do curso de um rio), uma mudança social (como a organização de um coletivo), ou ainda o resultado de um serviço (como os cuidados a pacientes). Assim, evidencia-se que a pesquisa tecnológica objetiva a solução de problemas específicos, tendo foco no artefato a ser desenvolvido, lembrando-se de que este artefato não necessariamente será algo material, mas um projeto ou uma intervenção artificial sobre um sistema.

Conforme Buss (2015), a metodologia empregada na pesquisa tecnológica tem como diferencial o fato de objetivar a criação de uma inovação em resposta a uma demanda já conhecida. A partir dessa afirmação é possível elencar as etapas que compõe a estrutura da metodologia de pesquisa tecnológica:

1. Identificação da necessidade: É necessário o conhecimento das necessidades do mercado ou mesmo de ideias originais do pesquisador, para que através destas seja possível a definição do tema que será abordado, originando objetivos e metas para supri-las;

2. Objetivo: De modo geral têm-se como objetivos uma coisa concreta, que possua parâmetros numéricos de controle e medida;

3. Metas: são segmentações do objetivo, etapas que devem ser completadas para o cumprimento do objetivo;

4. Projeto: Nas etapas anteriores definiu-se a necessidade, ou seja, o problema à investigar, o que será feito e como será feito. A fase de projeto constitui a complementação dos passos anteriores com a definição de quando fazer, e com o que fazer. É o levantamento dos recursos necessários à plena execução das atividades previstas e a sua determinação de viabilidade;

5. Modelagem e Desenvolvimento: compreende o desenvolvimento de um protótipo, sua avaliação, otimização e aplicação.

A pesquisa tecnológica, portanto, deve valer-se cada vez mais de enunciados e métodos científicos para dar-lhe a segurança necessária para o avanço consciente da inovação e da própria tecnologia (FREITAS JUNIOR ET AL., 2014). Neste aspecto, apresenta-se como opção metodológica para o desenvolvimento de uma pesquisa tecnológica, os princípios e fundamentos de Design Science Research Methodology (DSRM).

1.7.3 METODOLOGIA DA PESQUISA

A metodologia Design Science Research (DSR) recentemente vem se destacando como um método de investigação adequado a trabalhos na área científica. Ela tem sua origem na diferenciação entre os ambientes naturais e artificiais proposta por Herbert Simon. Para o autor, a ciência natural seria aquela que se ocupa de descrever e ensinar como os fenômenos naturais funcionam, interagem com o mundo. Entretanto, em contraponto ao mundo natural, Simon (2019) apresenta uma discussão acerca da necessidade, igualmente, de se estudar o universo daquilo que é considerado “artificial”, afirmando que as “ciências do artificial se ocupam da concepção de artefatos que realizem objetivos” (SIMON, 2019).

A DSR tem sido amplamente adotada em muitos campos de pesquisa, principalmente nas áreas ligadas a sistemas e ciências da informação, como forma de modelar as pesquisas, destacando tanto o problema abordado quanto a intervenção proposta, ao mesmo tempo em que auxilia o pesquisador a transmitir como ele desenvolve sua contribuição para a base de conhecimento existente (STOREY et al. 2017).

A DSR consiste em buscar e modelar artefatos que interajam com o contexto do problema e venham a atuar nesse contexto, interagindo de forma a se atingir o objetivo proposto (WIERINGA 2014). Por meio da DSR, busca-se através das experiências adquiridas para a concepção do artefato, e conseqüente resolução do problema de base, fornecer acréscimos ao conhecimento até então existente, por exemplo, estendendo teorias e métodos utilizados durante a pesquisa (HEVNER E CHATTERJEE 2010).

Embora não haja um método consensual e consagrado sobre como fazer pesquisa na abordagem DSR, Dresch, Lacerda e Antunes Jr (2015) identificam algumas etapas recorrentes entre os diferentes métodos: Definição do problema; Revisão da literatura e busca por teorias existentes; Sugestões de possíveis soluções; Desenvolvimento; Avaliação; Decisão sobre a melhor solução; Reflexão e aprendizagens; e Comunicação dos resultados. Seguindo esta linha, a condução da DSR nesta tese se dará conforme Peffers et al., (2007), chamada de Design Science Research Methodology (DSRM). O DSRM fornece uma metodologia na forma de roteiro para apresentar e conduzir a DSR, clarificando assim o enquadramento metodológico das pesquisas conduzidas sobre esta metodologia conforme será detalhado no capítulo 4.

1.8 DELIMITAÇÕES

O escopo deste trabalho baseia-se na proposição e desenvolvimento de um framework capaz de guiar a transformação digital na gestão de dados pessoais de saúde e apoiar a criação de conhecimento que naturalmente se dá na execução de ensaios clínicos, com o objetivo de aumentar a confiabilidade, a rastreabilidade e a proteção dos dados pessoais dos participantes. A tese não pretende evoluir nas questões relacionadas aos critérios e protocolos próprios para o delineamento e execução dos ensaios clínicos.

A fim de clarificar a abordagem deste trabalho, o conceito de framework adotado é o de uma estrutura que suporta a compreensão e a construção de algo sobre sua base inicial, podendo representar um problema e fornecer as bases para a resolução de um propósito em específico. Um framework pode também ser denominado como híbrido, quando se preocupa em realizar a compreensão de um determinado fenômeno, e aplicá-lo a um contexto real, orientando inclusive, sua própria implementação. Assim, a construção de frameworks híbridos está intimamente relacionada a Design Science Research, na qual frameworks podem ser artefatos propostos para a resolução de um problema real, ao mesmo tempo em que ao pesquisador é permitido fazer uma contribuição científica sobre a compreensão do fenômeno que se estuda (LACERDA et al., 2013; SHEHABUDDEEN et al., 1999; TOMHAVE, 2005)

Apesar de possuir entre seus objetivos o desenvolvimento de um artefato, não se tem como propósito nesta tese prover um produto final, mas sim um protótipo na forma de software para a avaliação do framework por meio de cenários embasados em exemplos de ensaios clínicos reais. Por isso não haverá preocupação em relação ao desempenho ou outros critérios tipicamente comerciais de uma solução computacional.

As bases de dados utilizadas no protótipo serão populadas com informações geradas por máquina, com base em regras definidas para representar, com a maior similaridade viável, as circunstâncias encontradas nas práticas das pesquisas clínicas na realidade.

1.9 ORGANIZAÇÃO DO TRABALHO

Esta seção apresenta a forma como o trabalho está estruturado, ao longo de seus cinco capítulos. O primeiro capítulo foi dedicado à introdução, tendo por objetivo a contextualização do tema, caracterização do problema, definição de objetivos, justificativa, aderência do assunto

da tese ao programa de Pós-Graduação em Engenharia e Gestão do Conhecimento, um apanhado sobre os procedimentos metodológicos aplicados e as delimitações do trabalho.

O segundo capítulo aprofunda a conceituação de pesquisa clínica, em particular daquela definida como ensaio clínico. Apresenta seu histórico e sua contemporaneidade, buscando demonstrar o quão fundamental é este tipo de ciência para o avanço da medicina e, por conseguinte, da saúde da raça humana. Ao caracterizar as etapas e componentes da pesquisa clínica, apresenta os resultados da pesquisa na literatura e também ilustra o quanto ainda se pode avançar, principalmente do ponto de vista da segurança dos pacientes participantes, focando na privacidade e segurança dos dados sensíveis pessoais destes.

O terceiro capítulo contextualiza e apresenta as ferramentas base utilizadas para a criação do framework, demonstrando a organização dos conceitos que o compõe.

O quarto capítulo relata a proposição do trabalho por meio da construção do framework utilizando a metodologia DSRM, a definição de sua estrutura e funcionamento e uma prova de conceito evidenciando a arquitetura e os módulos que compõem o sistema computacional desenvolvido com base no framework proposto.

Por fim, o quinto capítulo, traz as considerações finais sobre o trabalho. São apresentadas, ainda, possibilidades e sugestões para trabalhos futuros que possam ser desenvolvidos a partir desta tese.

2 REFERENCIAL TEÓRICO

Este capítulo tem como objetivo evidenciar os conceitos que compõe a área temática desta tese. Busca-se contextualizar as pesquisas clínicas, ilustrando suas características e particularidades frente a outras áreas da ciência. Apresenta-se também, os desafios encontrados através da revisão da literatura, que atualmente afetam esse campo de estudos, classificando-os e apresentando as técnicas e abordagens até então utilizadas para seu tratamento.

As bases conceituais que sustentam a construção do framework e a escolha das tecnologias aplicadas, bem como seu inter-relacionamento, são demonstradas fornecendo um embasamento científico para a proposta da pesquisa.

2.1 TRABALHOS RELACIONADOS

Para verificar a originalidade e ineditismo da abordagem proposta, foi realizada uma revisão da literatura no sentido de averiguar o estado da arte por meio de trabalhos relacionados ou similares ao proposto. Foram selecionadas duas amplas bases integradoras de publicações periódicas, o Portal CAPES que, segundo seu site “...oferece acesso a textos completos disponíveis em mais de 45 mil publicações periódicas, internacionais e nacionais, e a diversas bases de dados que reúnem desde referências e resumos de trabalhos acadêmicos e científicos até normas técnicas, patentes, teses e dissertações dentre outros tipos de materiais, cobrindo todas as áreas do conhecimento. Inclui também uma seleção de importantes fontes de informação científica e tecnológica de acesso gratuito na web...” e o portal EBSCOHOST, que também funciona como um grande host agregador de periódicos, e possui vínculo com sistemas de referência clínica, o que amplia o aprofundamento da pesquisa na área específica da saúde. A estratégia de busca final está descrita na Tabela 1, abaixo.

Tabela 1 - Estratégia de Busca
PORTAL CAPES / EBSCOHOST

1	Qualquer campo contém “ <i>personal health data</i> ”
2	AND Qualquer campo contém “ <i>health data security</i> ”
3	AND Qualquer campo contém “ <i>health data privacy</i> ”
4	AND Título contém “ <i>clinical research</i> ”
5	2011 < DATE > 2022

6	Inglês e Português
7	Apenas artigos avaliados por pares

Como resultado foram encontrados um total de 180 artigos. Desta amostra foi efetuada a leitura dos títulos e dos resumos, a fim de garantir a associação da publicação com o tema da tese, resultando em 38 artigos selecionados. Destes, apenas 1 utiliza como embasamento de sua proposta para a garantia de privacidade das informações pessoais de saúde, a utilização da descentralização de dados por meio da perspectiva de dados sociais ligados (*social linked data*), o que contribui para reforçar a para a atualidade e a originalidade do tema, corroborando também com o ineditismo da abordagem.

Frente aos achados na literatura relacionados aos desafios atualmente encontrados na condução dos ensaios clínicos, buscou-se efetuar uma classificação, levando-se em conta características comuns entre os apontamentos, seja na sua origem ou na consequente implicação de sua existência. Os achados foram agrupados em 6 (seis) categorias conforme o Quadro 2 abaixo.

Quadro 2 - Categorização dos achados na bibliografia

Categoria	Características da categoria
Problemas de ordem financeira	Foram incluídas nessa categoria as citações na literatura relacionadas ao custo de execução e manutenção dos ensaios clínicos, custos relacionados a obtenção e atualização de consentimentos, seguros, ressarcimentos e custos de infraestrutura física, entre outros.
Problemas de ordem tecnológica	Foram incluídos nessa categoria as citações na literatura relacionadas a falta de utilização de registros eletrônicos de saúde (<i>EHR - electronic health record</i>), soluções de softwares proprietários e não intercambiáveis, falta de integração de sistemas, entre outros.
Problemas de ordem social ou políticos	Foram incluídos nessa categoria as citações na literatura relacionadas a concentração das

	pesquisas em poucos <i>players</i> de mercado, conflitos de interesse, vieses de publicação entre outros.
Problemas de privacidade	Foram incluídos nessa categoria as citações na literatura relacionadas a cruzamento de dados de diferentes sistemas, relatos de problemas de privacidade (nestes termos), venda de dados, repasse dos dados a terceiros, falhas de consentimento, entre outros.
Problemas de ordem científica	Foram incluídos nessa categoria as citações na literatura relacionadas a vieses de publicação, modelagens estatísticas não apropriadas, estudos não escaláveis, estudos não reproduzíveis, entre outros.
Problemas no campo do tratamento clínico	Foram incluídos nessa categoria as citações na literatura relacionadas ao descompasso das informações das pesquisas com o desfecho clínico, falta de colaboração entre os pesquisadores e a equipe médica responsável pelo paciente, falha na comunicação de efeitos adversos, entre outros.

Fonte: autor.

No Quadro 3, busca-se representar a existência de conceitos de diferentes categorias que transpassam vários trabalhos da literatura, apontando sua relevância enquanto temática científica e a contemporaneidade das abordagens discutidas nesta tese.

Quadro 3 - Relação das categorias e o aparecimento de suas características na literatura

	(MELLO; LIEOU; GOODMAN, 2018)	(EAGLESON et al., 2017)	(HUTCHINGS et al., 2020)	(HUTCHINGS et al., 2018)	(YAAKOV et al., 2021)	(GELINAS et al., 2021)	(OMAR et al., 2020)							Totais
Problemas de ordem financeira	x	x	x		x	x			x			x	x	8
Problemas de ordem tecnológica	x	x	x		x	x	x		x	x	x		x	10
Problemas de ordem social ou políticos	x	x	x		x	x			x	x	x			8
Problemas de privacidade	x	x		x	x	x	x	x	x	x		x	x	11
Problemas de ordem científica			x	x	x	x	x	x	x	x	x	x	x	11
Problemas no campo do tratamento clínico	x	x	x		x	x						x		6

Fonte: autor.

O quadro acima evidencia a atualidade e a relevância científica relacionadas aos objetivos deste trabalho. A crescente preocupação com os problemas correlacionados com o gerenciamento dos dados nos ensaios clínicos, justifica e demonstra o alinhamento da metodologia empregada neste trabalho com os anseios deste campo de pesquisa pois, trata não somente da explicitação e modelagem do problema abordado, mas também de uma proposta de intervenção neste contexto, interagindo através das experiências adquiridas para fornecer acréscimos ao conhecimento até então disponível (HEVNER E CHATTERJEE 2010).

2.2 PESQUISAS CLÍNICAS

2.2.1 CONTEXTUALIZANDO A PESQUISA CLÍNICA

Pesquisa Clínica é aquela que envolve direta ou indiretamente seres humanos, particularmente pacientes de instituições públicas ou privadas. As formas de realizar pesquisa clínica são variadas. Podemos referir como exemplo: pesquisas envolvendo pacientes; ensaios clínicos e pesquisa translacional (HULLEY et al., 2013).

Uma classificação prática é aquela que divide os estudos em dois grandes grupos: experimentais e não experimentais. Os estudos experimentais são sempre analíticos, longitudinais e prospectivos. Dos vários tipos de estudos experimentais, o de uso mais frequente, uma vez que proporciona evidências mais fortes, e atualmente considerado como o padrão ouro em investigação clínica é o ensaio clínico randomizado (ECR). Diferente dos estudos observacionais em que o pesquisador não interfere na exposição, nesse tipo de estudo o pesquisador planeja e intervém ativamente nos fatores que influenciam os indivíduos da amostra, observando os seus efeitos e desfechos (HULLEY et al., 2013; SHARMA; SRIVASTAV; SAMUEL, 2020).

A Resolução CNS 466/2012, Título II, item II.14, define a pesquisa envolvendo seres humanos como a “pesquisa que, individual ou coletivamente, tenha como participante o ser humano, em sua totalidade ou partes dele, e o envolva de forma direta ou indireta, incluindo o manejo de seus dados, informações ou materiais biológicos”(CONSELHO NACIONAL DE SAÚDE, 2012; SEEFELD, 2020).

No Documento das Américas, o ensaio clínico é definido como “estudo sistemático de medicamentos e/ou especialidades medicinais em voluntários humanos que seguem estritamente as diretrizes do método científico”, tendo como objetivo “descobrir ou confirmar os efeitos e/ou identificar as reações adversas ao produto investigado e/ou estudar a farmacocinética dos ingredientes ativos, de forma a determinar sua eficácia e segurança”(SEEFELD, 2020).

Na Tabela 1 são listados os documentos que norteiam os regramentos e a execução das pesquisas clínicas conforme (FEIJÓ et al., 2018).

Tabela 2 - Documentos norteadores das pesquisas clínicas

Documento	Contribuição
Código de Nüremberg - 1946	Cenário pós 2ª Guerra Mundial. O documento frisa a necessidade do consentimento voluntários e traz a importância da análise de riscos e benefícios de qualquer pesquisa a ser realizada com seres humanos.
Declaração de Helsinki - 1964	Partindo dos princípios estabelecidos no Código de Nüremberg, a Declaração de

	<p>Helsinki se tornou padrão internacional de conduta nessa área. Encontra-se nesse documento ênfase à prioridade do bem-estar humano sobre a ciência e a sociedade, assim como a ressalva sobre a importância do consentimento livre a ser dado pelos participantes de pesquisa. Possui as questões relacionadas aos medicamentos experimentais e sua relação com o uso de placebo. É importante salientar que os representantes do Brasil divergiram veementemente da maneira como o documento apresenta critérios tanto para acesso ao produto em análise após o estudo como para o uso de placebo (“Ministério da Saúde Conselho Nacional de Saúde Comissão Nacional de Ética em Pesquisa,” [s.d.]).</p>
<p>Relatório Belmont - 1978</p>	<p>Esse relatório veio em resposta a constatação da existência de abusos na condução de alguns estudos clínicos nos Estados Unidos. Salienta três princípios básicos que devem ser levados em conta quando se trata de pesquisa com seres humanos:</p> <p>1) respeito, que engloba tanto o respeito à autonomia da pessoa como o cuidado com aqueles que possuem sua autonomia diminuída; 2) beneficência, que defende não causar danos e minimizar riscos, maximizando benefícios; e 3) justiça, que defende tratamento igual para todos.</p>

Guia de boas práticas clínicas - 1995	<p>Esse conjunto de normas foi reconhecido e adotado pela Organização Mundial da Saúde, a partir de 1995, como recomendação formal aos países filiados. E, em 1996, a partir da Conferência Internacional de Harmonização (ICH) e editoração do “Manual tripartite harmonizado” da ICH 15, marco na evolução da pesquisa clínica, oficializou-se padrão único de condutas tanto para os Estados Unidos como para União Europeia e Japão. A adesão ao GCP é universalmente aceita como requisito fundamental para a investigação científica envolvendo seres humanos. Objetiva-se com as boas práticas clínicas zelar pela segurança e integridade das pessoas que vão se envolver na investigação como participantes de pesquisa, assim como buscar obter excelentes resultados que gerarão contribuição real aos indivíduos e à sociedade</p>
Documento das Américas - 2005	<p>Esse documento foi proposto por países do continente americano que não fizeram parte da ICH onde foi proposto o GCP, mas que sentiam necessidade de estabelecer princípios éticos e científicos universalmente aceitos em seus territórios, também a partir da Declaração de Helsinki. Teve como objetivo propor manuais de boas práticas clínicas que deveriam nortear tanto agências regulatórias como pesquisadores,</p>

	comitês de ética e demais órgãos ou indivíduos envolvidos em ensaios clínicos.
Resolução CNS 466/2012	Esse documento salienta o reconhecimento e a afirmação da dignidade, liberdade e autonomia do ser humano e fundamenta seus princípios na Constituição Federal do Brasil e em conhecidos documentos internacionais. São eles: Código de Nuremberg, Declaração Universal dos Direitos Humanos, Pacto Internacional sobre Direitos Econômicos, Sociais e Culturais, Pacto Internacional sobre os Direitos Civis e Políticos, Declaração Universal sobre o Genoma Humano, Declaração Internacional sobre Dados Genéticos Humanos e Declaração Universal sobre Bioética e Direitos Humanos.

Fonte: adaptado de (FEIJÓ *et al.*, 2018)

A execução de uma pesquisa clínica está baseada no cumprimento de diretrizes contidas no documento denominado de Protocolo de Pesquisa, definido pela Resolução CNS 466/2012, Título II, item II. 17, como um conjunto de documentos que contempla a descrição da pesquisa em seus aspectos fundamentais como também as informações relativas ao participante da pesquisa, a qualificação dos pesquisadores e de todas as instâncias responsáveis (BRASIL, 2012).

O protocolo de pesquisa ou desenho do estudo, é aprovado pelas autoridades regulatórias e éticas brasileiras. A Agência Nacional de Vigilância Sanitária e a Comissão Nacional de Ética em Pesquisa (CONEP) também analisam e aprovam o modelo de Termo de Consentimento Livre e Esclarecido (TCLE) (DALLARI, 2019, p. 204).

Importa também destacar o papel dos Comitês de Ética em Pesquisa (CEP), os quais, de acordo com a Resolução, devem ter formação interdisciplinar e independente de relevância pública, de caráter consultivo, deliberativo e educativo, criado para defender os interesses dos participantes da pesquisa em sua integridade e dignidade, e contribuem com o desenvolvimento

da pesquisa dentro de padrões éticos. Já a CONEP é uma instância colegiada, de natureza consultiva, deliberativa, normativa, educativa e independente, vinculada ao Conselho Nacional de Saúde/MS.

Ao falar sobre o papel dos Comitês de Ética em Pesquisa (CEP), destacamos que a atuação dos CEP, da CONEP e do CNS, educando e fiscalizando as pesquisas, garante a manutenção dos direitos humanos como uma prerrogativa de todos os membros da sociedade, e que esse sistema precisa de critérios de avaliação e capacidade operacional (BATISTA; ANDRADE; BEZERRA, 2012).

No que diz respeito à pesquisa envolvendo o estudo de um medicamento, vacina ou procedimento diagnóstico até a sua efetiva liberação para o uso assistencial, exigem-se inúmeros estudos que geram as informações necessárias para comprovar a sua segurança e eficácia, enquanto que às agências reguladoras cabem as responsabilidades pela elaboração de normas técnicas e pela avaliação das novas solicitações de uso específico, pela fiscalização e pelo acompanhamento do uso de novos medicamentos dentro de suas indicações (GOLDIM, 2007, p. 66).

Os estudos clínicos são classificados em duas etapas, uma, que é pré-clínica, realizada em modelos celulares e animais, e a outra, que é a clínica, realizada em seres humanos (GOLDIM, 2007, p. 67).

A pesquisa pré-clínica deve gerar informações que permitam justificar a realização de pesquisas em seres humanos. Os estudos experimentais devem especificar o método utilizado, os modelos celulares, teciduais ou animais utilizados, os testes laboratoriais, os dados de farmacocinética e toxicologia, de forma a permitir que os resultados pré-clínicos demonstrem a relevância dos achados e as possíveis aplicações terapêuticas (GOLDIM, 2007, p. 67).

No que se refere à etapa clínica das pesquisas de novas drogas, esta é realizada em seres humanos e é subdividida em quatro diferentes fases, denominadas de 1 a 4, de acordo com o nível crescente de conhecimento que se tem sobre os efeitos desta substância em modelos celulares, animais e em seres humanos, sendo que essas fases se diferenciam pelos objetivos específicos de cada tipo de estudo, pelo tipo de delineamento utilizado e principalmente pelo número e características dos participantes (GOLDIM, 2007, p. 68).

