

***Blockchain* para gestão de dados científicos: um olhar sobre Archangel e dARK**

Carlos Victor de Oliveira

Tese de doutorado

Abril de 2024



UNIVERSIDADE FEDERAL DO RIO DE JANEIRO – UFRJ
ESCOLA DE COMUNICAÇÃO – ECO
INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA –
IBICT
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO – PPGCI

CARLOS VICTOR DE OLIVEIRA

BLOCKCHAIN PARA GESTÃO DE DADOS CIENTÍFICOS: UM OLHAR SOBRE
ARCHANGEL E dARK

Rio de Janeiro

2024

CARLOS VICTOR DE OLIVEIRA

BLOCKCHAIN PARA GESTÃO DE DADOS CIENTÍFICOS: UM OLHAR SOBRE
ARCHANGEL E dARK

Tese de Doutorado apresentada ao Programa de Pós-Graduação em Ciência da Informação, convênio entre o Instituto Brasileiro de Informação em Ciência e Tecnologia e a Universidade Federal do Rio de Janeiro / Escola de Comunicação, como requisito parcial à obtenção do título de Doutor em Ciência da Informação.

Orientadores:

Prof. Dr. Luis Fernando Sayão

Profa. Dra. Luana Farias Sales Marques

Rio de Janeiro

2024

O48b Oliveira, Carlos Victor de
BLOCKCHAIN PARA GESTÃO DE DADOS CIENTÍFICOS: UM
OLHAR SOBRE ARCHANGEL E dARK / Carlos Victor de
Oliveira. -- Rio de Janeiro, 2024.
169 f.

Orientador: Luis Fernando Sayão.
Coorientadora: Luana Farias Sales Marques.
Tese (doutorado) - Universidade Federal do Rio
de Janeiro, Escola da Comunicação, Instituto
Brasileiro de Informação em Ciência e Tecnologia,
Programa de Pós-Graduação em Ciência da Informação,
2024.


1. Blockchain. 2. Gestão de Dados de Pesquisa.
3. Preservação Digital. 4. Ciência da Informação. 5.
Tecnologia da Informação. I. Sayão, Luis Fernando,
orient. II. Marques, Luana Farias Sales, coorient.
III. Título.

CARLOS VICTOR DE OLIVEIRA

**BLOCKCHAIN PARA GESTÃO DE DADOS CIENTÍFICOS: UM OLHAR SOBRE
ARCHANGEL E dARK**

Tese de Doutorado apresentada ao Programa de Pós-Graduação em Ciência da Informação, convênio entre o Instituto Brasileiro de Informação em Ciência e Tecnologia e a Universidade Federal do Rio de Janeiro / Escola de Comunicação, como requisito parcial à obtenção do título de Doutor em Ciência da Informação.


Rio de Janeiro, 25 de abril de 2024.

Documento assinado digitalmente
 **LUIS FERNANDO SAYAO**
Data: 26/04/2024 16:19:16-0300
Verifique em <https://validar.iti.gov.br>


Prof. Dr. Luis Fernando Sayão
(IBICT/ UFRJ)
Orientador

Documento assinado digitalmente
 **MILTON SHINTAKU**
Data: 29/04/2024 10:14:27-0300
Verifique em <https://validar.iti.gov.br>


Prof. Dr. Milton Shintaku
(UFPR)
Membro externo

Documento assinado digitalmente
 **LUANA FARIAS SALES MARQUES**
Data: 26/04/2024 17:10:00-0300
Verifique em <https://validar.iti.gov.br>


Prof.^a Dr.^a Luana Farias Sales Marques
(IBICT/ UFRJ)
Coorientadora

Documento assinado digitalmente
 **MOISES ANDRE NISENBAUM**
Data: 29/04/2024 11:15:19-0300
Verifique em <https://validar.iti.gov.br>


Prof. Dr. Moisés André Nisenbaum
(IFRJ)
Membro externo

Documento assinado digitalmente
 **FABIO CASTRO GOUVEIA**
Data: 10/06/2024 10:52:23-0300
Verifique em <https://validar.iti.gov.br>

Prof. Dr. Fabio Castro Gouveia
(IBICT/ UFRJ)
Membro interno

Documento assinado digitalmente
 **WASHINGTON LUIS RIBEIRO DE CARVALHO SEG**
Data: 03/05/2024 18:03:42-0300
Verifique em <https://validar.iti.gov.br>

Prof. Dr. Washington Luís Ribeiro de
Carvalho Segundo
(IBICT / MCTI)
Membro externo

Documento assinado digitalmente
 **JORGE CALMON DE ALMEIDA BIOLCHINI**
Data: 28/05/2024 16:01:57-0300
Verifique em <https://validar.iti.gov.br>

Prof. Dr. Jorge Calmon de Almeida
Biolchini
(IBICT/ UFRJ)
Membro interno

AGRADECIMENTOS

Este trabalho só foi possível graças à contribuição de muitas pessoas, algumas diretamente e outras indiretamente, sem mesmo perceberem.

Primeiramente, agradeço a Deus.

Aos meus pais, a quem devo tudo (in memoriam).

À minha querida esposa Juliana e aos nossos filhos Vinícius e Théo, pelo suporte, amor e paciência incondicionais.

Aos orientadores Luis Fernando Sayão e Luana Farias Sales Marques, por sua direção e sabedoria.

Ao amigo Moisés André Nisenbaum que ao longo desta jornada, descobri ser não apenas um pilar de suporte, mas também uma fonte inesgotável de inspiração e amizade.

À minha amiga, Rita de Cássia Cordeiro de Castro, por ser uma parceira valiosa nas discussões científicas.

À minha amiga Eliane Gomes, por sua orientação e clarificações sobre a revisão sistemática.

Ao Instituto Federal de Educação, Ciência e Tecnologia do Rio de Janeiro (IFRJ), por fornecer as ferramentas e o ambiente necessários para a realização desta pesquisa.

A todos vocês, expresso minha profunda gratidão!

“A mente que se abre a uma nova ideia jamais
volta ao seu tamanho original”

(Albert Einstein)

RESUMO

OLIVEIRA, Carlos Victor de. ***Blockchain para gestão de dados científicos: um olhar sobre Archangel e dARK***. Orientadores: Luis Fernando Sayão e Luana Farias Sales Marques. 2024. 172 f. Tese (Doutorado em Ciência da Informação) Universidade Federal do Rio de Janeiro, Escola de Comunicação, Instituto Brasileiro de Informação em Ciência e Tecnologia, Programa de Pós-Graduação em Ciência da Informação, Rio de Janeiro, 2024.

Esta pesquisa de doutorado investiga a aplicação da tecnologia *blockchain* na gestão e preservação de dados digitais de pesquisa científica, um desafio emergente no contexto da rápida evolução tecnológica e da expansão da era digital. Com a crescente produção de dados científicos e as limitações das tecnologias de armazenamento tradicionais, a *blockchain* se destaca por sua capacidade de oferecer registros descentralizados, imutáveis e seguros. Por meio de uma abordagem metodológica que inclui revisão sistemática da literatura e estudos de caso, como o Projeto Archangel e o Projeto dARK, esta tese explora a potencialidade da *blockchain* para superar os desafios atuais na preservação digital e na gestão de dados de pesquisa. O estudo introduz uma discussão sobre identificadores persistentes e fornece uma visão geral sobre o conceito de objeto digital FAIR. Contribui para o domínio da Ciência da Informação por meio da sugestão de modelos semânticos para a implementação prática da *blockchain*, abrindo novas perspectivas para o manejo eficiente de dados de pesquisa. As conclusões apontam para a relevância da tecnologia *blockchain* como uma solução promissora para os desafios contemporâneos de preservação e gestão de dados digitais, marcando um passo significativo em direção à integração efetiva desta tecnologia nas práticas de gestão de dados de pesquisa.

Palavras-chave: *Blockchain*, Gestão de Dados de Pesquisa, Preservação Digital, Ciência da Informação, Tecnologia da Informação.

ABSTRACT

OLIVEIRA, Carlos Victor de. ***Blockchain para gestão de dados científicos: um olhar sobre Archangel e dARK***. Orientadores: Luis Fernando Sayão e Luana Farias Sales Marques. 2024. 172 f. Tese (Doutorado em Ciência da Informação) Universidade Federal do Rio de Janeiro, Escola de Comunicação, Instituto Brasileiro de Informação em Ciência e Tecnologia, Programa de Pós-Graduação em Ciência da Informação, Rio de Janeiro, 2024.

This doctoral research investigates the application of *blockchain* technology in the management and preservation of digital research data, an emerging challenge in the context of rapid technological evolution and the expansion of the digital age. With the increasing production of scientific data and the limitations of traditional storage technologies, *blockchain* stands out for its ability to offer decentralized, immutable, and secure records. Through a methodological approach that includes a systematic literature review and case studies, such as the Archangel Project and the dARK Project, this thesis explores the potential of *blockchain* to overcome current challenges in digital preservation and research data management. The study introduces a discussion on persistent identifiers and provides an overview of the FAIR digital object concept. It contributes to the field of Information Science by suggesting semantic models for the practical implementation of *blockchain*, opening new perspectives for the efficient handling of research data. The conclusions point to the relevance of *blockchain* technology as a promising solution for contemporary challenges in the preservation and management of digital data, marking a significant step towards the effective integration of this technology into research data management practices.

Keywords: *Blockchain*, Research Data Management, Digital Preservation, Information Science, Information Technology.

LISTA DE FIGURAS

Figura 1 - Representação de um livro-razão centralizado e distribuído	25
Figura 2 - Cadeia de blocos	26
Figura 3 - Conceitos da tecnologia <i>blockchain</i>	31
Figura 4 - <i>Criptografias</i> de chave simétrica e assimétrica	34
Figura 5 - Benefícios da tecnologia <i>blockchain</i>	36
Figura 6 - Como funciona a <i>blockchain</i>	37
Figura 7 - O <i>hash</i> criptográfico	38
Figura 8 - Desafios da tecnologia <i>blockchain</i>	40
Figura 9 - Aplicações da <i>blockchain</i> na Ciência da Informação	47
Figura 10 - Síntese do capítulo <i>blockchain</i>	47
Figura 11 - Estratégias de preservação digital	53
Figura 12 - Repositórios digitais confiáveis	59
Figura 13 - Síntese do capítulo preservação digital	60
Figura 14 - Armazenamento e segurança de dados	67
Figura 15 - Compartilhamento e reuso de dados na pesquisa científica.....	69
Figura 16 – Tópicos apresentados no capítulo gestão de dados de pesquisa	72
Figura 17 - Características dos dados.....	75
Figura 18 - Correlação entre os conceitos da tecnologia <i>blockchain</i> e os da gestão de dados.....	82
Figura 19 - Modelo semântico para uso da <i>blockchain</i> na gestão de dados de pesquisa.....	85
Figura 20 - Modelo semântico prático para uso da <i>blockchain</i> na gestão de dados de pesquisa	87
Figura 21 - Protocolo da metodologia	88
Figura 22 - Protocolo de pesquisa inicial.....	93
Figura 23 - Protocolo de pesquisa da revisão sistemática	96
Figura 24 - Ciência da Informação, subáreas e áreas interdisciplinares	99
Figura 25 - Síntese da revisão sistemática.....	117
Figura 26 - A arquitetura da plataforma Archangel proposta.....	127
Figura 27 - Estudo de caso do projeto Archangel	132
Figura 28 - Estudo de caso do projeto dARK	150

LISTA DE QUADROS

Quadro 1 - Revisão comparativa entre <i>blockchain</i> e sistemas de bancos de dados convencionais.....	42
Quadro 2 - Características de proveniência, rastreabilidade, confiabilidade e autenticidade dos dados.....	74
Quadro 3 - Correlação entre os conceitos da tecnologia <i>blockchain</i> e os da gestão de dados.....	79
Quadro 4 - Percurso metodológico.....	91
Quadro 5 - Artigos selecionados na pesquisa inicial	94
Quadro 6 - Artigos selecionados para revisão sistemática.....	97
Quadro 7 - Artigos da revisão sistemática - autor/ano, título e questão/objetivo dos artigos.....	100
Quadro 8 - Identificadores persistentes URN, ARK, PURL, Handle System e DOI.....	135
Quadro 9 - Identificadores persistentes ORCID, ResearcherID e ScopusID	136
Quadro 10 - Vantagens dos FDOs	141
Quadro 11 - Características dos identificadores persistentes para análise e aplicabilidade na organização e acesso à informação científica	146
Quadro 12 - Proposta de trabalho do Projeto dARK	148

LISTA DE SIGLAS E ABREVIATURAS

AIP - Pacote de Informação de Arquivamento
AIS - Sistema de Identificação Automática
AMI - Arquivo e Instituição de Memória
ARK - Archival Resource Key
ARK - Archival Resource Key
CDL - California Digital Library
CODE - Centro para a Economia Digital
CVSSP - Centro de Visão por Computador e Processamento de Sinais Multimodais
DIP - Pacote de Informação de Disseminação
DLT - Distributed Ledger Technology
DOI - Digital Object Identifier
FAIR – Findable, Accessible, Interoperable, Reusable
FASTDB - Fast Array Storage and Transmission DataBase
FDO - Objeto Digital FAIR
GDP - Gestão de Dados de Pesquisa
GPS - Global Positioning System
HTML - HyperText Markup Language
HTMT - Heterotrait-Monotrait Ratio of Correlations
IA – Inteligência Artificial
IBICT - Instituto Brasileiro de Informação em Ciência e Tecnologia
IBM - International Business Machines
ID - Identificador Único
ISTA - Information Science & Technology Abstracts
LISTA - Library, Information Science & Technology Abstracts
OAIS - Open Archive Information Systems
ODI - Instituto de Dados Abertos
ORCID - Open Researcher and Contributor ID
PGD - Plano de Gestão de Dados
PID - Identificadores Persistentes
PPD - Política de Preservação Digital
PPDBN - Política de Preservação Digital da Biblioteca Nacional
PREMIS - Preservation Metadata: Implementation Strategies

PURL - Persistent Uniform Resource Locator

RDC - Rede de Documentação e Conservação de Arquivos

RDC - Repositórios Digitais Confiáveis

RSL - Revisão Sistemática da Literatura

SCOSS - Global Sustainability Coalition for Open Science Services

SIP - Pacote de Informação de Submissão

SWOT - Strengths, Weaknesses, Opportunities & Threats

TRAC - Trustworthy Repository Audit & Certification

URI - Uniform Resource Identifier

URL - Uniform Resource Locator

URN - Uniform Resource Name

SUMÁRIO

1 INTRODUÇÃO	15
1.1 O problema da pesquisa	19
1.2 Justificativa	21
1.3 Objetivos	23
1.3.1 Objetivo geral	23
1.3.2 Objetivos específicos	23
2 BLOCKCHAIN	24
2.1 Problemas que a tecnologia blockchain visa a resolver	27
2.2 Conceitos fundamentais da tecnologia blockchain	29
2.3 Benefícios da tecnologia blockchain.....	31
2.4 Como funciona a tecnologia blockchain: tecnologias e padrões subjacentes	36
2.5 Desafios em relação à tecnologia blockchain.....	39
2.6 Relações da tecnologia blockchain com outras tecnologias.....	41
2.7 Aplicações da tecnologia blockchain	43
2.8 Possíveis aplicações da tecnologia blockchain na área da Ciência da Informação	45
3 PRESERVAÇÃO DIGITAL	48
3.1 Arquitetura de objetos digitais	48
3.2 Modelo de referência OAIS	49
3.3 Estratégias de preservação digital.....	51
3.4 Metadados de preservação digital.....	53
3.5 Política e Plano de preservação digital.....	54
3.6 A confiança na preservação digital.....	56
3.7 Repositórios digitais confiáveis	57
4 GESTÃO DE DADOS DE PESQUISA	61
4.1 O planejamento da gestão de dados.....	62
4.2 A organização e documentação de dados	63
4.3 Armazenamento e segurança de dados.....	65
4.4 A Ética e a legalidade na gestão de dados de pesquisa	67
4.5 Compartilhamento e reuso de dados na pesquisa científica	68
4.6 A Preservação de dados na pesquisa científica	69
4.7 Ferramentas e recursos para a gestão eficiente de dados de pesquisa	71
4.8 Características de proveniência, rastreabilidade, confiabilidade e autenticidade dos dados de pesquisa.....	73

5 MODELOS SEMÂNTICOS	76
5.1 A escolha das ferramentas de representação	77
5.2 Correlação entre os conceitos da tecnologia blockchain e os da gestão de dados	78
5.3 Proposta de esquema de modelo semântico para uso da blockchain na gestão de dados de pesquisa	82
5.4 Proposta de esquema de modelo semântico prático para a implementação da blockchain na gestão dos dados de pesquisa	86
6 METODOLOGIA	88
6.1 Caracterização metodológica	89
6.2 Percorso metodológico.....	91
6.2.1 Mapeamento de domínio	92
6.2.2 Revisão bibliográfica da literatura	92
6.2.3 Revisão sistemática de literatura.....	95
7 A REVISÃO SISTEMÁTICA	98
8 PROJETO ARCHANGEL (ARQUIVO NACIONAL DO REINO UNIDO)	117
8.1 O estudo de caso do Projeto Archangel (Arquivo Nacional do Reino Unido)	118
9 PROJETO dARK (INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA E DA REDE LA REFERENCIA).....	132
9.1 O estudo de caso do Projeto dARK (Instituto Brasileiro de Informação em Ciência e Tecnologia e da Rede LA Referencia).....	134
10 CONSIDERAÇÕES FINAIS	151
REFERÊNCIAS.....	155
APÊNDICE – Memorial	164
ANEXO.....	166

1 INTRODUÇÃO

No atual contexto de rápida evolução tecnológica e expansão da era digital, a sociedade encara desafios sem precedentes concernentes à gestão e preservação de informações. Dentro desse cenário dinâmico, emerge a necessidade premente de explorar soluções inovadoras para a gestão e preservação de dados digitais de pesquisa científica, um imperativo vital para a contínua progressão da ciência.

Esta pesquisa de doutorado planeja verificar a possibilidade de aplicação da tecnologia *blockchain* (a qual é um livro-razão compartilhado que registra permanentemente suas transações anexando blocos) na gestão e preservação digital de dados de pesquisa (que são resultados de investigações metodológicas envolvendo observação, experimentação ou simulação nos processos científicos), considerando a crescente quantidade de dados gerados pelas pesquisas científicas e reconhecendo as limitações das tecnologias tradicionais de armazenamento e gestão.

Nesse sentido a tecnologia *blockchain*, conhecida por sua natureza descentralizada, imutabilidade e segurança, emerge como uma promissora solução para enfrentar as preocupações inerentes à gestão e preservação de dados digitais de pesquisa científica.

As transformações ocorridas nos últimos séculos nos padrões e dinâmicas da sociedade trouxeram consigo a comodidade da vida moderna, mas também a necessidade de enfrentar a avalanche de informações inerentes à pesquisa científica.

Surge, portanto, a necessidade premente de encontrar abordagens eficazes para a gestão dos dados de pesquisa nos bancos de dados de informação científica.

Nesse contexto, emergem fenômenos informacionais contemporâneos, como a tecnologia *blockchain*, cuja inovação tecnológica está intrinsecamente vinculada às novas demandas, impactando, ademais, a própria prática científica.

A preocupação é a gestão desta informação, conceito que, segundo Saracevic (1996), está no cerne da Ciência da Informação desde suas origens. Segundo o autor, a Ciência da Informação nasceu em um cenário que inclui, por um lado, a produção e

organização da informação e, por outro, a transmissão da informação mediada pela tecnologia visando preservá-la e utilizá-la.

O próprio Saracevic (1996) já delineava a ideia de que a Ciência da Informação está inexoravelmente ligada à tecnologia da informação, definindo-a como:

Um campo dedicado às questões científicas e à prática profissional voltadas para os problemas da efetiva comunicação do conhecimento e de seus registros entre os seres humanos, no contexto social, institucional ou individual do uso e das necessidades de informação. No tratamento destas questões são consideradas de particular interesse as vantagens das modernas tecnologias informacionais (Saracevic, 1996, p. 47).

Além disso, a definição de Capurro e Hjørland apoia esta visão e afirma que a

Ciência da Informação se ocupa com a geração, coleta, organização, interpretação, armazenamento, recuperação, disseminação, transformação e uso da informação, destacando particular ênfase na aplicação de tecnologias modernas nesses domínios (Capurro; Hjørland, 2007, p. 186).

No contexto da Ciência da Informação, as mudanças sociais e tecnológicas redefinem os desafios relacionados à comunicação, organização e gestão da informação. A definição de Saracevic (1996) enfatiza a natureza da Ciência da Informação na preservação e transferência do conhecimento, reforçando a sua inerente ligação com a tecnologia.

De maneira complementar, Capurro e Hjørland (2007) realçam o papel das tecnologias modernas na geração, armazenamento e disseminação da informação. Entretanto, uma inovação contemporânea, como a tecnologia *blockchain*, demanda uma análise mais profunda sobre como ela pode abordar os desafios do armazenamento e gestão de dados científicos (os quais são sustentados por documentos acadêmicos que incluem dissertações, teses, artigos, patentes, entre outros trabalhos de pesquisa).

Dentro dessas atuais inovações, notadamente a tecnologia *blockchain*, preenche um papel crucial ao enfrentar os desafios intrínsecos à Ciência da Informação. Pesquisamos que a tecnologia *blockchain*, com sua capacidade de criar

registros imutáveis e descentralizados, oferece soluções concretas para a gestão e preservação de informações científicas.

Ao investigar a interação entre essas inovações tecnológicas e os princípios e objetivos da Ciência da Informação, é possível ampliar o entendimento sobre como a disciplina pode eficazmente enfrentar as complexidades da era digital.

Esta pesquisa traz uma visão geral da tecnologia *blockchain*, dos problemas que a *blockchain* planeja resolver, conceitos envolvidos, benefícios, como funciona, relações com outras tecnologias, aplicações e possíveis aplicações na área da Ciência da Informação e da gestão e preservação digital.

Além disso, por meio da análise conceitual da tecnologia *blockchain*, do delineamento de sua trajetória, aplicações e possíveis novos usos relacionados à gestão e preservação digital de informações científicas, esta pesquisa poderá se tornar uma contribuição para o campo teórico da Ciência da Informação.

Esta tese de doutorado aborda a interseção inovadora entre a tecnologia *blockchain* e a preservação digital, especificamente no contexto da gestão de dados de pesquisa. A estrutura da tese é delineada em capítulos que progressivamente constroem um argumento compreensivo, metodológico e prático para a utilização da *blockchain* nesta área emergente.

O primeiro capítulo estabelece o cenário para a pesquisa, apresentando o tema e sublinhando sua relevância no contexto digital contemporâneo da Ciência da Informação. Este segmento expõe o problema da pesquisa, detalhando o desafio central que esta tese visa abordar, e lança as bases para a investigação subsequente. A necessidade do estudo é defendida ao destacar a importância do trabalho, ressaltando o papel significativo que a tecnologia *blockchain* pode desempenhar na preservação digital e na gestão de dados de pesquisa. A seção de objetivos é dividida entre o objetivo principal e os específicos, delineando claramente as metas que se pretende atingir.

A fundamentação teórica inicia-se no capítulo 2, onde exploramos os fundamentos e princípios desta tecnologia disruptiva, cuja aplicação transcendeu o campo das criptomoedas, influenciando diversas áreas, incluindo a gestão de dados de pesquisa. Em seguida, o capítulo 3 aborda a necessidade crescente de preservar conteúdos digitais de maneira segura, confiável e acessível ao longo do tempo, incluindo uma discussão sobre repositórios confiáveis, essenciais para a compreensão da aplicação da *blockchain* nessa área.

O capítulo 4 foca em um aspecto crucial da Ciência da Informação: a organização, armazenamento, preservação e compartilhamento de dados gerados durante atividades de pesquisa. Destaca as características de proveniência, rastreabilidade, confiabilidade e autenticidade, elementos essenciais para a gestão eficaz dos dados.

No capítulo 5, abordamos os modelos semânticos e as ferramentas de representação, focando na correlação entre os conceitos da tecnologia *blockchain* e os da gestão de dados. Propomos esquemas de modelos semânticos para o uso e aplicação prática da *blockchain* nesse domínio.

A metodologia é exposta no capítulo 6, detalhando a abordagem metodológica adotada na pesquisa, delineando os procedimentos e técnicas utilizados para investigar a aplicação da tecnologia *blockchain* na gestão e preservação de dados digitais de pesquisa científica.

No capítulo 7, apresentamos uma revisão sistemática da literatura sobre as tendências atuais do uso da tecnologia *blockchain* na gestão de dados de pesquisa.

Os estudos de caso do Projeto Archangel e do Projeto dARK são apresentados nos capítulos 8 e 9, respectivamente, oferecendo uma visão prática das aplicações da *blockchain* na gestão de dados de pesquisa. Inclui também uma discussão sobre identificadores persistentes e uma introdução ao conceito de objeto digital FAIR.

A tese culmina no capítulo 10, onde sintetizamos os achados da pesquisa, refletimos sobre as implicações teóricas e práticas do estudo, e apontamos direções futuras para a pesquisa nessa área promissora.

Cada capítulo desta tese não apenas avança nosso entendimento sobre a aplicação potencial da tecnologia *blockchain* na preservação digital e gestão de dados de pesquisa, mas também contribui para o campo emergente de estudo com percepções teóricas e práticas. Este trabalho espera inspirar futuras investigações e aplicações práticas, marcando um passo significativo em direção à efetiva integração da tecnologia *blockchain* em práticas de preservação digital e gestão de dados de pesquisa. Sendo assim, apresentamos a seguir o problema da pesquisa.

1.1 O problema da pesquisa

Nos últimos anos, os avanços em vários dispositivos digitais e serviços de computação em nuvem geraram enormes quantidades de dados (Mariotto *et al.*, 2021). No entanto, uma gestão de dados efetiva que possibilite o acesso contínuo a objetos digitais, produzidos, selecionados e arquivados por instituições diversas enfrenta como desafio a obsolescência tecnológica, que “é uma consequência da evolução das tecnologias, e que os documentos digitais estão condicionados aos seus efeitos” (Santos; Flores, 2017, p. 29).

Para Alzahrani *et al.* (2022) o armazenamento de dados em ambientes de nuvem é um grande desafio para questões de confiabilidade e confiança de dados.

Nesse contexto, a aplicação da tecnologia *blockchain* na esfera da Ciência da Informação propõe uma visão abrangente das possibilidades que essa tecnologia traz para a gestão da informação nos contextos da preservação digital de dados de pesquisa.