A pesquisa clínica de Fase 1 é o primeiro tipo de estudo em seres humanos com um novo princípio ativo ou com uma nova formulação de uma substância, cujo propósito é estabelecer uma evolução preliminar da segurança e da tolerabilidade. A Fase 2 da pesquisa avalia a eficácia terapêutica e a segurança de um novo produto farmacêutico como também

pode ser utilizada para avaliar novas drogas ou novas indicações para drogas já liberadas para outros usos. A Fase 3 busca avaliar se os efeitos terapêuticos demonstrados nos estudos de Fase 2 têm significância estatística e relevância clínica, para uma indicação e para um grupo específico de pacientes. Por fim, a Fase 4, é a etapa de realização de pesquisas depois que a droga experimental se torna um medicamento comercializável, tendo como finalidade a confirmação do valor terapêutico do novo medicamento em grandes grupos de pacientes que utilizam este produto em situação menos controlada que a estabelecida nos estudos de fase imediatamente anterior, a Fase 3 (GOLDIM, 2007, p. 70).

Os ensaios clínicos, antes de serem iniciados, devem ter o seu protocolo definido e aprovado pelas agências reguladoras, documento este que define o planejamento do estudo e seus objetivos além de uma metodologia e considerações estatísticas que possibilitem a avaliação e o acompanhamento dos testes. A elaboração do estudo envolve equipes multidisciplinares e experientes, podendo o desenho ser determinante para o sucesso ou o fracasso de uma nova molécula, por exemplo (GOMES et al., 2009).

Ainda observando os termos e definições da Resolução CNS 466/2012, destacamos que a realização da pesquisa clínica exige um patrocinador, público ou privado; um centro de pesquisa, legitimamente constituído e habilitado, no qual alguma das fases ou etapas da pesquisa se desenvolve; uma equipe de profissionais treinados, sob a responsabilidade do pesquisador responsável pela coordenação da pesquisa e corresponsável pela integridade e bem-estar dos participantes da pesquisa, que passamos a analisar.

2.2.2 O PROTOCOLO DOS ENSAIOS CLÍNICOS

Os ensaios clínicos são conduzidos de acordo com um protocolo (ou desenho do estudo) para sua execução. O protocolo demonstra as diretrizes para condução do estudo. Ele é o documento que declara explicitamente o raciocínio por trás do estudo e o que será feito explicando cada parte essencial dele e como esta parte deve ser realizada. Ele também descreve os detalhes relacionados a elegibilidade dos participantes, a duração do estudo, os medicamentos (se o caso) e todos os exames relacionados (CIPRIANI; BARBUI, 2010; O'BRIEN; WRIGHT, 2002).

O protocolo do ensaio clínico é o documento mais importante para garantir que os elementos críticos do estudo de pesquisa sejam comunicados aos investigadores que conduzem o estudo, o Comitê de Ética Independente (Institutional Review Boards – IRB, em inglês) e

agências reguladoras federais. O formato e o conteúdo dos protocolos de ensaios clínicos são padronizados e seguem as diretrizes de Boas Práticas Clínicas (BPC) (publicadas pela primeira vez em 1995, atualizadas pela última vez em 2016) de acordo com a Conferência Internacional sobre Harmonização de Requisitos Técnicos para Registro de Produtos Farmacêuticos para Uso Humano (ICH) (CHAN; HRÓBJARTSSON, 2018; MARTIN, 2020).

O protocolo é uma ferramenta fundamental de controle de qualidade para todos os aspectos de um ensaio clínico, sendo necessário por vários motivos:

- Garante a saúde e segurança de todos os participantes do estudo
- Ele fornece um plano de estudo preciso
- Ele define e gerencia o estudo e, portanto, deve ser seguido rigorosamente por todos os investigadores do estudo.
- Garante a integridade dos dados, permitindo a combinação e comparação de dados em todos os investigadores e/ou centros de estudo
- Informa os administradores do estudo, que frequentemente são uma organização de pesquisa contratada (CRO)
- É necessário obter aprovação ética de um Comitê de Ética em Pesquisa ou *Institutional Review Board* (IRB)

A ICH recomenda os tópicos que devem ser incluídos em um protocolo conforme resumido pelo autor na tabela 01. Informações específicas relacionadas as particularidades locais, e desenho do estudo previsto podem ser fornecidas em um contrato separado referenciado no protocolo (MARTIN, 2020).

Tabela 3 - Tópicos que devem constar em um protocolo de estudo clínico

Informações Gerais	Incluindo o título do protocolo; número de identificação do protocolo; data; alterações potenciais (mais informações sobre alterações na seção seguinte); nome e detalhes de contato do patrocinador, monitor, especialista médico, investigador e outras instituições envolvidas.
Informações Básicas	Incluindo nome e descrição do(s) produto(s) sob investigação; um resumo dos resultados de estudos não clínicos e outros estudos clínicos; riscos e benefícios conhecidos e potenciais; descrição da via de administração e dosagem; declaração garantindo a conformidade com protocolo, GCP e requisitos regulatórios; descrição da população; referências à literatura relevante.

Objetivos / Finalidade	Uma descrição detalhada dos objetivos e da finalidade do estudo.
Design do estudo	Incluindo informações sobre os desfechos primários e secundários, descrição do desenho do estudo, descrição das medidas para minimizar viés (incluindo randomização e cegamento), descrição do tratamento e regime de dosagem, duração esperada da participação do sujeito, descrição dos critérios de descontinuação, procedimentos de responsabilização, etc.
Seleção e exclusão de participantes	Explicação dos critérios de inclusão, exclusão e retirada, bem como os procedimentos de retirada.
Tratamento dos participantes	Informações sobre o tratamento (nome do produto, dose, esquema de dosagem, via, período de tratamento, acompanhamento), outros medicamentos/tratamentos permitidos ou não, procedimentos para monitorar a adesão.
Avaliação da eficácia	Especificação de parâmetros de eficácia, bem como métodos e tempo para sua avaliação, registro e análise.
Avaliação de segurança	Especificação de parâmetros de segurança; métodos e tempo para avaliação, registro e análise de parâmetros de segurança; notificação e acompanhamento após eventos adversos.
Procedimentos estatísticos	Descrição dos métodos estatísticos a serem empregados; tamanho da amostra; critérios de rescisão; procedimentos para contabilização de dados ausentes, não utilizados e espúrios; procedimentos para reportar qualquer desvio; seleção dos participantes a serem incluídos nas análises.
Acesso direto aos dados / documentos de origem	O patrocinador deve assegurar que está especificado no protocolo (ou em outro acordo por escrito) que o(s) investigador(es) / instituição(ões) permitirá(ão) monitoramento(s) relacionado(s) ao estudo, auditorias, revisão do IRB e inspeções regulatórias, fornecendo acesso a dados e documentos de origem.

Controle de qualidade e garantias	Descrição dos sistemas de garantia de qualidade e controle de qualidade implementados para garantir a qualidade dos dados (incluindo qualquer sessão de treinamento, monitoramento de investigadores, manuais de instruções, uso de procedimentos independentes de auditoria interna ou externa, etc.).
Ética	Descrição das considerações éticas relacionadas ao estudo, explicando o cumprimento dos padrões éticos aplicáveis.
Manipulação de dados e manutenção de registros	Esta seção garante que o(s) sistema(s) de processamento eletrônico de dados estejam em conformidade com os requisitos estabelecidos pelo patrocinador para integridade, precisão, confiabilidade e desempenho consistente pretendido. Deve mencionar se um comitê independente de monitoramento de dados foi estabelecido para avaliar o progresso do estudo.
Financiamento e seguro	Detalhes sobre o financiamento do estudo e o seguro do participante do estudo clínico que é obrigatório em muitos países, inclusive no Brasil.
Política de Publicação	Descrição das políticas de publicação de dados clínicos (por exemplo, manuscritos, resumos, pôsteres, apresentações orais e artigos de revisão publicados por periódicos científicos revisados por pares ou apresentados em reuniões e congressos científicos), declarando adesão às BPC e às diretrizes aplicáveis.
Cronograma/fluxograma do projeto	Descrição da linha do tempo, incluindo o início e o fim planejados do julgamento.
Referências	
Suplementos / Apêndices	

Fonte: próprio autor. Adaptado de (MARTIN, 2020)

2.2.3 DESVIOS, VIOLAÇÕES E ALTERAÇÕES DE PROTOCOLO

Desvios de protocolo são definidos como mudanças acidentais ou não intencionais no protocolo, ou não conformidade com o protocolo de pesquisa, que não aumentam o risco ou diminuem o benefício, ou não têm um efeito significativo sobre os direitos, segurança ou bem-estar do participante, e /ou na integridade dos dados. Os desvios de protocolo resultam da ação do sujeito ou equipe de pesquisa. Exemplos de desvios de protocolo incluem uma visita de estudo reagendada ou falha em coletar um questionário auxiliar de autorrelato. Um exemplo de desvio de protocolo é medir os sinais vitais antes de obter o consentimento informado (“Clinical trial documentation,” 2021; MARTIN, 2020)Clique ou toque aqui para inserir o texto..

Desvios de protocolo devem ser diferenciados de violações de protocolo, que são mudanças acidentais ou não intencionais no protocolo, ou não conformidade com o protocolo aprovado pelo IRB, sem aprovação prévia do patrocinador e do IRB. As violações geralmente aumentam o risco ou diminuem o benefício, afetam os direitos, a segurança ou o bem-estar do sujeito ou a integridade dos dados. Exemplos de violações de protocolo incluem o uso de um consentimento informado inadequado, a inscrição de sujeitos que não atendem aos critérios de inclusão e/ou exclusão, o uso de medicamentos proibidos etc.(BHATT, 2012; MARTIN, 2020).

Modificações de protocolo que potencialmente impactam o desenvolvimento do estudo, o benefício potencial do paciente ou podem afetar a segurança do paciente (incluindo mudanças nos objetivos do estudo, desenho do estudo, população de pacientes, critérios de elegibilidade, tamanhos de amostra, procedimentos do estudo ou aspectos administrativos significativos) requerem uma alteração formal do protocolo. Exemplos de alterações de protocolo que devem ser relatadas incluem qualquer aumento na dosagem do medicamento ou duração da exposição ao medicamento, qualquer aumento significativo no número de indivíduos a serem inscritos, novos procedimentos de teste ou a inclusão de um novo investigador em estudos multicêntricos (BHATT, 2012; MARTIN, 2020).

2.2.4 PROBLEMAS DE INTEGRIDADE NOS PROTOCOLOS

Erros evitáveis na condução do estudo, também conhecidos como violações de protocolo (VP), podem levar à introdução de vieses, erros e erros de design em um ensaio clínico que, em última análise, resulta em conclusões falhas. Por exemplo, a má conduta do estudo levando a erros na dosagem do tratamento do estudo pode causar danos aos pacientes

inscritos. Danos evitáveis decorrentes de cálculos de dose incorretos diluirão quaisquer benefícios reais do tratamento, reduzindo assim o poder estatístico geral, levando a um resultado falso negativo. Uma compreensão das taxas de VP pode, portanto, melhorar a interpretação dos resultados de um ensaio clínico (SWEETMAN; DOIG, 2011).

Não existe, no entanto, um acordo geral sobre o quais seriam os limites para níveis de VP aceitáveis e excessivos (SWEETMAN; DOIG, 2011).

Sweetman E. e Doig G (2011)(SWEETMAN; DOIG, 2011), classificaram as violações de protocolos (VPs) em 5 categorias:

1) Os VPs de inscrição, que ocorrem quando um membro da equipe de pesquisa não aplica adequadamente os critérios de elegibilidade do estudo, resultando na inscrição de um paciente inadequado no estudo.

2) Um VP de randomização, que foi definido como erro técnico ou humano que leva à violação da sequência de randomização pretendida ou qualquer tentativa de subverter o sigilo da alocação.

3) Um VP de intervenção do estudo, que foi definido como um erro de dosagem, tempo ou entrega na intervenção do estudo atribuível a membros da equipe de pesquisa. A equipe de pesquisa pode ser entendida como os membros do centro coordenador do estudo, pesquisadores do centro participante, coordenadores de pesquisa e membros da equipe de saúde que cuida dos participantes.

4) Um VP de adesão do paciente, que é caracterizado quando participantes do estudo não cumprem com o protocolo do estudo em relação a uma intervenção do estudo ou outros requisitos para a participação no estudo (por exemplo, pular consultas agendadas). A retirada formal do consentimento para participar do estudo não é considerada uma VP de adesão do paciente.

5) Os VPs de coleta de dados, que incluem erros nos quais a equipe de pesquisa não cumpriu as diretrizes pré-específicas do estudo para coleta de dados e/ou avaliação de resultados devido a razões evitáveis.

A integridade das pesquisas tem se tornado, especialmente após a declaração do período de pandemia, uma grande preocupação, e conferências sobre o tema são cada vez mais comuns. A World Conferences on Research Integrity Foundation (WCRIF) definiu integridade de pesquisa como "os princípios e normas que têm por objetivo garantir validade e confiabilidade da pesquisa" (INTEGRITY, 2022).

Para discutir a integridade da pesquisa, Steneck categoriza a conduta de pesquisa em três categorias:

- 1) fabricação, falsificação e plágio (FFP),
- 2) prática de pesquisa questionável (PPQ) e,
- 3) condução responsável da pesquisa.

O Escritório de Política Científica e Tecnológica dos EUA definiu FFP como má conduta de pesquisa em 2000, e muitos países adotaram a mesma definição (RESNIK; RASMUSSEN; KISSLING, 2015).

Uma meta-análise recente descobriu que de 2% a 14% dos cientistas localizados em países de alta renda podem ter fabricado ou falsificado dados. A fraude em ensaios clínicos é especialmente grave pois, além de envolver diretamente vidas humanas, ameaça a segurança e a eficácia das práticas clínicas para outros pacientes, além do que, muitos dos responsáveis pelas pesquisas clínicas são médicos, que deveriam defender os padrões éticos do cuidado ao paciente e da ciência. Apesar dessa estatística, o comportamento antiético da pesquisa não se limita aos FFPs. Os QRP são comportamentos considerados prejudiciais devido ao risco de consequências enganosas e também estão incluídos no comportamento de pesquisa antiético (NISHIMURA et al., 2021).

2.2.5 O TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO

O Termo de Consentimento Livre e Esclarecido (TCLE) é documento de caráter explicativo, onde são abordadas todas as questões relativas ao estudo clínico que possam estar relacionadas à decisão do sujeito da pesquisa e, assim, garantir sua participação voluntária. Ao pesquisador, juntamente com o coordenador do estudo, cabe o cuidado com o processo de consentimento livre e esclarecido, que deve estar registrado nos documentos-fonte e ser renovado a cada visita, pois o processo de anuência não pode se restringir a simples assinatura de termo de consentimento livre e esclarecido (TCLE) na primeira visita do participante ao centro (CEZAR, 2015; FEIJÓ et al., 2018; SEEFELD, 2020).

Ao participante de pesquisa é assegurado o direito ao esclarecimento sobre a natureza da pesquisa, seus objetivos, métodos, benefícios previstos, potenciais riscos e incômodo que esta possa lhe acarretar, na medida de sua compreensão e respeito com relação às suas singularidades. Qualquer alteração no protocolo do estudo deve ser encaminhada para

aprovação dos órgãos competentes e refletida no TCLE que deve ser novamente submetido ao participante para consentimento (CEZAR, 2015; FEIJÓ et al., 2018; SEEFELD, 2020).

A Constituição Federal, ao prever como norma fundamental a dignidade da pessoa, a proteção da integridade, da vida e da liberdade; o Estatuto da Criança e do Adolescente, ao assegurar que o direito ao respeito consiste na inviolabilidade da integridade física, psíquica e moral da criança e do adolescente; o Código Civil, ao dispor sobre as pessoas incapazes; a Lei Geral de Proteção de Dados, ao regulamentar o tratamento de dados pessoais com o objetivo de proteger os direitos fundamentais de liberdade e privacidade; e as normas de boas práticas reconhecem, nas situações de pesquisa, a acentuada vulnerabilidade do sujeito, condição esta que determina um comportamento de solidariedade, que impõe deveres de proteção (CEZAR, 2015; FEIJÓ et al., 2018; SEEFELD, 2020).

2.2.6 O DESAFIO DO CONSENTIMENTO NAS PESQUISAS CLÍNICAS

A exigência de que os pesquisadores obtenham o consentimento informado dos potenciais participantes antes de incluí-los na pesquisa é um princípio fundamental da pesquisa médica, consagrado na Declaração de Helsinki e nos instrumentos legais subsequentes. A exigência de consentimento é sustentada por princípios éticos de respeito pelas pessoas e autonomia individual. O consentimento é também a base para a proteção dos dados pessoais e leis de respeito a privacidade na maioria dos países (KAYE et al., 2015).

O formulário ou termo de consentimento tornou-se o principal meio de registrar o envolvimento individual na pesquisa e para determinar se *é necessário consentimento adicional ou se o consentimento existente abrange uma nova pesquisa ou mesmo uma nova abordagem no estudo em andamento*. Como tal, formaliza parte do contrato social implícito entre os participantes e os pesquisadores. Novas formas de pesquisa biomédica e a facilidade com que vários dados de diversas fontes podem agora ser coletados, armazenados, analisados e compartilhados em volumes maiores do que nunca antes fora viável, desafiam o significado do consentimento informado e questionam se os processos existentes de envolvimento com os participantes são ainda apropriados para essa nova era, com formas muito mais dinâmicas de se realizar pesquisas (KAYE et al., 2015).

Um exemplo bastante atual que se aplica ao contexto abordado, é o dos biobancos, onde vários pesquisadores e projetos de pesquisa fazem uso dos mesmos materiais, e torna-se

difícil, se não impossível, obter consentimento informado para todos os futuros usos de pesquisa no momento do recrutamento para o biobanco ou antes do início de tal pesquisa, conforme exigido nas formulações originais da Declaração de Helsinki (WORLD MEDICAL ASSOCIATION, 2013). Nestes casos, novos consentimentos tornam-se caros e demorados, principalmente devido a dificuldade em localizar pessoas, resultando em altas taxas de desistência. Como solução prática, frequentemente é solicitado aos participantes a concordância com um consentimento amplo, onde aceita-se que o material pode ser amplamente compartilhado e usado por muitos pesquisadores. No entanto, pode-se levantar uma série de razões, nas quais essa abordagem é inadequada para atender às demandas de consentimento informado de forma significativa. Por exemplo, ao contrário da pesquisa tradicional, não existe um único procedimento experimental sendo conduzido, com duração fixa, pontos finais ou resultados, no qual os indivíduos estão sendo convidados a participar. Em vez disso, os indivíduos estão sendo solicitados a consentir com o seu envolvimento em complexas linhas de pesquisas com possíveis variadas questões e métodos de pesquisa emergentes, onde os riscos potenciais à privacidade de informações de saúde pessoais e de identificação individual podem se tornar uma preocupação maior do que o risco de danos físicos. À medida que a natureza das pesquisas médicas evolui, o contrato social entre participantes e pesquisadores precisa evoluir com ela. Se a pesquisa é aberta e contínua, as tecnologias de informação atuais já oferecem a possibilidade de envolvimento dos participantes de forma semelhante para se estender ao longo do tempo. Os participantes não precisam mais ser considerados “sujeitos” humanos passivos, mas podem ser engajados ao longo do tempo e reconhecidos como participantes de pesquisa ativos, interessados, respeitados e valorizados (CAULFIELD; KAYET, 2009; KAYE et al., 2015; MESLIN; CHO, 2010).

2.3 FERRAMENTAS

Como já apresentado anteriormente, vivemos em um mundo que vem continuamente se digitalizando. Soma-se a isso, uma grande aceleração nesse processo vivida nos últimos dois anos, principalmente causada pelos novos desafios apresentados pela pandemia da COVID-19 e, até devido a isso, no campo das ações relacionadas a saúde, o impacto é ainda mais notório (AMANKWAH-AMOAHA et al., 2021).

A adoção de soluções mais tecnológicas para os serviços de saúde, trouxeram entre outros legados, a substituição dos prontuários, receitas e documentos em papel por registros de saúde eletrônicos (Electronic Medical Record – EMR), que permitem a coleta, extração, gerenciamento, compartilhamento e busca de informações. Isso abre uma série de novas possibilidades para a saúde, pois torna viável que várias instituições compartilhem e colaborem nos registros de saúde pessoais com maior facilidade, tanto para o tratamento dos pacientes quanto para a realização de pesquisas clínicas. É também evidente, no entanto, que quando os dados se tornam mais fáceis de serem criados e compartilhados, muitas outras possibilidades emergem, entre elas questões éticas, de privacidade e de segurança, aumentando a necessidade de segurança da informação (confidencialidade, integridade e disponibilidade) (FERREIRA et al., 2007).

O ano de 2018, foi decisivo para que se ampliasse as atenções relacionadas a proteção das informações disponíveis no ambiente digital e, em especial, nas redes sociais, pois foi quando se descobriu que os dados de 87 milhões de usuários do Facebook foram utilizados para traçar perfis de comportamento e influenciar na eleição americana e no plebiscito que acabou por separar o Reino Unido da União Europeia. Também no ano de 2018, o Facebook divulgou ter sofrido um ataque hacker que alcançou 50 milhões de usuários em todo o mundo. Em razão desse fato, vários perfis foram desconectados. Uma nova falha, ocorrida em dezembro, possibilitou a exposição das imagens postadas por 6,8 milhões de usuários, no mesmo ano, o jornal The New York Times expôs o fornecimento pelo Facebook, sem autorização, dos dados de seus usuários para outras empresas (AZAMBUJA; GRANVILLE; SARMENTO, 2019).

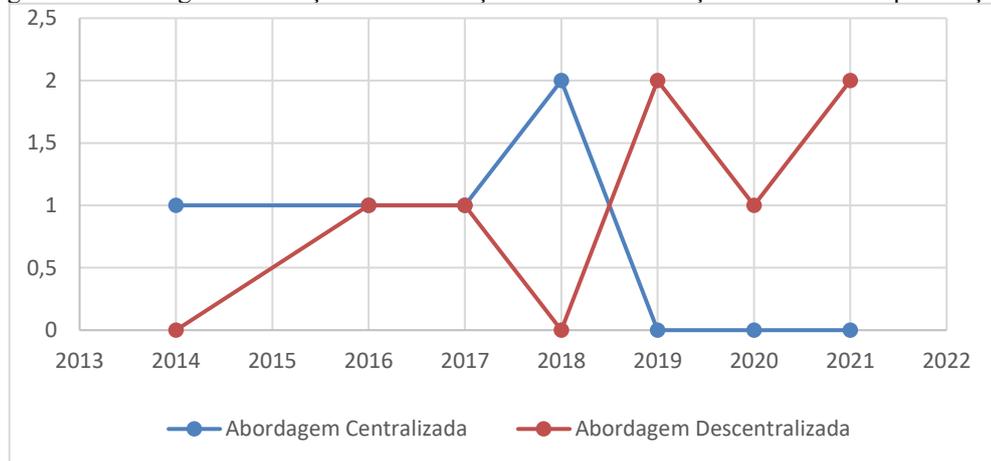
Quando o foco é a área da saúde, estima-se que entre 2011 e 2021, ocorreram por volta de 4200² violações de dados de saúde, que resultaram no roubo ou perda de mais de 240 milhões

² Considerando apenas casos envolvendo ao menos 500 registros (“Healthcare Data Breach Statistics,” 2022).

de registros de saúde apenas nos Estados Unidos. No Brasil, em apenas um episódio relatado em 2020, foram expostos dados de 243 milhões de cidadãos³.

No levantamento efetuado na literatura, dos 38 (trinta e oito) artigos selecionados, 18 (dezoito) continham propostas ligadas a frameworks, modelos, desenvolvimento ou utilização de soluções de software para atuar na mitigação dos problemas relacionados à confiabilidade, rastreabilidade e portabilidade dos dados de ensaios clínicos ou à meios de garantir a privacidade dos participantes. Destes, foi possível notar uma correlação positiva entre as publicações mais recentes e proposições relacionadas a adoção de soluções baseadas em descentralização de dados.

Figura 2 - Abordagem em relação a centralização ou descentralização dos dados nas publicações



Fonte: autor

A opção pela adoção de um modelo descentralizado de dados é também uma resposta aos desafios que a grande concentração de dados pessoais em silos de dados ou sistemas de big data oferece ao direito à privacidade, à liberdade de expressão e, em especial na área da saúde, a segurança das informações e suas finalidades de uso (AZAMBUJA; GRANVILLE; SARMENTO, 2019).

Soluções baseadas em descentralização de dados não são necessariamente uma novidade no ambiente acadêmico e até mesmo em soluções comerciais. Uma variedade de sistemas tem sido proposta com o objetivo de resolver este problema. Porém, na maioria das vezes, eles apresentam uma solução altamente técnica para tornar os dados de saúde

³ <https://g1.globo.com/economia/tecnologia/noticia/2020/12/02/nova-falha-do-ministerio-da-saude-expoe-dados-de-243-milhoes-de-brasileiros-na-internet-diz-jornal.ghtml>

amplamente disponíveis, mas ainda seguros, envolvendo o uso extensivo de blockchains, chaves criptográficas ou ambos (PAULOS, 2020).

Blockchains, são estruturas do tipo livro-razão público, descentralizado e distribuído em rede ponto a ponto que tem recebido uma grande atenção recentemente. Trata-se da aplicação de uma estrutura de blocos vinculados para verificar e armazenar dados com a aplicação de mecanismos de consenso confiável para sincronizar as alterações nos blocos, o que possibilita a criação de uma plataforma digital à prova de adulteração para armazenamento e compartilhamento de dados. Acredita-se que o blockchain pode ser utilizado em diversos sistemas interativos da Internet (por exemplo, Internet das Coisas, sistemas de cadeia de suprimentos, gerenciamento de identidade e assim por diante) (FENG et al., 2019).

Embora essas soluções possam abordar a questão da disponibilidade e segurança dos dados, elas não necessariamente tornam esses dados fáceis e simples de usar, pois pode ser difícil para pacientes e instituições navegar com eficiência em sistemas tão complexos. Ainda, essa abordagem apesar de garantir a inviolabilidade dos dados, possui desafios na questão de privacidade e sérias restrições em relação a anonimização quando esta se faz necessária, uma constante no caso das pesquisas clínicas (FENG et al., 2019; PAULOS, 2020).

Uma transação básica no padrão blockchain contém os endereços de seus participantes, valores transacionais, carimbo de data/hora e assinatura de seu remetente. Devido à natureza pública da rede blockchain, é possível rastrear o fluxo de transações para extrair as identidades físicas dos usuários ou outras informações adicionais por meio de mineração de dados (FENG et al., 2019)

Existem outras arquiteturas descentralizadas que se mostram de mais fácil compreensão e aplicação prática. SAMBRA et al. (2016) apresentaram o projeto Solid como uma plataforma de dados distribuídos, que permite aos usuários exibir a propriedade real de seus dados sem a necessidade de blockchain ou tecnologias relacionadas. Isso se dá por meio do uso de *pods* Solid, que são servidores individuais de armazenamento de dados pessoais, que os indivíduos podem hospedar e disponibilizá-los na web por conta própria ou delegar essa tarefa ao serviço de um terceiro de sua confiança, semelhante ao que acontece com a hospedagem de websites. O padrão Solid foi projetado para permitir que os usuários decidam exatamente quem mais deve ter acesso a esses dados em um nível granular, de tal modo que conceder o acesso aos dados a alguém não seja uma questão de tudo ou nada. O protocolo Solid vem sendo testado em um número crescente de aplicativos da Web que protegem a privacidade

dos usuários, aproveitando suas possibilidades de gestão da propriedade de dados (PAULOS, 2020; SAMBRA et al., 2016; “Solid,” 2022).

Nesta tese, apresento um framework para aplicações intensivas em conhecimento, utilizando bases descentralizadas para gerenciar, registrar e compartilhar dados, e tendo como prova de conceito, sua aplicação na resolução dos desafios encontrados nas pesquisas clínicas atuais. Para tal, os registros dos usuários são formatados usando um padrão aberto para registros de saúde, e é dada a eles capacidade de controlar o compartilhamento de seus dados com outras pessoas, como médicos e pesquisadores caso a caso.

2.4 LINKED DATA

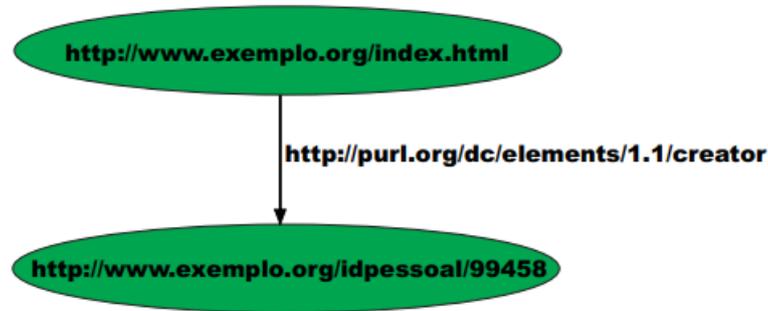
Linked Data é um conjunto de princípios que podem ser usados para transformar a web de um servidor de arquivos distribuído em um banco de dados distribuído. Atualmente, é comum que recursos na Web façam referência a outros recursos por seu URI. Esse esquema permite que a Web seja vista como um grafo de recursos interconectados, onde cada nó é um URI que representa um recurso recuperável. Este grafo não fornece muitas informações sobre a relação entre os recursos, apenas que eles existem ou não. Em contraste, os recursos *Linked Data* descrevem exatamente seus relacionamentos com outros recursos. Esses relacionamentos também são definidos por URIs. Ontologias definem os URIs que podem ser usados como relacionamentos entre recursos, bem como o significado desses relacionamentos. O *Resource Description Framework* - RDF é a linguagem utilizada para descrever esses conceitos e seus relacionamentos (“Data - W3C,” 2022).

2.4.1 RESOURCE DESCRIPTION FRAMEWORK (RDF)

O Resource Description Framework (RDF) é uma linguagem de propósito geral para representação de recursos na Web. Ele foi projetado para situações onde as informações necessitam ser processadas por aplicações, em lugar de somente serem visualizadas por pessoas. Os documentos RDF contêm um conjunto de declarações, onde cada declaração consiste em um sujeito, um predicado e um objeto. Essas declarações são chamadas de “triplas RDF”. A presença de uma tripla RDF é uma afirmação de que o sujeito tem uma relação de mão única com o objeto, e essa relação é descrita pelo predicado. Especificamente, a parte que identifica o objeto da declaração (uma página Web, por exemplo) é chamada de sujeito (recurso). A parte que identifica uma propriedade ou uma característica (o criador de uma página Web, por

exemplo) de um recurso é chamada de predicado (propriedade), e a parte que identifica o valor de uma propriedade é chamada de objeto (valor de propriedade).

Figura 3 - Declaração RDF simples.



Fonte: (LIMA; CARVALHO, 2005)

Um predicado deve ser um URI, enquanto assuntos e objetos podem ser um URI, um literal ou um nó em branco. Um literal é um valor definido, como um número, string ou data. Um nó em branco é um documento RDF incorporado que contém mais triplas que o descrevem. O poder do *Linked Data* está justamente no fato de que cada URI usado como assunto ou objeto representa um documento na Web que pode ser recuperado para expandir ainda mais o gráfico de informações criado pelo documento RDF (LIMA; CARVALHO, 2005).

A especificação RDF descreve o tipo de dados contido em um documento RDF, mas não o formato exato desses dados. Isso ocorre porque o RDF pode ser usado em muitos contextos diferentes, cada um dos quais pode ter uma maneira conveniente, mas diferente, de representar dados RDF. A representação RDF utilizada nesta tese será o formato Turtle, já que essa é a serialização RDF mais comum utilizada no Solid. No Turtle, as triplas são formadas por declarações que terminam com um ponto conforme o exemplo abaixo.

```

<#me> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type>
<http://xmlns.com/foaf/0.1/Person>.
<#me> <http://xmlns.com/foaf/0.1/name> "João da Silva".
<#me> <http://xmlns.com/foaf/0.1/image> <foto.jpeg>.
  
```

2.5 ONTOLOGIAS

A busca por interoperabilidade semântica em domínios complexos depende da capacidade de identificar e distinguir os elementos heterogêneos que compõem estes domínios, bem como as práticas de interoperabilidade utilizadas para garantir a recuperação da informação. O desenvolvimento de estratégias para a articulação destes elementos visa o acesso consensual à dados igualmente heterogêneos, gerados durante processos organizacionais, favorecendo o surgimento de inovações, sejam elas tecnológicas ou não. Uma alternativa viável à organização do conhecimento nestes domínios remete às ontologias. (EMYGDIO; ALMEIDA, 2019).

Modelos ontológicos são concebidos com significativo grau de formalidade, em função do uso de axiomas e lógica formal, sendo úteis para o refinamento de modelos de representação do conhecimento menos precisos. Sua conversão em artefatos computacionais, os caracterizam como complexos informacionais de alto nível, aptos portanto, a prover comunicação unívoca e autônoma entre humanos e máquinas, a despeito do domínio de conhecimento para o qual tenham sido projetados (EMYGDIO; ALMEIDA, 2019; SCHULZ; JANSEN, 2013).

Adotar-se-á neste trabalho a conceituação de ontologias conforme o arranjo trazido por BUSS (2015) em que uma ontologia compõe uma estrutura através da qual se torna possível a construção de uma base de conhecimento. A ontologia descreve um determinado domínio de conhecimento através do emprego de um conjunto de conceitos e termos, enquanto a base de conhecimento usa estes termos para descrever uma determinada realidade momentânea. Caso esta realidade se modifique, a base de conhecimento também será modificada; porém a ontologia permanecerá inalterada, enquanto o domínio seja o mesmo.

A concepção de estratégias baseadas no uso de ontologias para a articulação dos elementos heterogêneos presentes em ecossistemas de saúde, em específico no âmbito deste estudo, no domínio dos ensaios clínicos, viabiliza a conceituação consensual dos termos da área, mitigando obstáculos à construção de sistemas que façam uso desses dados, promovendo interoperabilidade (EMYGDIO; ALMEIDA, 2019).

Em termos práticos no modelo que se propõe, a ontologia é empregada como o vocabulário que especifica os predicados e as classes que os documentos RDF podem conter. Diferentes ontologias podem ser utilizadas ao mesmo tempo em documentos RDF.

2.5.1 FOAF

A ontologia FOAF (Friend of a Friend) define relações e classes de forma muito útil para aplicações de redes sociais descrevendo o mundo usando ideias inspiradas na web. Nas descrições do FOAF, existem apenas “tipos de coisas” que são as classes e links, chamados de propriedades. FOAF é, portanto, definido como um dicionário de termos, cada um dos quais é uma classe ou uma propriedade. As descrições FOAF são publicadas como documentos vinculados na Web usando a sintaxe RDF, por exemplo. O resultado do projeto FOAF é uma rede de documentos que descrevem uma rede de pessoas (e outras coisas). Cada documento FOAF é em si uma codificação de uma estrutura de rede descritiva. Embora esses documentos nem sempre concordem, eles têm a característica útil de poderem ser facilmente mesclados, permitindo que descrições parciais e descentralizadas sejam combinadas. Diferentes tipos de aplicações podem usar ou ignorar diferentes partes do FOAF. Projetos que estendem FOAF fornecem outros conjuntos de classes e propriedades, muitas das quais estão vinculadas às já definidas no FOAF (“FOAF Vocabulary Specification,” 2022).

2.5.1.1 *Fast Healthcare Interoperability Resources*

A especificação Health Level Seven Fast Healthcare Interoperability Resources (HL7 FHIR, ou apenas FHIR) é propagada como um padrão com alto alinhamento com a Web Semântica utilizada para representar, recuperar, compartilhar e trocar registros eletrônicos de saúde (BERNERS-LEE; HENDLER; LASSILA, 2001; MARTINEZ-COSTA; SCHULZ, 2017). O HL7 FHIR destina-se a facilitar a troca eletrônica de informações de saúde entre todos os agentes envolvidos nos serviços de saúde, incluindo as ações na área de pesquisa. Os recursos FHIR podem ser representados como um gráfico RDF serializado no formato Turtle. O formato RDF é definido para auxiliar o processo de ponte entre a troca de dados operacionais e os sistemas formais de processamento de conhecimento. O uso de RDF permite que dados FHIR sejam usados com aplicativos compatíveis com RDF para suportar inferência, semântica compartilhada em vários padrões e formatos de dados, integração de dados, validação de dados semânticos, aplicação de conformidade, consultas SPARQL entre outros usos (PAULOS, 2020).

2.6 SNOMED-CT

Integrar uma terminologia clínica controlada com a capacidade de especificar claramente a linguagem médica em um registro eletrônico de saúde é um desafio contínuo. Uma

terminologia padronizada suporta a indexação e o processamento eficientes de dados de pacientes e é um elemento essencial para a implementação de suporte à decisão clínica, recuperação e agregação de dados baseados em conhecimento. Isso é feito explorando relacionamentos semânticos predefinidos, tanto de natureza hierárquica quanto não hierárquica. Para colher os benefícios desse vocabulário clínico, várias terminologias personalizadas foram sendo criadas ao longo do tempo. A Nomenclatura Sistematizada de Termos Clínicos de Medicina (*Systematised Nomenclature of Medicine Clinical Terms* SNOMED CT) foi designada como a terminologia de referência clínica recomendada para uso em sistemas de informação clínica em todo o mundo e é relatada como sendo usada em mais de 50 países (LEE et al., 2013; WASSERMAN; WANG, 2003).

A *SNOMED International* é uma organização sem fins lucrativos que possui e mantém a SNOMED CT que, segundo a própria, é a terminologia clínica mais abrangente do mundo.

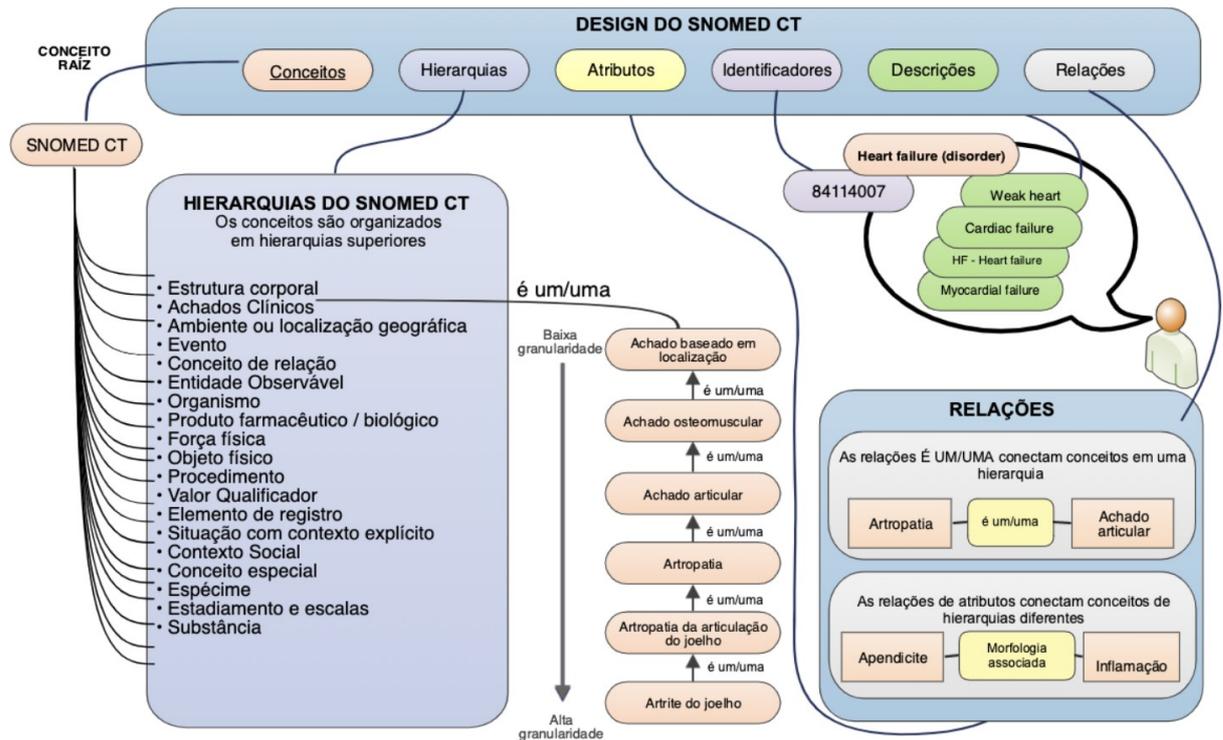
A SNOMED CT permite que informações de saúde sejam trocadas globalmente para o benefício de pacientes/cidadãos, prestadores de cuidados e outras partes interessadas. Com o SNOMED CT, os usuários podem registrar os dados do paciente com mais precisão, trocar dados do paciente tanto dentro da equipe de saúde quanto com os pacientes, localmente e além-fronteiras, para melhorar os resultados dos pacientes. Além disso, as partes interessadas podem usar o SNOMED CT em plataformas de análise e dados de saúde para análise clínica, análise populacional, análise de gerenciamento, pesquisa clínica, pesquisa aplicada e outras atividades de pesquisa para melhorar os cuidados de saúde (traduzido pelo autor) (SNOMED, 2022a).

As informações clínicas que empregam SNOMED CT beneficiam pacientes e profissionais de saúde individuais, assim como populações, e apoiam a prestação de serviços de saúde baseada em evidências. O uso de um Registro Eletrônico de Saúde (EHR, na sigla em inglês) melhora a comunicação e aumenta a disponibilidade de informações relevantes. Se as informações clínicas forem armazenadas de uma maneira que permita a recuperação baseada no significado, os benefícios aumentam significativamente. Os benefícios adicionais vão de oportunidades incrementadas para o apoio a decisões em tempo real a relatos retrospectivos mais precisos para pesquisa e gerenciamento (SNOMED, 2022b).

O SNOMED CT é uma terminologia clínica de escopo global que cobre uma ampla gama de especialidades, disciplinas e requerimentos clínicos. Como resultado do amplo escopo, um dos benefícios do SNOMED CT é a redução de efeitos limitantes entre especialidades que surgem do uso de diferentes terminologias ou sistemas de codificação por profissionais de saúde ou departamentos distintos. Isso permite um compartilhamento e reutilização de informações clínicas estruturadas mais amplos. Outro benefício do SNOMED CT é que os mesmos dados podem ser processados e apresentados de maneiras que servem

diferentes propósitos. Por exemplo, registros clínicos representados usando SNOMED CT podem ser processados e apresentados de maneiras diferentes para facilitar o atendimento direto aos pacientes, auditorias clínicas, pesquisa, epidemiologia, gestão e planejamento de serviços. Adicionalmente, o escopo global do SNOMED CT reduz efeitos geográficos de fronteiras que surgem do uso de terminologias ou sistemas de codificação diferentes em organizações ou

Figura 4 - Estrutura da SNOMED CT



países distintos (SNOMED, 2022b).

Fonte: (SNOMED, 2022b)

2.7 SOLID

Conforme a documentação do projeto, “os objetivos do projeto Solid estão alinhados com os da própria Web: empoderamento para uma sociedade equitativa, informada e interconectada. O Solid se soma aos padrões da Web existentes para criar um espaço onde os indivíduos possam manter sua autonomia, controlar seus dados e privacidade e escolher aplicativos e serviços para atender às suas necessidades”. Solid, acrônimo para *Social Linked Data*, pode ser resumido como um conjunto proposto de convenções e ferramentas para construir aplicações sociais descentralizadas. Solid baseia-se nos padrões e protocolos da Web semântica para oferecer aos usuários uma forma segura de armazenamento e compartilhamento de informações. No ecossistema Solid, um usuário é representado por um WebID, que é um

Uniform Resource Identifier (URI) que pode ser acessado usando o protocolo HTTP (Hypertext Transfer Protocol). O WebID de um usuário representa um documento recuperável contendo o perfil público desse usuário, que pode incluir seu nome, foto e outras informações públicas sobre ele. Este documento também indica a localização de seu Personal Online Datastore, ou pod, que é um servidor que armazena todas as informações desse usuário no ecossistema Solid. Por exemplo, meu WebID é <https://pod.inrupt.com/maico/profile/card#me> e o pod Solid criado para esta tese está localizado em <https://maico.inrupt.net/>. O projeto Solid define os protocolos que determinam como os servidores de pods Solid e os aplicativos clientes devem interagir (SAMBRA et al., 2016).

É possível que um usuário tenha mais de um pod. Um usuário pode escolher entre diferentes provedores de pod, pois os aplicativos Solid podem funcionar com qualquer servidor de pod, independentemente de sua localização ou provedor de serviços. Diferentes provedores de pod podem oferecer diferentes graus de privacidade, confiabilidade (por exemplo, garantias de disponibilidade ou latência) ou proteção legal (por exemplo, a legislação específica do país onde o pod está hospedado) (SAMBRA et al., 2016).