A *blockchain*, ao contrário dos bancos de dados na nuvem, garante descentralização, imutabilidade e segurança dos dados, usando criptografia especial, enquanto dados na nuvem são mutáveis (Altamira, 2024).

Nesse cenário, a questão problematizadora desta pesquisa é: **Como a tecnologia *blockchain* pode fornecer uma abordagem confiável na gestão e preservação de dados digitais de pesquisa científica?**

Para responder a essa questão, esta investigação se alinhará à abordagem interdisciplinar adotada na Ciência da Informação, conforme delineado por Capurro e Hjørland (2007), proporcionando um meio de enfrentar desafios multifacetados relacionados à tecnologia e preservação da informação.

A abordagem interdisciplinar na Ciência da Informação, conforme descrita por Capurro e Hjørland (2007), destaca a importância de enfrentar desafios multifacetados na tecnologia e preservação da informação por meio da integração de diferentes campos do saber. Esta visão alinha-se à análise de Pinheiro (2006), que argumenta sobre a especialização do conhecimento em áreas específicas da ciência, enfatizando que cada uma delas possui um enfoque particular com métodos e objetos de estudo próprios.

No entanto, Pinheiro (2006) também ressalta a necessidade de ampliar esses ângulos de visão para fomentar o desenvolvimento científico, sugerindo que a interconexão entre diferentes áreas de conhecimento é crucial.

Em uma visão mais atual, Paletta e Silva (2020) destacam que

No contexto da Ciência da Informação, Biblioteconomia, Museologia, Arquivologia e áreas afins, os avanços impostos pela tecnologia da informação e das comunicações estão a provocar profundas mudanças nos entendimentos sobre as técnicas e procedimentos da organização e gestão da informação e dos recursos de tecnologia da informação em Repositórios e Bibliotecas Digitais (Paletta; Silva, 2020, p. 10).

Essa perspectiva é reforçada pela definição de interdisciplinaridade de Pombo (2004), que a descreve como uma combinação de duas ou mais disciplinas para compreender um objeto de estudo a partir de diversos pontos de vista, culminando em uma síntese comum. Portanto, estudos e medições da interdisciplinaridade oferecem uma contribuição significativa para o entendimento de como um campo de conhecimento se desenvolve e se transforma, com base nas relações epistemológicas entre diferentes áreas (Pinheiro, 2018). Essa abordagem é fundamental na Ciência da Informação ao permitir a integração de conhecimentos e métodos de diferentes disciplinas para enfrentar os complexos desafios presentes na gestão, tecnologia e preservação da informação.

O cerne desta investigação reside na suposição de que a adoção da tecnologia *blockchain* em sistemas de gestão de dados digitais de pesquisa científica desempenha um papel crucial no aprimoramento da segurança, da integridade e da capacidade de rastreamento dos dados. Esta inovação tecnológica introduz uma forma descentralizada e imutável de registrar informações, assegurando a proteção dos dados de pesquisa. Ademais, a *blockchain* possibilita estabelecer um registro transparente e auditável das modificações feitas nos dados, elemento fundamental para manter a integridade dos dados ao longo do tempo e promover a cooperação e o compartilhamento de informações entre os pesquisadores.

Apresentamos a seguir a justificativa da aplicação da tecnologia *blockchain* na área da Ciência da Informação, especificamente na gestão e preservação digital dos dados de pesquisa.

1.2 Justificativa

Na era contemporânea, a informação, potencializada pela digitalização e interconexão global, assume um papel estratégico para o avanço científico, tecnológico e social. De acordo com Santos e Flores (2017), além de implementar estratégias para migrar, emular, encapsular, refrescar e preservar a tecnologia, é preciso também “que os documentos digitais sejam armazenados em sistemas informatizados de gestão e preservação” (Santos; Flores, 2017, p. 34).

Neste cenário, a Ciência da Informação e a tecnologia da informação são fundamentais para gerir esse novo paradigma informacional. A rápida evolução tecnológica proporciona uma convergência única entre domínios acadêmicos e científicos, transcendendo as fronteiras tradicionais entre disciplinas.

Complementando, Santos e Flores (2017), explicam que os sistemas de gerenciamento e preservação têm a função de garantir a autenticidade dos documentos digitais, sendo assim necessário devido à sua vulnerabilidade inerente em termos de facilidade de adulteração.

Emergem tecnologias disruptivas como a *blockchain*, promovendo novas formas de garantir a confidencialidade e a segurança dos dados em um ambiente digital complexo. Afirma Oliveira (2023) que o uso dessa tecnologia garante amplamente a confidencialidade e a segurança dos dados transacionados de forma distribuída.

Este contexto de inovação encoraja a exploração da interseção entre a Ciência da Informação e a tecnologia da informação, principalmente no que concerne à aplicação da *blockchain* para contribuir para a gestão e preservação da informação científica.

O Projeto Archangel do Arquivo Nacional do Reino Unido exemplifica a aplicação bem-sucedida da *blockchain* para garantir a autenticidade de registros digitais. Esse projeto demonstra o potencial da *blockchain*, especificamente a *distributed ledger technology* (DLT) de Green *et al.* (2018), para verificar a integridade e proveniência de registros digitais, impactando diretamente a confiabilidade dos arquivos digitais. Da mesma forma, o Projeto dARK, atualmente em andamento, representa um avanço significativo nessa área, buscando expandir os limites da tecnologia *blockchain* para a gestão de registros digitais em uma escala ainda maior. Com um foco particular em aprimorar as metodologias de segurança e eficiência na

verificação de dados, o Projeto dARK visa fornecer uma nova camada de confiabilidade e acessibilidade para o futuro dos arquivos digitais, complementando e estendendo os esforços iniciados pelo Projeto Archangel.

Além de sua associação inicial com criptomoedas, a tecnologia *blockchain* possui diversas potencialidades. Segundo Tapscott e Tapscott (2016), essa nova maneira de lidar com o armazenamento de dados e de transações financeiras consegue ser programada para guardar virtualmente praticamente qualquer coisa que tenha valor e relevância para a humanidade.

Esta pesquisa pode ainda ajudar na compreensão da aplicação da tecnologia *blockchain* na Ciência da Informação e fornecer informações valiosas para as comunidades acadêmicas e profissionais, considerando o interesse quanto às vantagens das modernas tecnologias (Saracevic, 1996). Ao demonstrar os benefícios práticos da *blockchain* na gestão de dados científicos, esta investigação pode influenciar a adoção de novas práticas e tecnologias no campo, tendo um impacto positivo na forma como os dados de investigação são geridos, compartilhados e preservados.

Isto não só enriquecerá o campo acadêmico, mas também permitirá que profissionais e instituições adotem estratégias mais eficazes para gerenciar e preservar a informação científica no atual cenário de crescimento exponencial de dados.

Por fim, para o Programa de Pós-Graduação em Ciência da Informação, a exploração da *blockchain* é relevante devido ao seu potencial inovador na gestão da informação. A pesquisa proposta desempenha um papel crucial na concretização dos objetivos do programa, ao combinar os princípios interdisciplinares da Ciência da Informação com as inovações tecnológicas da tecnologia da informação, visando a aprimorar as práticas de preservação digital, gestão de dados e colaboração entre pesquisadores. A seguir destacamos o objetivo geral e os específicos dessa pesquisa.

1.3 Objetivos

1.3.1 Objetivo geral

Diante do problema colocado e sua justificativa, o objetivo geral desta pesquisa é **explorar como a tecnologia *blockchain* pode fornecer uma abordagem confiável na gestão e preservação de dados digitais de pesquisa científica.**

1.3.2 Objetivos específicos

- Investigar a tecnologia *blockchain* desde a sua origem, conceitos, benefícios e sua relação com outras tecnologias;
- Mapear e analisar as tendências emergentes e consolidadas na aplicação da *blockchain* para a gestão de dados de pesquisa;
- Apresentar possíveis aplicações da *blockchain* na Ciência da Informação, com ênfase em sua utilização para gestão e preservação digital de dados científicos.

No capítulo 2, iniciamos a fundamentação teórica sobre a tecnologia *blockchain*. Este capítulo visa proporcionar uma compreensão básica desta tecnologia, abordando sua origem, conceitos fundamentais, benefícios, relação com outras tecnologias, desafios e possíveis aplicações na área da Ciência da Informação.

2 BLOCKCHAIN

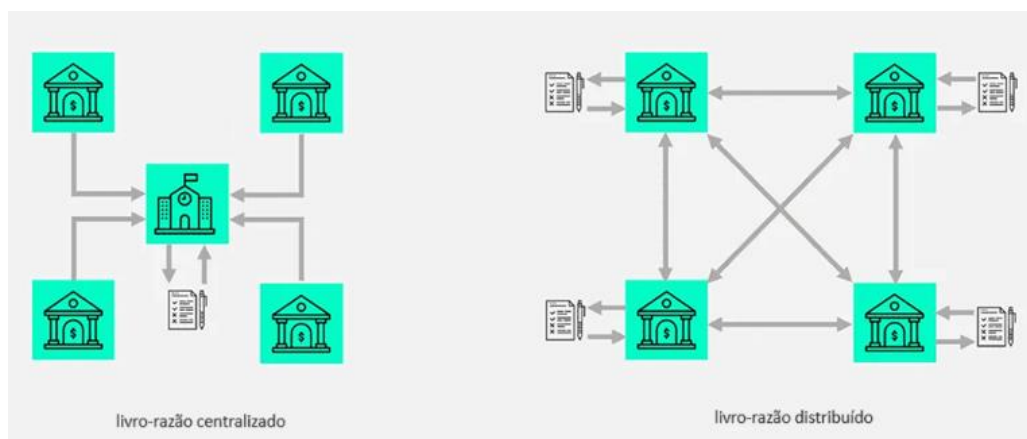
A tecnologia *blockchain*, uma inovação paradigmática originada no âmbito das criptomoedas, transpõe as fronteiras inicialmente previstas, ou seja, ela não só tem potencial para desencadear inúmeras novas aplicações, mas também tem a capacidade iminente de transformar muitas coisas (Tapscott; Tapscott, 2016).

Inicialmente concebida como a estrutura subjacente ao *Bitcoin* por uma entidade anônima conhecida como Satoshi Nakamoto (2008), a *blockchain* passou por uma evolução significativa, catalisando uma era de transformações digitais com sua estrutura descentralizada, sua natureza imutável e sua transparência. Estas propriedades fazem da *blockchain* um sistema confiável para a realização de transações e para o armazenamento de dados, eliminando a necessidade de intermediários (Mougayar, 2016).

O núcleo operacional da *blockchain* reside em um livro-razão (onde os dados são armazenados em uma rede de nós descentralizados e quando um bloco de informações é construído, ele é registrado em uma espécie de livro, e uma cópia dele é compartilhada com todos os membros da rede) digital de transações, distribuído entre vários nós participantes.

Para Swan (2015), a *blockchain* é o livro-razão que serve como registro em blocos de todas as transações já executadas em um determinado domínio. A integridade linear da informação é assegurada por processos *criptográficos* que vinculam cada bloco ao anterior na cadeia, tornando as transações praticamente imutáveis uma vez registradas e validadas pela rede. A figura 1 demonstra a representação de um livro-razão centralizado de outro livro-razão distribuído.

Figura 1 - Representação de um livro-razão centralizado e distribuído



Fonte: Bit2Me Academy¹, 2019, tradução nossa.

A relevância desta investigação é amplificada pela promessa da *blockchain* de renovar setores críticos, incluindo finanças, saúde, logística, governança, entre outros. Assim, antecipa-se que a tecnologia *blockchain* transforme radicalmente as práticas no comércio, na indústria e na educação (Chen *et al.*, 2018), bem como tenha impactos significativos na ciência, tecnologia e, particularmente, na Ciência da Informação. Esta tecnologia pode ser considerada um marco decisivo na comunicação científica e na gestão e preservação de dados de pesquisa científica.

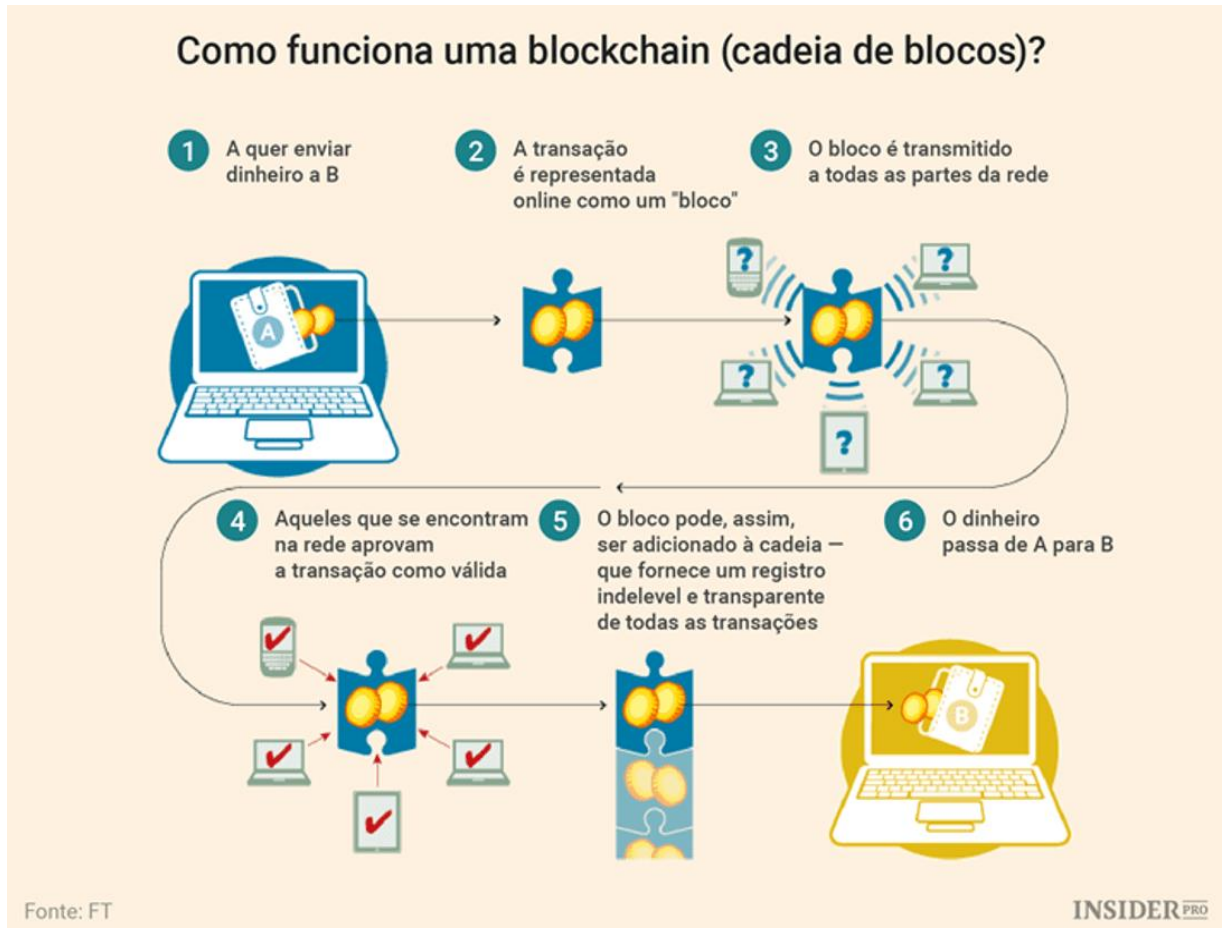
A tecnologia *blockchain*, emergindo como um avanço revolucionário, captura um interesse considerável tanto no cenário tecnológico quanto na academia nos últimos anos. Segundo Chen *et al.* (2018) esta tecnologia disruptiva terá um impacto significativo na governança nacional, nas funções institucionais, nos negócios, operações, educação e nossa vida diária no século XXI, representando muito mais do que uma mera inovação tecnológica.

A *blockchain* é uma estrutura de dados distribuída que consiste em uma cadeia de blocos interligados. Cada bloco (onde os registros são inseridos) contém um conjunto de transações, validadas por meio de um processo de consenso (implementado quando todos os participantes da rede concordam com a validade de uma transação, garantindo que seus livros-razão sejam cópias exatas uns dos outros) entre os participantes da rede (Hwang; Shon; Park, 2020). Esta validação ocorre de

¹ <https://academy.bit2me.com/pt/qual-%C3%A9-o-livro-distribu%C3%ADdo-do-livro/> acessado em 02/12/2023.

maneira descentralizada, eliminando a necessidade de uma autoridade central. A figura 2 exemplifica como funciona a cadeia de blocos.

Figura 2 - Cadeia de blocos



Fonte: Ihodl, 2017.

O conceito de *blockchain* foi introduzido por uma entidade anônima sob o pseudônimo Satoshi Nakamoto em 2008, como um “sistema de dinheiro eletrônico ponto a ponto”, (é um sistema que funciona como uma organização coletiva, permitindo que cada indivíduo interaja diretamente com os outros) em que transações são registradas de forma transparente e imutável. Este foi um marco que proporcionou um registro público e seguro de todas as transações, incutindo uma camada adicional de confiança em ambientes digitais (Nakamoto, 2008).

A narrativa histórica da *blockchain* vai além do seu criador misterioso e do advento do *bitcoin*. Suas raízes podem ser traçadas a partir de tecnologias e conceitos

anteriores, como os livros-razão de contabilidade distribuídos, que formaram a base para o desenvolvimento subsequente da *blockchain* (Mori *et al.*, 2020).

Originalmente concebida para sustentar o *bitcoin*, a *blockchain* capturou rapidamente a atenção pela sua capacidade de proporcionar segurança e transparência em um mundo cada vez mais digitalizado. A história da *blockchain* é marcada por mitos e lendas, incluindo a identidade misteriosa de Nakamoto e os primeiros mineradores de *bitcoin*, fundamentais na validação de transações e manutenção da rede.

A capacidade da *blockchain* de transformar uma ampla gama de setores é notável. Conforme observado por Tapscott e Tapscott (2016), a *blockchain* é um facilitador-chave deste novo sistema financeiro, permitindo a transferência segura e direta de ativos digitais, sem a necessidade de intermediários. Esta tecnologia tem implicações profundas para a eficiência e a transparência nas transações comerciais e financeiras.

A *blockchain* também detém a promessa de criar sistemas de gerenciamento de conteúdo descentralizados e plataformas de publicação acadêmica, garantindo a transparência e a autenticidade das informações científicas, oferecendo processos de publicação mais rápidos e eficientes, aumentando a acessibilidade e reduzindo o potencial de fraude (Skala *et al.*, 2023). Ela tem o potencial de melhorar a rastreabilidade das pesquisas e aumentar a confiança no processo de revisão por pares.

A tecnologia *blockchain* sinaliza uma mudança substancial na maneira como a informação é gerenciada, compartilhada e protegida. No tópico a seguir, serão explorados os problemas que a tecnologia *blockchain* aspira a resolver.

2.1 Problemas que a tecnologia blockchain visa a resolver

A expansão exponencial do volume de transações em escala global apresenta desafios crescentes relacionados à complexidade, vulnerabilidade, ineficiência, riscos e custos associados aos sistemas de transações tradicionais. Para Williams (2023) a *blockchain*, reconhecida como a tecnologia mais disruptiva do século XXI, ultrapassa seu papel amplamente conhecido nas criptomoedas, aplicada de maneiras inovadoras em diversos outros setores. Exemplificando, “os direitos autorais de um artista musical, *royalties* e informações de licenciamento podem ser armazenadas usando

contratos inteligentes, uma tecnologia apoiada por *blockchain*" (Williams, 2023, p. 213).

A complexidade das transações on-line torna-se evidente com o aumento constante do volume de operações, sobrecarregando os sistemas existentes e, como apontado pela IBM (2017), o volume de transações e o número de interações entre sistemas estão crescendo de maneira exponencial. Essa complexidade pode resultar em atrasos, erros e custos adicionais, refletindo a necessidade de sistemas mais eficientes e seguros.

De acordo com Washington Segundo *et al.* (2023), é comum que sistemas computacionais e aplicativos funcionem baseados na ideia de que uma entidade central é responsável pela gestão dos dados e/ou funcionalidades desses sistemas, resultando em uma estrutura de gestão centralizada, a exemplo de um banco. Entretanto, essa estratégia apresenta como inconveniente a formação de um ponto único de falha.

Dessa forma, os sistemas descentralizados para Washington Segundo *et al.* (2023)

Surgiram para mitigar as limitações dos sistemas centralizados. Em vez de usar uma autoridade central para executar e orquestrar operações, essas operações são realizadas pelos nós de uma rede descentralizada. Nesse cenário, o controle do sistema é igualmente compartilhado entre os nós da rede, usando protocolos criptográficos e de segurança (Washington Segundo *et al.*, 2023, p. 152).

O ataque de 51%, por exemplo, é um tipo de ataque que pode ocorrer em sistemas, *blockchain* que usam o algoritmo de consenso *Proof of Work* (PoW). Esse ataque ocorre quando um único minerador ou grupo de mineradores controla mais de 50% da capacidade total de mineração da rede. Isso lhes permite obter o controle da *blockchain* e realizar várias atividades maliciosas.

A *blockchain* propõe uma solução para essas vulnerabilidades. Segundo Bell *et al.* (2019), o princípio fundamental da *blockchain* é que seria necessário pelo menos 51% do poder de computação para reescrever códigos na cadeia principal de registros, fortalecendo a segurança e integridade dos dados.

A tecnologia *blockchain*, conforme discutido por Lemieux (2017), é uma resposta promissora para os desafios apresentados pelos sistemas de registros atuais. Por meio de um livro-razão distribuído, a *blockchain* desloca a necessidade de intermediários e centralização. Como destacado por Gangadevi e Renuka Devi (2021,

p. 4, tradução nossa), a “*blockchain* é um livro-razão público digital descentralizado para armazenamento distribuído e comunicação ponto a ponto entre nós”, fomentando a eficiência, transparência e confiabilidade nas transações.

Nakamoto (2008) propôs um sistema descentralizado que utiliza criptografia e redes de computadores distribuídas para realizar transações digitais sem intermediários, como, por exemplo, os bancos.

A tecnologia *blockchain* não apenas revoluciona a maneira como as transações são conduzidas, mas também propõe reformar a gestão de informações e registros, tornando-os mais seguros e eficientes em diversos setores. No tópico seguinte, serão explorados os conceitos fundamentais da tecnologia *blockchain*.

2.2 Conceitos fundamentais da tecnologia *blockchain*

A tecnologia *blockchain*, ao revolucionar a gestão da informação, introduz conceitos cruciais que facilitam transações seguras e eficazes de dados. *Blockchains* podem ser integradas com os sistemas de gestão de informação para armazenamento de registros vitais seguros e imutáveis (Akoramurthy *et al.*, 2022).

O consenso é um pilar central da *blockchain*, sendo o processo que permite aos participantes da rede concordarem sobre a validade das transações e o estado atual do livro-razão. Como elucidado por Narayanan *et al.* (2016), cada nó na rede detém uma cópia do livro-razão e verifica as transações conforme as regras do protocolo *blockchain*, realizando essa concordância de forma descentralizada, eliminando intermediários e fortalecendo a confiabilidade do sistema.

Woodal e Ringel (2019) destacam a natureza distribuída da *blockchain*, em que os registros são mantidos em uma rede *peer-to-peer*, sem a dominância de uma única entidade. Esta distribuição facilita a criação de um sistema de registros resiliente, que, conforme explicado por Tapscott e Tapscott (2016), mantém um registro imutável e público de todas as transações na rede, proporcionando uma trilha de auditoria confiável, um dos pontos-chave da tecnologia *blockchain*.

A imutabilidade, um conceito central sublinhado por Gangadevi e Renuga Devi (2021), significa que as transações, uma vez registradas, permanecem inalteradas, garantindo a integridade dos dados. Esta característica é reforçada por Feng *et al.* (2019), que destacam a resiliência dos registros contra alterações mal-intencionadas devido à verificação coletiva realizada pelos nós.

O compartilhamento do livro-razão é outro aspecto essencial, em que o livro-razão é acessível a todos os participantes da rede, mantendo a transparência e a integridade dos dados (Mougayar, 2016). Além disso, a proveniência, destacada por Swan (2015), permite o rastreamento da origem e do histórico das transações, uma funcionalidade crucial na gestão eficaz da informação.

A confiança é estabelecida por meio da descentralização e imutabilidade, criando um ambiente onde os participantes podem confiar nas transações e registros (SWAN, 2015). Este ambiente de confiança é sustentado pelo algoritmo de consenso, que, conforme explicado por Gangadevi e Renuga Devi (2021), assegura a precisão e a confiabilidade das informações armazenadas na rede.

A sustentabilidade da *blockchain* é alcançada por meio da descentralização e distribuição da rede, tornando-a resistente a falhas e ataques (Androulaki *et al.*, 2018). A confidencialidade também é abordada por Mougayar (2016), explicando como *blockchains* permissivas podem controlar o acesso aos dados, garantindo que apenas partes autorizadas tenham acesso a informações sensíveis.

As assinaturas digitais (são códigos digitais gerados por *criptografia* de chave pública, sendo anexado a documentos transmitidos eletronicamente para verificar seu conteúdo e a identidade do remetente. É um mecanismo matemático que permite a alguém provar que é o proprietário) e multiassinaturas são instrumentos cruciais para a confirmação de identidade e integridade dos dados (Bralić; Stančić; Stengård, 2020; Narayanan *et al.*, 2016). Lemieux (2017) reitera a importância das assinaturas digitais na autenticação de mensagens ou documentos digitais, solidificando ainda mais a segurança e a governança nas transações *blockchain*.

Em suma, este tópico delineou os conceitos fundamentais da tecnologia *blockchain* que incluem consenso, sistemas de registros distribuídos, compartilhamento do livro-razão, proveniência, imutabilidade, confiança, sustentabilidade, confidencialidade e assinaturas digitais. A exploração subsequente se concentrará nos benefícios derivados da implementação da tecnologia *blockchain* em diversos domínios. Na figura 3 apresentamos um esquema com os conceitos discutidos nesse tópico.

Figura 3 - Conceitos da tecnologia *blockchain*



Fonte: Elaborado pelo autor.

No próximo tópico, serão explorados os benefícios da tecnologia *blockchain*.

2.3 Benefícios da tecnologia *blockchain*

Ao explorar conceitos cruciais como confiança, verificabilidade, privacidade e transparência, este tópico delinea como a *blockchain* está remodelando nossa interação com dados em âmbitos acadêmicos e profissionais.

De acordo com Oliveira (2023), a *blockchain* no meio científico

Vem de encontro às necessidades e aspectos críticos da comunicação da ciência (Scholarly Communication), e pesquisa científica, o que inclui transparência, confiança, reprodutibilidade e crédito, resolvendo alguns dos problemas como custos, abertura e acessibilidade universal a informações científicas, oferecendo a geração descentralizada e autorregulada de dados,

a partir de uma infraestrutura compartilhada, onde todas as transações são salvas e armazenadas (Oliveira, 2023, p. 10 e 11).

Destacando as potencialidades das soluções baseadas em *blockchain*, a necessidade emergente de adotar a tecnologia *blockchain* para a preservação segura de registros eletrônicos é evidente. Os profissionais necessitam de confiança na credibilidade desta tecnologia para registros eletrônicos (Deng *et al.*, 2019).

Ren *et al.* (2021) elucidam que a continuidade digital, como um método de gestão e preservação de informação digital, reforça a gestão e preservação dos registros eletrônicos originais, atendendo às exigências de instituições, governos e comunidades, assegurando a disponibilidade necessária da informação digital.

A verificabilidade, sendo um benefício substancial da *blockchain*, permite a qualquer participante da rede validar a autenticidade de uma transação. Tapscott e Tapscott (2016) salientam que a *blockchain* proporciona uma trilha de auditoria completa, possibilitando a verificação de cada transação por qualquer pessoa autorizada, uma relevância notável em contextos acadêmicos que valorizam a autenticidade das informações.

A *blockchain* incorpora selos de tempo em cada transação, registrando sua ocorrência. Swan (2015) observa que os selos de tempo fornecem uma evidência irrefutável da ordem das transações, prevenindo manipulação retroativa dos registros, um aspecto crucial para a gestão de informações com histórico cronológico, como documentos acadêmicos.

Lemieux (2017) acrescenta que cada bloco na *blockchain* contém uma lista de transações e um carimbo de data/hora, entre outras informações. O *timestamps* ou carimbo de data/hora é vital, ao estabelecer a posição de uma transação ou registro em uma lista cronológica. Gangadevi e Renuga Devi (2021) corroboram que a funcionalidade de *timestamps* (é uma cadeia de caracteres denotando a hora ou data que certo evento ocorreu) na *blockchain* é instrumental para reconhecer e armazenar cada transação, auxiliando o nó participante a preservar a ordem das transações para dados rastreáveis.

A *blockchain* permite a configuração de níveis de permissão e privacidade, controlando o acesso e validação de transações. Segundo Mougayar (2016), isso capacita o compartilhamento de informações confidenciais apenas com partes autorizadas, preservando a privacidade quando necessário, mediante técnicas de *criptografia* (a qual é a codificação de uma mensagem para evitar que outras pessoas

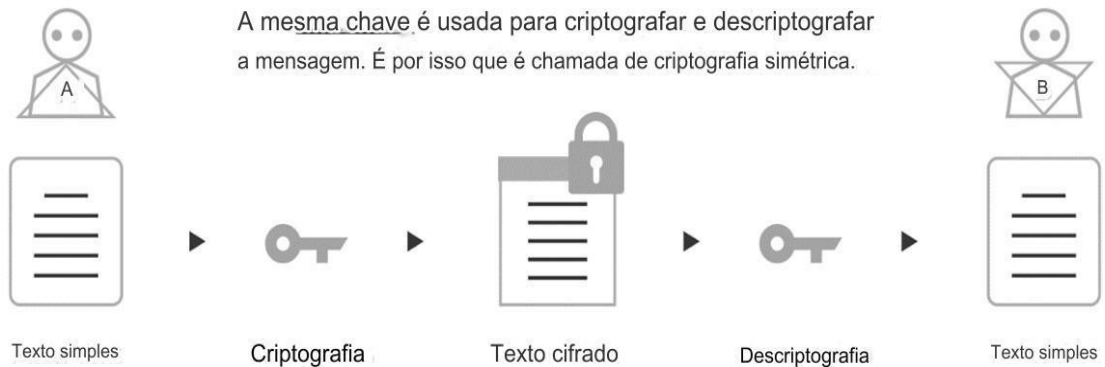
possam decifrar seu conteúdo) para proteger os dados. Para compreender melhor o compartilhamento seguro, é essencial conhecer a *criptografia* simétrica, que utiliza a mesma chave para *criptografar* e *descriptografar* a informação, e a *criptografia* assimétrica, que emprega uma chave privada para *descriptografia* e uma chave pública para *criptografia*.

Swan (2015) também enfatiza esse impacto significativo da *blockchain* nas redes de negócios, facilitando a cooperação e compartilhamento de informações entre instituições acadêmicas e empresas.

A figura 4 descreve os conceitos fundamentais por trás da *criptografia* de chave simétrica e assimétrica. Na *criptografia* simétrica, a mesma chave é utilizada tanto para *criptografar* quanto para *descriptografar* a mensagem, o que pode ser um desafio para a comunicação segura em larga escala devido à necessidade de compartilhar a chave seguramente. Já na *criptografia* assimétrica, são utilizadas duas chaves diferentes: uma chave privada para criar uma assinatura digital que não pode ser forjada e uma chave pública que qualquer pessoa pode usar para verificar a autenticidade da assinatura. Isso resolve o problema de como compartilhar chaves seguramente, pois a chave pública pode ser distribuída livremente, enquanto a chave privada permanece segura com o usuário.

Figura 4 - *Criptografias* de chave simétrica e assimétrica

Criptografia de chave simétrica



Para comunicar-se em particular com alguém, é necessário comunicar a chave por meio de reuniões secretas, envelopes lacrados ou correios confiáveis. Não escala bem.

Criptografia de chave assimétrica



Resolveu o problema de coordenação introduzindo duas chaves, uma chave pública e uma chave privada. A chave privada só é conhecida pelo proprietário e precisa ser mantida privada, mas a chave pública pode ser dada a qualquer pessoa. Os remetentes podem combinar uma mensagem com sua chave privada para criar uma assinatura digital na mensagem. Qualquer pessoa com a chave pública correspondente pode agora verificar se a assinatura é válida.

Narayanan *et al.* (2016) evidenciam que a característica distintiva da *blockchain* é a eliminação de intermediários, permitindo transações diretas entre pares, o que diminui custos de transação e aumenta a eficiência, ao passo que simplifica e acelera a troca de informações.

A *blockchain*, por sua natureza imutável e consenso descentralizado, atenua riscos e fraudes. Androulaki *et al.* (2018) ressaltam que a imutabilidade dos registros torna fraudes extremamente desafiadoras, já que uma transação registrada não pode ser alterada, e qualquer manipulação fraudulenta nos dados seria prontamente perceptível aos participantes da rede.

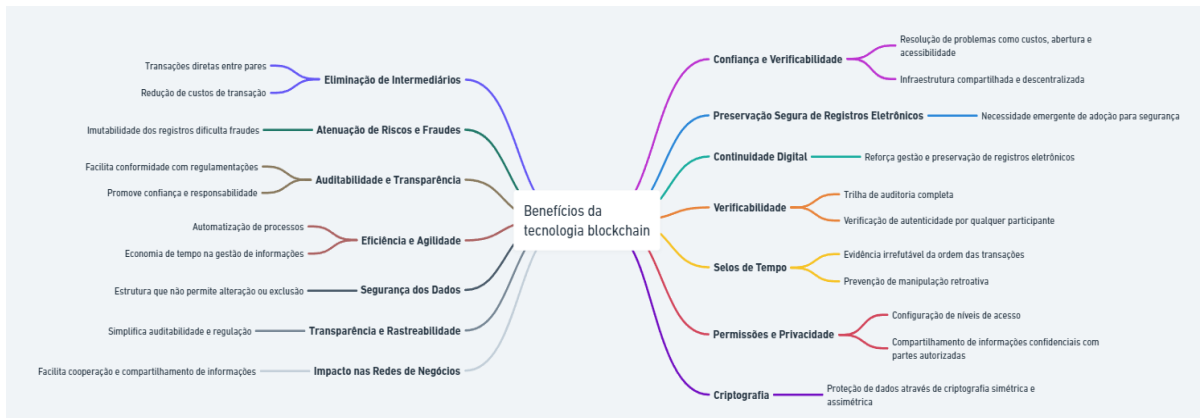
A *blockchain*, com suas características de auditabilidade, transparência, imutabilidade e descentralização, atribui solidez às transações (Shamshad *et al.*, 2020). Enquanto isso, Swan (2015) sublinha que a transparência inerente à *blockchain* promove confiança e responsabilidade, fatores essenciais em ambientes acadêmicos que prezam pela integridade dos dados.

A eficiência da *blockchain* economiza tempo na gestão e verificação de informações. Tapscott e Tapscott (2016) notam que a automatização de processos agiliza a troca de informações, liberando tempo para atividades acadêmicas mais produtivas.

Mori *et al.* (2020) explicam que a estrutura da *blockchain* não permite alteração ou exclusão, mas apenas a adição de blocos, tornando a adulteração uma tarefa árdua, reforçando a segurança dos dados.

A transparência e rastreabilidade da *blockchain* simplificam a auditabilidade e a regulação. Mougayar (2016) enfatiza que as transações na *blockchain* são facilmente auditáveis, facilitando a conformidade com regulamentações acadêmicas e legais. Por exemplo, a *blockchain* pode ser utilizada para armazenar certificados acadêmicos, agilizando a verificação da veracidade desses registros, reduzindo a falsificação de certificados e outras atividades fraudulentas, e melhorando a integridade do sistema. A figura 5 apresenta um esquema com os benefícios da *blockchain* discutidos nesse tópico.

Figura 5 - Benefícios da tecnologia *blockchain*



Fonte: Elaborado pelo autor.

Este tópico elucidou os benefícios primordiais da tecnologia *blockchain*. No tópico seguinte, será explorado o funcionamento da tecnologia *blockchain*.

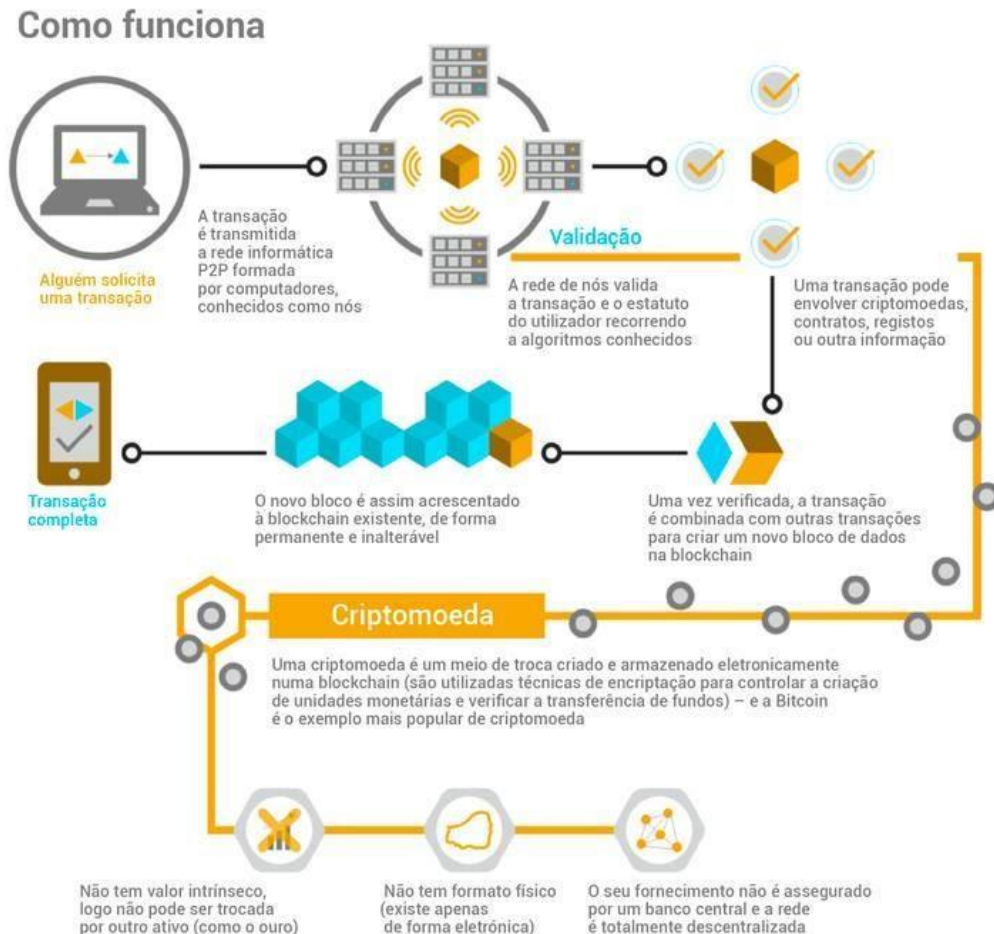
2.4 Como funciona a tecnologia *blockchain*: tecnologias e padrões subjacentes

Neste tópico, será aprofundada a compreensão sobre a tecnologia *blockchain* por meio da exploração das tecnologias e padrões subjacentes que a tornam viável. A complexidade da *blockchain* é derivada da interação entre elementos-chave como arquitetura, criptografia, bancos de dados distribuídos e mecanismos de consenso, tais como *proof of work* (é um algoritmo de consenso distribuído que precisa desempenhar um papel ativo na mineração de blocos de dados, muitas vezes consumindo recursos como eletricidade, confirmando que as transações foram verificadas e são legítimas) (PoW) e *proof of stake* (é um sistema de verificação de transações baseado em grupos de validadores, ou seja, entidades que chegam a um acordo a cada novo bloco por meio de seus nós) (PoS).

A arquitetura da *blockchain*, conforme enfatizado por Ouchi e Arakaki (2020), consiste em uma estrutura de dados sustentada por uma rede ponto a ponto distribuída. Em uma rede do tipo ponto a ponto, os computadores, referidos como “nós”, compartilham recursos e tarefas sem necessidade de um controle centralizado. Gonçalves e Rodrigues (2020) destacam características cruciais da *blockchain*, incluindo a cadeia de blocos, identificador único, *criptografia* assimétrica, validação por consenso, distribuição de dados e controle descentralizado.

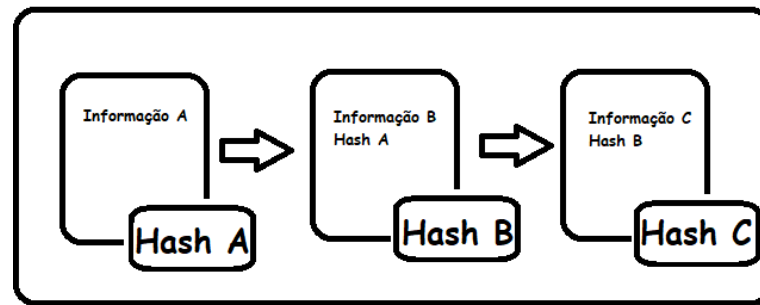
Esses componentes formam a base de uma rede descentralizada segura para transações e registros. A figura 6 ilustra o processo de funcionamento da *blockchain*.

Figura 6 - Como funciona a *blockchain*



Fonte: Livti, 2017.

Um *hash* é um código criado a partir de um bloco de dados usando um algoritmo de *criptografia* e usado para identificar blocos de dados, bem como criar “*links*” entre cada bloco de dados na *blockchain*. As funções *hash* são essenciais na *blockchain*, com cada bloco identificado por seu *hash* criptográfico, e cada novo bloco fazendo referência ao *hash* do bloco anterior (Ouchi; Arakaki, 2020). Esta ligação entre blocos assegura a integridade e sequência das transações, como exemplificado na figura 7.

Figura 7 - O *hash* criptográfico

Fonte: Elaborado pelo autor.

Gonçalves e Rodrigues (2020) elucidam que os valores *hash*, gerados por meio de cálculos, são resistentes contra a decomposição por engenharia reversa, mesmo com métodos matemáticos avançados, assegurando a segurança das transações.

A *criptografia* é um pilar para a segurança da *blockchain*. Oliveira (2023) destaca que a *blockchain* garante a confidencialidade e segurança dos dados transmitidos de forma descentralizada por protocolos complexos, *criptografia* e poder computacional, enquanto Liu, He e Xuan (2021) reiteram a essencialidade da *criptografia* na garantia da segurança.

A *blockchain* opera como um banco de dados distribuído, com várias cópias sincronizadas em diferentes localidades, sendo mantido por nós na rede, como destacado por Narayanan *et al.* (2016). Cada transação confirmada e registrada em toda a rede assegura a disponibilidade e a confiabilidade dos registros.

A consistência e confiabilidade dos dados são garantidas pelos nós na rede *blockchain* mediante um mecanismo de consenso (Liu; He; Xuan, 2021). Bell *et al.* (2019) fazem uma analogia interessante com os escribas da Idade Média, contrastando com os métodos digitais contemporâneos de verificação de dados.

A rede *peer-to-peer* (P2P) subjacente à *blockchain* elimina a necessidade de intermediários, promovendo a descentralização (Swan, 2015). Nesse contexto, os autores Woodall e Ringel (2019) destacam como a confiança é fortalecida pela eliminação de uma autoridade central.

Os mecanismos de consenso PoW e PoS são explorados, com Androulaki *et al.* (2018) descrevendo o processo de mineração (ato de validar transações na *blockchain*) no PoW, enquanto Mougayar (2016) explica a validação de transações no PoS com base na quantidade de criptomoeda detida pelos participantes.

Lemieux (2017) esclarece o processo de *hashing* criptográfico dos blocos, destacando seu papel na proteção contra fraudes e falsificação de dados.

Este tópico proporcionou uma visão detalhada das tecnologias e padrões que sustentam a *blockchain*. No próximo tópico, serão explorados os desafios em relação à tecnologia *blockchain*.

2.5 Desafios em relação à tecnologia *blockchain*

A tecnologia *blockchain*, apesar de revolucionária, enfrenta desafios significativos que afetam sua segurança, privacidade, escalabilidade, interoperabilidade e conformidade regulatória. Estes desafios destacam a complexidade da implementação e da adoção generalizada da *blockchain* em diferentes domínios. Alguns destes desafios são apresentados neste tópico.

Embora a *blockchain* ofereça um nível elevado de segurança nas transações, preocupações persistem quanto à segurança e privacidade dos dados. Problemas com questões de segurança na rede *peer-to-peer* do *bitcoin* evidenciam vulnerabilidades. Além disso, a transparência inerente às *blockchains* públicas, como *bitcoin* e *ethereum*, levanta questões de privacidade, já que todas as transações são rastreáveis e públicas, potencialmente expondo informações sensíveis dos usuários (Zaghloul *et al.*, 2019).

A escalabilidade é um dos obstáculos mais significativos para a *blockchain*, especialmente para aquelas que utilizam o mecanismo de consenso *Proof of Work* (PoW). O crescimento no número de transações pode causar atrasos e aumentar as taxas de transação, criando gargalos que afetam o desempenho e a eficiência da rede (Bhutta *et al.*, 2021).

A transparência das *blockchains* públicas, apesar de ser uma vantagem para a segurança, também gera preocupações com a privacidade. A visibilidade do histórico de transações pode não ser adequada em todos os contextos, especialmente onde a privacidade é uma exigência crucial (Feng *et al.*, 2019).

A falta de interoperabilidade entre diferentes *blockchains* restringe sua funcionalidade, dificultando ou impossibilitando realizar transações diretas entre redes distintas sem recorrer a intermediários. Essa limitação reduz a utilidade da tecnologia *blockchain* em aplicações que necessitam de interação entre diferentes plataformas (Singh *et al.*, 2020).

A natureza descentralizada e global da *blockchain* apresenta desafios significativos para a regulamentação e conformidade legal. As diferenças nas regulamentações entre jurisdições podem complicar a adoção e implementação de soluções baseadas em *blockchain*, criando um ambiente complexo para a navegação legal e regulatória (Golosoza; Romãnovs, 2018).

Esses desafios não só evidenciam as limitações atuais da tecnologia *blockchain*, mas também sublinham a necessidade de pesquisa contínua, desenvolvimento e colaboração regulatória para superá-los. A busca por soluções, como a adoção de mecanismos de consenso mais eficientes, o desenvolvimento de padrões de interoperabilidade e a colaboração com órgãos regulatórios, é crucial para maximizar o potencial transformador da *blockchain*, mitigando seus impactos negativos e promovendo sua aplicação sustentável e eficaz em uma variedade de contextos. Na figura 8 é apresentado um esquema com os desafios referentes a tecnologia *blockchain*.

Figura 8 - Desafios da tecnologia *blockchain*



Fonte: Elaborado pelo autor.

Após discutir alguns desafios em relação à tecnologia *blockchain*, no próximo tópico vamos verificar a relação da tecnologia *blockchain* com outras tecnologias.

2.6 Relações da tecnologia *blockchain* com outras tecnologias

A *blockchain*, sendo uma inovação disruptiva, tem o potencial de redefinir a maneira como interagimos com sistemas de informação e gerenciamos dados de pesquisa. Este tópico examina as interações da *blockchain* com outras tecnologias, como: bancos de dados, armazenamento/processamento em nuvem, *web* semântica, dados ligados e computação em grade.

A complexidade da relação entre a *blockchain* e bancos de dados convencionais é evidente. Ao contrário da *blockchain*, um banco de dados é um registro centralizado gerenciado por um administrador. Conforme elucidado por Tapscott e Tapscott (2016), a *blockchain*, manifestando-se como um banco de dados distribuído, proporciona imutabilidade e descentralização, características vitais onde a confiabilidade e segurança dos dados são primordiais. Lemieux (2017) complementa que a *blockchain* opera como um banco de dados de transações distribuído onde diferentes nós colaboram para armazenar sequências de bits *criptografadas* em blocos, sendo subsequentemente encadeados. No quadro 1 apresentamos uma revisão comparativa entre a *blockchain* e os bancos de dados convencionais.

Quadro 1 - Revisão comparativa entre *blockchain* e sistemas de bancos de dados convencionais

Blockchain	Bancos de dados convencionais
Atributos/Recursos	
<p>Blockchain é um livro-razão distribuído inviolável que funciona em toda a rede. É a aplicação de criptografia para armazenar informações para maior segurança, transparência, confiança e disponibilidade.</p> <p>Características do Blockchain:</p> <ol style="list-style-type: none"> 1. Estrutura de armazenamento descentralizada - as informações são distribuídas em vários nós da rede. 2. Imutabilidade das informações armazenadas - as informações armazenadas não podem ser alteradas, são lacradas com assinaturas criptográficas. 3. As informações/transações são armazenadas em blocos - as informações são armazenadas em blocos sequencialmente com um registro de data e hora. Assim, uma transação pode ser facilmente rastreada com os dados e a hora da ocorrência. 4. Blocos de informação são criptografados/hashead - esta aplicação de criptografia em blockchain garante a integridade dos dados. 5. Alta confidencialidade - os dados são criptografados com uma chave pública e privada, dependendo da arquitetura do sistema. A criptografia impede o acesso não autorizado ao sistema. 6. Disponibilidade - a estrutura de armazenamento descentralizado de informações em vários nós da rede garante que as informações estejam disponíveis para todos os participantes quando necessário. 7. Transparência - as informações armazenadas são visíveis para todos os participantes 8. Validade da informação/transação realizada - toda transação é validada pela maioria dos participantes antes de ser adicionada ao bloco 9. Privacidade e anonimato - o sistema incorpora modos de proteção como privacidade e anonimato. Além disso, a identidade de um usuário pode ser anônima ou pseudônima. 10. Alta segurança - a segurança das transições e informações armazenadas é garantida 11. Responsabilidade e consistência de dados - onde uma ramificação/informação maliciosa não é consistente com o protocolo padrão, cada nó na rede contabiliza sua própria cópia dessa transação para validar adequadamente. 12. Unânime - todos ou a maioria dos nós da rede devem concordar com a validade de um registro antes do processamento posterior. 13. Longevidade dos registros históricos - o sistema é durável. 14. Escalabilidade - a escalabilidade em modelos recentes de blockchain melhorou consideravelmente. 15. Confiança - há confiança entre os usuários/participantes nas transações ou informações armazenadas porque passam pela validação da maioria dos usuários. 16. Controle distribuído de informações via protocolo de consenso - cada nó da rede valida uma transação ou informação antes de ser adicionada ao bloco. 	<p>Uma coleção organizada e estruturada de dados arranjados em alguma ordem definida para facilitar a acessibilidade e recuperação. Um hub de dados centralizado de dados logicamente relacionados, acessados e compartilhados por vários sistemas. Gerenciado e manipulado através de um Sistema de Gerenciamento de Banco de Dados (DBMS).</p> <p>Características do banco de dados/DBMS:</p> <ol style="list-style-type: none"> 1. Estrutura de armazenamento centralizado - as informações são armazenadas em um servidor e acessadas por vários nós/terminais. 2. Permite atualização/edição de informações - as informações armazenadas no banco de dados/tabelas podem ser editadas a qualquer momento. 3. As informações são armazenadas de forma tabular, em linhas (registros) e colunas (campos). Todo o banco de dados é chamado de esquema. 4. Controle de redundância. 5. Consistência de dados 6. Segurança - os registros do banco de dados são acessados por meio de nome de usuário e senhas 7. Backup e restauração de informações 8. Reversão da transação em caso de falha do sistema, etc. 9. Controle central do sistema pelo administrador do banco de dados. Somente o administrador pode registrar os usuários do banco de dados.

Fonte: Omame; Alex-Nmecha, 2021.

A integração da *blockchain* com o armazenamento em nuvem pode oferecer uma camada adicional de segurança e rastreabilidade. A *blockchain* para Oliveira (2023) garante a auditabilidade, autenticidade e integridade armazenadas em um livro-razão distribuído, sendo ideal, por exemplo, para aplicações acadêmicas que demandam alta integridade dos registros. A integridade dos dados, como explicado por Gangadevi e Renuga Devi (2021), implica manter a estabilidade, precisão e confiabilidade dos dados.

A interação emergente entre *blockchain* e *web* semântica tem potencial para fornecer registros confiáveis para dados semânticos. Mougayar (2016) sugere que essa combinação pode facilitar a criação de ontologias e registros confiáveis, melhorando a interoperabilidade de dados acadêmicos.

No contexto dos dados ligados, a *blockchain* pode proporcionar uma base segura para a criação de gráficos de dados distribuídos. Androulaki *et al.* (2018) destacam que a *blockchain* pode validar a autenticidade e a origem dos dados, fortalecendo os princípios dos dados ligados em ambientes acadêmicos, e facilitando a partilha segura de dados entre diferentes entidades.