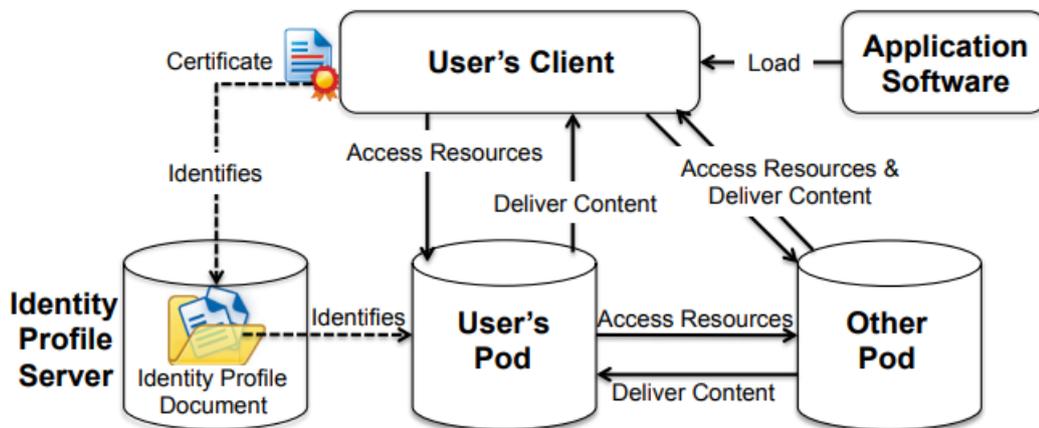
As aplicações Solid podem ser implementados como aplicativos WEB *client-side* ou aplicativos *mobile* que lêem e gravam dados diretamente dos pods. O protocolo Solid facilita o desenvolvimento e o uso de recursos sociais, pois os dados do aplicativo estão sempre acessíveis na Web, sob a governança de um mecanismo de controle de acesso. Os aplicativos podem agregar dados de diferentes fontes na Web acessando o pod do usuário que executa o aplicativo e os pods pertencentes a outros usuários. O Solid permite que vários aplicativos reutilizem os mesmos dados em um pod. Os usuários podem escolher entre diferentes aplicativos que oferecem funcionalidades semelhantes. Quando um usuário muda para um novo aplicativo, esse aplicativo pode acessar todos os dados existentes do usuário, pois os aplicativos são, por design, desacoplados dos dados que eles usam (SAMBRA et al., 2016).

Os provedores de pod também podem se beneficiar do Solid porque os usuários podem trocar de provedor facilmente, viabilizando uma menor barreira à entrada de novos provedores. Além disso, como os usuários têm um alto grau de controle sobre seus dados, é mais provável que eles armazenem dados com maior volume e variedade em seus pods. Um maior volume e variedade de dados por usuário é bom para provedores de pod. Por exemplo, um provedor de pod que é suportado por publicidade pode executar análises na ampla variedade de dados armazenados em cada pod (com o consentimento do usuário, é claro) para decidir o melhor anúncio a ser exibido a esse usuário (SAMBRA et al., 2016).

2.7.1A ARQUITETURA SOLID

A arquitetura Solid depende o máximo possível dos padrões W3C existentes conforme exibido na Figura 5. A arquitetura especifica todos os protocolos necessários como autenticação, comunicação do aplicativo para pod e de pod para pod.

Figura 5 - Arquitetura Solid



Fonte: (SAMBRA et al., 2016)

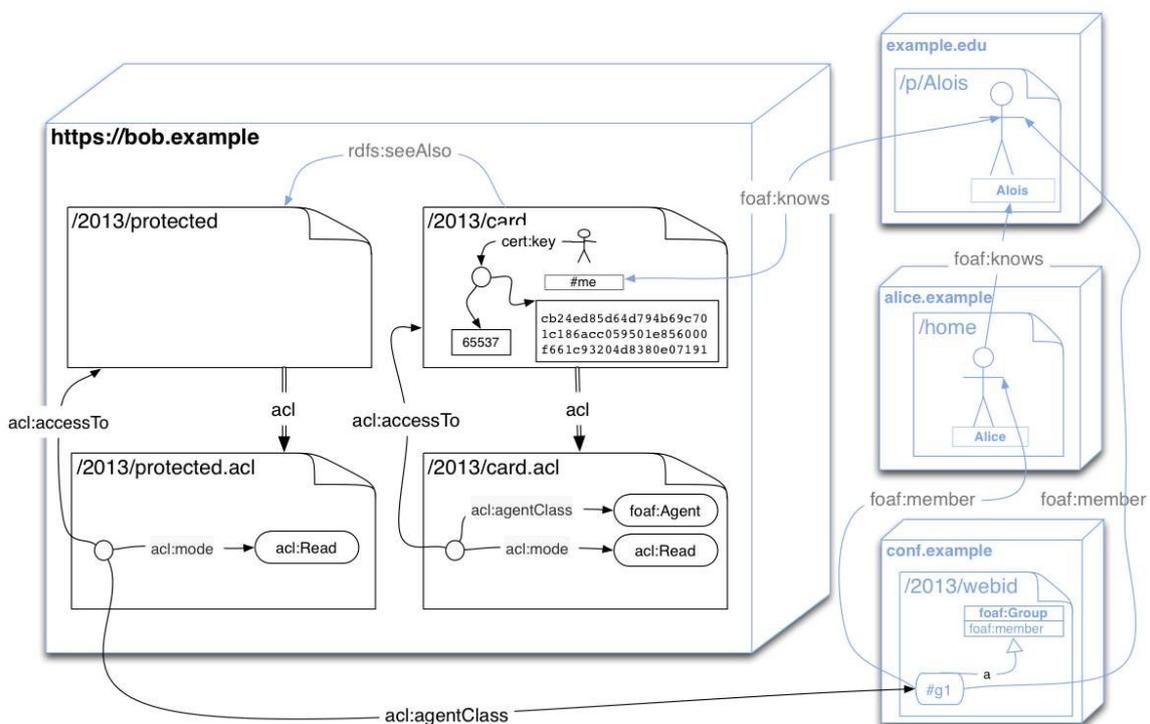
A Figura 5 ilustra a arquitetura da especificação Solid onde o usuário controla sua identidade usando um documento de perfil RDF, geralmente armazenado em um servidor pod. O usuário carrega um aplicativo Solid de um provedor de aplicativos. O aplicativo obtém o pod do usuário por meio de seu perfil de identidade. Em seguida, segue links do perfil para descobrir dados no pod do usuário, bem como em outros pods, realizando autenticação quando necessário.

2.7.1.1 Servidores pod Solid

A característica de descentralização do Solid está baseada na capacidade de que os usuários podem hospedar seu pod Solid em qualquer lugar que seja acessível na Web. Isso contrasta frontalmente com as plataformas existentes, como o Facebook, por exemplo, que armazenam todos os dados do usuário em um banco de dados centralizado, inviabilizando qualquer tipo de controle do usuário sobre como seus dados são gerenciados. Tendo como única opção para a utilização do serviço, a concordância com os termos de uso impostos. Independentemente de onde um pod está hospedado, esse pod deve se comunicar com as aplicações clientes de acordo com protocolos bem definidos (PAULOS, 2020; SAMBRA et al., 2016).

O pod do usuário atua como um repositório de dados para esse usuário e seus aplicativos Solid. Os dados em um pod do Solid são organizados como um sistema de arquivos, onde cada diretório e cada arquivo tem permissões que especificam quais usuários do Solid podem executar determinadas ações, como ler ou modificar um arquivo. Os pods Solid usam a especificação Web Access Control (WAC), que é um sistema de controle de acesso de entre domínios descentralizado que fornece uma maneira para sistemas de *Linked Data* definirem condições de autorização em recursos HTTP usando o modelo de lista de controle de acesso (*Access Control List - ACL*) (“Web Access Control,” 2022).

Figura 6 - Ilustração exemplificando o funcionamento do WAC



Fonte: WebAccessControl 2022 (<https://www.w3.org/wiki/WebAccessControl>)

A figura ilustra o funcionamento de um sistema que utiliza o WAC. Esse sistema é semelhante ao protocolo de controle de acesso usado em muitos sistemas de arquivos, exceto que os documentos controlados, os usuários e os grupos são todos identificados por URIs. Os usuários são identificados por WebIDs. Grupos de usuários são identificados pelo URI de uma classe de usuários que, se você pesquisar, retorna uma lista daqueles que a compõe. Isso significa que uma “pessoa” hospedada em qualquer site pode ser membro de um grupo hospedado em qualquer outro site.

É esse sistema descentralizado de controle de acesso entre domínios que permite ao proprietário de um pod ter o controle preciso sobre quem tem acesso a cada parte de seu pod. Esse uso de controle de acesso estrito é como a Solid protege os dados de um usuário e permite que os usuários controlem exatamente com quem escolhem compartilhar seus dados (PAULOS, 2020).

2.7.1.2 Aplicativos Clientes Solid

Os benefícios da segurança em armazenar dados em um pod Solid não teriam sentido sem uma maneira de gerar e aplicar esses dados. Isso é possível através do uso da especificação de autenticação WebID-OIDC criada pelo projeto Solid. Essa especificação é baseada nos protocolos OAuth 2.0 e OpenID Connect, que são padrões abertos para verificação de identidade na Web. O WebID-OIDC se difere porque foi projetado para operar em um ecossistema totalmente descentralizado como o Solid, onde existem muitos provedores de identificação e servidores de recursos diferentes que devem trabalhar juntos (PAULOS, 2020; SAMBRA et al., 2016).

Para autenticar um usuário, um aplicativo Solid Web oferece aos usuários a capacidade de fazer login usando o Solid, semelhante à forma como sites como Google e Facebook podem atuar como provedores de identidade OAuth para sites de terceiros. No entanto, como o Solid é um ecossistema distribuído, não há um único servidor Solid que possa autenticar todos os usuários. Em vez disso, os usuários também devem especificar o local do provedor de pod que poderá realizar a autenticação para eles. Depois que um usuário especifica e é autenticado por seu provedor de pod, o aplicativo recebe dados de sessão para o usuário atual. Esses dados incluem o WebID do usuário, bem como os tokens criptográficos usados pelo solid-auth-client para provar a identidade do usuário e do aplicativo ao acessar recursos em um pod Solid.

Quando o pod de um usuário é o único destino de armazenamento para um aplicativo, esse usuário pode controlar todos os dados que o aplicativo criou. Isso dá ao usuário a capacidade de exportar e importar facilmente dados de aplicativos a qualquer momento entre dois aplicativos compatíveis, evitando assim ecossistemas de *Walled Garden*, ou jardim murado na tradução literal, que é uma prática utilizada para manter o usuário dentro de determinada plataforma a fim de coletar seus dados. Estes tipos de práticas privam a liberdade de escolher aplicativos ou dispositivos preferidos. Por exemplo, considere a situação em que dois aplicativos de mensagens que utilizam o modelo Solid usam a mesma ontologia para armazenar dados. Como resultado, se um usuário decidir que não prefere mais seu aplicativo de mensagens

atual, ele poderá alternar para o outro sem perder nenhum contato, histórico de mensagens ou conversas existentes. Essa mudança nem seria perceptível para outros usuários em suas conversas (WebID-OIDC, 2022; Solid, 2022; WAC, 2022).

2.8 CONSIDERAÇÕES FINAIS DO CAPÍTULO

Este capítulo trouxe uma contextualização acerca da importância das pesquisas clínicas e, em especial, dos ensaios clínicos, e sua contribuição histórica para com as ciências da saúde. Ao se aprofundar através da literatura, nos métodos e materiais para a condução desses estudos, foi possível acompanhar a constante evolução e adequação destes às novas exigências e realidades de um mundo em constante transformação.

A digitalização das ferramentas e processos e a constante busca por respostas cada vez mais velozes trouxeram, em conjunto com suas vantagens, uma série de novos desafios, que perpassam diferentes níveis e estruturas envolvidas na condução de ensaios clínicos. O reconhecimento e classificação desses desafios, se mostra um passo importante para viabilizar o desenvolvimento de intervenções para dirimi-los.

A recente preocupação, apontada nas publicações científicas, com a busca de se incluir os participantes de ensaios clínicos como legítimos agentes ativos desses estudos, ao mesmo passo em que novas regulações pelo mundo exigem uma maior preocupação com os dados pessoais e, por consequência, com a privacidade dos cidadãos, reforça a atualidade e relevância da área de aplicação deste trabalho.

Embasando-se no que de mais moderno os trabalhos relacionados apresentam, foram evidenciadas técnicas para a descentralização dos dados e salientados os fundamentos que justificaram a escolha da especificação Solid, com a utilização de padrões abertos, para nortear a criação do framework, constructo desta tese. Também foram apresentados os conceitos sobre as tecnologias e padrões de modo a embasar cientificamente sua aplicação na criação do framework que passa a ser descrito no próximo capítulo.

3 DESENVOLVIMENTO DO FRAMEWORK

Este capítulo busca demonstrar o alinhamento da condução, plano de ação, recursos e estratégias utilizadas nos estágios que decorrem o trabalho com a metodologia selecionada, procurando confirmar a adesão dos resultados à questão de pesquisa do estudo e sua validade científica (SAMPIERI; COLLADO; LUCIO, 2013).

3.1 CONSIDERAÇÕES INICIAIS

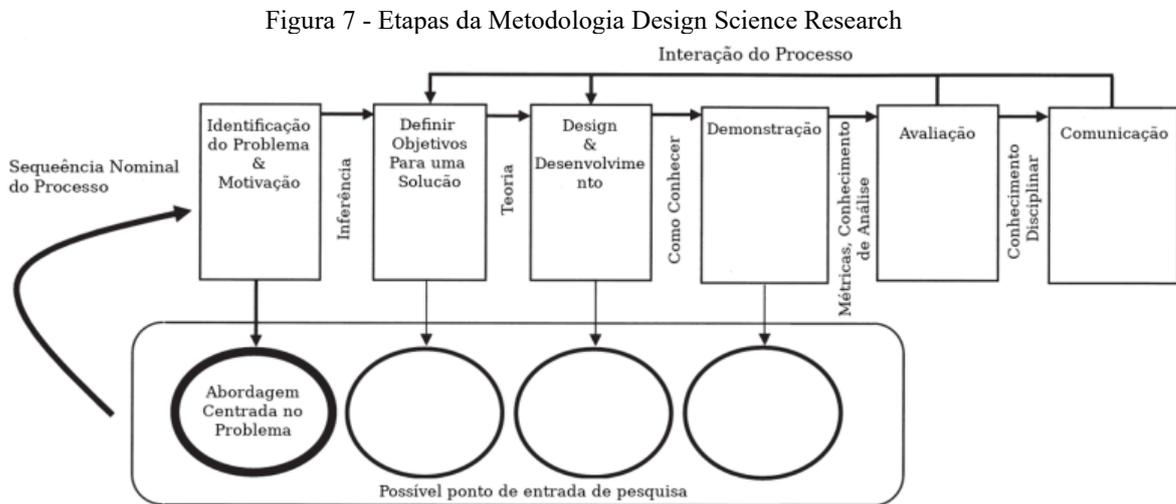
Conforme já explicitado, esta tese caracteriza-se por ser uma pesquisa tecnológica, que se utilizará de um conjunto de construtos, protocolos, e instrumentos teórico-práticos para originar um framework que colabore para a criação de conhecimento por meio da execução de ensaios clínicos com um maior grau de confiabilidade e integridade, ao mesmo tempo em que se amplia o nível de participação dos usuários ao dá-los domínio sobre seus dados. Este estudo apoia-se na abordagem metodológica Design Science Research (DSR). Uma característica da DSR é sua orientação à solução de problemas específicos, não necessariamente buscando uma solução final, mas sim uma solução satisfatória para a situação. Pode-se dizer que a DSR tem sob premissa que as soluções desenvolvidas sejam passíveis de generalização para uma determinada classe de problemas, permitindo que outros pesquisadores e profissionais, em diversas situações, também possam fazer uso do conhecimento/solução gerada (LACERDA et al., 2013; RIBEIRO JUNIOR, 2010).

“A Design Science Research tem sido apontada como uma abordagem de pesquisa adequada quando pesquisadores necessitam trabalhar de forma colaborativa com as organizações para testar novas ideias em contextos reais. Assim, pode ser usada como uma forma de produção de conhecimento para alcançar dois diferentes propósitos em projetos de pesquisa: produzir conhecimento científico e ajudar as organizações a resolver problemas reais” (LACERDA et al., 2013)

Nos termos de relevância, a DSR se interessa pela validade pragmática, ou seja, pela sua utilidade no campo prático. A validade pragmática busca assegurar que a solução proposta para resolver determinado problema de pesquisa de fato funcione, garantindo que os resultados esperados sejam alcançados (LACERDA et al., 2013).

Em relação a condução da Design Science Research nesta tese, a aplicação se dará conforme Peffers et al., (2007), chamada de Design Science Research Methodology (DSRM). O DSRM fornece um roteiro para apresentar e conduzir a DSR, clarificando assim o enquadramento metodológico das pesquisas conduzidas sobre esta metodologia. Esta abordagem é caracterizada por possuir seis etapas assim divididas: (1) identificação do

problema e motivação; (2) definição dos objetivos da solução; (3) design/projeto e desenvolvimento; (4) demonstração; (5) avaliação; e (6) comunicação. A Figura 7 ilustra as etapas da DSRM.



Cada etapa exposta acima apresenta objetivos e atividades distintas. Cada etapa e sua aplicação neste estudo são alvo dos subcapítulos 4.1 a 4.6.

A DSR, assim como a Pesquisa-Ação, está alinhada à Pragmática, um (meta)paradigma epistemológico em que se objetiva mudar a realidade, projetá-la, e não apenas tentar compreendê-la e explicá-la. Na Tabela 4, é apresentada uma comparação entre os (meta)paradigmas epistemológicos: Positivista, predominante nas Ciências Naturais; interpretativo, predominante nas Ciências Sociais; e Design (baseado em projeto), que fundamenta as Ciências Artificiais como Engenharia e Sistemas de Informação (JAQUES et al., 2020).

Tabela 4 - Comparação entre (meta)paradigmas de pesquisa

Crença Básica	Perspectivas de Pesquisa		
	Positivista (objetivo)	Interpretativo (subjetivo)	Design (projeto)
Ontologia	Uma única realidade. Conhecível, probabilística	Múltiplas realidades socialmente construídas	Múltiplas alternativas de estados do mundo contextualmente situadas, possibilitadas sociotecnicamente
Axiologia: qual é o valor (ética)	Verdade: universal e bonita; predição	Compreensão: situada e descrição	Controle; criação; progresso (melhoria, aperfeiçoamento); compreensão
Epistemologia	Objetiva, imparcial. Observador separado da verdade (neutro)	Subjetiva, i.e., valores e conhecimento emergem da interação pesquisador- participante.	<i>Conhecimento pelo fazer:</i> construção objetivamente restrita dentro de um contexto. A circunscrição iterativa revela significados.
Abordagens/ Métodos de Pesquisa	Experimento; <i>survey</i> Estudo de Caso	Etnografia	DSR; Pesquisa-Ação

Fonte: (JAQUES et al., 2020)

3.1.1 IDENTIFICAÇÃO DO PROBLEMA E MOTIVAÇÃO

Considerando como ponto de entrada da pesquisa a abordagem centrada no problema, a primeira atividade do processo de Peffers et al. (2007) é a definição do problema de pesquisa e sua motivação, sendo esta alcançada pelo pesquisador, por meio da justificativa do valor de uma solução. Assim, os recursos necessários para esta atividade são o conhecimento do estado do problema e a importância de solucioná-lo (PEFFERS et al., 2007).

Durante a experiência profissional do autor na área de gestão de pesquisas em saúde, pode-se observar alguns pontos onde as pesquisas clínicas, em especial àquelas definidas como ensaios clínicos, mesmo sendo atualmente o padrão ouro no apoio ao desenvolvimento de novas intervenções no campo da saúde, ainda podem evoluir. Essas observações até então empíricas, foram confirmadas por meio da literatura e apresentadas neste trabalho nos subcapítulos 1.2 IDENTIFICAÇÃO DO PROBLEMA DA PESQUISA e 1.5 JUSTIFICATIVA DA PROPOSTA como motivação desta pesquisa, e justificativa de sua elaboração.

3.1.2 DEFINIÇÃO DOS OBJETIVOS PARA UMA SOLUÇÃO

Na segunda etapa, espera-se do pesquisador a percepção dos objetivos da solução proposta a partir do problema definido na etapa anterior e do conhecimento do que é viável e exequível. Estes objetivos podem ser quantitativos, como os termos em que uma solução

desejável seria melhor do que as atuais, ou qualitativos, como uma descrição do propósito do artefato (PEFFERS et al., 2007).

De acordo com o levantamento efetuado na literatura, das categorias de problemas relacionadas a execução dos ensaios clínicos, àquelas mais citadas são as relacionadas a privacidade dos participantes e a confiabilidade científica dos dados utilizados nas pesquisas. Em atenção a isso nesta tese, apresenta-se um framework para aplicações intensivas em conhecimento, utilizando bases descentralizadas para gerenciar, registrar e compartilhar dados, e tendo como prova de conceito, sua aplicação na resolução dos desafios encontrados no desenvolvimento dos ensaios clínicos atuais. Para tal, os registros dos usuários são formatados usando um padrão aberto para registros de saúde, e é dada a eles capacidade de controlar o compartilhamento de seus dados com outras pessoas, como médicos e pesquisadores caso a caso.

Com essa arquitetura, o framework, artefato esta tese, serve como base científica e tecnológica de apoio à criação de conhecimento em ensaios clínicos, ampliando a confiabilidade destes por meio da rastreabilidade e proteção dos dados dos participantes aumentando também o grau de privacidade enquanto sujeito do ensaio.

3.1.3 DESIGN/PROJETO E DESENVOLVIMENTO

Na terceira etapa, é realizado o projeto do artefato, definindo-se as funcionalidades desejadas, sua arquitetura e efetuando o seu desenvolvimento. Para passar dos objetivos ao projeto e desenvolvimento, os recursos necessários incluem principalmente o conhecimento da teoria que pode ser utilizada na solução. Conceitualmente, um artefato de DSR pode ser qualquer objeto projetado para resolver um problema que possua uma contribuição de pesquisa incorporada ao desenho (PEFFERS et al., 2007).

Para esta tese, optou-se pela criação de um framework híbrido conforme a definição de Tomhave (2005). Para tanto, considerando os processos de negócio e a compreensão do ambiente dos ensaios clínicos, fez-se a exposição das diretrizes e requisitos que orientam o artefato para a solução do problema do qual se propõe.

3.2 LEVANTAMENTO DOS REQUISITOS

Algumas abordagens de elicitação de requisitos propõem utilizar os modelos de processos de negócios como base para os requisitos funcionais, porém, além das funcionalidades, faz-se necessário identificar também os requisitos não-funcionais e regras de negócio que refletem as reais necessidades presentes (VIEIRA et al., 2012). Os requisitos foram elaborados no sentido de oferecer respostas para os problemas identificados na etapa inicial da metodologia, de forma não ambígua, estão explicitamente declarados na tabela abaixo e serão discutidos nesta seção. É importante ressaltar que o levantamento de requisitos apresentado, está vinculado ao objetivo principal da proposta, que consiste na preservação da propriedade dos dados dos participantes de ensaios clínicos, por esta razão, questões também encontradas na literatura relacionadas a outros desafios no gerenciamento dos ensaios clínicos, não estão contemplados, conforme explicitado no item 1.8 delimitações da pesquisa.

Tabela 5 - Requisitos do Framework

REQUISITOS DO FRAMEWORK

REQ01	O artefato deve ampliar as possibilidades de privacidade dos participantes
REQ02	O artefato deve permitir o compartilhamento granular das informações dos participantes
REQ03	O artefato deve garantir o cumprimento do termo de consentimento
REQ04	O artefato deve utilizar padrões e tecnologias abertas
REQ05	O artefato deve permitir escalabilidade

Fonte: autor.

Requisito 01 - O artefato deve ampliar as possibilidades de privacidade dos participantes.