A integração da *blockchain* com a computação em grade pode promover a segurança e a verificabilidade das tarefas executadas numa rede distribuída (tipo de rede no qual o poder de processamento e os dados são distribuídos pelos nós em vez de centro de dados centralizados). Narayanan *et al.* (2016) propõem que a *blockchain* pode servir como uma camada adicional de segurança, protegendo a integridade dos resultados acadêmicos e assegurando sua imutabilidade.

Em conclusão, este tópico proporcionou uma exploração das relações da tecnologia *blockchain* com outras tecnologias, ilustrando o potencial da *blockchain* em aprimorar a segurança, verificabilidade e gestão de dados em diversos domínios tecnológicos, principalmente na Ciência da Informação. No tópico a seguir, serão exploradas as aplicações práticas da tecnologia *blockchain*.

2.7 Aplicações da tecnologia *blockchain*

A tecnologia *blockchain*, uma inovação multifacetada, vai além de sua aplicação inicial em *criptomoedas*, desempenhando um papel crucial em diversos setores. Esta seção delinea suas implementações em *criptomoedas*, rastreamento de ativos, contratos inteligentes, redes de negócios, registros de negócios, pagamentos internacionais, identidade digital confiável, comércio global, registros médicos eletrônicos e na plataforma *Hyperledger*, destacando o impacto significativo na gestão e compartilhamento de informações.

Inicialmente, as *criptomoedas* emergiram como a primeira aplicação da *blockchain*. Nakamoto (2008) realçou em seu *whitepaper* do Bitcoin que a *blockchain* serve como um registro público para transações de *bitcoin*, eliminando a necessidade de intermediários, o que revolucionou o conceito de moeda e transações financeiras.

A *blockchain* também facilita o rastreamento eficiente de ativos, sejam tangíveis, sejam intangíveis. Como Tapscott e Tapscott (2016) apontam, a *blockchain* proporciona um registro transparente e imutável, melhorando o rastreamento ao longo da cadeia de suprimentos, permitindo a monitorização em tempo real das transações.

Para Gomes e Coutinho (2022), no contexto da *blockchain*, um contrato inteligente, ou *smart contract* é um fluxo de valor baseado em termos e condições específicos, tal como um contrato físico. A única diferença é que é totalmente digital.

São criados usando uma linguagem de programação específica e executados em uma *blockchain*. Quando uma transação é iniciada, o *smart contract* é ativado, e ele executa automaticamente o código contido nele. Isso pode incluir a transferência de ativos digitais, a validação de dados ou a execução de outras funções definidas pelas partes envolvidas.

Tapscott e Tapscott (2016) argumentam que os *smart contracts* são uma das principais inovações da *blockchain*, permitindo a criação de novos modelos de negócios e a transformação de indústrias inteiras.

Uma das principais vantagens dos *smart contracts* é que eles são imutáveis e seguros. Eles são executados em uma *blockchain* descentralizada e distribuída, o que significa serem resistentes a ataques cibernéticos e falhas de sistema.

A adoção de *blockchain* em redes de negócios é vista como um meio de aprimorar a transparência e eficiência. Mougayar (2016) sugere que a *blockchain* favorece a troca de informações confiáveis entre empresas, criando oportunidades para colaboração e crescimento.

Além disso, a *blockchain* tem o potencial de simplificar os registros de negócios. Casey e Vigna (2018) notam que ela torna os registros mais acessíveis e rastreáveis, reduzindo a burocracia.

A criação de identidades digitais confiáveis é outra aplicação valiosa da *blockchain*. Swan (2015) observa que a *blockchain* oferece uma base segura para identidades digitais, garantindo autenticidade e controle do usuário sobre seus dados.

No comércio global, a *blockchain* pode simplificar processos de importação/exportação e garantir a integridade das informações, conforme explicado por Narayanan *et al.* (2016), eliminando fraudes e atrasos no comércio global.

Deng *et al.* (2019) reiteram que a integridade, referindo-se a registros completos e inalterados, é uma das vantagens centrais da *blockchain*.

Finalmente, o Projeto Hyperledger é uma plataforma de *blockchain* de código aberto que oferece soluções flexíveis para diversas aplicações. Androulaki *et al.* (2018) salientam que o *hyperledger* proporciona um ambiente de desenvolvimento seguro e escalável para aplicações empresariais baseadas em *blockchain*.

Este tópico elucidou algumas aplicações da tecnologia *blockchain*, ressaltando sua transformação na gestão e compartilhamento de informações em variados contextos. No tópico subsequente, serão exploradas as potenciais aplicações da tecnologia *blockchain* no domínio da Ciência da Informação.

2.8 Possíveis aplicações da tecnologia *blockchain* na área da Ciência da Informação

Este tópico explora as aplicações da tecnologia *blockchain* na Ciência da Informação, Biblioteconomia e Arquivística, enfatizando seu papel além da preservação digital. A *blockchain* promove aprimoramentos na gestão de dados, publicações, arquivos e colaboração em projetos de pesquisa.

A *blockchain* favorece a criação de redes de colaboração seguras e transparentes na Ciência da Informação, como destacado por Swan (2015), proporcionando uma plataforma de colaboração aberta que permite participação segura de múltiplas partes sem intermediário centralizado.

Os laboratórios de pesquisa podem utilizar a *blockchain* para registrar e compartilhar dados de forma confiável, conforme observado por Androulaki *et al.* (2018). A tecnologia cria uma trilha de auditoria à prova de falsificação, onde todas as partes podem confiar, registrando quem criou os dados, quando estes foram criados e as modificações ou atualizações efetuadas.

A comunidade arquivística e outras organizações patrimoniais beneficiam-se da *blockchain* como uma plataforma descentralizada, onde a colaboração é fundamental para o sucesso (Green *et al.*, 2018).

Em projetos distribuídos em grande escala, a *blockchain*, conforme apontado por Mougayar (2016), oferece uma camada extra de segurança e verificabilidade. Os dados são criptografados e distribuídos em várias máquinas, mitigando alterações ou exclusões indevidas.

A *blockchain* serve como um diário imutável para pesquisadores, permitindo registros permanentes de pesquisas, como explica Narayanan *et al.* (2016). Isso assegura a autenticidade e integridade dos registros ao longo do tempo e as publicações podem ser enriquecidas com dados adicionais e informações sobre proveniência usando *blockchain*, conforme afirmam Tapscott e Tapscott (2016). A interoperabilidade de dados e publicações é facilitada, promovendo uma troca eficiente entre sistemas e pesquisadores, como observam Casey e Vigna (2018).

Mori *et al.* (2020) afirmam que a tecnologia garante a autenticidade dos dados e retenção de um histórico operacional, estabelecendo repositórios digitais confiáveis. Narayanan *et al.* (2016) explicam que a *blockchain* proporciona um registro imutável, oferecendo um registro seguro de transações.

Este tópico abordou como a *blockchain* pode ser significativamente aplicada na Ciência da Informação, Biblioteconomia e Arquivística. As aplicações discutidas podem transformar a gestão de dados, publicações e documentos de pesquisa, apresentando um caminho promissor para aprimorar a confiabilidade e rastreabilidade das informações.

Ao longo deste capítulo, desvendamos a tecnologia *blockchain* e seu leque de aplicações e aprofundamos os fundamentos da *blockchain*, abordando sua estrutura, algoritmos de consenso e *criptografia* subjacente.

É evidente que a tecnologia *blockchain* pode reformular radicalmente nossa abordagem no manejo de informações e dados em diversos domínios. Seus atributos de imutabilidade, descentralização e rastreabilidade promovem segurança, transparência e confiabilidade nas operações de informação.

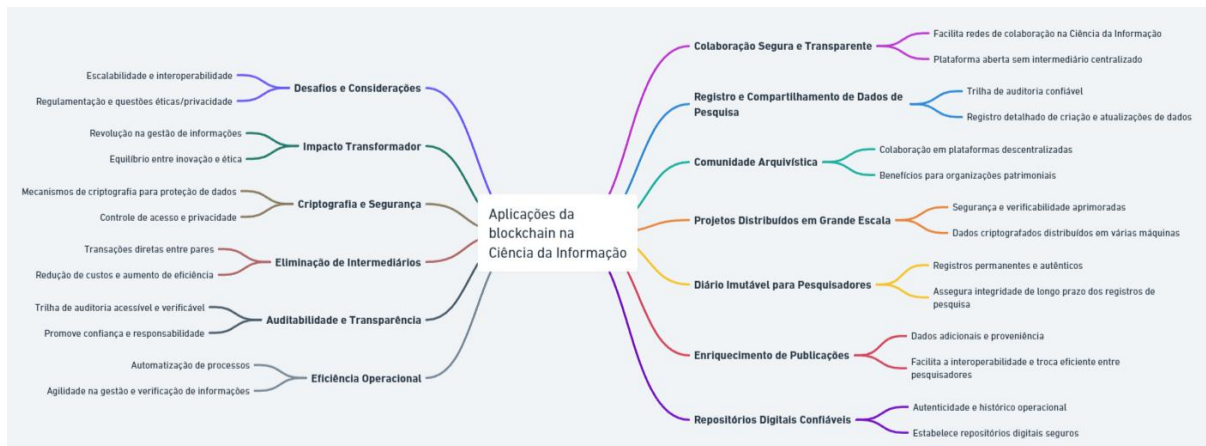
Na esfera de preservação digital, a *blockchain* apresenta soluções auspiciosas para assegurar a autenticidade e integridade de objetos e metadados digitais, criando registros imutáveis que fornecem uma trilha de auditoria confiável.

Identificamos também desafios significativos que necessitam ser superados para uma implementação bem-sucedida da *blockchain*. As questões de escalabilidade, interoperabilidade e regulamentação emergem como preocupações centrais que demandam atenção.

Além disso, o uso da *blockchain* suscita considerações éticas e de privacidade. É vital equilibrar a transparência e segurança proporcionadas pela *blockchain* com a imperativa proteção de dados sensíveis e a privacidade dos usuários.

À medida que a tecnologia *blockchain* evolui, é crucial manter-se atualizado com seu desenvolvimento. Pesquisas adicionais são necessárias para enfrentar os desafios e capitalizar as oportunidades que a *blockchain* oferece. Na figura 9 apresentamos um esquema com as possíveis aplicações da *blockchain* na área da Ciência da Informação.

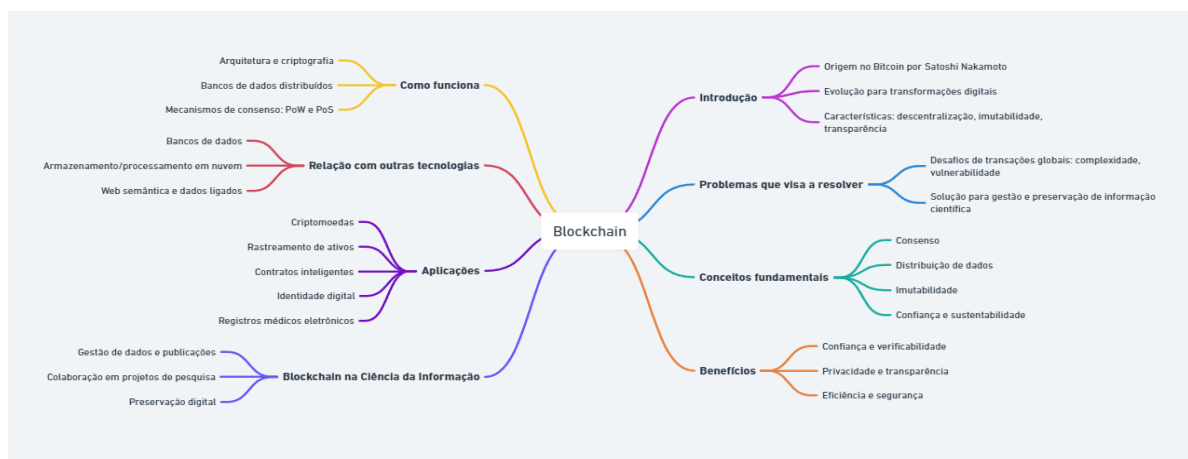
Figura 9 - Aplicações da *blockchain* na Ciência da Informação



Fonte: Elaborado pelo autor.

A *blockchain* emerge como uma ferramenta poderosa com o potencial de revolucionar a Ciência da Informação em diversos aspectos. Sua contribuição para garantir a autenticidade e integridade dos dados é inegável. No entanto, a implementação cuidadosa, ponderando os desafios e questões éticas, é crucial para maximizar seu potencial transformador. Conforme exploramos as aplicações da *blockchain* na Ciência da Informação, torna-se imperativo manter um equilíbrio entre a inovação tecnológica e a responsabilidade ética. Na figura 10 apresentamos um esquema com os tópicos apresentados no capítulo. Após a figura 10, prosseguimos para o capítulo 3, que aborda a preservação digital. Este capítulo também inclui uma discussão aprofundada sobre os repositórios confiáveis.

Figura 10 - Síntese do capítulo *blockchain*



Fonte: Elaborado pelo autor.

3 PRESERVAÇÃO DIGITAL

No cenário contemporâneo, caracterizado por uma marcante transição rumo ao universo digital, evidencia-se progressivamente a relevância da preservação digital no âmbito da Ciência da Informação. Ferreira (2006) ressalta que a preservação digital vai além da simples manutenção técnica de informações em formato digital. Trata-se, na verdade, de uma combinação de arte e ciência dedicada a garantir a disponibilidade futura dessas informações, mantendo sua integridade interpretativa mesmo quando transferidas para ambientes tecnológicos distintos dos originais.

Madio e Grácio (2023) explicam que tal abordagem demanda mais do que meras práticas de *backup* de dados; requer uma perspectiva holística que acompanhe os objetos digitais desde sua gênese até seu arquivamento ou descarte definitivo. Sayão (2010) reitera essa concepção ao descrever a preservação digital como um conjunto de estratégias projetadas para assegurar o acesso duradouro aos conteúdos digitais, um desafio que transcende o tempo e o espaço.

Nesse contexto, Woodall e Ringel (2019) introduzem a tecnologia *blockchain* como um recurso promissor, enfatizando sua capacidade de garantir segurança, autenticidade, confiabilidade e confiança. Ouchi e Arakaki (2020) complementam que a *blockchain* serve como um mecanismo de confiança ao registrar cada alteração de maneira transparente, constituindo um registro histórico confiável, fundamental para futuras pesquisas sobre reprodutibilidade. Desse modo, a *blockchain* tem um potencial de promover uma gestão de dados mais transparente e verificável, contribuindo de maneira significativa para a preservação do legado digital para futuras gerações.

Diante desses desafios e possibilidades, emerge como próximo tema de investigação a arquitetura de objetos digitais, um elemento fundamental que detalha a estrutura e organização dos dados digitais, visando otimizar sua preservação e acessibilidade a longo prazo.

3.1 Arquitetura de objetos digitais

A arquitetura de objetos digitais representa um componente essencial na preservação efetiva de ativos digitais. Márdero Arellano (2008) destaca a crucialidade de identificar e adaptar-se às diversas dependências tecnológicas imprescindíveis para a interpretação adequada de objetos digitais.

Tais dependências, incluindo *hardware*, *software*, sistemas e ferramentas, são fundamentais para a correta leitura e compreensão desses objetos, uma visão compartilhada por Santos; Flores (2019). A preservação digital enfrenta o desafio significativo de garantir a longevidade dessas tecnologias.

Sayão (2015) sublinha a importância de uma documentação metódica dessas dependências para assegurar a acessibilidade dos objetos digitais frente a evoluções tecnológicas.

Adicionalmente, Márdero Arellano (2008) introduz o conceito de camadas de abstração, que delineiam os distintos níveis dos objetos digitais, desde sua existência física até sua representação e autenticidade. Este princípio é crucial para manter a integridade dos objetos digitais, garantindo sua conformidade com o design original. De maneira complementar, Sayão (2015) aborda a importância das propriedades significativas, atributos fundamentais dos objetos digitais vitais para sua preservação.

A vulnerabilidade dos objetos digitais a alterações e à deterioração da mídia, que pode resultar na perda parcial ou total das informações, é uma preocupação expressa por Sayão (2010).

Márdero Arellano (2008) amplia esta discussão ao examinar os diferentes níveis de preservação, que variam desde a conservação bit a bit até estratégias conceituais destinadas a manter a essência ou funcionalidade do objeto.

Em síntese, a arquitetura de objetos digitais abrange uma ampla gama de conceitos e práticas voltadas para a salvaguarda do patrimônio digital para o futuro.

Com esta fundamentação, direcionamos nossa análise para o próximo tema crítico em preservação digital: o modelo de referência para um sistema aberto de arquivamento de informações (OAIS).

3.2 Modelo de referência OAIS

O Modelo de Referência para um Sistema Aberto de Arquivamento de Informações (OAIS) constitui um marco conceitual crucial, estabelecendo os elementos funcionais indispensáveis para os sistemas de informação comprometidos com a preservação digital de longo prazo. Thomaz e Soares (2004) sublinham que o OAIS vai além de um simples arcabouço teórico, emergindo como um padrão essencial no campo da preservação digital ao definir terminologias, conceitos

fundamentais e fluxos de informações cruciais para a gestão eficaz de um repositório digital.

No contexto dos avanços tecnológicos contemporâneos, a tecnologia *blockchain* surge como uma inovação compatível com os princípios do modelo OAIS. Green *et al.* (2018) argumentam que a incorporação da abordagem descentralizada da *blockchain* ao modelo OAIS pode impulsionar o desenvolvimento de uma infraestrutura de arquivamento digital mais robusta, resiliente e colaborativa, introduzindo métodos inovadores de gestão de risco.

A interação entre a *blockchain* e o modelo OAIS é notavelmente sinérgica, à medida que a *blockchain* proporciona um registro imutável e auditável das transações associadas aos objetos digitais. Sayão (2010) amplia essa perspectiva, observando que o OAIS especifica o ambiente operacional, as interfaces externas, os componentes funcionais e os objetos de informação de sistemas dedicados à preservação de conteúdos digitais a longo prazo.

Na estrutura do OAIS, os metadados ocupam uma posição central, e o modelo de informação relacionado a eles é essencial para a caracterização e preservação adequadas dos objetos digitais. Santos e Flores (2019) enfatizam que, no contexto OAIS, os participantes — produtores, arquivistas ou usuários finais — têm papéis essenciais.

Thomaz e Soares (2004) identificam três componentes principais no modelo OAIS: o Pacote de Informação de Submissão (SIP), o Pacote de Informação de Arquivamento (AIP) e o Pacote de Informação de Disseminação (DIP). A *blockchain* desempenha um papel fundamental na rastreabilidade e autenticação desses pacotes ao longo de seu ciclo de vida, garantindo a integridade e a autenticidade dos objetos digitais.

O modelo OAIS oferece uma estrutura detalhada para a realização eficiente, sustentável e transparente da preservação digital. Segundo Green *et al.* (2018), a integração da abordagem distribuída da *blockchain* com o modelo OAIS tem o potencial de criar uma infraestrutura de arquivamento digital mais cooperativa e resiliente, abrindo caminhos para novas práticas de gerenciamento de riscos. Esta inovação responde às demandas da era digital moderna e facilita a transição para um futuro onde o patrimônio digital seja preservado e valorizado adequadamente pelas futuras gerações.

À medida que nos debruçamos sobre os desafios inerentes à preservação digital, somos convocados a explorar, no próximo tópico, as estratégias de preservação digital e as metodologias específicas que podem ser adotadas para assegurar a manutenção efetiva do patrimônio digital para o benefício das gerações vindouras.

3.3 Estratégias de preservação digital

A preservação digital surge como uma necessidade premente na contemporaneidade, uma era definida por avanços tecnológicos acelerados, e desempenha um papel crucial na garantia da integridade, autenticidade e acessibilidade de objetos digitais para as gerações futuras. Márdero Arellano (2008) distingue duas diretrizes principais nas estratégias de preservação digital: a conservação do ambiente tecnológico existente e a adaptação às rápidas e constantes mudanças tecnológicas.

Na vertente da preservação do ambiente tecnológico, a atenção é voltada para a manutenção física dos objetos digitais. Márdero Arellano (2008) explora duas abordagens fundamentais: a constituição de museus tecnológicos, que atuam como cápsulas do tempo da evolução tecnológica, e a emulação, técnica que preserva dados em sua forma original por meio de *softwares* que replicam sistemas de *hardware* obsoletos. Essas metodologias possibilitam o acesso e a interpretação autêntica de objetos digitais históricos.

Márdero Arellano (2008) recomenda práticas como a normalização, o refrescamento, a migração e o encapsulamento para navegar por essas transformações. A *blockchain* pode servir como uma base confiável ao oferecer um registro imutável que valida a autenticidade e a integridade dos objetos digitais ao longo do tempo.

Souza *et al.* (2012) expandem essa perspectiva, argumentando que a preservação digital transcende a simples proteção de documentos, envolvendo tecnologias avançadas para pesquisa de mídia, *backup*, migração e autenticação. Este campo interdisciplinar requer a colaboração entre profissionais da informação e tecnologia, promovendo um diálogo constante para uma gestão eficaz da preservação digital. Tal colaboração fomenta um ecossistema de conhecimento e inovação, essencial para o desenvolvimento de estratégias de preservação robustas.

A fragilidade dos objetos digitais a modificações, tanto intencionais quanto acidentais, ressalta a necessidade de estratégias de preservação eficazes. Bell *et al.* (2019) apontam para os desafios impostos pelo avanço implacável das tecnologias digitais. A *blockchain*, com sua capacidade de criar registros imutáveis e verificáveis, emerge como uma solução potencial para atenuar os desafios da preservação digital, proporcionando uma salvaguarda resiliente contra adversidades.

A criação de repositórios digitais confiáveis, o desenvolvimento de capacidades técnicas, políticas e infraestruturas institucionais adequadas são cruciais para a manutenção, armazenamento, gerenciamento de dados, acesso, disseminação e transição para novas mídias e formatos (Deng *et al.*, 2019). O domínio complexo da preservação digital demanda uma abordagem multifacetada, com as estratégias propostas servindo como alicerce para enfrentar os desafios em um ambiente tecnológico em constante evolução.

A figura 11 apresenta um esquema com as estratégias apresentadas nesse tópico. À medida que prosseguimos nesta jornada de preservação digital, o próximo tema de interesse é o dos metadados de preservação digital. Este conceito refere-se à informação estruturada que é vital para o gerenciamento efetivo dos objetos digitais ao longo do tempo, desempenhando um papel crucial na facilitação do acesso, na garantia da autenticidade e na implementação de estratégias de preservação apropriadas.

Figura 11 - Estratégias de preservação digital



Fonte: Elaborado pelo autor.

3.4 Metadados de preservação digital

Os metadados de preservação digital emergem como componentes cruciais no cenário complexo da preservação digital, atuando como portadores de informações essenciais que garantem a continuidade do acesso e interpretação de objetos digitais ao longo do tempo. Sayão (2010) defende a necessidade de que os objetos armazenados sejam auto descritivos, integrados em estruturas físicas ou lógicas que encapsulem todas as informações necessárias para sua futura recuperação e compreensão. Desta forma, os metadados de preservação se posicionam como elementos centrais, servindo como guias essenciais nas estratégias de preservação digital.

A iniciativa PREMIS (*Preservation Metadata: Implementation Strategies*), estabelecida em 2003, marca um avanço significativo na consolidação dos metadados de preservação, reunindo um consórcio de especialistas internacionais dedicados a fundamentar as atividades de preservação digital nos metadados. Sayão (2010)

reconhece o PREMIS como uma síntese das experiências acumuladas por uma diversidade de instituições, como museus, bibliotecas e arquivos, tanto do setor público quanto do privado, refletindo a colaboração de profissionais de vários países, incluindo Austrália, Nova Zelândia e Estados Unidos, entre outros.

Um dos contributos inovadores do PREMIS foi a introdução das unidades semânticas, que criam um vocabulário comum para descrever os atributos dos objetos digitais. Márdero Arellano (2008) enfatiza que essas unidades encapsulam dados cruciais referentes à proveniência, autenticidade, integridade e contexto dos objetos digitais, aspectos fundamentais para a efetividade dos metadados de preservação.

Em termos de consistência e interoperabilidade, o PREMIS promove uma linguagem padronizada e um conjunto de normas que facilitam a interoperabilidade entre diferentes sistemas e contextos (Santos *et al.*, 2019).

Os metadados de preservação representam um alicerce vital na manutenção do legado digital da humanidade, garantindo que seja preservado meticulosamente e permaneça acessível para gerações futuras. Para Deng *et al.* (2019), os metadados constituem uma solução primordial para proteger a integridade dos registros digitais.

Essa sinergia representa um progresso significativo, alinhando as práticas de preservação digital às exigências e oportunidades da era digital atual. A discussão proposta oferece uma análise coesa e aprofundada sobre o papel estratégico dos metadados de preservação digital, incentivando a reflexão sobre sua importância e impacto na salvaguarda do patrimônio digital em um mundo cada vez mais digitalizado.

Prosseguindo nesta exploração da preservação digital, o próximo tópico a ser abordado envolve a política e o plano de preservação digital. Este segmento crucial da preservação digital engloba o desenvolvimento de diretrizes estratégicas e operacionais que norteiam as instituições na gestão eficiente de seus acervos digitais.

3.5 Política e Plano de preservação digital

A Política de Preservação Digital (PPD) surge como um instrumento estratégico fundamental para as instituições dedicadas à proteção de seu patrimônio digital. Lima e Rabello (2020) salientam que a formalização de uma política de preservação é essencial para fomentar a integração e a transparência dos repositórios, além de

apoiar os autores e procedimentos que garantem a acessibilidade e a confiabilidade contínua dos objetos digitais armazenados.

No Brasil, instituições renomadas como a Biblioteca Nacional e o Arquivo Nacional lideram o desenvolvimento e a implementação de políticas de preservação digital. A Política de Preservação Digital da Biblioteca Nacional (PPDBN), por exemplo, define diretrizes e princípios estratégicos para a conservação preventiva e a gestão eficaz de objetos digitais (Fundação Biblioteca Nacional, 2020).

Para a eficácia de uma política de preservação digital, uma abordagem holística é indispensável. Souza e Aganette (2022) argumentam que a formulação e aprovação de uma política de preservação são etapas cruciais para estabelecer os objetivos e diretrizes necessários para implementar processos de gestão eficientes.

Mori *et al.* (2020) aconselham que, antes de integrar inovações tecnológicas como, por exemplo, a *blockchain*, as instituições devem primeiro estabelecer políticas ou programas de gestão claros. Essa integração pode reforçar a integridade, transparência e autenticidade dos dados nas políticas de preservação digital.

Dessa forma, Deng *et al.* (2019) registram que

Para garantir a autenticidade dos registros eletrônicos, políticas e procedimentos correspondentes precisam ser formulados para controlar a formação, coleta, circulação, custódia e descarte de registros eletrônicos, de modo a garantir que os criadores, participantes e guardiões dos registros eletrônicos sejam confiáveis, e impedir a adulteração não autorizada do registro, excluir, usar e ocultar, para garantir que o registro também seja preciso e confiável ao mesmo tempo (Deng *et al.*, 2019, p. 139).

A integração de tecnologias nas políticas de preservação digital marca um avanço significativo, promovendo a sustentabilidade e a confiabilidade na gestão da informação digital. Este progresso indica um caminho promissor para o futuro da preservação digital, destacando a importância de explorar e estudar a gestão de registros eletrônicos confiáveis (Deng *et al.*, 2019).

Avançando na discussão sobre preservação digital, o próximo tema de importância é a confiança na preservação digital. Este conceito enfoca a necessidade de estabelecer e manter a confiança nos sistemas e processos de preservação digital, um elemento chave para o sucesso duradouro dessas iniciativas.

3.6 A confiança na preservação digital

A confiança é um elemento fundamental para a eficácia e harmonia dos sistemas em qualquer domínio, notadamente na preservação digital, onde a autenticidade, integridade e acessibilidade das informações detêm importância primordial. Márdero Arellano (2008) destaca que a confiança é fomentada por meio do estabelecimento de políticas bem estruturadas, a criação de repositórios digitais confiáveis e a adesão a padrões de certificação rigorosos.

Santos *et al.* (2018) reforçam a importância de estratégias de preservação contínuas para garantir o acesso de longo prazo às informações. A iniciativa da RDC-Arq (Rede de Documentação e Conservação de Arquivos) ilustra um esforço significativo na gestão de práticas de preservação, enfatizando a relevância da interoperabilidade entre diversas estratégias para manter documentos autênticos acessíveis ao longo do tempo.

As normas de certificação são essenciais para estabelecer critérios claros que asseguram a confiabilidade dos repositórios digitais (Márdero Arellano, 2008). A interoperabilidade entre esses repositórios cria uma rede colaborativa entre instituições encarregadas de preservar materiais digitais de valor cultural. Tais práticas colaborativas permitem a troca de conhecimentos entre diferentes repositórios, enriquecendo o ecossistema de preservação digital (Santos *et al.*, 2018).

Assim, a confiança constitui o alicerce fundamental para o progresso no ambiente digital. Com estruturas bem definidas, padrões estabelecidos e a adoção de inovações tecnológicas como, por exemplo, a *blockchain*, estamos criando um ambiente onde a confiança na preservação digital é reforçada. Isso não apenas previne pontos únicos de falha, mas também eleva a confiança nos registros, assegurando uma preservação íntegra e acessível do legado digital da humanidade para as gerações futuras (Bralić; Stančić; Stengård, 2020).

Esse diálogo enriquecido entre tecnologia e preservação digital pavimenta o caminho para uma compreensão aprofundada de como assegurar que o patrimônio digital seja preservado com integridade e acessibilidade para futuras gerações.

Ao nos dedicarmos a fortalecer esses vínculos de confiança, o próximo tópico de discussão se volta para os repositórios digitais confiáveis. Exploramos as características, critérios e práticas que definem um repositório digital como confiável, e como essas instituições são vitais para a preservação digital de longo prazo.

3.7 Repositórios digitais confiáveis

Os Repositórios Digitais Confiáveis (RDC) desempenham um papel central na preservação digital, funcionando como guardiões na proteção da integridade, autenticidade e acessibilidade das informações digitais ao longo do tempo. Santos e Flores (2015) argumentam que um repositório digital autêntico deve ser projetado para a preservação de longo prazo, equipado com recursos adequados para implementar estratégias de preservação e integrar padrões de metadados. Destaca-se a importância de documentar meticulosamente todas as intervenções realizadas nos documentos digitais, tais como migrações, para assegurar um registro histórico fiel de cada objeto digital armazenado, reforçando a confiança nos conteúdos preservados.

A confiabilidade de tais repositórios é fundamentada em sua conformidade com padrões e práticas rigorosos, assegurando que o patrimônio digital seja mantido de forma segura e acessível para as gerações vindouras. Gava e Flores (2020) caracterizam os RDCs como repositórios que mantêm materiais digitais de maneira autêntica, garantindo sua preservação e acesso pelo tempo requerido.

Para além da armazenagem de dados, os RDCs são valorizados por sua gestão meticulosa de riscos, transparência operacional e comprometimento com a responsabilidade e sustentabilidade de longo prazo. A implementação de um sistema confiável é, portanto, essencial para a gestão efetiva e a preservação de documentos arquivísticos digitais (Santos; Flores, 2015).

O modelo OAIS desempenha um papel crucial ao identificar os componentes funcionais necessários de um sistema de informação para preservação digital (Gava; Flores, 2020). A adesão a este modelo é imperativa para o estabelecimento e manutenção de RDCs robustos.

A clareza e abrangência das políticas e procedimentos são fundamentais para a confiabilidade dos repositórios. Márdero Arellano (2008) ressalta a importância de políticas de preservação digital bem articuladas para garantir o armazenamento seguro e o acesso contínuo a longo prazo dos ativos digitais.

A comunidade de preservação digital tem desenvolvido critérios de avaliação para validar a confiabilidade dos RDCs. Um repositório que atende a esses critérios pode ser considerado confiável, fornecendo um ambiente adequado para a

preservação de dados a longo prazo (Márdero Arellano, 2022). O modelo Hipátia², desenvolvido pelo Instituto Brasileiro de Informação em Ciência e Tecnologia — IBICT, é um exemplo de uma abordagem para a preservação em repositórios arquivísticos digitais confiáveis.

Iniciativas como a *Trustworthy Repository Audit & Certification* (TRAC) fornecem diretrizes para a auditoria e certificação de RDCs, estruturando um processo de certificação e estabelecendo metodologias para avaliar a sustentabilidade dos repositórios digitais (Sayão, 2010). Na figura 12 apresentamos um esquema referente aos repositórios confiáveis.

² <https://hipatia.ibict.br/o-modelo/> acessado em 13/12/2023.

Figura 12 - Repositórios digitais confiáveis



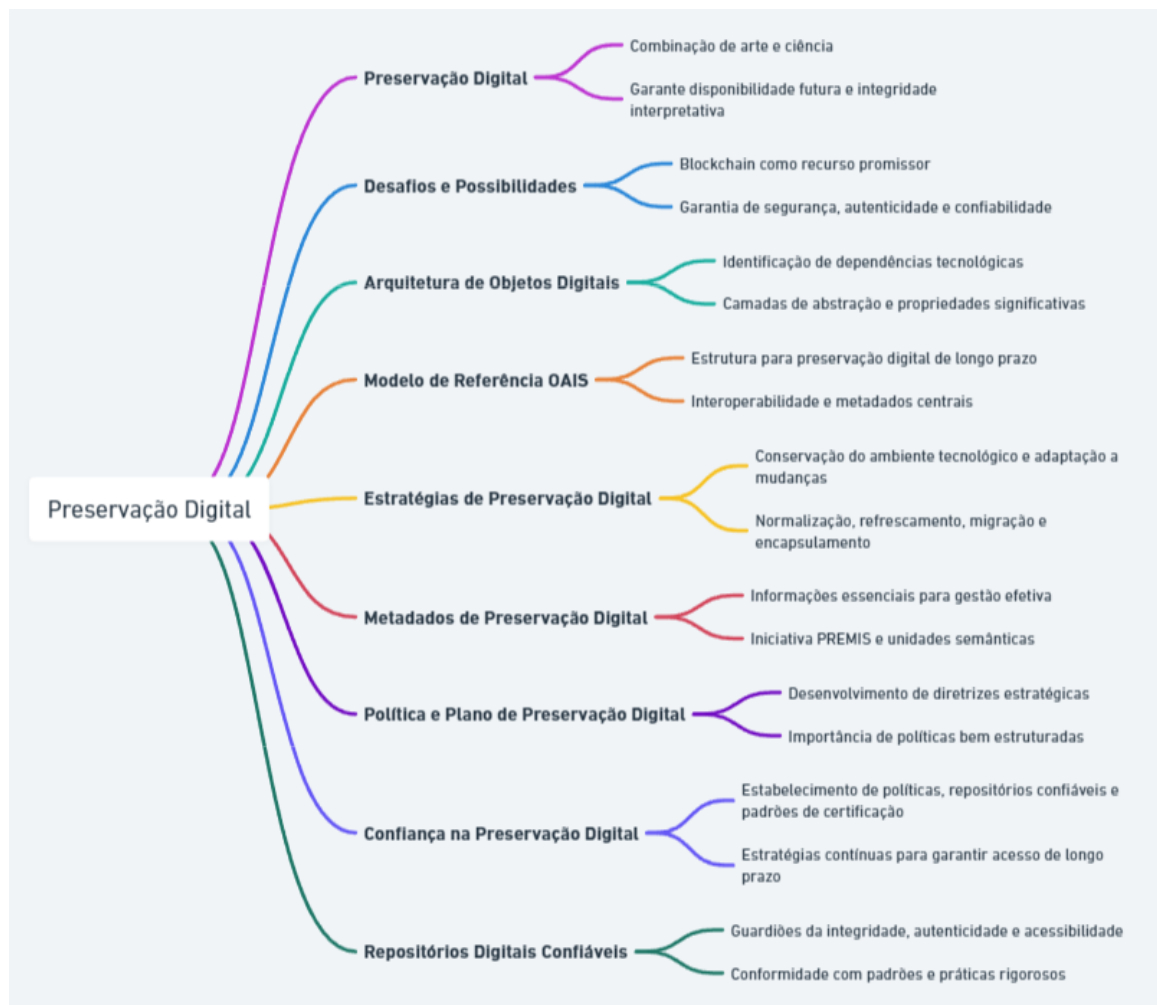
Fonte: Elaborado pelo autor.

Em resumo, os RDCs desempenham um papel crucial na salvaguarda do legado digital para futuras gerações. A confiança depositada nesses repositórios se baseia na adesão a padrões internacionais, na implementação de políticas e procedimentos claros, na eficaz gestão de riscos, no compromisso com a transparência e na dedicação à sustentabilidade de longo prazo. Conforme avançamos para o futuro da preservação digital, a importância dos RDCs se amplifica, exigindo investimentos contínuos em recursos, inovação e práticas rigorosas para assegurar a preservação e acessibilidade do patrimônio digital.

Na figura 13 apresentamos um esquema com os tópicos apresentados no capítulo.

Após a apresentação da figura 13, que finaliza a fundamentação teórica, avançamos para o capítulo 4, que trata da gestão de dados de pesquisa. Este capítulo destaca as características de proveniência, rastreabilidade, confiabilidade e autenticidade dos dados.

Figura 13 - Síntese do capítulo preservação digital



Fonte: Elaborado pelo autor.

4 GESTÃO DE DADOS DE PESQUISA

A pesquisa científica é fundamentada em dados, os quais constituem as evidências necessárias para a validação de conclusões derivadas de investigações e esses dados, conforme definido por Sales e Sayão (2020), abrangem qualquer tipo de registro coletado, observado, gerado ou utilizado no contexto da pesquisa científica. Eles são interpretáveis, tratáveis e reconhecidos como evidência pela comunidade científica, desempenhando um papel crucial na análise, validação e produção de resultados de pesquisa. A importância desses dados de pesquisa transcende o desenvolvimento científico, estendendo-se à validação, replicação e reprodução de pesquisas, tornando-os produtos de pesquisa valiosos (Ouchi; Arakaki, 2020).

Surgindo como uma área vital na Ciência da Informação, a gestão de dados de pesquisa (GDP) ressalta a complexidade e o crescente volume de dados produzidos em pesquisas científicas. Zhang (2021) destaca que o objetivo principal da gestão eficiente de dados é garantir a integridade, eficiência e acessibilidade dos dados digitais, englobando questões relacionadas à segurança, armazenamento e uso dos dados. Zhang (2021) enfatiza que

A segurança da informação de dados refere-se à segurança dos dados digitais, para não serem roubados e destruídos. A segurança do armazenamento de dados refere-se ao meio de armazenamento de informações, bem como à segurança do sistema de software e hardware, garantindo que os dados eletrônicos não sejam roubados e danificados. A segurança de utilização deve se referir aos dados digitais no uso de segurança e estabilidade no processo de transmissão. Não vai vazar (Zhang, 2021, p. 607, tradução nossa).

Sendo assim, a importância da gestão de dados de pesquisa (GDP) transcende a simples organização e preservação da integridade dos dados, estendendo-se ao potencial de aumentar o impacto e a visibilidade da pesquisa. Rekunenko *et al.* (2022) destacam a relevância de se estabelecer padrões acordados para o tratamento de dados ao longo de seu ciclo de vida, promovendo a acessibilidade universal, interoperabilidade e a capacidade de reutilização dos dados tanto por pessoas quanto por sistemas automatizados.

Um planejamento eficaz é crucial para a gestão dos dados de pesquisa, iniciando-se na fase de delineamento do projeto e abordando o gerenciamento dos dados ao longo do desenvolvimento do projeto e o compartilhamento posterior dos

mesmos. A formalização das ações e compromissos relacionados aos dados desde os estágios iniciais é fundamental para a replicabilidade dos estudos (Sales; Sayão, 2015). Dessa forma, a tecnologia *blockchain*, por exemplo, pode assegurar a continuidade e funcionamento dos repositórios de dados indeterminadamente, promovendo a descentralização e o compartilhamento de dados.

Sales e Sayão, (2015) esclarecem que a adoção de um plano de gestão de dados (PGD) é essencial para detalhar os dados a serem processados, coletados ou gerados; as metodologias e padrões a serem empregados; as condições de compartilhamento e disponibilização dos dados; e as estratégias de curadoria e preservação dos mesmos.

Em resumo, a gestão de dados de pesquisa é fundamental para a ciência contemporânea, tratando os dados gerados pela pesquisa como um recurso valioso e sustentável. A implementação de práticas robustas de GDP não apenas reforça a integridade e confiabilidade da pesquisa, mas também fomenta uma cultura de compartilhamento e colaboração, cruciais para o progresso do conhecimento científico. Avançar para o planejamento estratégico em GDP é, portanto, um passo crítico para assegurar a coleta, processamento e gestão eficazes dos dados ao longo do ciclo de vida da pesquisa.

4.1 O planejamento da gestão de dados

O planejamento da gestão de dados (PGD) é essencial para a implementação de práticas eficazes na gestão de dados de pesquisa e Sales e Sayão (2015) destacam que o PGD estabelece o ciclo de vida dos dados que serão coletados, processados ou gerados em um projeto de pesquisa. Este documento estabelece uma promessa de cuidado apropriado com os dados durante e após o desenvolvimento do projeto. Funciona como uma orientação para pesquisadores e entidades, com o propósito de garantir que os dados sejam tratados de forma ética, conforme a legislação e eficazmente, maximizando seu valor e a possibilidade de serem reutilizados.

A importância do PGD é amplificada pelos requisitos de financiadores e editores de periódicos, especialmente em um contexto caracterizado pela proliferação de dados. No cenário atual, marcado pela abundância de dados, o plano de gestão de dados (PGD) rapidamente se estabelece como um documento fundamental na

rotina dos pesquisadores. Isso se deve ao fato de que, nos últimos anos, diversas agências financiadoras de pesquisa incluem em suas exigências para o financiamento de projetos de pesquisa a necessidade de apresentação de um plano de gestão e compartilhamento de dados nos pedidos de auxílio (Sales; Sayão, 2015). Além disso, os repositórios de dados desempenham um papel crucial na preservação de metadados e documentos digitais, facilitando a gestão de dados a longo prazo (Santos; Flores, 2015).

Assim, o planejamento eficaz na gestão de dados vai além da simples adesão a diretrizes externas, sendo uma medida crucial para aumentar a eficiência da pesquisa, promover a inovação e assegurar a integridade e o impacto na ciência. Neste contexto, Santos e Flores (2015) defendem que a implementação de padrões em gestão de dados potencializa a troca de conhecimentos entre as comunidades científicas e a propagação de metodologias, contribuindo significativamente para o desenvolvimento de novas pesquisas. Então, práticas robustas de planejamento em gestão de dados asseguram que os conhecimentos produzidos pela pesquisa sejam acessíveis, reutilizáveis e preservados para futuras gerações.

Esses requisitos visam promover a transparência, replicabilidade e reutilização de dados, princípios centrais da ciência aberta. Para Barrozo e Almeida (2021) No âmbito da ciência aberta, além da disponibilização de periódicos, a prática científica envolve a abertura da pesquisa de maneira ampla. Isso inclui a gestão ou curadoria de dados científicos, as anotações e a revisão por pares, facilitando assim o reuso de dados e a reprodutibilidade das pesquisas.

A crescente geração e utilização de dados em atividades acadêmicas e de pesquisa sublinham a necessidade premente de infraestruturas gerenciais e tecnológicas adequadas para gerir o ciclo de vida dos dados, desde o planejamento até o arquivamento seguro (Sales; Sayão, 2016). Com um PGD bem elaborado, o passo subsequente envolve a organização e documentação dos dados, garantindo sua acessibilidade e compreensibilidade para todos os envolvidos.

4.2 A organização e documentação de dados

A organização e a documentação de dados constituem pilares fundamentais para a gestão eficiente de dados de pesquisa, garantindo que os dados sejam não apenas acessíveis e compreensíveis, mas também reutilizáveis tanto pelo

pesquisador original quanto pela comunidade científica, em geral. Uma metodologia eficaz inicia com a descrição detalhada do projeto de pesquisa, facilitando a contextualização e justificativa para a coleta de dados específicos. Segue-se a descrição dos dados, abordando suas características técnicas e estruturais, conforme destacado por Sales e Sayão (2015). Práticas rigorosas em nomeação de arquivos, estruturação de pastas e versionamento de dados, complementadas pela elaboração de metadados detalhados, são essenciais para esses procedimentos.

Nesse sentido, o versionamento de dados emerge como uma prática crítica, possibilitando o acompanhamento de modificações e a recuperação de versões anteriores dos arquivos de dados. Mori *et al.* (2020) salientam que, na ausência de controle rigoroso sobre a produção de informação digital, a autenticidade dos dados pode ser comprometida, dificultando o rastreamento do produtor original. Assim, o uso de ferramentas de versionamento, como sistemas de controle de versão, automatiza esse processo e fornece um histórico detalhado das alterações realizadas nos dados.

Metadados, definidos como “dados sobre dados”, são cruciais para a documentação de conjuntos de dados, oferecendo informações vitais sobre a origem, contexto, qualidade e condições de uso dos dados. Na gestão de dados de pesquisa, é imperativo que os dados sejam compreensíveis e interpretáveis por qualquer usuário, atual ou futuro. Isso exige uma descrição precisa e abrangente dos dados, incluindo anotações e informações contextuais que permitam a transmissão de informação e conhecimento por meio do tempo e espaço, e enfatizam a importância da documentação acompanhante dos dados para atender a essa necessidade (Sales; Sayão, 2015).

A elaboração e manutenção de metadados detalhados não somente facilitam a acessibilidade e reutilização dos dados, mas também são fundamentais para sua preservação a longo prazo. Mori *et al.* (2020) argumentam que para preservar a integridade dos documentos digitais, é essencial armazenar adequadamente os metadados, destacando a importância do papel do profissional arquivista na identificação dos metadados relevantes para estabelecer a conexão do documento digital com sua entidade proprietária.

Em síntese, a organização e documentação de dados são etapas vitais no ciclo de vida da gestão de dados de pesquisa, assegurando a preservação significativa dos dados e sua disponibilidade para futuras investigações. A adoção de práticas recomendadas, como estratégias eficazes de nomeação de arquivos, estruturação de

pastas, versionamento de dados e a criação de metadados ricos, reforça a integridade, transparência e eficiência da pesquisa científica. Após a organização e documentação adequadas dos dados, o foco se desloca para o armazenamento e a segurança dos dados, visando proteger a integridade dos dados contra perdas e acessos não autorizados.

4.3 Armazenamento e segurança de dados

O armazenamento e a segurança dos dados representam aspectos críticos na gestão eficiente de dados de pesquisa, necessitando de uma abordagem meticulosa para prevenir perdas, corrupção ou acessos não autorizados. Gangadevi e Renuka Devi (2021) ressaltam que os serviços de armazenamento em nuvem terceirizados introduzem avanços significativos em desempenho, escalabilidade, custo-benefício e independência geográfica para o gerenciamento de dados. A diversidade de opções de armazenamento em nuvem, como *One Drive*, *Amazon Web Services*, *Google Drive* e *Dropbox*, exige assim a seleção cuidadosa de soluções de armazenamento e a adoção de práticas de segurança recomendadas para garantir a integridade e confidencialidade dos dados ao longo de seu ciclo de vida.

As opções para a preservação segura dos dados variam desde alternativas locais, como servidores próprios e dispositivos de armazenamento físico, até opções em nuvem, oferecidas por empresas especializadas que permitem o acesso aos dados remotamente. Neste aspecto, Deng *et al.* (2019) destacam a necessidade de criar repositórios digitais seguros para o armazenamento, gerenciamento, acesso e disseminação de dados, além da sua transição para novos formatos e mídias, abrangendo aspectos técnicos, políticas de gestão e suporte institucional.

O armazenamento em nuvem oferece vantagens notáveis em termos de escalabilidade, flexibilidade e acessibilidade, permitindo aos pesquisadores acessar dados de qualquer local e a qualquer momento. Ren *et al.* (2021) destacam que a continuidade digital é essencial para manter a informação completa e acessível, sem exceder o período necessário para sua preservação. A usabilidade é, portanto, o objetivo principal do planejamento de continuidade digital. Contudo, a dependência de serviços terceirizados para a gestão de dados sensíveis ou confidenciais exige uma análise detalhada das políticas de privacidade e segurança dos provedores.

Para assegurar a segurança dos dados, é fundamental a implementação de *backups* regulares, *criptografia* de dados sensíveis e uma gestão de acessos rigorosa. Sales e Sayão (2015) sublinham a importância de realizar *backups* atualizados regularmente como uma medida preventiva contra perdas acidentais ou intencionais, facilitando a restauração dos dados originais. A frequência e o método de *backup* devem ser determinados com base na criticidade e na volatilidade dos dados.

A *criptografia* de dados serve como uma barreira contra o acesso não autorizado, transformando informações sensíveis em formatos ilegíveis sem a chave de *descriptografia* apropriada. Shamshad *et al.* (2020) complementam que a assinatura digital e a *criptografia* são métodos eficazes para garantir a integridade e a confidencialidade dos dados.

Além disso, a gestão de acessos desempenha um papel crucial na segurança dos dados, envolvendo a atribuição de permissões de usuário baseadas no princípio do menor privilégio, limitando o acesso às funcionalidades estritamente necessárias para a execução de suas tarefas. A segurança é uma preocupação preeminente na computação em nuvem, abrangendo questões como acessibilidade, autenticação, integridade de dados, recuperação de dados e confidencialidade (Gangadevi; Renuga Devi, 2021).

Em síntese, o armazenamento e a segurança dos dados de pesquisa exigem uma estratégia integrada, que combine soluções tecnológicas avançadas com práticas de segurança rigorosas. A decisão entre armazenamento local e em nuvem deve ser baseada em uma avaliação criteriosa das necessidades específicas do projeto, enquanto a implementação de *backups* regulares, *criptografia* e gestão de acessos são indispensáveis para a proteção dos dados contra uma variedade de riscos.

A discussão sobre armazenamento e segurança dos dados conduz naturalmente à consideração de questões éticas e legais, abordando a coleta, compartilhamento e uso dos dados de maneira a respeitar a privacidade dos participantes e cumprir com as regulamentações aplicáveis. Na figura 14 apresentamos um esquema com a discussão sobre armazenamento e segurança de dados apresentadas nesse tópico.

Figura 14 - Armazenamento e segurança de dados



Fonte: Elaborado pelo autor.

4.4 A Ética e a legalidade na gestão de dados de pesquisa

A ética e a legalidade representam pilares fundamentais na gestão de dados de pesquisa, abrangendo a responsabilidade e a integridade na coleta, armazenamento, processamento e compartilhamento dos dados. Sales e Sayão (2015) sublinham a importância de um comportamento ético rigoroso por parte dos pesquisadores, especialmente quando a pesquisa envolve a obtenção de dados de pessoas. Este comportamento deve estar alinhado com os códigos de ética, padrões e protocolos recomendados por entidades profissionais, instituições de pesquisa e organizações financiadoras, além de estar consoante com a legislação aplicável do país.

A observância desses princípios éticos e legais é crucial em todas as fases do ciclo de pesquisa, com ênfase particular na fase de compartilhamento de dados, exatamente para assegurar o respeito pelos direitos e pela privacidade dos participantes e para cumprir com as legislações nacionais e internacionais relevantes (Sales; Sayão, 2015).

A gestão ética e legal dos dados de pesquisa requer uma abordagem metódica e deliberada, que harmonize o imperativo do avanço científico com o respeito pelos direitos e dignidade dos participantes da pesquisa. A implementação de práticas éticas robustas e a adesão à legislação vigente são essenciais para sustentar a confiança pública na ciência e garantir a integridade da pesquisa.

Ao navegar pelas complexidades éticas e legais, adentramos no território do compartilhamento e reuso de dados, investigando as maneiras pelas quais os dados podem ser disponibilizados de forma responsável, visando beneficiar a comunidade científica de forma mais ampla. Este aspecto destaca a importância de estratégias de gestão de dados que não apenas cumpram com os requisitos éticos e legais, mas também promovam a colaboração e o avanço do conhecimento científico.

4.5 Compartilhamento e reuso de dados na pesquisa científica

O compartilhamento e o reuso de dados de pesquisa constituem práticas essenciais na ciência moderna, essenciais para promover a transparência, a replicabilidade e a eficiência na geração de conhecimento científico. A prática de disponibilizar os dados gerados e coletados em repositórios que garantam sua preservação ativa e de longo prazo é crucial. Isso envolve manter as características de autenticidade, integridade e proveniência dos dados, assegurando que permaneçam acessíveis e aptos para uso futuro (Sales; Sayão, 2015). Tais práticas são fortemente incentivadas por financiadores, editores de periódicos e pela comunidade científica, em geral, visando maximizar o valor dos dados de pesquisa e estimular novas descobertas.

O compartilhamento e o reuso de dados de pesquisa trazem consigo oportunidades significativas para o avanço do conhecimento científico e nesse sentido Zhang (2021) destaca que a tecnologia *blockchain* é caracterizada por apresentar funcionalidades fundamentais, incluindo o compartilhamento de informações. O compartilhamento efetivo de dados conduz à questão da preservação de dados,

enfatizando a importância de assegurar que conjuntos de dados valiosos permaneçam acessíveis e utilizáveis para pesquisas futuras e para gerações vindouras.