Os estudos da literatura realizados durante o projeto apontaram para a existência de potenciais riscos à privacidade dos voluntários participantes de ensaios clínicos. A centralização excessiva de dados pelos detentores/patrocinadores, a possibilidade de cruzamento, manipulação deliberada e o compartilhamento desses dados sem o efetivo controle ou consentimento dos reais proprietários, estão entre os principais itens encontrados na literatura que evidenciam a necessidade de se dar maior controle sobre a privacidade dos dados pessoais dos participantes.

Requisito 02 - O artefato deve permitir o compartilhamento granular das informações dos participantes.

Embasando-se nos achados da literatura e nas mais recentes legislações publicadas mundialmente acerca da proteção dos dados pessoais, como a Lei Geral de Proteção de Dados (LGPD)⁴ no Brasil e o Regulamento Geral sobre a Proteção de Dados (GPDR, do inglês *General Data Protection Regulation*)⁵, o consumidor/titular (termo das leis) ou, no caso específico deste estudo, o participante de estudos clínicos, deve possuir a possibilidade de escolher – de forma livre – quais operações pretende autorizar que sejam efetuadas com os seus dados, e também de compartilhar um cadastro completo ou somente determinadas informações. Em suma, a escolha não pode se resumir ao acesso total e ilimitado a todos os dados ou nada, no momento de solicitar a autorização do titular.

Requisito 03 - O artefato deve garantir o cumprimento do termo de consentimento.

Um dos principais apontamentos encontrados no levantamento bibliográfico em se tratando de falhas na condução de ensaios clínicos, é aquele relacionado a falhas na obtenção do consentimento e a quebra do protocolo do estudo. A alteração desses documentos/acordos durante a execução das pesquisas e a não obtenção de novo consentimento do participante feito a este acontecimento, devido a diferentes razões, é um fato relatado em diferentes publicações. Da mesma forma, o uso e compartilhamento dos dados para outros fins, sem o cuidado de se efetuar a comunicação a quem de direito deveria consentir para tal. Tal preocupação também encontra amparo nas legislações vigentes já citadas (LGPD e GPDR), que exigem o processo de consentimento granular compartilhado. O consentimento por parte do voluntário (titular dos dados) deve ocorrer de forma inequívoca, livre e informada e com o próprio sabendo que o uso dos dados ocorrerá para determinada finalidade. Para que esse consentimento ocorra em consonância com a lei brasileira, as partes consumidoras dos dados devem possibilitar ao titular formas de personalizar essa privacidade de dados, como *checkboxes* desmarcadas por exemplo, em que ele possa decidir clicar ou não. Cabe a parte consumidora dos dados (pesquisador ou patrocinador) o ônus da prova em relação a obtenção do consentimento. Devido a isso, o framework precisa atuar garantindo os limites acordados por meio do consentimento.

⁴ Lei n. 13.709, de 14 de agosto de 2018

⁵ Regulamento (UE) 2016/679 do Parlamento Europeu e do Conselho, de 27 de abril de 2016

Requisito 04 - O artefato deve utilizar padrões e tecnologias abertas.

Este requisito é decorrente aos achados categorizados entre os problemas de ordem tecnológica nesta pesquisa, que são, entre outros, a utilização de soluções de softwares proprietários, não intercambiáveis e a falta de integração de sistemas para ensaios clínicos. Padrões abertos, ou seja, aqueles que estão publicamente disponíveis e não são controlados por nenhum governo ou corporação, tornam possível que quaisquer empresas, cidadãos e países se conectem e troquem informações com autonomia. O uso de tecnologias de informação e comunicação de padrões abertos traz benefícios a todos, potencializando a interoperabilidade e a colaboração. A falta de interoperabilidade e de padronização de formatos se torna uma barreira de entrada para novas opções no mercado, causando um certo nível de aprisionamento dos usuários as plataformas já existentes. Por outro lado, padrões proprietários criam barreiras econômicas e dificultam a competitividade e a interação entre pessoas, empresas, governos.

Requisito 05 - O artefato deve permitir escalabilidade.

Este requisito está relacionado as categorias de problemas de ordem tecnológica e de ordem financeira na execução de ensaios clínicos. Os ensaios clínicos, principalmente aqueles de fase III e IV, necessitam de uma grande abrangência geográfica e numérica para alcançar resultados significantes de validação. Isso resulta em pesquisas de dimensões globais, que tratam dados em diferentes países, com diferentes níveis de exigência legal. A manutenção de grandes estudos, com avaliação e acompanhamento por longos períodos de uma quantidade cada vez maior de diferentes parâmetros dos participantes, exige dos sistemas que exercem ações sobre esses dados, a capacidade de acompanhar as exigências de confiabilidade e escalabilidade.

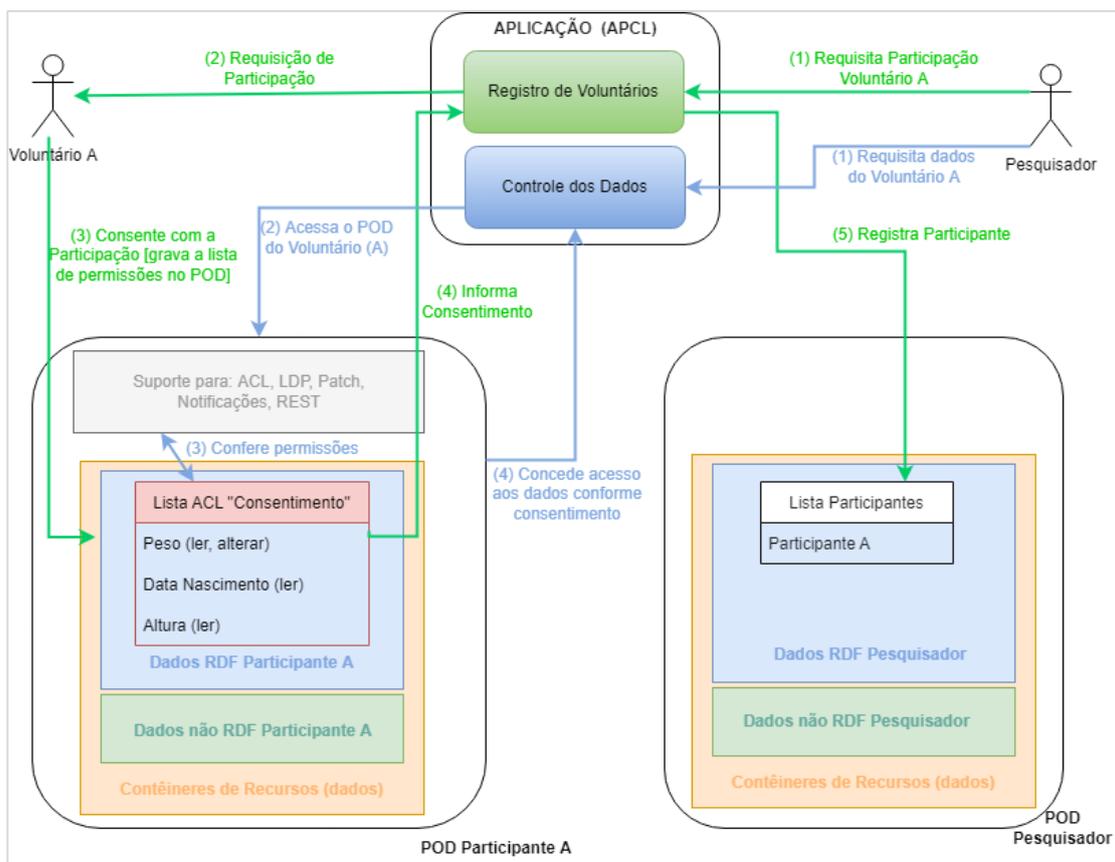
3.3 ARQUITETURA DO FRAMEWORK

Nesta seção é apresentada a arquitetura do framework desenvolvido e aspectos relacionados ao seu fluxo de atividades. Uma das premissas empregadas na estruturação do framework, é o total desacoplamento entre a lógica das aplicações e os dados por elas consumidos. Isso possibilita, por exemplo, a alternância de um software que esteja sendo utilizado para auxílio a gestão de um ensaio clínico por outro, sem que seja necessária nenhuma preocupação com exportação de dados ou tarefas similares. Por não possuir nenhum dado atrelado a aplicação e, por seguirem os mesmos padrões, inclusive de identificação e formato

de dados, é possível inclusive que diferentes pesquisadores que atuem na mesma pesquisa, utilizem aplicações distintas conforme sua preferência pessoal, um cenário até então inexecuível. Isso só é possível devido a adoção de tecnologias e padrões abertos.

Em relação a abordagem adotada para a proteção dos dados e consequente preservação da privacidade dos voluntários nos ensaios clínicos, embasando-se nos achados da literatura, uma perspectiva que vem sendo adotada nos trabalhos mais recentemente publicados, é a de se praticar a descentralização dos dados pessoais, a fim de permitir uma maior autonomia quanto ao seu uso e compartilhamento. A abordagem consiste em, ao invés de se modificar o paradigma cliente-servidor centralizado, buscar uma melhora na rede ponto a ponto de maneira tal a adicionar mais recursos de controle e desempenho do que o conceito tradicional de comunicação. Essa descentralização, mais do que prover o desacoplamento entre as soluções de software desenvolvidas dos dados que consomem, deve empoderar o proprietário desses dados, de forma que lhe seja conferida a opção de definir quais os usos passíveis e de que forma essas informações podem ou não ser compartilhadas.

Figura 8 – Ilustração do fluxo de atividades Framework



Fonte: Autor.

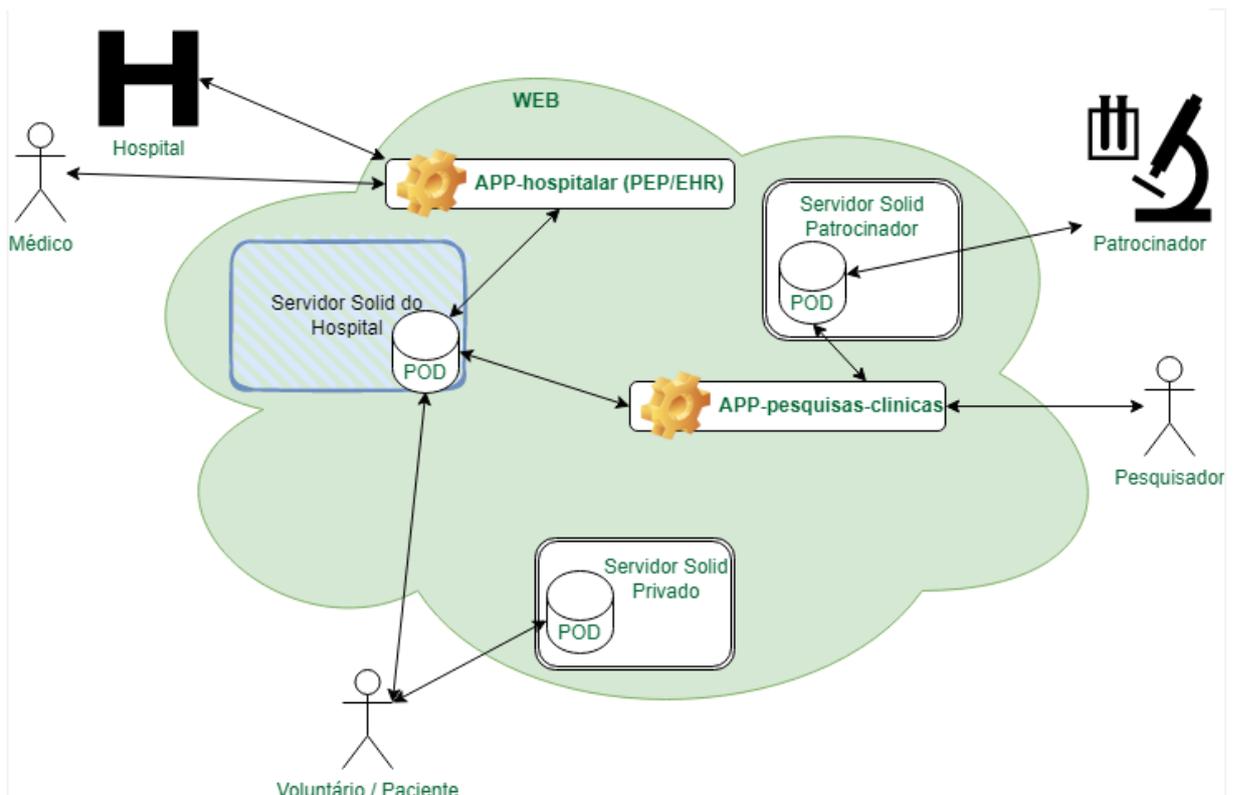
A Figura 8 apresenta um recorte esquemático de um fluxo de atividades proposto para a tarefa de recrutar um voluntário para a pesquisa (identificada pelas atividades na cor verde) e obter seus dados pessoais (atividades identificadas em azul). O processo se inicia com o pesquisador, por meio de uma aplicação web, requisitando a participação do Voluntário A (atividade 1). Para tanto o pesquisador deve conhecer o WebID desse voluntário A. Em um cenário real, essa informação pode ser obtida de várias formas: em contato direto com a pessoa (numa consulta médica por exemplo), por meio dos dados do prontuário em posse da instituição, em uma lista de contatos do próprio pesquisador etc. Com essa informação, a aplicação web, que passará a ser denominada como APCL (Aplicação Pesquisas CLínicas) para fins de simplificação, envia uma notificação para o Voluntário A (atividade 2). Seguindo a notificação o Voluntário A tem a possibilidade de consentir ou não com a participação na pesquisa (atividade 3). Para auxiliá-lo a tomar esta decisão, todos os dados requisitados pelo pesquisador, assim como o nível de acesso a esses dados (leitura, escrita, atualização, exclusão), são exibidos individualmente ou em uma granularidade maior definida pela aplicação. No caso de uma aprovação por parte do Voluntário A, um registro é efetuado no Solid Pod do Voluntário A, identificando todos os acessos concedidos ao aplicativo APCL e também um registro é efetuado no Solid Pod do Pesquisador, identificando o Voluntário A como integrante da pesquisa (atividade 5). Esse duplo registro é essencial pois faz o papel do Termo de Consentimento Livre e Esclarecido, onde as duas partes anuem as regras do compartilhamento de dados para a pesquisa em específico. É proposto que, na arquitetura do framework a aplicação não deve permanecer com nenhum dado, assim todos os registros, tanto os pessoais, quanto os relativos à execução da pesquisa, são armazenados nos Solid Pods aplicáveis. A Figura 9 apresenta uma representação que esclarece quem são os nós ou entidades, que podem compor um ensaio clínico.

Buscando também ilustrar como se dá uma rotina de atualização de dados relativos a um participante de pesquisa segundo o framework, pode-se observar na Figura 8 o fluxo representado pelas linhas azuis. A partir da solicitação do Pesquisador (atividade 1), a APCL busca os dados do Voluntário A em seu Solid Pod (atividade 2). A verificação de conformidade da solicitação frente ao consentimento é feita pelo próprio Solid Pod do Voluntário A, garantindo assim a conformidade com o que foi previamente acordado (atividade 3). Com a verificação efetuada, os acessos são liberados para a APCL.

Os ensaios clínicos não são estudos com um formato totalmente fechado. Apesar de não ocorrerem sem a existência do pesquisador principal e dos voluntários, uma série de outros agentes podem ou não se fazerem presentes. Essa variação ocorre devido aos desenhos dessas pesquisas conforme aprofundado no Capítulo 2. Estudos multicêntricos possuem vários pesquisadores principais, e várias instituições executoras, o que não ocorre em estudos mais simples. Por reconhecer cada nó ou entidade do framework da mesma forma, como sendo uma fonte de recursos/dados que compartilham do mesmo vocabulário ontológico, cada modelo instanciado pode assumir a arquitetura mais adequada para o cenário em que será utilizado. Em outras palavras, aplicações derivadas do conceito do framework, não precisam seguir uma determinada estrutura de classes ou módulos.

Hospitais, participantes(voluntários), médicos, pesquisadores locais, pesquisadores responsáveis (PIs), patrocinadores e entidades fiscalizadoras são vários nós do sistema. Todos os nós estão conectados aos seus respectivos Solid pods. Os dados são transferidos e acessados de acordo com os protocolos da especificação Solid.

Figura 9 - Visão ilustrativa do funcionamento do modelo



Fonte: autor.

O modelo contém as seguintes entidades para manter os dados seguros e disponíveis:

Participante: Um participante é um usuário central do sistema proposto, possuindo um WebID exclusivo. Os usuários do tipo participante podem decidir onde armazenar os PODs, optando por manter seu POD em uma estrutura própria fornecida pelo hospital ou pelo patrocinador ou alternativamente em outro serviço de armazenamento em nuvem.

Hospital: Um hospital é outro usuário crucial com um WebID exclusivo. Os hospitais ou centros de saúde podem gerar registros eletrônicos de saúde, informar e atualizar sobre o estado de saúde dos indivíduos.

Médico: Um usuário com um WebID exclusivo. O médico é responsável por criar e atualizar registros do paciente e pode ser um solicitante frequente de dados do paciente.

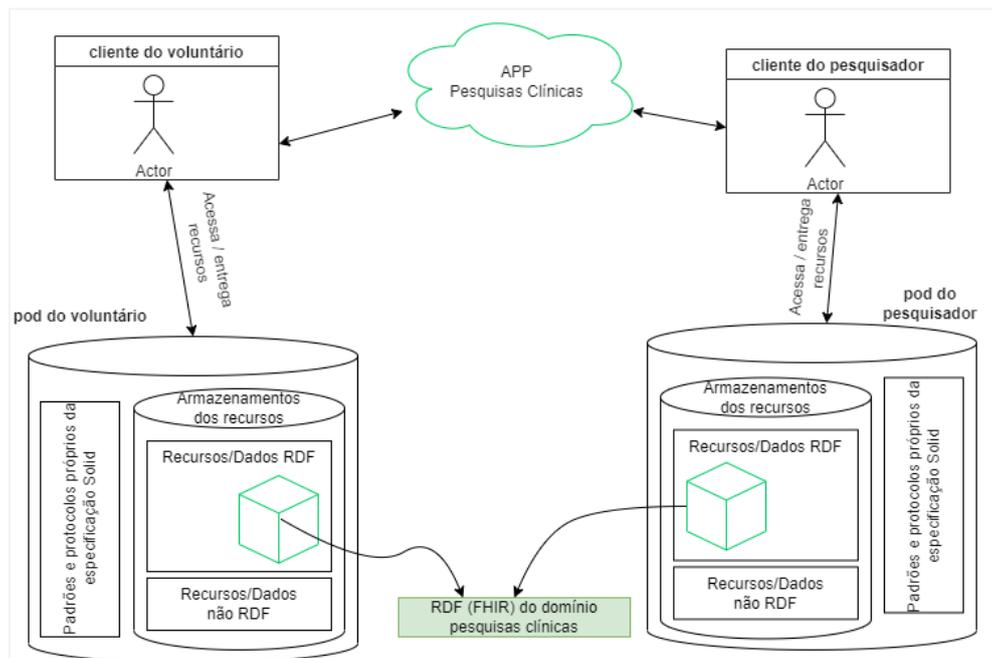
Pesquisador: Um usuário com um WebID exclusivo. O pesquisador é responsável por atualizar registros da pesquisa e pode ser um solicitante de dados do paciente.

Pesquisador Responsável ou PI (*principal investigator*): Um usuário com um WebID exclusivo. O PI é responsável por criar e atualizar registros da pesquisa e pode ser um solicitante de dados do paciente.

Patrocinador: Outro usuário do sistema que possui atributos de acesso aos dados da pesquisa.

Entidade Fiscalizadora: Autoridades com acesso aos dados das pesquisas, dos hospitais e dos pacientes em determinadas circunstâncias, variando conforme o estudo ou tipo entidade.

Figura 10 - Recorte ilustrando a estrutura dos pods utilizados no Framework



A Figura 10, ilustra por meio de uma simplificação, a composição da estrutura dos pods, responsáveis pelo gerenciamento dos dados e recursos vinculados aos seus proprietários. Os recursos referentes aos ensaios clínicos foram mapeados por meio de uma ontologia de domínio e são mantidos por meio de arquivos RDF nos pods.

Como já mencionado anteriormente, o conceito de framework adotado neste trabalho se preocupa em realizar a compreensão de um determinado fenômeno, e aplicá-lo a um contexto real, orientando inclusive, sua própria implementação. A implementação de forma harmônica dos conceitos da Web Semântica com base na especificação Solid é o núcleo do Framework proposto devolvendo aos usuários seus direitos de proprietários de seus dados por meio da descentralização destes. Essa abordagem foi construída de forma a ser considerada suficientemente genérica para ser aplicável a diferentes contextos, mas também completa a um ponto em que possa ser implementada em modelos (considerando modelo como uma instanciação de um framework) baseados em domínios de conhecimento específicos. Como já delimitado ao longo do texto, o domínio de conhecimento determinado para a aplicação do construto é, dentro dos ecossistemas de saúde, aquele concernente ao de ensaios clínicos.

Os ecossistemas de saúde constituem-se em um exemplo de domínios complexos e, assim como qualquer domínio dessa natureza, possuem duas dimensões básicas: a dimensão interna (microambiente) e a dimensão externa (macroambiente). No microambiente as comunidades de profissionais de saúde e demais profissionais técnicos realizam as práticas estabelecidas a partir de processos, visando uma convergência para atendimento aos propósitos das instituições de saúde (EMYGDIO; ALMEIDA, 2019).

No macroambiente encontram-se toda a gama composta pelos consumidores dos serviços de saúde, os fornecedores de produtos e serviços de apoio às práticas nas instituições dedicadas ao tratamento, pesquisa e desenvolvimento em saúde, os parceiros, os concorrentes e os órgãos reguladores destas instituições, além dos atores políticos, econômicos e científicos cujas atividades e descobertas provocam impactos nas práticas do microambiente. Tais impactos ocorrem por exemplo, através das modificações legais e tributárias, nos aspectos de adoção de normas internacionais de regulamentação ou inovações (EMYGDIO; ALMEIDA, 2019; ESLAMI ANDARGOLI et al., 2017; FREITAS; SCHULZ; MORAES, 2009).

A comunicação dentro do ecossistema é constituída tanto pela linguagem formal, legislação e contratos, quanto pelos jargões clínicos e padrões terminológicos adotados. A comunicação técnica, por sua vez é norteadada por linguagens as vezes artificiais e especializadas,

além de padrões técnicos, modelos de representação, acordos organizacionais e acordos sobre conteúdo. Como resultado da interação entre os atores durante as práticas de saúde, um grande volume de dados heterogêneos é gerado (EMYGDIO; ALMEIDA, 2019).

A busca por uniformização da comunicação no contexto da saúde é um campo de pesquisa antigo e consolidado, que já conta com diversas organizações internacionais provendo padrões para diferentes usos ou circunstâncias, podendo-se destacar algumas como a International Organization for Standardization (ISO)⁶, a Health Level Seven International (HL7)⁷, a openEHR Foundation⁸, entre outras. Em decorrência deste fato, são comuns sobreposições semânticas e sintáticas, às quais, associadas à natureza distinta de cada padrão, comprometem a interoperabilidade no contexto. Apenas segundo o projeto BioPortal, mantido pelo National Center for Biomedical Ontology (NCBO)⁹, até a atual data, existem 1002 ontologias diferentes cadastradas em sua base. Por essa razão, diversas ontologias acabam concorrendo para os mesmos fins (EMYGDIO; ALMEIDA, 2019).

Figura 11 - Estatísticas BioPortal

BioPortal Statistics	
Ontologies	1,002
Classes	14,555,494
Properties	36,286
Mappings	79,636,946

Fonte: <https://bioportal.bioontology.org/>

Devido à impossibilidade de um único padrão atender a todas as abordagens possíveis do ecossistema de saúde, abre-se uma demanda natural de estratificação dos domínios, criação

⁶ Disponível em: <https://www.iso.org/home.html>

⁷ Disponível em: <https://www.hl7.org/>

⁸ Disponível em: <https://www.openehr.org/>

⁹ Disponível em: <https://bioportal.bioontology.org/> (acessado em 06/2022)

e integração de padrões com os já existentes. As ontologias oferecem alternativas viáveis para soluções de interoperabilidade semântica em contextos diversos. Por incluírem relações hierárquicas, ou seja, taxonomias, classificações e vocabulários, sustentam numerosas aplicações como a busca semântica, integração de dados, mineração de textos, inferências e anotação semântica de recursos, elementos imprescindíveis às práticas de interoperabilidade esperadas (EMYGDIO; ALMEIDA, 2019). Para o recurso RDF representativo do domínio selecionado para a aplicação do framework, estendeu-se a ontologia FHIR adicionando classes e propriedades próprios dos ensaios clínicos, apoiando-se também no tesauro SNOMED CT para a definição dos termos a serem utilizados.

Para armazenar esse recurso RDF, com base na especificação Solid, o usuário cria um subdiretório em seu diretório público ou privado para compartilhá-lo com outras partes. O usuário é responsável por gerenciar o acesso a qualquer recurso no pod, garantindo assim total transparência em referência a níveis de privacidade no Framework.

A organização do nó do participante e o compartilhamento e acessibilidade de registros podem ser descritos como uma estrutura que inclui o id público do participante, a URL exclusiva que o representa (WebId), a URL do recurso (Linked Data) e a lista ACL que representa a lista de acesso criada pelo participante na qual muitos usuários podem ser definidos com seus respectivos limites de acesso.

3.4 CONSIDERAÇÕES FINAIS DO CAPÍTULO

Neste capítulo foi explicitada a concepção do framework, objetivo principal desta pesquisa. Seguindo a metodologia proposta deste trabalho, partindo-se da abordagem centrada no problema, buscou-se respaldo na literatura, tanto para a especificação dos problemas que motivaram a pesquisa quanto para a identificação dos requisitos aos quais o framework deve atender na etapa de avaliação.

Observando a etapa de projeto e desenvolvimento da DSRM, foi elaborada a arquitetura do construto, identificando os objetivos e as iterações entre os conceitos e ferramentas apresentados no capítulo anterior. No próximo capítulo, faz-se a demonstração da arquitetura do framework proposto por meio da instanciação de um modelo, que se dá através da implementação de um aplicativo web denominado APCL que instancia os seus conceitos simulando tarefas reais de um ensaio clínico. Segue-se desta forma, o método de avaliação em Design Science que é identificado por Lacerda et al. (2013) como forma de avaliação experimental.

4 AMBIENTE EXPERIMENTAL

Neste capítulo é exposto o ambiente experimental onde se realiza a prova de conceito. Espera-se nesta etapa, o uso do artefato resolvendo uma ou mais instâncias do problema por meio de um experimento, simulação, estudo de caso, prova ou outra atividade apropriada. Os recursos necessários para a demonstração incluem conhecimento efetivo de como usar o artefato para resolver o problema (PEFFERS, et. al, 2007).

A simulação, que dentro da DSRM consiste em executar o artefato com dados artificiais (LACERDA et al., 2013; PEFFERS et al., 2007) foi feita por meio da instanciação do framework através de um protótipo, que busca simular situações passíveis de ocorrer na execução de um ensaio clínico, confrontando situações alinhadas aos requisitos identificados na etapa de desenvolvimento do framework.

Uma instanciação é a concretização de um artefato em seu ambiente. Instanciações operacionalizam constructos, modelos ou métodos. Elas informam como implementar ou utilizar determinado artefato e seus possíveis resultados. Podem se referir a um determinado artefato ou à articulação de diversos artefatos para a produção de um resultado em um contexto. Instanciações demonstram a viabilidade e a eficácia dos modelos e métodos que elas contemplam. Instâncias fornecem artefatos funcionais, cujo estudo pode levar a avanços significativos tanto em Design Science quanto nas ciências naturais (MARCH; SMITH, 1995).

As seções a seguir apresentam o passo a passo da instanciação do framework. Além disso, os códigos-fonte e demais estruturas de dados estão disponíveis no repositório <https://github.com/omaicobuss/framework-tese>.

4.1 ONTOLOGIA DO DOMÍNIO ENSAIOS CLÍNICOS

A utilização de uma ontologia de domínio neste trabalho tem como principal objetivo prover um meio de representação e de facilitação da difusão das informações acerca do domínio de conhecimento representado, neste caso, os agentes envolvidos, termos, componentes e suas relações, envolvidos no processo de execução de ensaios clínicos.

Outras formas de representação como UML e diferentes tipos de fluxogramas, poderiam ter sido utilizados para tal fim, porém o poder de informação de uma ontologia, bem como a possibilidade de sua expansão e complementação, a conferem grau mais elevado de versatilidade e utilidade (BUSS, 2015).

Na área da saúde, o uso de ontologias tem destaque especial para auxiliar a comunicação de informações entre sistemas computacionais e no auxílio ao desenvolvimento de aplicações médicas, como sistemas de apoio a decisões, ou sistemas de telemedicina. Segundo (MANICA, 2009):

“As ontologias médicas são um recurso importante para o desenvolvimento da medicina baseada na evidência, pois além de incorporarem dados de saúde, introduzem especificações formais para representar relacionamentos estruturais entre os termos.”

O desenvolvimento de uma ontologia do domínio de ensaios clínicos, parte da necessidade de prover interoperabilidade entre os principais registros efetuados nesta categoria de estudos, servindo como um recurso estruturado que integra os termos e conceitos básicos no contexto de ensaios clínicos. Baseada em iniciativas já pré-existentes, a ontologia proposta é uma extensão que alinha e atualiza as nomenclaturas e conceitos dessas iniciativas anteriores, reestruturando sua organização de classes e heranças com aquelas adotadas pelo clinicaltrials.gov¹⁰ e a SNOMED-CT.

A SNOMED CT pode ser considerada como um thesouro ou dicionário. Ela traz consigo os significados dos conceitos médicos, mas, ao contrário de um dicionário, há apenas uma definição para cada termo. A SNOMED CT também mostra como os termos se relacionam entre si por meio de uma poli-hierarquia de relacionamentos "É UM" e atributos de conceito. No entanto, ela não informa como colocar esses conceitos em uma página (ou um registro médico) de uma forma que possa ser transmitida para outro sistema e compreendida. Para isso, requer um modelo de informação, um padrão de como os registros e mensagens devem ser dispostos e preenchidos. É para isso que se aplica HL7 FHIR. Ao combinar esses dois padrões de terminologia e modelo de informação, os sistemas que usam SNOMED CT e FHIR juntos serão capazes de comunicar um significado claro e inequívoco de uma maneira padrão que pode ser compreendida automaticamente em todo o mundo.

Este alinhamento, permite tanto sua utilização como base para a estruturação dos arquivos RDF utilizados como recursos armazenados nos Solid pods da aplicação proposta, quanto também para gerar novas versões estendidas desta ontologia, podendo ser reaproveitada em outras aplicações específicas já que permanecerá disponibilizada com licença *Creative Commons 4.0*, permitindo sua utilização, compartilhamento e adaptação.

¹⁰ Um banco de dados de estudos clínicos com financiamento público e privado realizados em todo o mundo. Disponível em: <https://www.clinicaltrials.gov/>

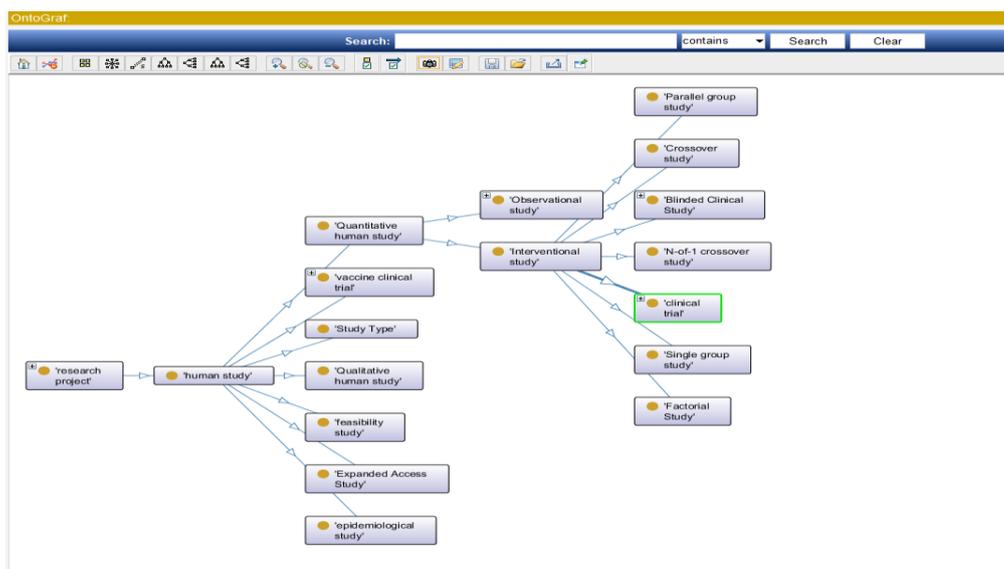
A ferramenta edição da ontologia se deu com a utilização do Protégé, software livre, gratuito e desenvolvido pelo Stanford Center for Biomedical Informatics Research na Stanford University School of Medicine. A Figura 12 exibe uma tela do software na qual se visualiza uma parte da lista de classes da ontologia, com destaque para o ordenamento das classes até a definição de *clinical trial*, que corresponde a tradução de ensaio clínico.

Figura 12 - Tela do Protégé exibindo classes da ontologia

The screenshot displays the Protégé interface with the 'Classes' tab selected. The left pane shows a class hierarchy starting from 'owl:Thing' and 'entity'. The 'planned process' class is expanded, showing subclasses like 'informed consent process', 'Intervention Type', 'medical intervention', 'research project', and 'human study'. The 'clinical trial' class is highlighted in blue. The right pane shows the 'Annotations' tab for 'clinical trial', listing annotations such as 'label', 'textual definition', 'example of usage', and 'has curation status'. The 'Description' tab shows the equivalent to 'Interventional study'.

Fonte: autor.

Figura 13 - Exemplo de visualização dos relacionamentos entre classes e



Fonte: autor.

Conforme apontado no Capítulo 3, de modo a prover o mais alto nível de interoperabilidade com outros aplicativos e dados de saúde, o framework serializa os dados dos ensaios clínicos como objetos que aderem ao padrão FHIR. Especificamente, cada informação coletada é transformada em uma instância da classe RDF correspondente na ontologia, conforme exigido pela especificação FHIR. Como resultado, outros aplicativos que utilizam as mesmas especificações FHIR poderão ler e operar com os dados gerenciados pelas aplicações desenvolvidas com base neste trabalho. Isso retira qualquer tipo de restrições sobre como o usuário pode compartilhar ou consumir seus dados pessoais de saúde fora do aplicativo para ensaios clínicos.

A especificação FHIR visa contemplar um conjunto razoável de cenários, mas não é possível abarcar todos os usos específicos existentes. Felizmente, o FHIR permite "personalizações" por meio de perfis (profiles). A definição de perfis é um esforço de modelagem de informação em saúde. Desenvolvedores precisam ter acesso aos perfis que personalizam os recursos a serem trocados. Afinal, toda a troca de dados deverá estar em conformidade com os perfis definidos.

Figura 14 - Recorte de um arquivo RDF representando um recurso do tipo Paciente

```
@prefix fhir: <http://hl7.org/fhir/> .

[ a fhir:Patient;
  fhir:nodeRole fhir:treeRoot; # if this is the parser root

  # from Resource: .id, .meta, .implicitRules, and .language
  # from DomainResource: .text, .contained, .extension, and .modifierExtension
  fhir:Patient.identifier [ Identifier ], ... ; # 0..* An identifier for this patient
  fhir:Patient.active [ boolean ]; # 0..1 Whether this patient's record is in active use
  fhir:Patient.name [ HumanName ], ... ; # 0..* A name associated with the patient
  fhir:Patient.telecom [ ContactPoint ], ... ; # 0..* A contact detail for the individual
  fhir:Patient.gender [ code ]; # 0..1 male | female | other | unknown
  fhir:Patient.birthDate [ date ]; # 0..1 The date of birth for the individual
  # Patient.deceased[x] : 0..1 Indicates if the individual is deceased or not. One of these 2
    fhir:Patient.deceasedBoolean [ boolean ]
    fhir:Patient.deceasedDateTime [ dateTime ]
  fhir:Patient.address [ Address ], ... ; # 0..* An address for the individual
  fhir:Patient.maritalStatus [ CodeableConcept ]; # 0..1 Marital (civil) status of a patient
  # Patient.multipleBirth[x] : 0..1 Whether patient is part of a multiple birth. One of these 2
    fhir:Patient.multipleBirthBoolean [ boolean ]
    fhir:Patient.multipleBirthInteger [ integer ]
  fhir:Patient.photo [ Attachment ], ... ; # 0..* Image of the patient
  fhir:Patient.contact [ # 0..* A contact party (e.g. guardian, partner, friend) for the patient
    fhir:Patient.contact.relationship [ CodeableConcept ], ... ; # 0..* The kind of relationship
    fhir:Patient.contact.name [ HumanName ]; # 0..1 A name associated with the contact person
    fhir:Patient.contact.telecom [ ContactPoint ], ... ; # 0..* A contact detail for the person
    fhir:Patient.contact.address [ Address ]; # 0..1 Address for the contact person
    fhir:Patient.contact.gender [ code ]; # 0..1 male | female | other | unknown
```

Fonte: HL7 FHIR v4.3(2022)

A Figura 14 apresenta alguns dos metadados que compõe o recurso do tipo *Patient* (paciente em português). Atualmente a FHIR possui 140¹¹ tipos de recursos que podem ser representados em XML, JSON ou RDF.

4.2 INSTANCIACÃO DO MODELO

Segundo apontam LACERDA et al. (2013), a escolha do método de avaliação dentro da DSRM depende do artefato desenvolvido e, por consequência, o método de avaliação deve estar alinhado diretamente ao artefato em si e a sua aplicabilidade. Entre os métodos aplicáveis para a avaliação do framework desenvolvido neste trabalho, HEVNER et al. (2004) propuseram a possibilidade de execução de um experimento controlado, que consiste em estudar o artefato em um ambiente controlado e a simulação com a utilização de dados artificiais. Para tanto, foi criado um protótipo, na forma de uma aplicação web, visando apresentar os passos para a instanciação de um modelo que siga o padrão do framework e avaliá-lo frente as propostas para o qual foi planejado. Neste capítulo se apresenta os passos executados para a implementação do protótipo e no item 4.3 são apresentados os resultados da avaliação em relação aos requisitos.

Antes do efetivo desenvolvimento do protótipo, o domínio de conhecimento “ensaio clínico” foi modelado através da extensão de uma ontologia, conforme descrito no item 4.2. Este passo foi importante pois é por meio da compreensão do domínio e, por conseguinte, da consolidação e alinhamento de seus conceitos com os padrões internacionais, que se faz possível criar soluções que propiciem a interoperabilidade.

A aplicação web APCL, foi desenvolvida com a utilização da linguagem de programação JavaScript, para tal, foram utilizadas bibliotecas de códigos fonte para criação de aplicações clientes Solid fornecidas e mantidas pela Inrupt Inc¹². Essa biblioteca (*@inrupt/solid-client*) faz parte de uma família de bibliotecas JavaScript de código aberto projetadas para oferecer suporte a desenvolvedores na criação de aplicativos Solid. Ela fornece uma camada de abstração sobre os princípios Solid e Resource Description Framework (RDF) e é compatível com a especificação RDF/JS.

Devido a característica da arquitetura do framework de total desacoplamento entre a parte lógica e os dados, antes que se possa executar a APCL, foi necessário criar Solid Pods

¹¹ Disponível em: <https://www.hl7.org/fhir/resourcelist.html>

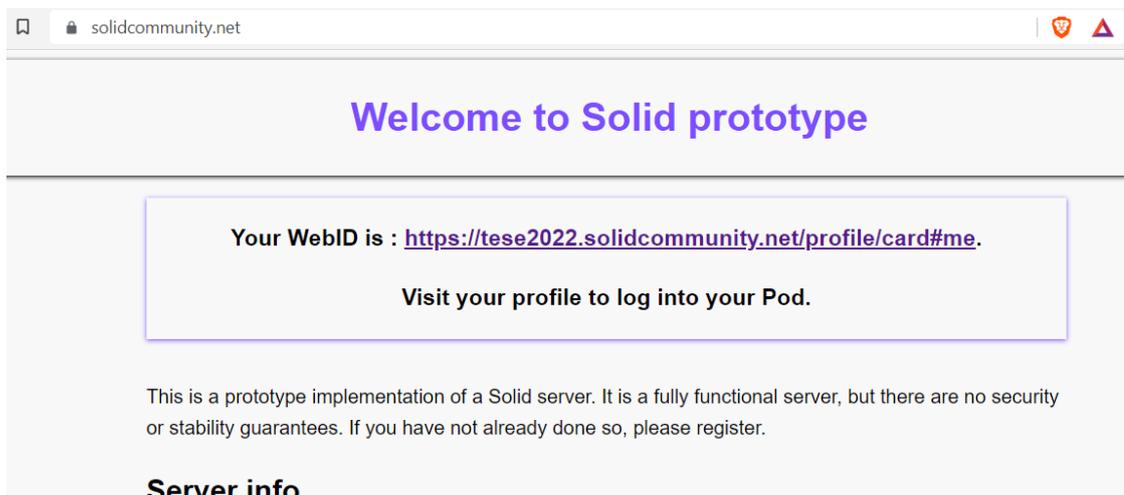
¹² Inrupt Inc, é uma empresa que fornece soluções baseadas na especificação Solid. Disponível em: <https://inrupt.com/about>

para atuarem como os nós de armazenamento. Para viabilizar a avaliação do modelo, foram criados 5 Solid Pods distintos, sendo 4 deles nomeados de v1apcl a v4apcl para atuarem como os nós de armazenamento de voluntários de um ensaio clínico e 1 Solid Pod nomeado tese2022 para atuar como sendo aquele associado ao pesquisador. Para facilitar a leitura, adotar-se-á como identificação dos Solid Pods a o termo Pod seguido de seu nome no servidor, por exemplo “Pod tese2022”.

A opção efetuada para a avaliação foi de se utilizar um provedor público de Solid Pods, permitindo assim que a comunicação entre a APCL e os Solid Pods não ocorresse em um ambiente totalmente controlado, porém, também não haveria impedimento em se criar um servidor próprio para a hospedagem dos Pods.

O Serviço utilizado foi o www.solidcommunity.net que oferece não só a estrutura para hospedar os recursos, mas também um navegador interativo que permite ao usuário visualizar a estrutura dos desses recursos. A Figura 15 apresenta a WebID correspondente ao Solid Pod criado para a utilização pelo papel do pesquisador (tese2022) na APCL.

Figura 15 - Criação do Solid Pod do Pesquisador



Fonte: do Autor.

A APCL foi instanciada em um servidor local conectado à internet, sendo acessada por meio de navegador web, de modo a permitir a conexão com os WebID do Pod tese2022 criado para o Pesquisador, conforme ilustrado na Figura 16.

Fonte: do Autor.

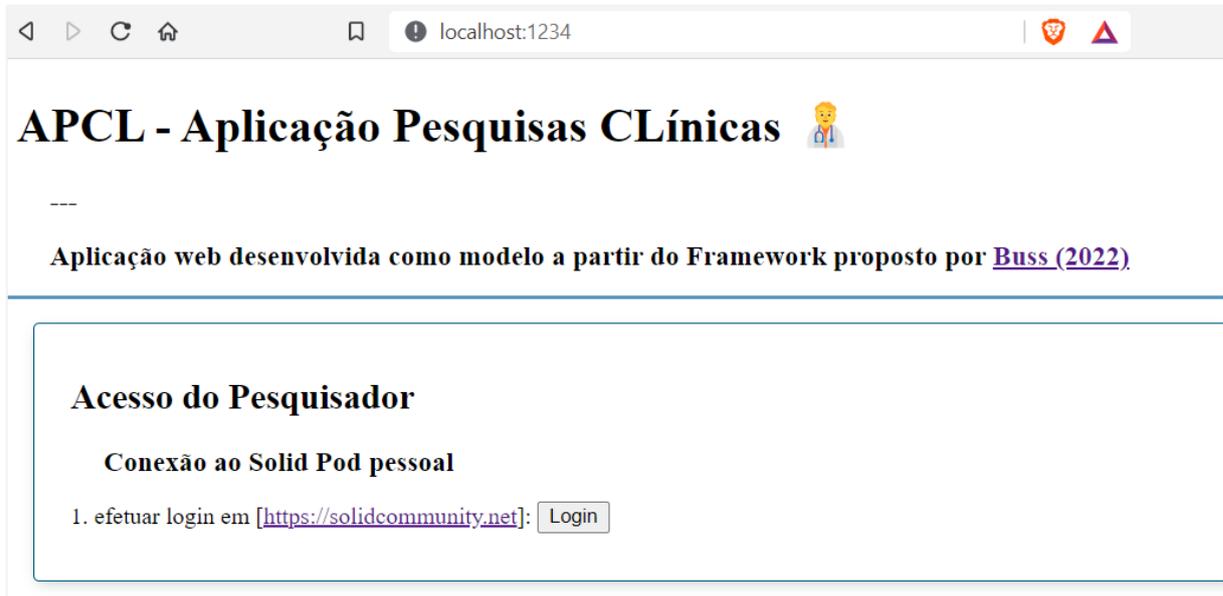


Figura 17 - Login no provedor

Conforme anteriormente mencionado, devido a especificação do framework, não há nenhum dado relacionado a qualquer usuário na mesma máquina/servidor onde a aplicação funciona, portanto, a primeira ação disponível ao acessar a APCL é a de se conectar ao provedor de pods onde o pesquisador armazena suas informações. É esse provedor que devolve para a APCL o WebID completo do pesquisador conforme exposto nas Figuras 17 e 18. A primeira imagem ilustra a tela de login fornecida pelo servidor, responsável por prover acesso ao Pod tese2022 e a segunda imagem ilustra a tela da APCL já vinculada a WebID do pesquisador.

Fonte: o Autor.