Esta etapa final no ciclo de vida dos dados de pesquisa reforça a necessidade de infraestruturas robustas e políticas claras que suportem tanto a preservação quanto o acesso contínuo aos dados científicos, garantindo que o legado do conhecimento científico seja mantido para o benefício da sociedade na totalidade. A figura 15 apresenta um esquema com a discussão sobre compartilhamento e reuso de dados na pesquisa científica apresentados nesse tópico.

Figura 15 - Compartilhamento e reuso de dados na pesquisa científica



Fonte: Elaborado pelo autor.

4.6 A Preservação de dados na pesquisa científica

A preservação de dados de pesquisa é um aspecto crítico da gestão de dados, vital para garantir que os dados permaneçam acessíveis, utilizáveis e compreensíveis a longo prazo, apesar da evolução tecnológica e da obsolescência dos formatos de arquivo. A obsolescência tecnológica e a fragilidade das mídias de armazenamento representam desafios significativos para a preservação da memória digital, podendo resultar na perda de documentos digitais (Sayão, 2010). Diante dessa vulnerabilidade em termos de *hardware*, *software* e suporte, a adoção de estratégias de preservação

digital torna-se imperativa para assegurar o acesso contínuo, a confiabilidade e a integridade dos documentos a longo prazo.

Considerando essa fragilidade ao nível de *hardware*, *software* e suporte, a aplicação de estratégias de preservação para documentos digitais é uma prioridade, pois sem elas não existiria nenhuma garantia de acesso, confiabilidade e integridade dos documentos de longo prazo. Então,

[...] o acesso e a usabilidade dos recursos informacionais digitais é impactado fortemente pela sua dependência a contextos tecnológicos específicos; esse fato gera uma área de tensão e complexidade na gestão de acervos digitais. A fragilidade estrutural da informação digital configura um dos maiores desafios a ser enfrentado pelos pesquisadores e profissionais das áreas de informação e de tantas outras áreas, neste começo de século (Sayão, 2010, p. 3).

Para sustentar o legado científico e facilitar o reuso de dados em futuras pesquisas, é essencial implementar estratégias eficazes de preservação de dados. Isso engloba a escolha de formatos de arquivo resilientes, assegurando a interoperabilidade, implementando *backups* periódicos e mantendo metadados minuciosos. No âmbito da gestão de dados de pesquisa, as práticas de garantia e controle de qualidade dizem respeito aos processos destinados a evitar e reduzir a incidência de falhas em conjuntos de dados, bem como à detecção de informações imprecisas. Para assegurar a alta qualidade dos dados, tornando-os mais úteis para aplicações presentes e futuras, diversas abordagens devem ser implementadas ao longo de todo o ciclo de desenvolvimento de um projeto de pesquisa (Sales; Sayão, 2015).

Essa interoperabilidade, facilitada pelo uso de padrões e protocolos comuns, é crucial para a acessibilidade e utilidade dos dados em diferentes sistemas e aplicações. *Backups* regulares e a replicação de dados em múltiplos locais físicos e digitais são medidas preventivas contra a perda de dados devido a falhas de hardware, desastres naturais ou obsolescência tecnológica. Dessa forma, Márdero Arellano (2022) ressalta a importância de ferramentas que protejam e assegurem a manutenção adequada dos documentos digitais e em papel. A seleção de repositórios de dados adequados é fundamental na preservação de dados de pesquisa. Repositórios de dados são infraestruturas especializadas no armazenamento, preservação e disseminação de conjuntos de dados científicos.

Em complemento, Márdero Arellano (2022) destaca a importância de os repositórios digitais incorporarem ferramentas específicas para preservação digital, as quais facilitam operações de identificação, validação, caracterização, arquivamento e garantia de acesso de longa duração a documentos digitais. Por outro lado, Santos e Flores (2015) defendem que estabelecer repositórios digitais de confiança, suportados por *softwares* especializados, é essencial para manter o acesso ininterrupto às informações ao longo do tempo.

Em síntese, a preservação de dados de pesquisa é uma responsabilidade científica fundamental, exigindo uma abordagem proativa e planejada. Por meio da implementação de estratégias de preservação a longo prazo e da escolha cuidadosa de repositórios de dados adequados, a comunidade científica pode garantir que os valiosos dados de pesquisa permaneçam acessíveis e úteis para futuras gerações, contribuindo assim para o progresso contínuo do conhecimento científico. Com os dados devidamente preservados, a atenção se volta para as ferramentas e recursos que facilitam todas as etapas da gestão de dados, apoiando os pesquisadores em suas práticas de gestão de dados eficazes.

4.7 Ferramentas e recursos para a gestão eficiente de dados de pesquisa

Para Ferreira (2006), uma parcela considerável da produção intelectual é realizada com o suporte de ferramentas digitais. A gestão eficaz de dados de pesquisa é apoiada por uma ampla gama de ferramentas e recursos, projetados para facilitar as diversas etapas do ciclo de vida dos dados, desde a coleta e armazenamento até a análise, compartilhamento e preservação.

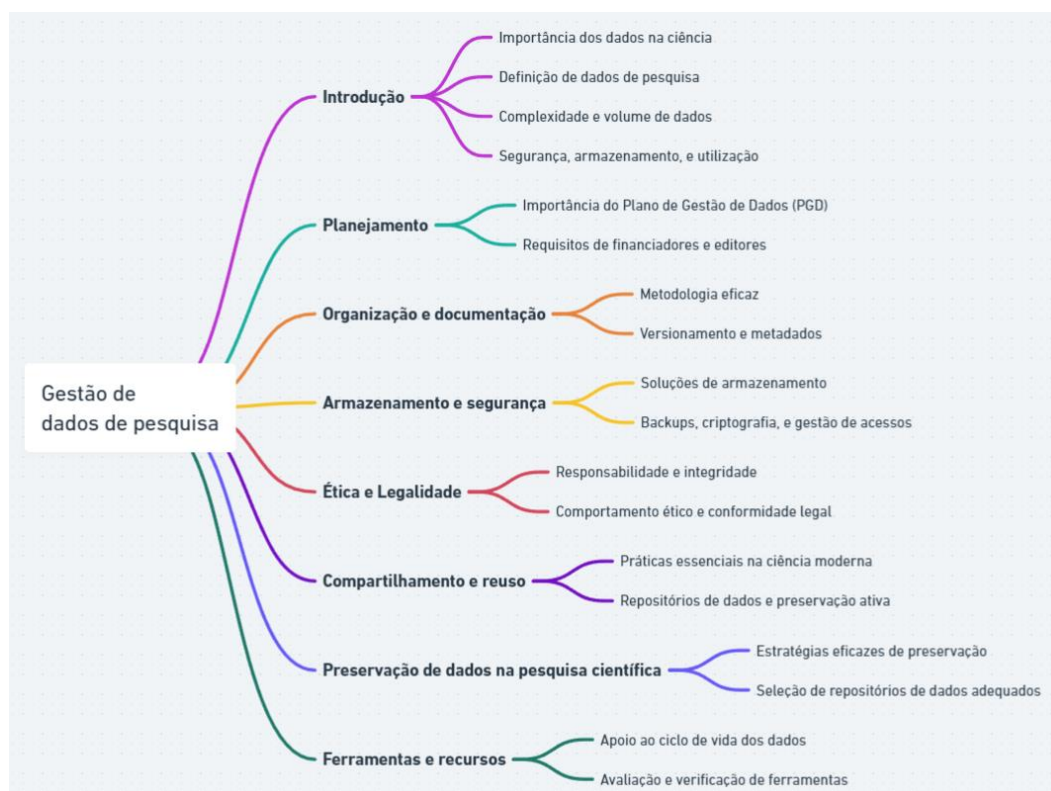
O repositório digital, por exemplo, deve constituir um ambiente autêntico voltado para a preservação de longo prazo, equipado, por exemplo, com ferramentas destinadas à implementação de estratégias de preservação e à inserção de padrões de metadados. Nesse contexto, todas as ações executadas sobre os documentos digitais, tais como migrações, devem ser devidamente registradas, construindo-se, assim, um histórico detalhado de cada objeto digital armazenado, o que adiciona uma camada de confiabilidade aos conteúdos (Santos; Flores, 2015). A escolha das ferramentas adequadas, juntamente com o acesso a recursos de treinamento e apoio, é fundamental para os pesquisadores poderem gerir seus dados de forma eficiente e conforme as melhores práticas e padrões da área.

A seleção criteriosa de ferramentas, aliada ao acesso a recursos de treinamento e suporte, é indispensável para os pesquisadores gerenciarem seus dados de maneira eficiente e alinhada às melhores práticas e padrões do campo. Santos e Flores (2015) ressaltam que a garantia de acesso a longo prazo aos dados está intrinsecamente ligada à eficácia das ferramentas de preservação utilizadas. Portanto, é essencial realizar avaliações periódicas e verificações constantes dessas ferramentas para assegurar sua funcionalidade e eficiência ao longo do tempo.

Em síntese, existe uma vasta gama de ferramentas e recursos disponíveis para facilitar a gestão de dados de pesquisa em todas as fases do seu ciclo de vida. A escolha cuidadosa dessas ferramentas, combinada com o acesso a treinamento de qualidade e suporte, é fundamental para os pesquisadores poderem gerenciar seus dados eficazmente.

Essa abordagem não só promove a integridade e acessibilidade dos dados de pesquisa, mas também potencializa sua reutilização, contribuindo significativamente para o avanço do conhecimento científico. Na figura 16 apresentamos um esquema com os tópicos apresentados até o momento.

Figura 16 – Tópicos apresentados no capítulo gestão de dados de pesquisa



Fonte: Elaborado pelo autor.

No próximo tópico apresentamos as características de proveniência, rastreabilidade, confiabilidade e autenticidade dos dados de pesquisa.

4.8 Características de proveniência, rastreabilidade, confiabilidade e autenticidade dos dados de pesquisa

A gestão de dados na Ciência da Informação, especialmente em face do aumento exponencial do volume de dados gerados por pesquisas, destaca a importância de características como proveniência, rastreabilidade, confiabilidade e autenticidade. Oliveira (2023) ressalta que a adequada gestão desses dados não apenas eleva a eficiência da pesquisa, mas também garante a integridade e a reprodutibilidade dos resultados científicos.

Neste contexto, a tecnologia *blockchain* surge como uma solução inovadora, oferecendo um registro descentralizado e imutável de transações. Tal característica se mostra fundamental para a melhoria da proveniência e da rastreabilidade de produtos e informações, baseando-se na segurança e transparência inatas ao sistema para fomentar a confiança digital (Omame; Alex-Nmecha, 2021). Desde sua inovação inicial por Nakamoto em 2008, a *blockchain* visava solucionar o problema do gasto duplo em sistemas de pagamento eletrônico, mas rapidamente seu potencial se estendeu para além das aplicações financeiras, destacando-se pela capacidade de verificar a autenticidade e o histórico de produtos ou informações sem dependência de autoridades centralizadas.

A incorporação da *blockchain* na pesquisa científica introduz um novo paradigma na gestão de dados, enfatizando a importância de assegurar a integridade, autenticidade e imutabilidade dos registros de dados. Isso não apenas melhora a compreensão do contexto de coleta e processamento dos dados, mas também facilita a verificação de sua autenticidade e integridade ao longo do tempo. De acordo com Bell *et al.* (2019), essa tecnologia inovadora amplia a confiança nos dados científicos, incentivando a colaboração e o compartilhamento de conhecimento entre pesquisadores.

A implementação da *blockchain* na gestão de dados de pesquisa científica, portanto, representa um passo significativo em direção à maior transparência, responsabilidade e confiabilidade na ciência. Tal abordagem não só reforça a

confiança nos resultados de pesquisas, mas também promove uma cultura de integridade e reprodutibilidade essencial para o avanço científico.

Apresentar as características de proveniência, rastreabilidade, confiabilidade e autenticidade dos dados de pesquisa científica é essencial para compreender a importância e o impacto dessas qualidades na integridade e eficácia da pesquisa. No Quadro 2, essas características são detalhadamente descritas, ressaltando-se a relevância de cada uma delas, juntamente com as contribuições dos autores que discutiram esses conceitos nesta tese.

Quadro 2 - Características de proveniência, rastreabilidade, confiabilidade e autenticidade dos dados

Característica	Descrição	Importância	Autores
Proveniência	Refere-se à origem e ao histórico de transformações dos dados.	Fundamental para compreender o contexto de coleta e processamento dos dados.	Bell <i>et al.</i> (2019); Collomosse <i>et al.</i> (2018); Green <i>et al.</i> (2018); Márdero Arellano (2008); Omame e Alex-Nmecha (2021); Sales e Sayão (2015); Washington Segundo <i>et al.</i> (2023); Swan (2015); Tapscott e Tapscott (2016)
Rastreabilidade	Capacidade de identificar e acompanhar os dados ao longo de seu ciclo de vida.	Indispensável para a verificação e validação dos dados.	Casey e Vigna (2018); Gangadevi e Renuga Devi (2021); Mori <i>et al.</i> (2020); Mougayar (2016); Omame e Alex-Nmecha (2021); Sales e Sayão (2016); Washington Segundo <i>et al.</i> (2023); Swan (2015); Tapscott e Tapscott (2016)
Confiabilidade	Refere-se à consistência e precisão dos dados ao longo do tempo.	Crucial para a tomada de decisões informadas.	Alzahrani <i>et al.</i> (2022); Deng <i>et al.</i> (2019); Gangadevi e Renuga Devi (2021); Gava e Flores (2020); Gil (1999); Green <i>et al.</i> (2018); Lima e Rabello; (2020); Liu, He e Xuan (2021); Márdero Arellano (2022); Narayanan <i>et al.</i> (2016); Santos e Flores (2015); Sayão (2010); Washington Segundo <i>et al.</i> (2023); Tapscott e Tapscott (2016); Woodall e Ringel (2019)
Autenticidade	Assegura que os dados são genuínos e não foram alterados indevidamente.	Essencial para a credibilidade da pesquisa.	Androulaki <i>et al.</i> (2018); Bell <i>et al.</i> (2019); Deng <i>et al.</i> (2019); Green <i>et al.</i> (2018); Márdero Arellano (2008); Mori <i>et al.</i> (2020); Narayanan <i>et al.</i> (2016); Oliveira (2023); Omame e Alex-Nmecha (2021); Sales e Sayão (2015); Santos e Flores (2017); Skala <i>et al.</i> (2023); Swan (2015); Tapscott e Tapscott (2016); Woodall e Ringel (2019)

Fonte: Elaborado pelo autor.

Em seguida, na figura 17 apresentamos um esquema com as características de proveniência, rastreabilidade, confiabilidade e autenticidade dos dados referente ao quadro 2.

Figura 17 - Características dos dados

Fonte: Elaborado pelo autor.

No capítulo 5, a discussão se concentra nos modelos semânticos e nas ferramentas de representação. Apresentamos a correlação entre os conceitos da tecnologia *blockchain* e os da gestão de dados, propondo esquemas de modelos semânticos para o uso e a aplicação prática da *blockchain*.

5 MODELOS SEMÂNTICOS

Modelos semânticos (um padrão de dados que auxilia na representação de objetos e suas conexões). A elaboração de modelos semânticos constitui um avanço notável no processamento e na gestão de vastos volumes de dados, particularmente relevante num contexto global crescentemente pautado por informações complexas e interligadas. A adoção de ferramentas de representação na construção desses modelos emerge como prática indispensável, viabilizando a organização e interpretação de dados de forma acessível tanto para humanos quanto para máquinas.

Vavilenkova (2015) ressalta a importância de superar as limitações das teorias preexistentes e de desenvolver modelos que integrem a estrutura semântico-sintática das sentenças, ilustrando a complexidade inerente à modelagem semântica. Complementarmente, Padó e Lapata (2007) enfatizam a necessidade de incorporar relações sintáticas na elaboração de espaços semânticos, visando aprimorar o processamento de linguagem natural.

Estudos como o de Wang, Wang e Acero (2004) demonstram a capacidade dos modelos de linguagem semântica em interpretar diálogos conversacionais com eficácia, enquanto Bellegarda (2000) investiga a utilização da análise semântica latente para aprimorar modelos de linguagem estatísticos, evidenciando a importância das relações semânticas de longo alcance. Adicionalmente, Sheth e Ranabahu (2010) exploram as vantagens dos modelos semânticos na computação em nuvem, desde a especificação de funcionalidades até a modelagem de dados e descrição de serviços.

Ferramentas de representação, incluindo linguagens de marcação e *frameworks* especializados, são fundamentais para possibilitar a construção precisa e a visualização desses modelos. Elas são cruciais para a elaboração de modelos que não somente armazenam informações eficientemente, mas também facilitam sua interpretação, análise e compartilhamento de maneira intuitiva e significativa. Este enfoque é corroborado por pesquisas recentes no campo dos modelos semânticos, que evidenciam o papel central dessas ferramentas na promoção de um entendimento mais profundo e na gestão eficaz de dados complexos.

Assim, o próximo tópico exemplifica as ferramentas de representação usadas nesta pesquisa.

5.1 A escolha das ferramentas de representação

A criação de mapas conceituais ou mentais é uma prática fundamental para a visualização e organização de ideias complexas, facilitando a compreensão e o compartilhamento de conhecimento (Eppler, 2006). Nessa tese optamos pela integração do uso de ferramentas como *Lucidchart*, *MindMup*, juntamente com o uso do código *Markdown* para a elaboração desses mapas, oferece uma variedade de vantagens que justificam sua escolha em ambientes acadêmicos, profissionais e de pesquisa.

A seleção dessas ferramentas baseia-se em suas características distintas e relevância no contexto de planejamento e organização de ideias em ambientes acadêmicos e profissionais. A investigação foi conduzida por meio de uma análise prática e reflexiva, empregando a metodologia de pesquisa-ação, onde o autor deste estudo imergiu na utilização dessas ferramentas para produzir e colaborar em mapas conceituais. O objetivo foi avaliar, de maneira empírica, as capacidades e limitações de cada plataforma, considerando, fatores como usabilidade, flexibilidade, integração com outras ferramentas, e a adequação ao processo criativo e colaborativo em contextos acadêmicos.

*Lucidchart*³ destaca-se pela sua interface intuitiva e recursos de colaboração em tempo real. Sua capacidade de integrar-se com diversas plataformas e a variedade de modelos oferecidos tornam a criação de mapas conceituais tanto eficiente quanto flexível. Essa ferramenta é ideal para equipes que buscam colaborar em projetos complexos, permitindo a edição simultânea e a partilha fácil de diagramas.

*MindMup*⁴ é outra ferramenta poderosa, conhecida por sua simplicidade e eficácia. Com *MindMup*, usuários podem criar rapidamente mapas mentais que são claros e visualmente atraentes, sem a necessidade de um treinamento extensivo na ferramenta. É especialmente útil para *brainstorming* individual ou em grupo, auxiliando na estruturação de ideias e na promoção do pensamento criativo.

Por outro lado, o código *Markdown*⁵ oferece uma abordagem mais textual e é altamente eficaz para quem prefere trabalhar com linguagem de marcação. Sua simplicidade e compatibilidade com uma variedade de plataformas digitais o tornam

³ <https://www.lucidchart.com/pages/examples/sitemap-generator> acessado em 15/10/2023.

⁴ <https://www.mindmup.com/> acessado em 15/10/2023.

⁵ <https://markmap.js.org/> acessado em 15/10/2023.

uma escolha versátil para a criação de mapas conceituais. Além disso, a natureza textual do *Markdown* facilita a integração desses mapas em documentos e apresentações, mantendo um formato claro e acessível.

A incorporação de ferramentas como *Lucidchart*, *MindMup* e o código *Markdown* na elaboração de figuras para esta tese demonstra um uso estratégico de recursos digitais modernos para aprimorar a representação e comunicação de conceitos complexos. Cada ferramenta, com suas características únicas, contribuiu de maneira significativa para a organização visual das ideias, facilitando a compreensão e o engajamento com o material apresentado.

A escolha dessas ferramentas reflete uma abordagem consciente na seleção de recursos que aliam eficiência, colaboração e flexibilidade, destacando a importância da representação visual no processo de disseminação do conhecimento acadêmico. A integração bem-sucedida de *Lucidchart*, *MindMup* e *Markdown* na elaboração das figuras desta tese não somente facilitou a organização e apresentação de informações complexas, mas também destacou o papel fundamental que ferramentas digitais avançadas desempenham na comunicação científica contemporânea. No próximo tópico elaboramos um esquema por meio do uso de ferramentas de representação para fazer a correlação entre conceitos e esquemas para uso da *blockchain* na gestão de dados de pesquisa.

5.2 Correlação entre os conceitos da tecnologia *blockchain* e os da gestão de dados

Visando simplificar a compreensão e estabelecer uma relação entre *blockchain* e gestão de dados, elaboramos o quadro 3, que explora como os princípios básicos da *blockchain* têm correspondências significativas na área de gestão de dados de pesquisa. Isso demonstra a conexão e a importância dessas tecnologias em variados aspectos do gerenciamento de informações.

A correlação entre os conceitos fundamentais da tecnologia *blockchain* e os princípios da gestão de dados oferece uma visão perspicaz sobre como as inovações em *blockchain* podem ser aplicadas para melhorar a gestão e a integridade dos dados.

Quadro 3 - Correlação entre os conceitos da tecnologia *blockchain* e os da gestão de dados

Conceitos	Correlação	Autores
Descentralização X Distribuição descentralizada	Na <i>blockchain</i> , a descentralização distribui autoridade e controle entre os participantes, eliminando a centralização. Isso é feito por meio de um consenso distribuído, com cada nó mantendo uma cópia do registro de transações. Na gestão de dados, a descentralização envolve a dispersão dos dados em vários locais para garantir segurança e redundância. Enquanto na <i>blockchain</i> foca-se em transparência e imutabilidade, na gestão de dados busca-se disponibilidade e resistência a falhas, visando maior segurança e eficiência em ambas as áreas.	Washington Segundo <i>et al.</i> (2023); Swan (2015); Androulaki <i>et al.</i> (2018); Collomosse <i>et al.</i> , 2018); Tapscott e Tapscott (2016); Gonçalves e Rodrigues (2020); Satoshi Nakamoto (2008); Narayanan <i>et al.</i> (2016); Oliveira (2023); Green <i>et al.</i> (2018)
Imutabilidade X Preservação e integridade	A imutabilidade na <i>blockchain</i> garante que transações registradas não podem ser alteradas, assegurada por um consenso que mantém cópias idênticas dos registros em cada nó, aumentando a transparência e confiabilidade. Em contraste, na gestão de dados, a ênfase é na preservação e integridade dos dados, utilizando <i>backups</i> e políticas de segurança para manter a precisão. Enquanto a <i>blockchain</i> foca na veracidade das transações, a gestão de dados visa manter a confiabilidade dos dados, ambas essenciais para segurança e eficiência.	Altamira (2024); Androulaki <i>et al.</i> (2018); Bell <i>et al.</i> (2019); Bralić; Stančić; Stengård, (2020); Capurro e Hjørland (2007); Chen <i>et al.</i> (2018); Collomosse <i>et al.</i> (2018); Deng <i>et al.</i> (2019); Ferreira (2006); Gabriel Junior <i>et al.</i> (2020); Gangadevi e Renuga Devi (2021); Gava e Flores (2020); Green <i>et al.</i> (2018); Lima e Rabello (2020); Márdero Arellano (2008); Narayanan <i>et al.</i> (2016); Ren <i>et al.</i> (2021); Sales e Sayão (2024); Santos e Flores (2017); Santos <i>et al.</i> (2018); Saracevic (1996); Satoshi Nakamoto (2008); Sayão (2015); Washington Segundo <i>et al.</i> (2023); Shamshad <i>et al.</i> (2020); Souza e Aganette (2022); Souza <i>et al.</i> (2012); Swan (2015); Tapscott e Tapscott (2016); Thomaz e Soares (2004)
Transparência X Transparência na pesquisa	Na <i>blockchain</i> , a transparência permite que todos verifiquem transações e históricos, graças à sua natureza distribuída e imutável, garantindo visibilidade completa. Na gestão de dados de pesquisa, a transparência envolve a divulgação completa de métodos, resultados e conclusões, permitindo replicação e validação. Essencial para a confiança nas transações na <i>blockchain</i> e para a credibilidade na pesquisa, ambos os conceitos	Barrozo e Almeida (2021); Gangadevi e Renuka Devi (2021); Hwang, Shon e Park (2020); Lima e Rabello (2020); Lima e Rabello (2020); Mori <i>et al.</i> (2020); Mori <i>et al.</i> (2020); Mougayar (2016); Oliveira (2023); Sales e Sayão (2015); Santos e Flores (2015); Satoshi Nakamoto (2008); Washington Segundo <i>et al.</i> (2023); Shamshad <i>et al.</i> (2020); Skala