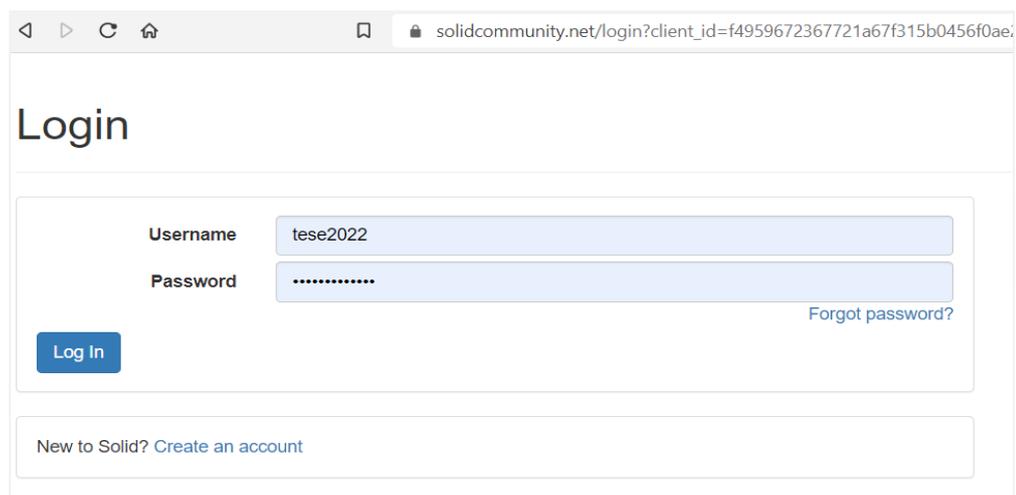


Figura 18 - APCL após a conexão com o Solid Pod do Pesquisador

The screenshot shows a web browser window with the address bar displaying 'localhost:1234'. The page title is 'APCL - Aplicação Pesquisas CLínicas' with a doctor icon. Below the title, there is a subtitle: 'Aplicação web desenvolvida como modelo a partir do Framework proposto por [Buss \(2022\)](#)'. The main content is divided into two sections:

Acesso do Pesquisador

Conexão ao Solid Pod pessoal

1. efetuar login em [\[https://solidcommunity.net/\]](https://solidcommunity.net/):

Sua sessão está logada com o WebID [\[https://tese2022.solidcommunity.net/profile/card#me\]](https://tese2022.solidcommunity.net/profile/card#me).

Edição e Leitura de dados no Solid Pod do Pesquisador

2. Escreva seu nome:

Status: ...ainda não foi escrito...

3. Leia o nome (de qualquer pessoa) pelo seu WebID:

Nome lido do pod: ...ainda não lido...

Fonte: o Autor.

Na Figura 18 é possível visualizar que após a vinculação da aplicação ao WebID do pesquisador, duas funções de testes podem ser efetuadas, uma para editar dados no próprio WebID do pesquisador e outra para ler os dados públicos de qualquer WebID disponível.

A Figura 19 exibe o resultado dos exemplos executados, sendo o primeiro relativo a atualização do nome do Pesquisador e o segundo uma consulta realizada em um WebID de um dos possíveis voluntários da pesquisa, simulando por exemplo, uma checagem de dados antes do envio de uma requisição de participação em um ensaio clínico.

Figura 19 - Exemplo de interação com dados privados e públicos

APCL - Aplicação Pesquisas CLínicas 🧑

Aplicação web desenvolvida como modelo a partir do Framework proposto por [Buss \(2022\)](#).

Acesso do Pesquisador

Conexão ao Solid Pod pessoal

1. efetuar login em [\[https://solidcommunity.net\]](https://solidcommunity.net):

Sua sessão está logada com o WebID [\[https://tese2022.solidcommunity.net/profile/card#me\]](https://tese2022.solidcommunity.net/profile/card#me).

Edição e Leitura de dados no Solid Pod do Pesquisador

2. Escreva seu nome:

Status: **Escreveu [Maico] como name com sucesso!**

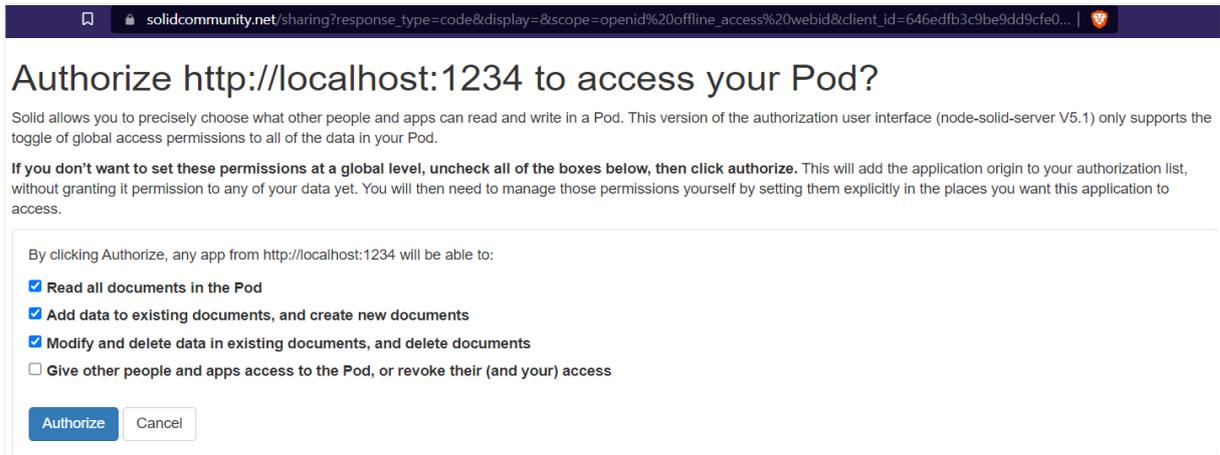
3. Leia o nome (de qualquer pessoa) pelo seu WebID:

Nome lido do pod: **[Voluntário 2]**

Fonte: o Autor.

A preservação da privacidade e do controle sobre os dados pessoais dos participantes dos ensaios clínicos, são objetivos centrais da concepção e desenvolvimento do framework. Para garantir o cumprimento desses objetivos, é parte obrigatória do procedimento de primeira conexão da APCL com o Pod v2apcl o processo de solicitação e aprovação de acessos e privilégios sobre os dados. Para tanto, a APCL encaminha ao Pod v2apcl uma lista identificando as concessões pretendidas e cabe ao voluntário a aprovação ou não dessas concessões por meio de um formulário próprio para tal, conforme pode ser observado na Figura 20. No caso de aprovação, critérios de acesso são registrados no Pod v2apcl como uma lista ACL, garantindo que qualquer novo acesso da mesma aplicação mantenha os privilégios já concedidos até que esta solicite uma alteração ou, que o Voluntário 2 revogue a totalidade ou parte dos privilégios.

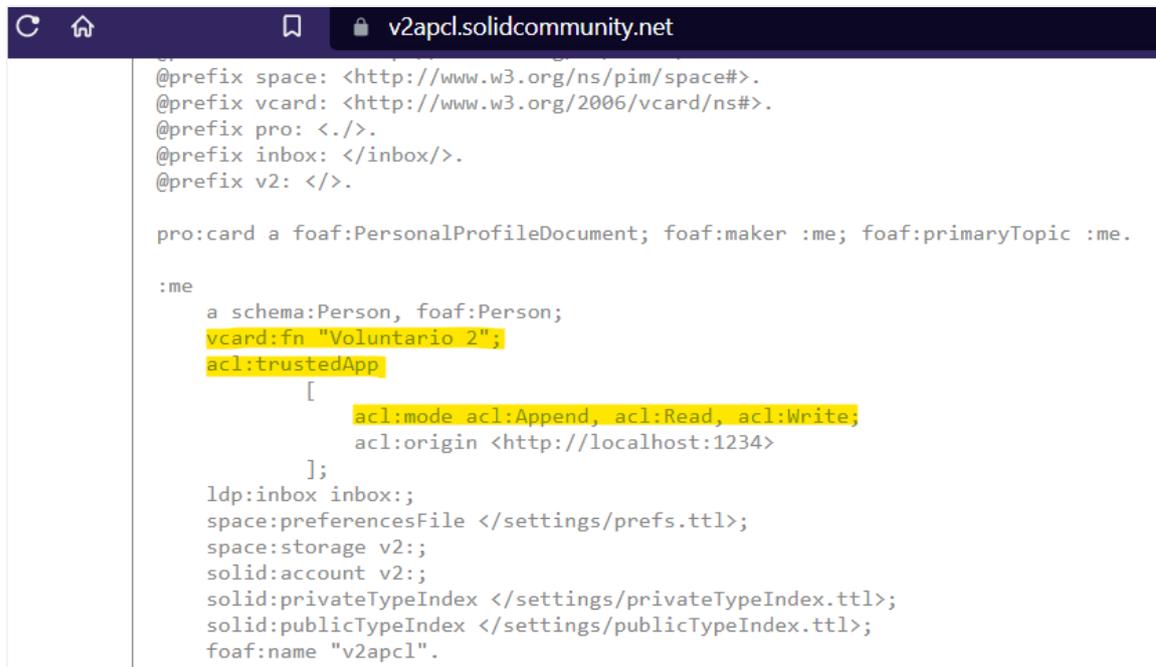
Figura 20 – Exemplo de página de autorização de acesso



Fonte: Autor.

Na Figura 21 abaixo é possível visualizar que a APCL possui acesso para atualizar, ler ou escrever o recurso Nome Formatado (fn, do inglês formatted name) no Pod v2apcl.

Figura 21 - Permissões da APCL no Pod v2apcl



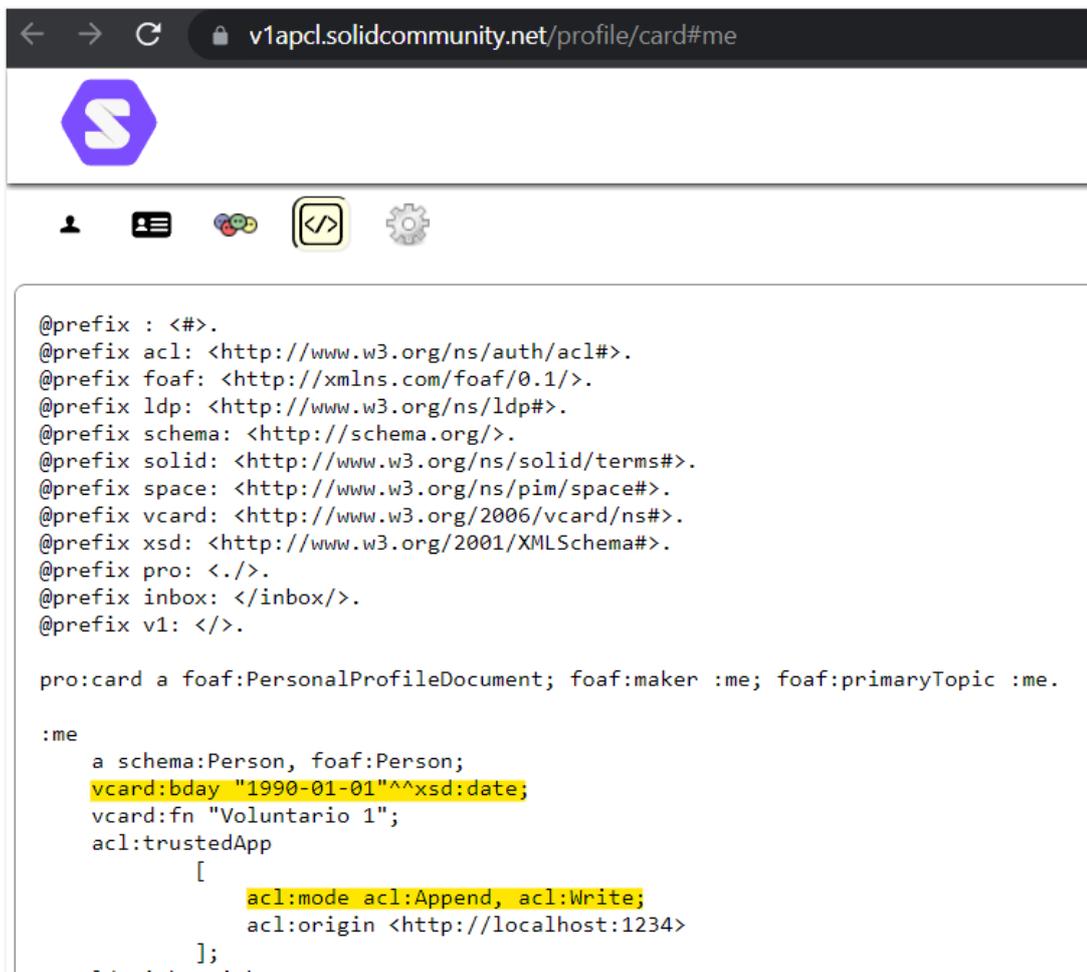
Fonte: o Autor.

Os recursos armazenados em Pods podem ser comumente acessados usando solicitações HTTP GET. Os provedores Solid são incentivados a realizar a negociação de conteúdo para recursos RDF, dependendo do valor do cabeçalho da requisição. Mas qualquer

tipo de arquivo (documentos, vídeos, fotos, etc.) pode ser considerado um recurso dentro da estrutura definida.

Na continuidade dos testes conduzidos por meio da APCL, efetuou-se a simulação de acessos a dados não constantes na lista ACL do Pod v1apcl. Buscou-se efetuar a leitura do valor correspondente a Data de Nascimento do Voluntário 1, que conforme a Figura 22, está gravada no Pod v1apcl, mas não corresponde a um item da lista que permite acesso de leitura pela APCL.

Figura 22 - Exibição da data de nascimento gravada no Pod v1apcl



```

@prefix : <#>.
@prefix acl: <http://www.w3.org/ns/auth/acl#>.
@prefix foaf: <http://xmlns.com/foaf/0.1/>.
@prefix ldap: <http://www.w3.org/ns/ldap#>.
@prefix schema: <http://schema.org/>.
@prefix solid: <http://www.w3.org/ns/solid/terms#>.
@prefix space: <http://www.w3.org/ns/pim/space#>.
@prefix vcard: <http://www.w3.org/2006/vcard/ns#>.
@prefix xsd: <http://www.w3.org/2001/XMLSchema#>.
@prefix pro: <./>.
@prefix inbox: </inbox/>.
@prefix v1: </>.

pro:card a foaf:PersonalProfileDocument; foaf:maker :me; foaf:primaryTopic :me.

:me
  a schema:Person, foaf:Person;
  vcard:bday "1990-01-01"^^xsd:date;
  vcard:fn "Voluntario 1";
  acl:trustedApp
    [
      acl:mode acl:Append, acl:Write;
      acl:origin <http://localhost:1234>
    ];
  ldap:inbox inbox:

```

Fonte: o Autor.

A Figura 23 ilustra a resposta recebida pela APCL ao requisitar a informação relativa a data de nascimento do Pod v1apcl, para a qual não tem permissão de leitura.

Figura 23 - Tentativa de acesso a dado do voluntário sem permissão



Fonte: o Autor.

Esse bloqueio do acesso a determinadas informações do voluntário é uma importante contribuição do framework. Essa garantia de granularidade no acesso a informações pessoais não é alcançada nos modelos atuais onde os voluntários só possuem a opção de anuir ou não com o compartilhamento de todas as suas informações armazenadas. A lista ACL corresponde a representação das condições de compartilhamento pactuadas no termo de consentimento e no protocolo da pesquisa.

É importante salientar que o protótipo não possui uma estrutura para a finalidade similar de gestão de um ensaio clínico, mas sim componentes que propiciam testes unitários de situações previstas na literatura de forma a viabilizar a realização de análises quanto a adequação das soluções apresentadas pelo framework desenvolvido.

4.3 AVALIAÇÃO

Espera-se nesta etapa, que o artefato seja observado e avaliado em relação a solução do problema que se propôs a solucionar. Esse procedimento de avaliação pode incluir qualquer evidência empírica apropriada ou prova lógica. No final desta atividade, o pesquisador pode avaliar se é necessário que se efetue uma nova iteração partindo da Etapa de Design e Desenvolvimento para tentar melhorar a eficácia do artefato ou se prossegue para a fase de Comunicação, permitindo deixar melhorias adicionais para trabalhos futuros (PEFFERS, et. al., 2007).

Seguindo a forma de avaliação experimental conforme Lacerda et al. (2013), buscou-se validar o modelo instanciado utilizando-se dos requisitos definidos para o framework na etapa de inferência da metodologia empregada neste trabalho. Assim, um aplicativo web foi produzido para executar testes específicos, alinhados a estes requisitos, dos quais os resultados são expressos a seguir.

Em relação ao requisito 01, “o artefato deve ampliar as possibilidades de privacidade dos participantes”, isso foi feito, conferindo a cada participante do ensaio clínico um Solid Pod, hospedado em um servidor Solid público, que pôde ser utilizado com sucesso para armazenar arquivos com informações pessoais. O acesso ao arquivo pela aplicação só foi possível por meio da permissão do proprietário do Pod, que possui a opção de apenas permitir a leitura dos dados ou conceder maiores direitos ao aplicativo.

A utilização da especificação Solid aplicada neste trabalho, se fez por estar alinhada aos objetivos propostos para o framework. Seu foco central é possibilitar a descoberta e o compartilhamento de informações de forma a preservar a confidencialidade. Ele permite, conforme já mencionado, que os usuários armazenem dados pessoais em Pods (um padrão para estruturas de armazenamento de dados online) hospedados no local definido pelos próprios. A forma como o Solid funciona também permite a flexibilidade de se distribuir dados entre vários Pods, permitindo que vários tipos de dados (pessoais, de contato, saúde, financeiros) sejam organizados em diferentes graus de controle de acesso. Em poucas palavras, a especificação Solid adotada, permite que os usuários mantenham a propriedade completa de seus dados, incluindo onde armazená-los e definindo quem possui permissão para acessá-los, fornecendo uma plataforma eficaz e segura para se executar aplicativos na web.

O requisito 02 traz como premissa que “o artefato deve permitir o compartilhamento granular das informações dos participantes”. Também derivada da utilização de um Solid Pod como estrutura para armazenamento dos dados pessoais, este requisito é satisfeito por meio da utilização na arquitetura proposta, de listas de controle de acesso (ACL, do inglês *Access Control List*), com definição de granularidade definida por design. A depender da implementação dada ao modelo instanciado do framework, a granularidade poderá ser maior ou menor, chegando inclusive a dados unitários. No exemplo ilustrado pela Figura 22 a APCL possui apenas acesso de escrita e atualização das características, não conseguindo efetuar a leitura da característica Data de Nascimento (item `vcard:bday` do recorte xml exibido) do voluntário 1.

Para satisfazer o requisito 03, “O artefato deve garantir o cumprimento do termo de consentimento”, o framework faz uso do conceito de WebID, combinando conceitos de web semântica com métodos de autenticação e autorização baseados em certificados para atender aos requisitos de segurança e confiança. No caso de uso de inscrição no estudo, por exemplo, usa-se as WebIDs das partes vinculantes (Pesquisador e Voluntário) do termo de consentimento para registrar em ambos os pods (Pod tese2022 e Pod v2apcl), os dados referentes ao acordo

firmado. O termo de consentimento, na leitura do framework proposto, é traduzido como uma lista ACL de permissões, concedidas pelo participante da pesquisa para o pesquisador, em relação ao acesso aos dados que estão registrados no seu pod pessoal. Um requisito imposto na condução dos ensaios clínicos pelos órgãos reguladores e que é facilitado pela arquitetura do framework, é a garantia da opção de *opt-out*(saída) do voluntário a qualquer tempo da pesquisa. No modelo apresentado, basta que o voluntário, a qualquer tempo, revogue as concessões fornecidas a APCL, para que seus dados não sejam mais acessados a partir deste ponto.

Em relação ao requisito 04, “O artefato deve utilizar padrões e tecnologias abertas”, este é seguramente o ponto central do emprego, na estruturação do framework como solução à pergunta de pesquisa, da especificação Solid, pois, em detrimento de outras soluções citadas na literatura, ela combina vários padrões abertos da Web, como WebID, Linked Data Platform, RDF e SPARQL, de forma coerente e coesa. Os aplicativos desenvolvidos segundo o framework, podem funcionar tanto em plataformas móveis quanto diretamente para a WEB, funcionando pelo lado do cliente (*client-side applications*) lendo e gravando dados diretamente dos pods. Nenhuma tecnologia proprietária foi empregada em qualquer passo da solução, permitindo que qualquer desenvolvedor crie novos produtos, retirando qualquer barreira de entrada para inovações futuras na área.

Em relação ao requisito 05, “o artefato deve permitir escalabilidade”, o framework emprega a dissociação entre a solução lógica da aplicação e as estruturas utilizadas para o armazenamento dos dados. A utilização de diferentes servidores, possivelmente também distribuídos geograficamente, tende a, além de conferir escalabilidade para o sistema, ampliar sua tolerância a falhas.

5 CONSIDERAÇÕES FINAIS

Nesse capítulo são apresentadas as considerações finais, a generalização dos resultados e as possibilidades de trabalhos futuros. Observando-se a metodologia DSRM, nesse capítulo é completada a etapa referente a comunicação da abordagem ao problema e sua importância, dando projeção ao desenvolvimento do artefato, sua aplicabilidade, utilidade, novidade, da sua relevância e eficácia para pesquisadores da área e até de outros públicos onde o artefato possa ser aplicado.

5.1 CONCLUSÃO

A constante necessidade de novos tratamentos, a velocidade com que as doenças têm se espalhado em um mundo cada vez mais interconectado, a evolução das tecnologias relacionadas a saúde, o aumento da expectativa de vida da população atrelada a uma exigência de que essa longevidade se dê com qualidade, entre outros fatores, pressionam continuamente a ciência a se expandir no campo da saúde. Ao longo dos anos, tem sido por meio das pesquisas clínicas que se obteve grande parte do conhecimento na área das ciências biomédicas que hoje está disponível. Em decorrência de tamanha significância, esse tipo de estudo vem continuamente sofrendo novas regulações e aperfeiçoamentos em seus protocolos e práticas. Apesar disso, ainda é possível a constatação da ocorrência de falhas e inconsistências que colocam em risco a segurança, a reprodutibilidade e até mesmo a qualidade desse tipo de estudos.

Neste mesmo campo há de se considerar a importância dos participantes voluntários dos ensaios clínicos. Deve-se garantir a estes entes, o mais alto grau de confiança que seja possível, de modo a conferir-lhes a certeza da relevância da sua contribuição nos estudos em que participam. Com a inexorável transformação digital nos ensaios clínicos, busca-se obter maior abrangência geográfica, maior velocidade na condução desses estudos e facilitar o recrutamento e o acompanhamento de participantes. Junto a isso tem-se o aumento da complexidade e da quantidade dos dados coletados que trouxeram consigo novos problemas. A compreensão desses problemas e uma adequada abordagem no seu tratamento é fundamental para o desenvolvimento de soluções tecnológicas bem-sucedidas.

Para a realização desta tese, efetuou-se uma revisão da literatura a fim de se reconhecer e classificar os desafios encontrados na condução de ensaios clínicos, principalmente aqueles relacionados a utilização dos dados pessoais dos participantes, frente aos recentes fatos

envolvendo a utilização indevida deste tipo de informação e a atual preocupação com a temática que tem sido demonstrada por vários países, que vem atualizando suas legislações para proteger os usuários no mundo digital. Com base nesse levantamento, foi possível categorizar os achados em 6 grupos, baseando-se nas suas características em comum para agrupá-los em problemas de ordem financeira, tecnológica, social e política, científica, clínica e de privacidade.

Com o intuito de se fornecer meios para apoiar a criação do conhecimento que se dá por meio dos ensaios clínicos, dirimindo os problemas relacionados a integridade dos dados utilizados e a privacidade dos participantes, estruturou-se em um framework híbrido, por meio da compreensão do domínio, as bases, critérios e padrões necessários para a realização do propósito.

A concepção do framework foi guiada por preceitos que buscaram, além de cumprir com o objetivo de preservar a propriedade e a privacidade dos dados dos participantes de ensaios clínicos, também intervir nos demais problemas levantados na literatura. O problema de interoperabilidade entre os sistemas existentes, foi abordado ao se adotar apenas tecnologias e padrões abertos e sem custos de aquisição para estruturar o framework. Também colabora para este ponto, a extensão, adaptação de nomenclatura ao preconizado pela SNOMED CT e disponibilização pública de uma ontologia específica para o domínio estudado. A opção pela descentralização dos dados seguindo a especificação Solid, ao mesmo tempo traz respostas para os problemas de propriedade, privacidade e rastreabilidade dos dados, busca atrelar o produto da tese a uma comunidade já consolidada e engajada em evoluir continuamente as tecnologias aplicadas, favorecendo futuras derivações e evoluções das abordagens efetuadas.

De modo a analisar a aplicabilidade do constructo desenvolvido, fez-se sua implementação em um cenário representativo daquele encontrado no dia a dia da condução de ensaios clínicos. Tendo como suporte e guia o framework apresentado, foi desenvolvido um aplicativo web que serviu como ferramenta para, dentro do que prega a metodologia que guiou este trabalho, validar os requisitos que correlacionavam os problemas relatados na literatura com as diretrizes para sua resolução ou atenuação.

A implementação demonstrou que o framework pode ser eficaz para garantir que o protocolo de pesquisa seja seguido e, por consequência que os procedimentos de consentimento informado sejam válidos durante toda a existência do ensaio clínico. Do ponto de vista do paciente, o desacoplamento entre a ferramenta de gestão do estudo e os seus dados pessoais e clínicos, lhe devolve o a propriedade e lhe oferece meios de definir o grau de compartilhamento ou de privacidade relacionado as suas informações. Pelo lado dos responsáveis pelas pesquisas,

tem-se a garantia da integridade dos dados e a promoção da transparência e rastreabilidade das informações entre as partes interessadas.

A adoção da especificação trazida pelo framework, produto desta tese, permite a adição ou retirada de diferentes nós especializados em sua estrutura, expandindo ou diminuindo sua complexidade, uma vez que as regras de negócio estão desatreladas das fontes de dados e estas, por sua vez, possuem a autonomia de decidir quanto a sua participação ou não a um nível tão granular quanto a modelagem do domínio implantada a permita. O framework foi desenvolvido de forma a orientar princípios a serem seguidos enquanto se atua na contínua transformação digital que se passa na área da saúde e, apesar de ter sido definido como uma contribuição no domínio dos ensaios clínicos, o artefato também pode ser útil para outras esferas, onde os requisitos de segurança para dados são tão ou menos severos do que os dados pessoais e de saúde.

5.2 TRABALHOS FUTUROS

Na realização deste trabalho, novas possibilidades de pesquisa e evolução do desenvolvimento foram percebidas. Há uma oportunidade importante agora de se aproveitar as novas tecnologias disponíveis para acelerar o ritmo em que geramos evidências por meio de ensaios clínicos. A digitalização proporciona a melhoria da eficiência do estudo, aprimorando e apoiando o papel dos investigadores e das equipes de apoio. A utilização de novos dispositivos digitais e de *internet das coisas*, possibilita uma nova forma muito rica de acompanhamento dos participantes de ensaios clínicos. Aproveitar a capacidade dessas ferramentas em coletar e transmitir dados continuamente aos pesquisadores, pode melhorar a detecção de eventos infrequentes ou aqueles que são específicos da situação e improváveis de ocorrer durante uma visita de acompanhamento. A vinculação destes dispositivos com sistemas mobile desenvolvidos com base no framework proposto, podem facilitar um compartilhamento granular de informações.

A possibilidade de se estender a ontologia, em conjunto com a comunidade médica e científica, pode colaborar com o desenvolvimento de novas ferramentas que sigam os principais padrões internacionais de nomenclatura e formato dos dados clínicos, cooperando com a interoperabilidade de dos sistemas na área da saúde.

A utilização de Solid pods para o armazenamento de informações médicas pessoais pode vir a permitir, com o avanço da adoção dessa especificação, que as pessoas tenham sempre

consigo seus históricos médicos, questão muito importante para um atendimento mais personalizado em situações de troca de prestadores ou mesmo emergências de saúde.

REFERÊNCIAS

AMANKWAH-AMOAHA, J. et al. COVID-19 and digitalization: The great acceleration. **Journal of Business Research**, v. 136, p. 602–611, 1 nov. 2021.

ANGELETTI, F.; CHATZIGIANNAKIS, I.; VITALETTI, A. Towards an architecture to guarantee both data privacy and utility in the first phases of digital clinical trials. **Sensors (Switzerland)**, v. 18, n. 12, 1 dez. 2018.

ANVISA. **RESOLUÇÃO DA DIRETORIA COLEGIADA - RDC Nº 9, DE 20 DE FEVEREIRO DE 2015**. [s.l: s.n.].

AZAMBUJA, A. J. G. DE; GRANVILLE, L. Z.; SARMENTO, A. G. M. A privacidade, a segurança da informação e a proteção de dados no Big Data. **Parcerias Estratégicas**, v. 24, n. 48, p. 9–32, 2019.

BENCHOUFI, M.; PORCHER, R.; RAVAUD, P. Blockchain protocols in clinical trials: Transparency and traceability of consent. **F1000Research**, v. 6, p. 66, 23 jan. 2017.

BERNERS-LEE, T.; HENDLER, J.; LASSILA, O. The Semantic WEB. **Scientific American**, v. 284, n. 5, p. 34–43, 2001.

BHATT, A. Quality of clinical trials: A moving target. **Perspectives in Clinical Research**, v. 2, n. 4, p. 124, 2011.

BHATT, A. Protocol deviation and violation. **Perspectives in Clinical Research**, v. 3, n. 3, p. 117, 2012.

BUNGE, M. A. **Emergence and convergence: qualitative novelty and the unity of knowledge**. [s.l.] University of Toronto Press, 2003.

BUSS, M. O. **MODELO DE SISTEMA DE CONHECIMENTO PARA GESTÃO DE LISTAS DE ESPERA PARA CIRURGIAS NO SISTEMA ÚNICO DE SAÚDE**. [s.l.] Universidade Federal de Santa Catarina, 2015.

CAULFIELD, T.; KAYET, J. **BROAD CONSENT IN BIOBANKING: REFLECTIONS ON SEEMINGLY INSURMOUNTABLE DILEMMAS** *Medical Law International*. [s.l: s.n.].

CECCHINI, M. et al. Challenges with Novel Clinical Trial Designs: Master Protocols. **Clinical Cancer Research**, v. 25, n. 7, p. 2049–2057, 1 abr. 2019.

CEDRO, L. F. D. A.; DUQUE, C. G. **Blockchain como tecnologia para transparência de dados de ensaios clínicos**. Vitória, ES: out. 2019. Disponível em: <<https://portal.febab.org.br/anais/article/view/2227>>. Acesso em: 1 nov. 2021

CEZAR, D. O. Pesquisa com Medicamentos, Aspectos Bioéticos. **Revista da Faculdade de Direito da Universidade Lusófona do Porto**, v. 7, n. 7, 2015.

CHAN, A. W.; HRÓBJARTSSON, A. Promoting public access to clinical trial protocols: Challenges and recommendations. **Trials**, v. 19, n. 1, 17 fev. 2018.

CIPRIANI, A.; BARBUI, C. What is a clinical trial protocol? **Epidemiologia e Psichiatria Sociale**, v. 19, n. 2, p. 116–117, 2010.

Clinical trial documentation. Disponível em: <<https://www.medtran.ru/eng/trials/trialdocumentation.htm>>. Acesso em: 26 abr. 2022.

CONSELHO NACIONAL DE SAÚDE. **RESOLUÇÃO Nº 466**. Brasil, 2012.

CRISIGIOVANNI, L. et al. **MY DATA IS MINE DECLARATION**. Bruxelas: [s.n.]. Disponível em: <<https://www.mydataismine.com/manifest>>.

Data - W3C. Disponível em: <<https://www.w3.org/standards/semanticweb/data>>. Acesso em: 10 maio. 2022.

DIMASI, J. A.; GRABOWSKI, H. G.; HANSEN, R. W. Innovation in the pharmaceutical industry: New estimates of R&D costs. **Journal of health economics**, v. 47, p. 20–33, maio 2016.

EMYGDIO, J. L.; ALMEIDA, M. B. Representações formais do Conhecimento aplicadas à interoperabilidade semântica de terminologias clínicas. **Múltiplos Olhares em Ciência da Informação**, v. 9, p. 1–15, 2019.

ESLAMI ANDARGOLI, A. et al. **Health information systems evaluation frameworks: A systematic review.** **International Journal of Medical Informatics** Elsevier Ireland Ltd, , 1 jan. 2017.

FEIJÓ, A. G. DOS S. et al. Pesquisa clínica sob a ótica da integridade. **Revista Bioética**, v. 26, n. 2, p. 172–182, jun. 2018.

FENG, Q. et al. **A survey on privacy protection in blockchain system.** **Journal of Network and Computer Applications** Academic Press, , 15 jan. 2019.

FERREIRA, A. et al. Access Control: how can it improve patients' healthcare? **Stud Health Technol Inform.**, p. 65–77, 2007.

FOAF Vocabulary Specification. Disponível em: <<http://xmlns.com/foaf/spec/20140114.html>>. Acesso em: 12 maio. 2022.

FREITAS, F.; SCHULZ, S.; MORAES, E. Survey of current terminologies and ontologies in biology and medicine. **RECIIS**, v. 3, n. 1, 11 mar. 2009.

GARDNER, W.; LIDZ, C. W.; HARTWIG, K. C. Authors' reports about research integrity problems in clinical trials. **Contemporary Clinical Trials**, v. 26, n. 2, p. 244–251, 2005.

GIL, A. C. **Como elaborar projetos de pesquisa**. [s.l.] Atlas, 2009.

GitHub - solid/webid-oidc-spec: WebID-OIDC Authentication Spec v0.1.0. Disponível em: <<https://github.com/solid/webid-oidc-spec>>. Acesso em: 10 maio. 2022.

GOLDENBERG, M. **A arte de pesquisar: como fazer pesquisa qualitativa em Ciências Sociais**. 8 ed. ed. Rio de Janeiro: Record, 2004.

HANG, L. et al. A Permissioned Blockchain-Based Clinical Trial Service Platform to Improve Trial Data Transparency. **BioMed Research International**, v. 2021, 2021.

HAY, M. et al. Clinical development success rates for investigational drugs. **nature biotechnology**, 2014.

Healthcare Data Breach Statistics. Disponível em: <<https://www.hipaajournal.com/healthcare-data-breach-statistics/>>. Acesso em: 6 maio. 2022.

HEVNER, A. et al. Design Science in Information Systems Research. **Management Information Systems Quarterly**, v. 28, p. 75, jul. 2004.

HIRANO, T. et al. Data validation and verification using blockchain in a clinical trial for breast cancer: Regulatory sandbox. **Journal of Medical Internet Research**, v. 22, n. 6, 2 jun. 2020.

HULLEY, S. B. et al. **Designing Clinical Research**. Four Editi ed. London: LIPPINCOTT WILLIAMS & WILKINS, 2013.

HUTCHINGS, E. et al. A systematic literature review of researchers' and healthcare professionals' attitudes towards the secondary use and sharing of health administrative and clinical trial data. **Systematic reviews**, v. 9, n. 1, p. 1–240, 2020.

INAN, O. T. et al. Digitizing clinical trials. **npj Digital Medicine**, v. 3, n. 1, p. 20, 2020.

INTEGRITY, W. C. ON R. **Mission - World Conferences on Research Integrity**. Disponível em: <<https://www.wcrif.org/foundation/mission>>. Acesso em: 27 abr. 2022.

IOANNIDIS, J. P. A. Why most published research findings are false. **Getting to Good: Research Integrity in the Biomedical Sciences**, p. 2–8, 23 jul. 2018.

JAQUES, P. et al. **Metodologia de Pesquisa Científica em Informática na Educação: Concepção de Pesquisa**. Porto Alegre: [s.n.]. v. 1

KARBWANG, J.; CRWLEY, F. **Ethical Review and the Globalization of Clinical Trials**. Disponível em: <<https://www.appliedclinicaltrials.com/view/ethical-review-and-globalization-clinical-trials>>. Acesso em: 14 fev. 2021.

KAYE, J. et al. Dynamic consent: A patient interface for twenty-first century research networks. **European Journal of Human Genetics**, v. 23, n. 2, p. 141–146, 20 fev. 2015.

LACERDA, D. P. et al. Design Science Research: método de pesquisa para a engenharia de produção. **Gestão & Produção**, v. 20, n. 4, p. 741–761, 2013.

LEE, D. et al. A survey of SNOMED CT implementations. **Journal of Biomedical Informatics**, v. 46, n. 1, p. 87–96, fev. 2013.

LIMA, J. C. DE; CARVALHO, C. L. DE. **Resource Description Framework (RDF)**. [s.l.: s.n.]. Disponível em: <www.inf.ufg.br>.

MANICA, H. P. T. **MODELO DE RECUPERAÇÃO E COMUNICAÇÃO DE UTILIZAÇÃO DE DISPOSITIVOS PORTÁTEIS**. [s.l.] Universidade Federal de Santa Catarina, 2009.

MARCH, S. T.; SMITH, G. F. **Design and natural science research on information technology Decision Support Systems**. [s.l.: s.n.].

MARTIN, L. M. **Writing a clinical trial protocol: Expert tips - The Kolabtree Blog**. Disponível em: <https://www.kolabtree.com/blog/how-to-write-a-clinical-trial-protocol/#The_SPIRIT_Statement>. Acesso em: 25 abr. 2022.

MARTINEZ-COSTA, C.; SCHULZ, S. HL7 FHIR: Ontological Reinterpretation of Medication Resources. **Studies in Health Technology and Informatics**, v. 235, p. 451–455, 2017.

MÉDICAS, F. DE C. **O que é pesquisa clínica**. Disponível em: <<https://www.fcm.unicamp.br/fcm/cpc-centro-de-pesquisa-clinica/pesquisa-clinica/o-que-e-pesquisa-clinica>>. Acesso em: 29 maio. 2021.

MESLIN, E. M.; CHO, M. K. **Research ethics in the era of personalized medicine: Updating science's contract with society**. **Public Health Genomics**, ago. 2010.

Ministério da Saúde Conselho Nacional de Saúde Comissão Nacional de Ética em Pesquisa. [s.d.].

NISHIMURA, R. et al. Experience and awareness of research integrity among Japanese physicians: A nationwide cross-sectional study. **BMJ Open**, v. 11, n. 10, 2021.

NOVITZKE, J. M. The significance of clinical trials. **Journal of Vascular and Interventional Neurology**, v. 1, n. 1, 2008.

O'BRIEN, K.; WRIGHT, J. How to write a protocol. **Journal of Orthodontics**, v. 29, p. 58–61, 2002.

OLSEN, R. et al. **The impact of clinical trial monitoring approaches on data integrity and cost - A review of current literature**. **European Journal of Clinical Pharmacology** Springer Verlag, , 1 abr. 2016.

PAULOS, J. **Investigating Decentralized Management of Health and Fitness Data**. [s.l.] MASSACHUSETTS INSTITUTE OF TECHNOLOGY, maio 2020.

PEFFERS, K. et al. A design science research methodology for information systems research. **Journal of Management Information Systems**, v. 24, n. 3, p. 45–77, dez. 2007.

PICKARD, K. T.; SWAN, M. **Big Desire to Share Big Health Data: A Shift in Consumer Attitudes toward Personal Health Information**. AAI 2014 Spring Symposium Series. **Anais...**2014. Disponível em: <<http://bit.ly/sharing-health-info-survey>>

PPEGC. **Programa de Pós-graduação em Engenharia e Gestão do Conhecimento (PPGEGC)**. Disponível em: <<https://ppgegc.paginas.ufsc.br/linhas-de-pesquisa/>>. Acesso em: 8 fev. 2022.

RAMACHANDRAN, M. et al. **Towards Complete Decentralised Verification of Data with Confidentiality: Different ways to connect Solid Pods and Blockchain**. Companion Proceedings of the Web Conference 2020. **Anais...**New York, NY, USA: ACM, 20 abr. 2020.

RANG, H. **Bad Pharma: how drug companies mislead doctors and harm patients** by Ben Goldacre. Published by Fourth Estate, London, 2012. 364 pp, ISBN: 978-0-00-735074-2. **British Journal of Clinical Pharmacology**, v. 75, n. 5, p. 1377–1379, 2013.

RESNIK, D. B.; RASMUSSEN, L. M.; KISSLING, G. E. An International Study of Research Misconduct Policies. **Accountability in Research**, v. 22, n. 5, p. 249–266, 2015.

RIBEIRO JUNIOR, D. I. **Modelo de Sistema Baseado em Conhecimento para Apoiar Processos de Tomada de Decisão Em Ciência e Tecnologia**. [s.l.] Universidade Federal de Santa Catarina, 2010.

SAMBRA, A. V. et al. **Solid: A Platform for Decentralized Social Applications Based on Linked Data**. [s.l: s.n.]. Disponível em: <<https://diasporafoundation.org/>>.

SBPPC. **SBPPC - Sociedade Brasileira de Profissionais em Pesquisa Clínica**. Disponível em: <<https://www.sbppc.org.br/pesquisa-clinica>>. Acesso em: 18 jun. 2022.

SCHULZ, S.; JANSEN, L. Formal ontologies in biomedical knowledge representation. **Yearbook of medical informatics**, v. 8, n. May 2014, p. 132–146, 2013.

SEEFELD, T. **PESQUISA CLÍNICA: A RESPONSABILIDADE CIVIL DO PESQUISADOR, DAS INSTITUIÇÕES E DO PATROCINADOR**. Porto Alegre: Hospital de Clínicas de Porto Alegre, 2020.

SHARMA, N.; SRIVASTAV, A. K.; SAMUEL, A. J. Randomized clinical trial: Gold standard of experimental designs-importance, advantages, disadvantages and prejudice. **Revista Pesquisa em Fisioterapia**, v. 10, n. 3, p. 512–519, 2020.

SHEHABUDDEEN, N. et al. **Centre for Technology Management Working Paper Series Representing and approaching complex management issues: Part 1-Role and definition**. [s.l: s.n.]. Disponível em: <<http://ssrn.com/abstract=1923155>>Electroniccopyavailableat:https://ssrn.com/abstract=1923155Electroniccopyavailableat:http://ssrn.com/abstract=1923155>.

SILVA, E. L. DA; MENEZES, E. M. **Metodologia da pesquisa e elaboração de dissertação**. 3. ed. Florianópolis: Laboratório de Ensino a Distância da UFSC, 2001.

SILVA, E. L.; MENEZES, E. M. **Metodologia da Pesquisa e Elaboração de Dissertação - 4a edição**. [s.l: s.n.].

SIMON, H. A. **The Sciences of the Artificial**. [s.l: s.n.].

SNOMED. **SNOMED - Who we are**. Disponível em: <<https://www.snomed.org/snomed-international/who-we-are>>. Acesso em: 8 jul. 2022a.

SNOMED. **SNOMED CT Starter Guide**. Disponível em: <<https://confluence.ihtsdotools.org/pages/viewpage.action?pageId=81268877>>. Acesso em: 8 jul. 2022b.

Solid. Disponível em: <<https://solidproject.org/about>>. Acesso em: 7 maio. 2022.

SWEETMAN, E. A.; DOIG, G. S. Failure to report protocol violations in clinical trials: A threat to internal validity? **Trials**, v. 12, p. 1–8, 2011.

TOMHAVE, B. L. **Alphabet Soup: Making Sense of Models, Frameworks, and Methodologies**. [s.l: s.n.].

TRIALS, A. C. **What is a clinical trial?** Disponível em: <<https://www.australianclinicaltrials.gov.au/what-clinical-trial>>. Acesso em: 23 maio. 2021.

VIEIRA, S. R. C. et al. Avaliando uma Técnica para Extrair Requisitos a partir de Diagramas de Processos de Negócios através de Estudos Experimentais. **CLEI 2012 - XXXVIII Conferencia Latinoamericana en Informática**, n. January 2012, p. 1–11, 2012.

WASSERMAN, H.; WANG, J. **An Applied Evaluation of SNOMED CT as a Clinical Vocabulary for the Computerized Diagnosis and Problem List**. [s.l: s.n.].

Web Access Control. Disponível em: <<https://solid.github.io/web-access-control-spec/>>. Acesso em: 10 maio. 2022.

WENG, C. et al. An Integrated Model for Patient Care and Clinical Trials (IMPACT) to support clinical research visit scheduling workflow for future learning health systems. **Journal of Biomedical Informatics**, v. 46, n. 4, p. 642–652, 2013.

WORLD MEDICAL ASSOCIATION. **World Medical Association declaration of Helsinki: Ethical principles for medical research involving human subjects**JAMA - **Journal of the American Medical Association**. [s.l.] American Medical Association, 27 nov. 2013.

APÊNDICE 1

Estratégia de busca:

Qualquer campo contém personal health data OU

Qualquer campo contém health data security OU

Qualquer campo contém Health data privacy OU

Qualquer campo contém clinical research

Apenas artigos revisados por pares

2012 - 2022

Resultados: 999.554 Resultados

Qualquer campo contém personal health data E

Qualquer campo contém health data security E

Qualquer campo contém Health data privacy E

Qualquer campo contém clinical research

Apenas artigos revisados por pares

2012 - 2022

Resultados: 6.363 Resultados

Qualquer campo contém personal health data E

Qualquer campo contém health data security E

Qualquer campo contém Health data privacy E

Título contém clinical research

Apenas artigos revisados por pares

2012 - 2022

Resultados: 180 Resultados

Filtro final:

Efetuada a leitura dos títulos e resumos: restaram 34 correlacionados com o tema da

tese.

APÊNDICE 2

Material complementar disponível no GitHub© sob *Licença Creative Commons Atribuição-CompartilhaIgual 4.0 Internacional*:

- Códigos fonte da aplicação: <https://github.com/omaicobuss/framework-tese>
- Ontologia de domínio: <https://github.com/omaicobuss/framework-tese/ontologia>
- Guia do desenvolvedor para FHIR: <https://github.com/omaicobuss/framework-tese/tree/main/Materiais%20Complementares/FHIR>
- Navegador gratuito SNOMED CT: <http://browser.ihtsdotools.org/>