Conceitos	Correlação	Autores
	reforçam a confiabilidade e integridade nas suas áreas.	<i>et al.</i> (2023); Swan (2015); Tapscott e Tapscott (2016)
Prova de Trabalho/Participação X Validação e verificação	Na <i>blockchain</i> , a prova de trabalho (PoW) e a prova de participação (PoS) validam transações e adicionam blocos à cadeia. O PoW resolve problemas matemáticos, garantindo segurança à custa de recursos computacionais, enquanto o PoS depende da posse de moeda para validar transações. Na gestão de dados, a validação assegura a correção e consistência dos dados por meio de controles de qualidade. PoW e PoS são vitais para a segurança e descentralização da <i>blockchain</i> , assim como a validação é essencial para a precisão dos dados, sendo ambos cruciais para a qualidade e confiabilidade nas suas áreas.	Androulaki <i>et al.</i> (2018); Bell <i>et al.</i> (2019); Collomosse <i>et al.</i> (2018); Gonçalves e Rodrigues (2020); Green <i>et al.</i> (2018); Márdero Arellano (2022); Mougayar (2016); Ouchi e Arakaki, (2020); Sales e Sayão (2020); Satoshi Nakamoto (2008); Tapscott e Tapscott (2016)
Contratos Inteligentes X Automação de políticas	Contratos inteligentes na <i>blockchain</i> são roteiros que automatizam os termos de um contrato, ativando-se quando condições pré-definidas são atendidas, eliminando intermediários e garantindo transparência e imutabilidade. Na gestão de dados, a automação de políticas aplica procedimentos automatizados para gerir o acesso e proteção de dados, promovendo segurança e conformidade regulatória. Enquanto contratos inteligentes simplificam a execução de contratos, a automação de políticas melhora processos e segurança. Ambos visam à eficiência e segurança nas suas respectivas áreas.	Collomosse <i>et al.</i> (2018); Deng <i>et al.</i> (2019); Gomes e Coutinho (2022); Márdero Arellano (2008); Mori <i>et al.</i> (2020); Ren <i>et al.</i> (2021); Tapscott e Tapscott (2016); Williams (2023)
Cadeia de Blocos X Cadeias de registros	A <i>blockchain</i> é uma estrutura composta por blocos conectados, contendo transações validadas e links para blocos anteriores, garantindo a integridade e imutabilidade das informações. Sua natureza descentralizada assegura a segurança e transparência das transações. Em contraste, cadeias de registros na gestão de dados são sequências cronológicas de registros que documentam atividades, facilitando auditorias e a manutenção de históricos para garantir precisão e confiabilidade. A <i>blockchain</i> é crucial para a segurança das transações digitais, enquanto cadeias de registros asseguram a	Androulaki <i>et al.</i> (2018); Bell <i>et al.</i> (2019); Bralić; Stančić; Stengård, (2020); Casey e Vigna (2018); Collomosse <i>et al.</i> (2018); Deng <i>et al.</i> (2019); Feng <i>et al.</i> (2019); Gabriel Junior <i>et al.</i> (2020); Gonçalves e Rodrigues (2020); Green <i>et al.</i> (2018); Hwang, Shon e Park (2020); Lemieux (2017); Mori <i>et al.</i> (2020); Narayanan <i>et al.</i> (2016); Oliveira (2023); Omame e Alex-Nmecha (2021); Ouchi e Arakaki (2020); Ren <i>et al.</i> (2021); Santos e Maimone (2023); Saracevic (1996); Washington Segundo <i>et al.</i> (2023); Swan (2015); Tapscott

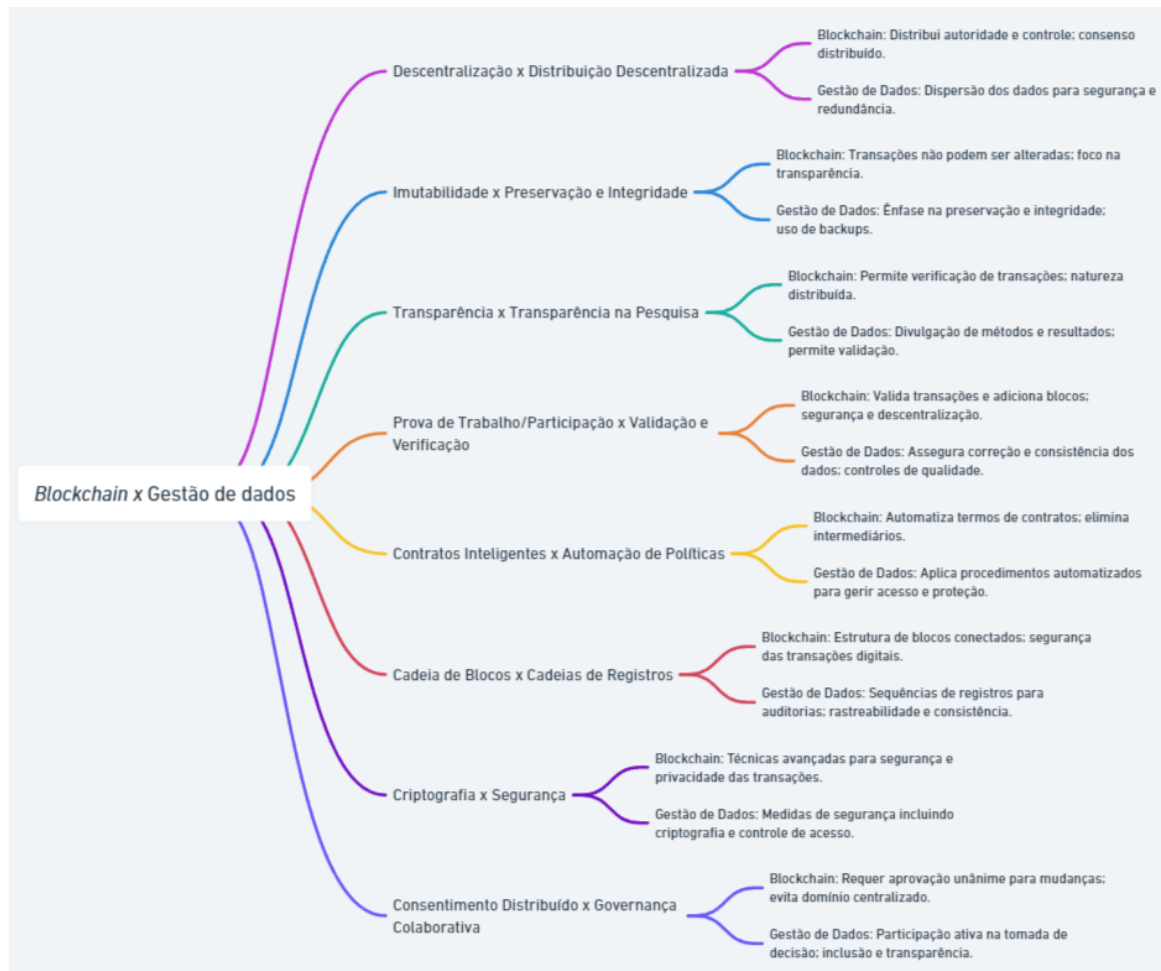
Conceitos	Correlação	Autores
	rastreabilidade e consistência dos dados, ambos promovendo confiabilidade e segurança em suas áreas.	e Tapscott (2016); Woodal e Ringel (2019)
Criptografia X Segurança	A <i>criptografia</i> na <i>blockchain</i> utiliza técnicas avançadas e algoritmos robustos para garantir a segurança, privacidade, autenticidade e integridade das transações. Essencial para a confidencialidade e proteção contra fraudes e ataques, complementa-se com a segurança na gestão de dados, que inclui medidas como <i>criptografia</i> , controle de acesso e estratégias de <i>backup</i> para proteger dados, assegurando conformidade regulatória. Ambas visam à proteção das informações, promovendo confidencialidade, integridade e disponibilidade em seus respectivos campos.	Altamira (2024); Bell <i>et al.</i> (2019); Collomosse <i>et al.</i> (2018); Gangadevi e Renuka Devi (2021); Gonçalves e Rodrigues (2020); Hwang, Shon e Park (2020); Lemieux (2017); Liu, He e Xuan (2021); Mori <i>et al.</i> (2020); Mougayar (2016); Nakamoto (2008); Narayanan <i>et al.</i> (2016); Oliveira (2023); Omame e Alex-Nmecha (2021); Washington Segundo <i>et al.</i> (2023); Shamshad <i>et al.</i> (2020); Tapscott e Tapscott (2016); Voshmgir (2020); Woodall e Ringel (2019); Zhang (2021)
Consentimento Distribuído X Governança colaborativa	O consentimento distribuído na <i>blockchain</i> requer a aprovação unânime dos participantes para mudanças no sistema, promovendo decisões transparentes e equitativas e evitando o domínio centralizado. Na gestão de dados, a governança colaborativa envolve a participação ativa de várias partes na tomada de decisão, visando a inclusão e transparência. Ambos enfatizam a democracia e justiça nos processos decisórios de seus campos.	Sales e Sayão (2024); Androulaki <i>et al.</i> (2018); Chen <i>et al.</i> (2018); Collomosse <i>et al.</i> (2018); Gangadevi e Renuka Devi (2021); Gonçalves e Rodrigues (2020); Green <i>et al.</i> (2018); Hwang, Shon e Park (2020); Lemieux (2017); Liu, He e Xuan (2021); Rekunen <i>et al.</i> (2022); Santos <i>et al.</i> (2018); Washington Segundo <i>et al.</i> (2023)

Fonte: Elaborado pelo autor.

O Quadro 3 ilustra a importância de cada uma dessas características para a gestão eficaz dos dados de pesquisa, conforme discutido pelos autores citados. A incorporação de tecnologias como a *blockchain* na pesquisa científica apresenta um avanço significativo na maneira como essas características são asseguradas, trazendo benefícios como maior transparência, confiabilidade e facilidade de verificação. Este entendimento é crucial para avançar na integridade e na eficácia da pesquisa científica, promovendo um ambiente de maior confiança e colaboração entre pesquisadores.

Na figura 18 apresentamos um esquema da correlação entre os conceitos de *blockchain* e da gestão de dados apresentados no quadro 3.

Figura 18 - Correlação entre os conceitos da tecnologia *blockchain* e os da gestão de dados



Fonte: Elaborado pelo autor.

No próximo tópico apresentamos uma proposta de esquema de modelo semântico para uso da *blockchain* na gestão de dados de pesquisa.

5.3 Proposta de esquema de modelo semântico para uso da *blockchain* na gestão de dados de pesquisa

Essa proposta de esquema de modelo semântico para uso da *blockchain* na gestão de dados de pesquisa tem uma abordagem inspirada em Beck *et al.* (2017) e Sfetcu (2023).

A tecnologia *blockchain* oferece um mecanismo confiável para o armazenamento seguro de dados, garantindo sua imutabilidade e rastreabilidade. Como discutido por Beck *et al.* (2017), a *blockchain* pode servir como um registro

digital confiável para várias formas de transações e dados, incluindo contratos inteligentes que podem automatizar a execução de acordos sem intervenção de terceiros.

Sfetcu (2023) destaca a importância da engenharia ontológica e das tecnologias da *Web* semântica para o desenvolvimento semântico e a modelagem do fluxo operacional necessário para o design do *blockchain*, permitindo a interoperabilidade entre diferentes aplicações e sistemas de informações.

1. Infraestrutura de *blockchain* para gestão de dados

Princípios básicos: Seguindo Beck *et al.* (2017), o modelo utiliza a *blockchain* como um registro digital imutável e seguro, essencial para armazenar, verificar e transferir dados de pesquisa de forma transparente e sem a necessidade de intermediários.

Implementação: A infraestrutura de *blockchain* armazenará metadados de pesquisa, publicações, conjuntos de dados e informações de propriedade intelectual. Contratos inteligentes automatizarão o controle de acesso, a atribuição de créditos e a execução de políticas de compartilhamento de dados.

2. Modelagem semântica para representação de dados

Base teórica: Inspirado por Sfetcu (2023), o modelo adota a engenharia ontológica e as tecnologias da *Web* semântica para desenvolver uma representação rica e interoperável dos dados de pesquisa.

Desenvolvimento: Serão criadas ontologias específicas de domínio que definem classes, propriedades e relações para representar abrangentemente os dados de pesquisa. Isso facilita a busca, a recuperação e a integração de dados em diversas plataformas.

3. Contratos inteligentes para governança de dados

Aplicação: Utilizando os princípios descritos por Beck *et al.* (2017), contratos inteligentes serão empregados para gerenciar automaticamente os direitos de acesso,

uso e distribuição dos dados de pesquisa, bem como para executar transações e acordos relacionados à pesquisa sem intervenção manual.

Funcionalidades: Incluem a gestão de licenças de uso de dados, a autenticação de usuários, o registro de citações e a distribuição de recompensas pela contribuição de dados.

4. Interoperabilidade e acesso aos dados

Estratégia: A modelagem semântica promove a interoperabilidade entre diferentes sistemas e plataformas, permitindo que os dados de pesquisa sejam facilmente compartilhados e reutilizados.

Implementação: A adoção de padrões semânticos globais e ontologias de domínio específicas facilita a integração de dados entre diferentes campos de pesquisa, bem como a comunicação entre diferentes tecnologias e plataformas de gestão de dados.

5. Interface de usuário semântica

Design: Interfaces de usuário serão desenvolvidas para permitir consultas semânticas aos dados, tornando a busca por informação de pesquisa mais intuitiva e eficaz.

Funcionalidades: A interface permitirá aos usuários formular questões complexas e receber respostas precisas, aproveitando a estrutura semântica dos dados para melhorar a descoberta e visualização de informações.

6. Validação e segurança dos dados

Segurança: A *blockchain* assegura que todos os dados de pesquisa e transações sejam registrados de maneira segura e imutável, protegendo contra fraudes e alterações não autorizadas.

Validação: Contratos inteligentes validarão a autenticidade dos dados e a autoridade dos usuários, garantindo que apenas informações confiáveis e autorizadas sejam acessadas e utilizadas.

Este modelo integra a tecnologia *blockchain* e a modelagem semântica para criar um sistema de gestão de dados de pesquisa, oferecendo uma solução inovadora para os desafios de segurança, transparência, interoperabilidade e acessibilidade dos dados de pesquisa. Alinhando-se com as contribuições de Beck *et al.* (2017) e Sfetcu (2023), a figura 19 apresenta o modelo proposto que representa uma evolução significativa na maneira como os dados de pesquisa são gerenciados, compartilhados e utilizados na era digital, promovendo um ecossistema de pesquisa mais colaborativo e eficiente.

No próximo tópico, após a apresentação da figura 19, apresentamos uma proposta de esquema de modelo semântico prático para a implementação da *blockchain* na gestão de dados de pesquisa.

Figura 19 - Modelo semântico para uso da *blockchain* na gestão de dados de pesquisa



Fonte: Elaborado pelo autor.

5.4 Proposta de esquema de modelo semântico prático para a implementação da *blockchain* na gestão dos dados de pesquisa

Essa proposta de esquema de modelo semântico prático para a implementação da *blockchain* na gestão dos dados de pesquisa tem uma abordagem inspirada em Beck *et al.* (2017) e Sfetcu (2023).

Foi concebido a partir dos alicerces teóricos e metodológicos sugeridos por esses autores, cujas ideias e métodos foram integrados para criar um modelo semântico focado no uso da tecnologia *blockchain* na gestão de dados científicos. A seleção dessas fontes bibliográficas se justifica pela importância de suas contribuições ao estudo da gestão de dados e à incorporação de inovações tecnológicas como a *blockchain*, visando superar problemas relacionados à integridade, segurança e interoperabilidade dos dados de pesquisa.

Início: Criação de dados de pesquisa

O processo começa com a criação de novos dados de pesquisa por pesquisadores.

1. Registro de identificador único na *blockchain*

Cada conjunto de dados de pesquisa recebe um identificador único (ID) sendo registrado na *blockchain*.

2. Armazenamento de metadados na *blockchain*

Metadados detalhados sobre os dados de pesquisa são armazenados na *blockchain*, incluindo informações sobre autores, afiliações, datas etc.

3. Definição de permissões de acesso via contratos inteligentes

Contratos inteligentes são usados para definir quem pode acessar, modificar ou compartilhar os dados de pesquisa.

4. Registro de histórico de modificações na *blockchain*

Todas as modificações nos dados ou nos metadados são registradas na *blockchain*, garantindo rastreabilidade e facilitando auditorias.

5. Adoção de padrões semânticos para interoperabilidade

O esquema utiliza padrões semânticos e ontologias para garantir que os dados de pesquisa sejam interoperáveis entre diferentes sistemas e plataformas.

6. Compartilhamento seguro de dados de pesquisa

A *blockchain* facilita o compartilhamento seguro de dados, permitindo que os autores controlem como seus dados são compartilhados.

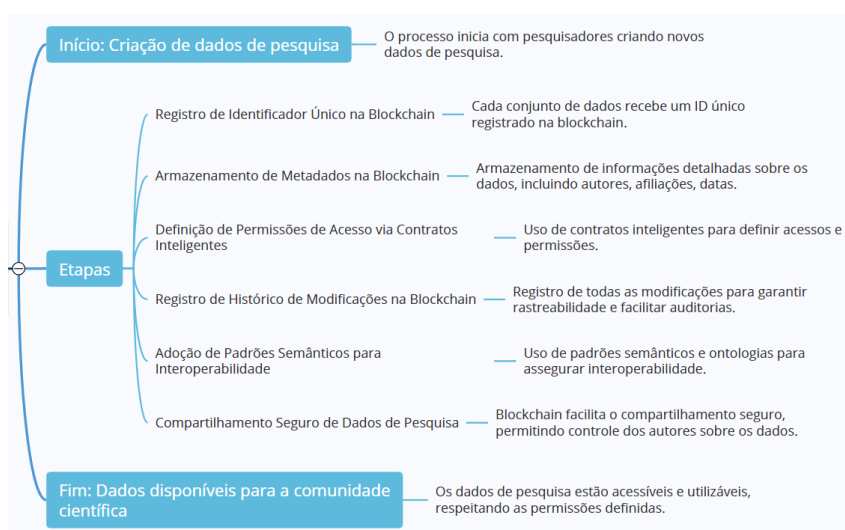
Fim: Dados disponíveis para comunidade científica

Os dados de pesquisa estão agora disponíveis para serem acessados e utilizados pela comunidade científica, respeitando as permissões definidas.

Este fluxo destaca como a tecnologia *blockchain* pode ser aplicada para melhorar a gestão, integridade, e compartilhamento de dados de pesquisa, promovendo a transparência e colaboração na comunidade científica. A figura 20 apresenta um esquema do modelo semântico prático para uso da *blockchain* na gestão de dados de pesquisa.

Após a apresentação da figura 20, introduzimos a metodologia, onde detalhamos a abordagem de pesquisa adotada no capítulo 6.

Figura 20 - Modelo semântico prático para uso da *blockchain* na gestão de dados de pesquisa



Fonte: Elaborado pelo autor.

6 METODOLOGIA

A metodologia é um componente essencial em qualquer pesquisa acadêmica, estabelecendo a estrutura sobre a qual o estudo é baseado e orientando o processo investigativo. Aqui, então, será apresentada a abordagem metodológica empregada, delineando cada passo crucial da investigação.

Inicialmente, na seção de caracterização metodológica, apresentamos a abordagem geral do estudo que tem como objeto desta pesquisa a tecnologia *blockchain* e sua aplicação na gestão e preservação de dados digitais de pesquisa científica. Explicamos a escolha dos métodos utilizados e como estes se alinham aos objetivos da pesquisa, estabelecendo um quadro de referência para as etapas subsequentes.

Prosseguimos com o percurso metodológico, dividido em quatro partes principais para facilitar a compreensão e a organização. São eles: o mapeamento de domínio com a revisão bibliográfica de literatura, a revisão sistemática de literatura, o estudo de caso do Projeto Archangel (Reino Unido) e do projeto dARK (Brasil) e, para finalizar, a elaboração de modelos semânticos utilizando ferramentas de representação específicas, o que contribuiu significativamente para a análise dos dados coletados. Após a figura 21, referente ao protocolo da metodologia, segue a caracterização metodológica.

Figura 21 - Protocolo da metodologia



Fonte: Elaborado pelo autor.

6.1 Caracterização metodológica

Empregamos uma abordagem exploratória, conforme definido por Gil (1999), essencial para desenvolver, esclarecer e modificar conceitos e ideias. Esta abordagem nos permitirá formular problemas mais precisos ou hipóteses pesquisáveis em estudos futuros. A escolha dessa metodologia é estratégica, considerando que, segundo Gil (2002), ela é ideal para trabalhar com material já disponível, como livros e artigos científicos.

1ª fase — Revisão bibliográfica: especificamente, aplicamos esta abordagem ao estudo da tecnologia *blockchain*, pesquisando problemas que ela visa a resolver, conceitos envolvidos, seus benefícios, funcionamento, relações com outras tecnologias, suas aplicações e principalmente suas possíveis aplicações na área da Ciência da Informação, utilizando-a para realizar um estudo bibliográfico aprofundado.

Seguindo a perspectiva de Gil (1999), a pesquisa científica é definida como um processo formal e sistemático de desenvolvimento do método científico. Assim, a revisão bibliográfica realizada foi rigorosa, garantindo a qualidade, relevância e confiabilidade das fontes. Esta revisão, adaptável tanto em formato sistemático quanto narrativo, será a espinha dorsal para entender as contribuições do uso da tecnologia *blockchain* na gestão e preservação de dados digitais de pesquisa científica.

2ª fase — Revisão sistemática: para complementar nosso entendimento, realizamos análises de conteúdo mediante uma revisão sistemática das tendências atuais no uso da tecnologia *blockchain* na gestão de dados de pesquisa. Essas análises, sugeridas pelos métodos de pesquisa qualitativa, enfatizarão a interpretação do significado e o papel do contexto.

Elas nos permitirão quantificar e analisar a presença de palavras-chave, frases e conceitos, assim como compreender o contexto, as práticas discursivas e os padrões comunicativos no material estudado. Esta etapa da metodologia é crucial para identificar padrões emergentes e nuances dentro do campo de estudo.

3ª fase — Estudos de casos: esta parte da metodologia estabelece uma base sólida para as fases seguintes, especialmente para os nossos estudos de casos. Incluímos, então, os estudos de casos sobre o Projeto Archangel (Reino Unido), que utilizou a tecnologia *blockchain* para a preservação de registros digitais e o projeto

dARK (Brasil), que utiliza a tecnologia *blockchain* e identificadores persistentes (PID) para criar uma rede descentralizada que facilita a gestão, preservação e acesso aberto a dados de pesquisa científica.

Estes estudos de casos não apenas oferecem um exemplo prático da aplicação da *blockchain*, mas também nos permite coletar e interpretar dados qualitativos, alinhando-se à afirmação de Gil (1999) sobre a pesquisa qualitativa ser subjetiva e focada na interpretação de sistemas complexos de significados. A análise desse estudo de caso nos ajudará a conectar teoria e prática, ampliando nossa compreensão da tecnologia *blockchain* em um contexto real.

4ª fase — Elaboração de modelos semânticos: usar a elaboração de modelos semânticos por ferramentas de representação na metodologia de pesquisa proporciona uma série de benefícios cruciais para o avanço do conhecimento científico e aplicado. Primeiramente, esses modelos permitem a captura e análise do significado contextual de palavras e frases de maneira quantitativa, revelando padrões ocultos em grandes volumes de texto que seriam difíceis de detectar manualmente (Landauer; Foltz; Laham, 1998). Além disso, ao empregar essas ferramentas, os pesquisadores podem melhorar a precisão e objetividade da análise semântica, reduzindo o viés interpretativo humano e aumentando a replicabilidade dos estudos. Por fim, a integração de modelos semânticos na metodologia de pesquisa facilita a colaboração interdisciplinar, permitindo que opiniões e métodos de um campo sejam aplicados a outro, promovendo uma compreensão mais rica e multifacetada dos fenômenos estudados.

Para assegurar a robustez e confiabilidade dos nossos resultados, utilizamos a técnica de triangulação de dados, conforme recomendado por Gil (1999). Esta técnica envolve a utilização de múltiplas fontes e métodos — neste caso, por meio de revisão bibliográfica, revisão sistemática, estudos de casos e representação de modelos semânticos.

Essa abordagem multifacetada garantirá que nossas conclusões sejam sólidas e confiáveis, reforçando a integridade da nossa pesquisa. O tópico seguinte apresenta o quadro com o percurso metodológico da pesquisa.

6.2 Percurso metodológico

Segundo Gil (2002), um percurso metodológico representa a rota planejada que o investigador adota para explorar seu objeto de estudo, incluindo a escolha minuciosa de métodos, técnicas e procedimentos. Esse processo começa com a definição do problema de pesquisa e continua até a divulgação dos resultados e conclusões, sendo essencial para garantir a validade científica da pesquisa.

Após a apresentação do quadro 4, com o percurso metodológico, será apresentado o tópico referente ao mapeamento de domínio.

Quadro 4 - Percurso metodológico

Objetivos específicos	Procedimento metodológico	Resultados obtidos
Investigar a tecnologia <i>blockchain</i> desde a sua origem até os conceitos, benefícios e sua relação com outras tecnologias.	Levantamento bibliográfico exploratório para mapeamento do domínio da tecnologia <i>blockchain</i> (conceitos e relações).	Conceitos e relações identificados no capítulo teórico da <i>blockchain</i> .
Mapear e analisar as tendências emergentes e consolidadas na aplicação da <i>blockchain</i> para a gestão de dados de pesquisa.	Revisão sistemática de literatura: "Quais são as tendências atuais do uso da tecnologia <i>blockchain</i> na gestão de dados de pesquisa?" Identificação de iniciativa para estudo de caso.	Apresentação das iniciativas de <i>blockchain</i> (a partir da revisão sistemática da literatura). Estudo de caso do projeto Archangel (Reino Unido) Estudo de caso do projeto dARK (Brasil)
Apresentar as possíveis aplicações da <i>blockchain</i> na Ciência da Informação, com ênfase em sua utilização para gestão e preservação digital de dados científicos.	Estudo sobre a adequação de modelos semânticos para representação do conhecimento sobre <i>blockchain</i> . Estudo sobre os impactos da instabilidade temporal dos dados de pesquisa que impactam contexto, proveniência, rastreabilidade, autenticidade etc. Elaboração de modelos semânticos por meio do uso de ferramenta de representação (<i>LucidChart</i> , <i>MindMup</i> etc).	Modelo semântico demonstrando as relações do conceito da <i>blockchain</i> aplicados aos constructos da Ciência da Informação, mais especificamente à Gestão e Preservação de Dados de Pesquisa Científica. Proposta de esquema de modelo semântico de uso da tecnologia <i>blockchain</i> na gestão dos dados de pesquisa. Proposta de esquema de modelo semântico prático para a implementação da <i>blockchain</i> na gestão dos dados de pesquisa.

Fonte: Elaborado pelo autor.

6.2.1 Mapeamento de domínio

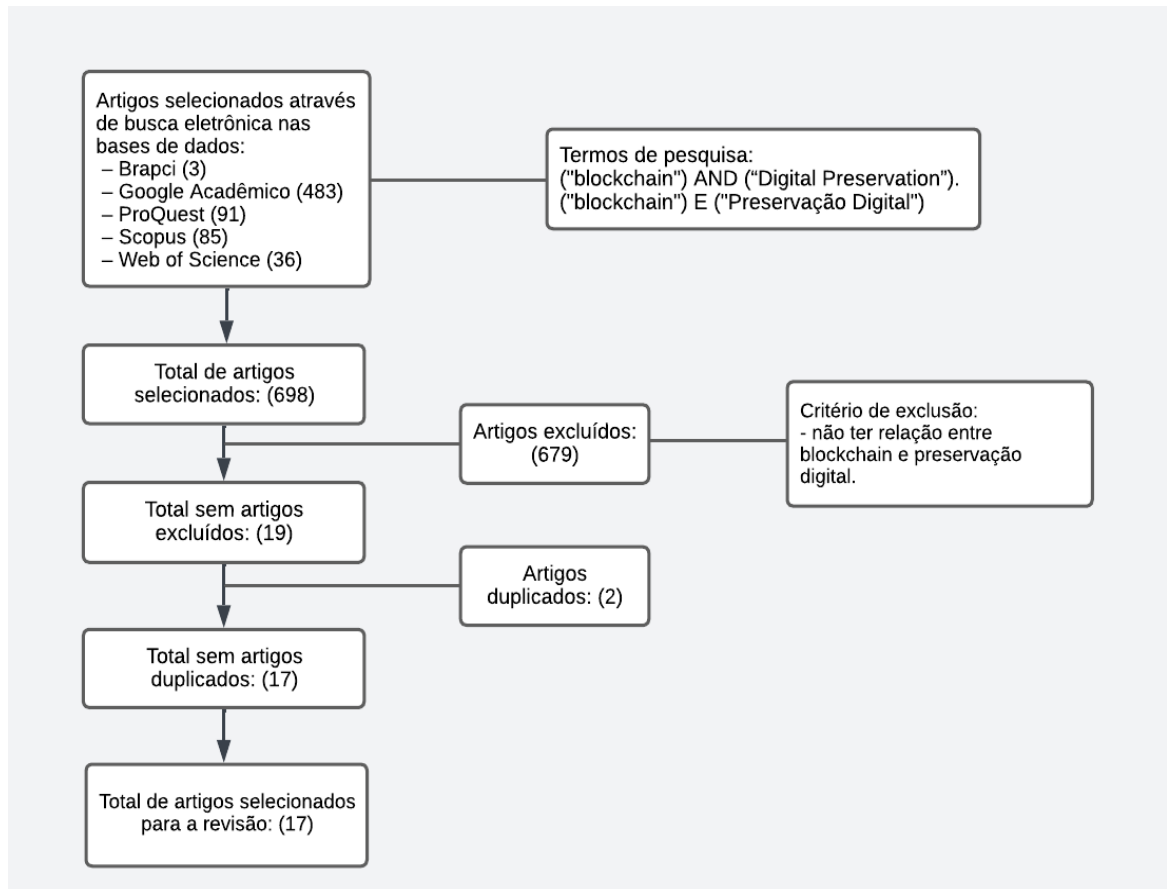
O objeto de estudo desta pesquisa é a tecnologia *blockchain* e sua aplicação na gestão e preservação de dados digitais de pesquisa científica. A investigação foca os conceitos fundamentais da *blockchain*, benefícios, desafios e potencial transformador em práticas de gestão de dados de pesquisa e estratégias de preservação digital. Começamos agora com a revisão bibliográfica da literatura.

6.2.2 Revisão bibliográfica da literatura

Inicialmente, o estudo se dedica a uma análise abrangente da tecnologia *blockchain*, baseando-se em um levantamento exaustivo da literatura existente. O objetivo é reunir informações relevantes sobre esta tecnologia, abarcando a investigação de seus conceitos básicos, benefícios, obstáculos e potenciais usos na área da Ciência da Informação, enfocando especialmente na preservação digital.

A pesquisa inicial foi realizada, durante o mês de maio de 2021, abrangendo publicações entre 2017 e 2021. A consulta foi feita com base na estratégia de busca *Topic*: (“*blockchain*” AND “*digital preservation*”) nas bases de dados científicas, Google Acadêmico, *ProQuest*, *Scopus* e *Web Of Science*. Na *Brapci* a estratégia de busca foi por (“*blockchain*” E “Preservação Digital”). A pesquisa foi realizada analisando-se os títulos, resumos e palavras-chave dos resultados encontrados, considerando sua relevância e alinhamento com a ideia inicial da minha pesquisa que era essa forte relação entre a *blockchain* e a preservação digital. O protocolo usado para a pesquisa inicial é apresentado na figura 22.

Figura 22 - Protocolo de pesquisa inicial



Fonte: Elaborado pelo autor.

Dos estudos recuperados, 17 foram selecionados para análise detalhada com base em critério previamente definido. Como vimos na figura 3, após a primeira análise, com avaliação dos títulos, 19 artigos foram considerados selecionáveis para a segunda fase desta revisão, que consistiu da leitura dos resumos.

Dos 698 artigos, 679 foram excluídos e 2 estavam duplicados. O critério usado para exclusão foi: não ter relação entre *blockchain* e preservação digital. O quadro 5 relaciona os títulos dos 17 artigos selecionados na pesquisa inicial.

Quadro 5 - Artigos selecionados na pesquisa inicial

Um estudo da <i>blockchain</i> aplicado ao contexto dos dados de pesquisa
Arquivologia e <i>blockchain</i> : discussão teórica sobre oportunidades e barreiras
Análise comparativa entre os requisitos da RDC-Arq e a tecnologia <i>blockchain</i> : uma perspectiva de profissionais arquivistas
<i>Blockchain and distributed ledgers as trusted recordkeeping systems: an archival theoretic evaluation framework</i>
<i>Using blockchain to engender trust in public digital archives</i>
<i>A survey on data integrity verification schemes using blockchain technology in cloud computing environment</i>
<i>Design of an enhanced web archiving system for preserving content integrity with blockchain</i>
<i>Design of secure protocol for cloud-assisted electronic health record system using blockchain</i>
<i>Long-term preservation of electronic record based on digital continuity in smart cities</i>
<i>Underscoring archival authenticity with blockchain technology</i>
<i>A secure blockchain-based e-health records storage and sharing scheme</i>
<i>A blockchain approach to digital archiving: digital signature certification chain preservation</i>
<i>Trusting records: is blockchain technology the answer?</i>
<i>A data preservation method based on blockchain and multidimensional hash for digital forensics</i>
<i>Blockchain-based trusted electronic records preservation in cloud storage</i>
<i>A survey on privacy protection in blockchain system</i>
<i>Blockchain archival discourse: trust and the imaginaries of digital preservation</i>

Fonte: Elaborado pelo autor.

No tópico seguinte apresentamos a metodologia usada na revisão sistemática da literatura que investiga quais são as tendências atuais do uso da tecnologia *blockchain* na gestão de dados de pesquisa.

6.2.3 Revisão sistemática de literatura

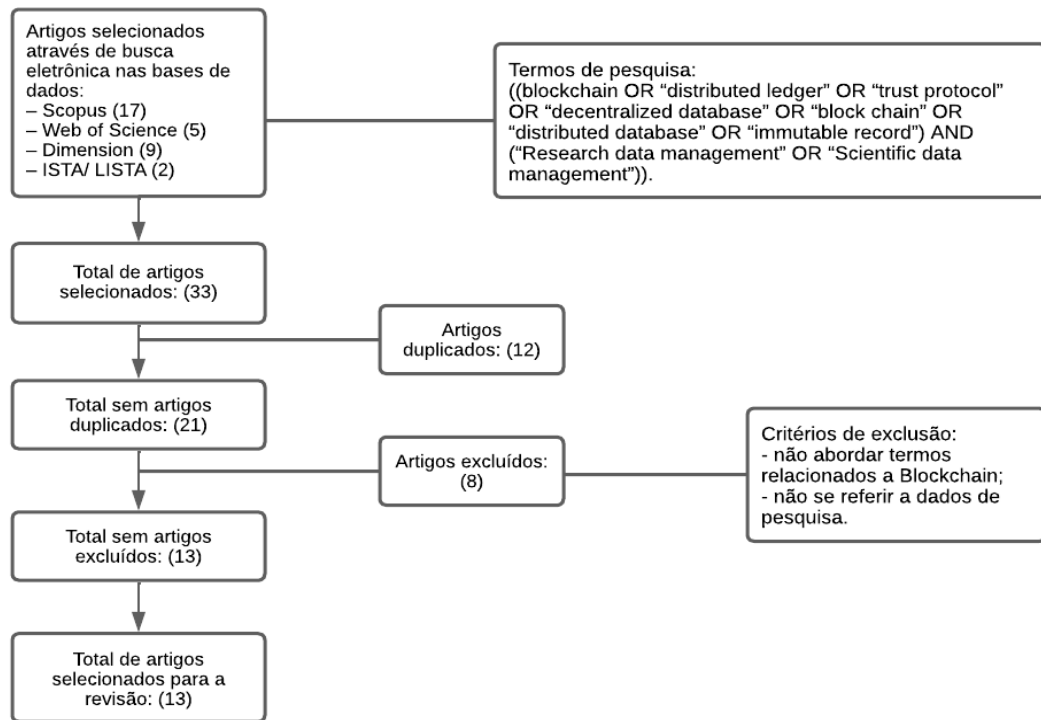
Posteriormente, após o amadurecimento da ideia, partimos para realizar uma revisão sistemática e analisar as tendências atuais do uso da *blockchain* na gestão de dados de pesquisa. Esta revisão foi estruturada com base na ideia da mandala proposta por Pinheiro (2018). Foram definidos o escopo, as perguntas de pesquisa, os critérios de inclusão/exclusão e o processo de análise, proporcionando uma compreensão clara e metodicamente sólida das tendências emergentes e consolidadas na aplicação da *blockchain* para a gestão de dados de pesquisa.

A metodologia adotada para esta revisão sistemática alinha-se ao rigor e à estrutura prescrita por Tranfield, Denyer e Smart (2003) e Denyer e Tranfield (2006), buscando proporcionar uma análise abrangente e imparcial das tendências atuais do uso da tecnologia *blockchain* na gestão de dados de pesquisa. Esta revisão sistemática delinea os procedimentos meticulosos empregados desde a coleta inicial dos dados até a análise final dos estudos selecionados, assegurando assim a reprodutibilidade e a validade dos achados.

A busca por literatura relevante foi realizada durante o mês de maio de 2023, abrangendo publicações entre 2010 e 2023. As bases de dados científicas consultadas foram *Scopus*, *Web of Science*, *Dimension*, *ISTA* e *LISTA*. Utilizando uma combinação de descritores e termos de pesquisa, a estratégia de busca foi estruturada para capturar uma gama ampla e diversificada de literatura sobre o tema proposto.

A estratégia de busca foi meticulosamente construída para garantir uma cobertura abrangente do tema em estudo. A estratégia de busca utilizada foi: ("*blockchain*" OR "*distributed ledger*" OR "*trust protocol*" OR "*decentralized database*" OR "*block chain*" OR "*distributed database*" OR "*immutable record*") AND ("*research data management*" OR "*scientific data management*"). Esta estratégia foi projetada para maximizar a recuperação de estudos relevantes, abrangendo uma variedade de termos e frases associadas à tecnologia *blockchain* e à gestão de dados de pesquisa. O protocolo usado para a revisão sistemática é apresentado na figura 23.

Figura 23 - Protocolo de pesquisa da revisão sistemática



Fonte: Elaborado pelo autor.

Dos estudos recuperados, 13 foram selecionados para análise detalhada com base em critérios previamente definidos. Como vimos na figura 4, após a primeira análise, com avaliação dos títulos, 33 artigos foram considerados selecionáveis para a segunda fase desta revisão, que consistiu da leitura dos resumos.

Desses 33 artigos, 12 estavam duplicados e 8 foram excluídos. Os critérios usados para exclusão foram: não abordar termos relacionados à tecnologia *blockchain* e não se referir a dados de pesquisa ou dados científicos. A análise dos estudos incluídos focou diversos aspectos como o fenômeno estudado, problema, perspectiva teórica, processo de coleta e validação dos dados, entre outros relevantes para a temática. O quadro 6 relaciona os títulos dos 13 artigos selecionados para a revisão sistemática.

Quadro 6 - Artigos selecionados para revisão sistemática

<i>"Prove it!" A user-centric client for the blockchain-based research lifecycle transparency framework</i>
<i>Blockchain for libraries</i>
<i>Blockchain study applied to the context of scientific research data</i>
<i>Exploring the determinants of blockchain acceptance for research data management</i>
<i>Towards archaeo-informatics: scientific data management for archaeobiology</i>
<i>Permissioned blockchain for data provenance in scientific data management</i>
<i>CryptSubmit: introducing securely timestamped manuscript submission and peer review feedback using the blockchain</i>
<i>Application of blockchain in libraries and information centers</i>
<i>Application of blockchain technology in data management of university scientific research</i>
<i>FASTDB: an array database system for efficient storing and analyzing massive scientific data</i>
<i>A performance evaluation of Hive for scientific data management</i>
<i>Merging file systems and data bases to fit the grid</i>
<i>Data management in healthcare research as a guarantee of its quality</i>

Fonte: Elaborado pelo autor.

Conforme orientações de Tranfield *et al.* (2003) e Denyer e Tranfield (2006), a revisão sistemática da literatura foi conduzida para evitar a convergência de resultados para uma mesma vertente (Salim *et al.*, 2019) e para detalhar de forma compreensível os passos aplicados, facilitando a replicabilidade dos procedimentos por outros pesquisadores (Nawaz; Koç, 2018; Tranfield *et al.*, 2003).

A análise e síntese dos achados foram conduzidas com um enfoque em identificar as tendências atuais e as implicações da aplicação da tecnologia *blockchain* na gestão de dados de pesquisa. A interpretação dos achados foi realizada à luz dos objetivos propostos para esta revisão, proporcionando uma compreensão abrangente do estado atual do campo.

A incessante busca por descobertas, inovações e avanços caracteriza a pesquisa científica em diversas disciplinas. No cerne desse processo está a gestão adequada e transparente dos dados de pesquisa. Uma *blockchain* representa uma tecnologia emergente que conecta um conjunto de dados, denominado bloco, por meio de chaves de *criptografia*, em um sistema descentralizado. Nesse sistema, qualquer pessoa pode compartilhar o livro-razão para revisar todos os conjuntos de dados (Hwang; Shon; Park, 2020).

Com a crescente proeminência da *blockchain* em várias áreas, sua aplicação na gestão de dados de pesquisa também tem sido explorada. O potencial dessa tecnologia para transformar a pesquisa científica é notável, oferecendo soluções para questões críticas como proveniência de dados, segurança, transparência e reprodutibilidade. A *blockchain* também pode ser aplicada em bibliotecas e centros de informação para o arquivamento e rastreamento de registros bibliotecários essenciais e vitais, onde se espera a proveniência, integridade e autenticidade. Essa tecnologia pode ser utilizada para verificar a origem e a integridade de todas as informações da biblioteca, especialmente no contexto de bibliotecas interconectadas em rede (Omame; Alex-Nmecha, 2021).

Contudo, conforme a utilização da *blockchain* avança na gestão de dados de pesquisa, emergem indagações sobre sua viabilidade, desafios técnicos e éticos, bem como as tendências atuais nesse domínio.

A revisão sistemática apresentada no capítulo 7 tem como objetivo elucidar a seguinte questão: **quais são as tendências atuais do uso da tecnologia *blockchain* na gestão de dados de pesquisa?** Para isso, será examinado um corpo diversificado de literatura científica e técnica, avaliando como a *blockchain* está sendo empregada na gestão de dados de pesquisa, bem como suas implicações, aplicações, limitações e resultados observados.

7 A REVISÃO SISTEMÁTICA

A motivação inicial para esta revisão sistemática foi inspirada na mandala da pesquisadora Lena Vania Ribeiro Pinheiro (2018), que enfatiza a interdisciplinaridade na Ciência da Informação. Para a autora, a mandala é um tipo de diagrama gráfico que representa as relações entre as partes de um todo, utilizado na Ciência da Informação como uma representação da interdisciplinaridade da área (Pinheiro, 2018).

Para esta pesquisa, destacamos na mandala subáreas como ciência aberta e dados de pesquisa, gestão da informação e tecnologia da informação e comunicação com suas devidas áreas interdisciplinares, conforme ilustrado na figura 24.

Quadro 7 - Artigos da revisão sistemática - autor/ano, título e questão/objetivo dos artigos

	Autores/Ano de publicação	Título	Questão/Objetivo de estudo
[1]	Yun-Chi, Li-Fei, Hsu-Chun e Wei (2022)	“Prove it!” A user-centric client for the <i>blockchain</i> -based research lifecycle transparency framework	Construir um ecossistema que promova a transparência segura do ciclo de vida da pesquisa com a ajuda da tecnologia.
[2]	Michael Meth (2019)	<i>Blockchain</i> for libraries	Apresentar a tecnologia de forma acessível e fornecer reflexão e conversa iniciais para ajudar as bibliotecas a examinar esse tópico complexo e se preparar para as mudanças futuras.
[3]	Ouchi e Arakaki (2020)	<i>Blockchain</i> study applied to the context of scientific research data	Analisar a tecnologia <i>blockchain</i> e suas possibilidades inseridas na comunicação científica e na gestão de dados de pesquisa. Os objetivos específicos são conceituar dados de pesquisa; identificar os principais conceitos, a estrutura e o funcionamento relacionados à tecnologia <i>blockchain</i> ; e apontar a tecnologia <i>blockchain</i> e suas potenciais aplicações na ciência.
[4]	Woo e Yoo (2023)	Exploring the determinants of <i>blockchain</i> acceptance for research data management	Analisar os determinantes que afetam a aceitação do serviço de <i>notebooks</i> de laboratório eletrônico baseado em <i>blockchain</i> como serviço de gerenciamento de dados de pesquisa.
[5]	Kriegel, Kroeger, van der Meijden, Obermaier, Peters e Renz (2010)	Towards archaeo-informatics: scientific data management for archaeobiology	Descrever um novo e rico domínio de aplicação para gerenciamento de dados científicos: Arqueobiologia.
[6]	Moeller, Froeschle e Hahn (2021)	Permissioned <i>blockchain</i> for data provenance in scientific data management	Fornecer uma solução descentralizada que preencha essas lacunas e se encaixe no ecossistema de dados científicos.
[7]	Gipp, Breitinger, Meuschke e Beel (2017)	CryptSubmit: introducing securely timestamped manuscript submission and peer review feedback using the <i>blockchain</i>	Apresentar o CryptSubmit – um sistema que cria automaticamente um carimbo de data/hora descentralizado, inviolável e publicamente verificável para cada manuscrito enviado, utilizando a <i>blockchain</i> da criptomoeda <i>Bitcoin</i> .
[8]	Omame e Alex-Nmecha (2021)	Application of <i>blockchain</i> in libraries and information centers	Explorar o conceito e a aplicação da tecnologia <i>blockchain</i> em bibliotecas e centros de informação.

	Autores/Ano de publicação	Título	Questão/Objetivo de estudo
[9]	Zhang (2021)	Application of <i>blockchain</i> technology in data management of university scientific research	Garantir a disponibilidade, autenticidade e validade dos dados da pesquisa científica e a aplicação da tecnologia <i>blockchain</i> para gerenciar todo o ciclo de vida desses dados e garantir a eficácia e segurança dos dados de pesquisa científica.
[10]	Li, Qiu, Chen, Li, Dai, Zhu e Huang (2015)	FASTDB: an array database system for efficient storing and analyzing massive scientific data	Apresentar o FASTDB, um sistema de banco de dados de matriz distribuída otimizado para gerenciamento de dados científicos massivos.
[11]	Liu, Liu, Liu e Li (2013)	A performance evaluation of Hive for scientific data management	Apresentar a implementação de um <i>benchmark</i> de gerenciamento de dados científicos, no Hive, um <i>data warehouse</i> baseado em <i>MapReduce</i> .
[12]	Denneulin, Labbé, D'Orazio e Roncancio (2010)	Merging file systems and data bases to fit the grid	Apresentar o Gedeon, um <i>middleware</i> que propõe uma abordagem híbrida para gerenciamento de dados científicos para infraestruturas de <i>grid</i> .
[13]	Rekunenko, Boiko, Kramarenko e Khan (2022)	Data management in healthcare research as a guarantee of its quality	Estudar a eficiência da gestão de dados no sistema de saúde.

Fonte: Elaborado pelo autor.

As análises a seguir englobam os 13 estudos revisados, considerando os seguintes aspectos metodológicos: o fenômeno investigado; a problemática; a abordagem teórica; os métodos de coleta de dados; a essência dos dados; sua validação; o procedimento de coleta e interpretação de descobertas; as respectivas descobertas; as considerações éticas; as implementações propostas nos estudos; e suas limitações. Os números entre colchetes [X] foram usados para facilitar a leitura das análises dos estudos selecionados e representa o artigo listado no quadro 7.

O fenômeno investigado

Compreender as tendências atuais do uso da tecnologia *blockchain* na gestão de dados de pesquisa é imperativo para avançar na era digital. A *blockchain*, conhecida por sua natureza imutável e descentralizada, emerge como uma base forte

para aprimorar o gerenciamento de dados de pesquisa, criando estruturas de transparência que validam e rastreiam cada etapa do ciclo de vida dos dados [1, 3, 4, 6].

A abordagem centrada no usuário, aliada a interfaces amigáveis, potencializa a experiência do pesquisador na gestão de dados, contribuindo para a eficiência e integridade do processo [1].

Um domínio notável de aplicação é o ambiente acadêmico, especialmente bibliotecas e centros de informação, onde a *blockchain* está sendo empregada para autenticar e catalogar obras, incrementando a confiabilidade das informações compartilhadas e assegurando uma distribuição confiável de conhecimento [2, 8].

A relevância desta tecnologia estende-se até a Arqueobiologia, sendo um ramo de pesquisa que estuda vestígios de animais domésticos e silvestres e outros vestígios faunísticos de épocas pré-históricas e históricas. A complexidade dos dados arqueológicos encontra nos bancos de dados distribuídos uma ferramenta valiosa para rastrear e autenticar descobertas, impulsionando a confiança nos resultados [5].

A interação entre a tecnologia *blockchain* e o gerenciamento de dados científicos é uma área de foco, com estudos dedicados a analisar os benefícios e implicações desta abordagem [6]. Seu potencial transversal para aprimorar a colaboração e a confiabilidade dos dados é evidente em diversos domínios.

Por exemplo, a integridade e aceitação de manuscritos científicos alcançam um novo patamar de segurança por meio do carimbo de data/hora e *feedback* de revisão por pares incorporados na tecnologia *blockchain*, não apenas fortalecendo o processo de publicação, mas também mitigando questões de plágio e falsificação de dados [7].

Além disso, a *blockchain* promove uma gestão robusta de dados de pesquisa científica em ambientes universitários, estabelecendo uma estrutura que mantém a transparência e a confiabilidade nas contribuições acadêmicas [9].

O desafio de lidar com grandes volumes de dados científicos encontra soluções promissoras na conjunção dos bancos de dados distribuídos e estratégias *MapReduce*, um modelo de programação utilizado para processamento distribuído de grandes conjuntos de dados em *clusters* de computadores, permitindo uma análise mais eficiente e escalonável. Tal abordagem se mostra fundamental em áreas de estudo como a Astronomia, essencial em campos como a Astronomia, exemplificado pelo *Sloan Digital Sky Survey*, o qual é um levantamento astronômico cujo objetivo é

criar um mapa digital de grande parte do universo, e a necessidade de gerenciar vastas quantidades de informações [10, 11].

As soluções para os desafios na gestão de dados também se estendem a ambientes complexos como o *Hive*, sendo um sistema de armazenamento de dados de código aberto baseado em *Hadoop* (uma estrutura de código aberto para armazenamento distribuído e processamento de grandes conjuntos de dados em *clusters* de computadores) e infraestruturas de *grid*. A combinação de diferentes tecnologias, incluindo os bancos de dados distribuídos, propõe uma abordagem híbrida que otimiza a capacidade de armazenamento e processamento, melhorando a acessibilidade e a análise dos dados [11, 12].

No setor de saúde, a tecnologia *blockchain* é vista como uma ferramenta vital para otimizar a eficiência da gestão de dados. A integridade dos registros médicos é reforçada, proporcionando um ecossistema seguro para o compartilhamento de informações confidenciais e melhorando a tomada de decisões clínicas [13].

Este tópico destaca o papel crescente da *blockchain* na transformação da gestão de dados de pesquisa em diversos domínios. O avanço contínuo nesta tecnologia promete revolucionar não apenas como os dados são gerenciados, mas como a pesquisa é conduzida e compartilhada, abrindo novas perspectivas para a colaboração e inovação científica.

Os problemas

A transição para novas tecnologias, como a *blockchain*, traz consigo desafios que devem ser cuidadosamente abordados. Este tópico desdobra os problemas identificados nos estudos analisados, lançando luz sobre as áreas de preocupação e as questões críticas emergentes.

O anseio dos pesquisadores por registrar suas pesquisas utilizando a tecnologia *blockchain* e as expectativas em relação às ferramentas de gerenciamento de registros de pesquisa refletem uma crescente demanda por transparência e rastreabilidade [1]. A *blockchain* surge como uma base sólida que tanto os pesquisadores quanto a comunidade científica almejam, contudo, traz consigo questões pertinentes de implementação e adoção.

A potencial revolução na gestão de coleções e na comunicação acadêmica nas bibliotecas por meio da *blockchain* é empolgante, mas também acarreta desafios de

integração, treinamento e harmonização das práticas tradicionais com as inovações tecnológicas [2]. Esta mudança paradigmática requer uma transição bem orquestrada, com adequado suporte educacional para as partes interessadas.

Enquanto a *blockchain* oferece um vislumbre de confiabilidade, a sua contribuição para a confiabilidade e reprodutibilidade dos resultados científicos também instiga questões críticas [3]. A tecnologia pode projetar um ambiente mais confiável, mas isso requer alinhamento com processos rigorosos de pesquisa e colaboração.

A investigação sobre os fatores que influenciam a intenção dos pesquisadores em adotar a tecnologia *blockchain* abre um campo intrigante de estudo [4]. Os benefícios são evidentes, mas as percepções de risco e a complexidade da transição para uma nova tecnologia são aspectos que permeiam o processo decisório.

Na Arqueobiologia, a lacuna na tecnologia da informação para o gerenciamento de dados evidencia preocupações sobre a heterogeneidade dos registros e análises [5]. A ausência de uma abordagem unificada culmina em dados fragmentados, criando obstáculos para uma interpretação global.

A questão da rastreabilidade e confiabilidade das informações de proveniência de dados é elevada à primazia em ambientes complexos [6]. A criação de um ecossistema de dados confiável, onde as fontes são rastreáveis e validadas, salienta a importância de tecnologias como a *blockchain*.

A necessidade de um mecanismo confiável e verificável para autores comprovarem a autenticidade de seus trabalhos é um desafio persistente que afeta a confiabilidade dos resultados científicos [7]. A *blockchain* apresenta-se como uma solução viável para criar um registro inalterável de contribuições.

A preservação da privacidade e a segurança dos dados são questões cruciais que emergem no debate [8, 9]. A *blockchain* promete um nível elevado de segurança, contudo, a sua implementação requer uma abordagem meticulosa para minimizar riscos e garantir a integridade e confidencialidade dos dados.

Com a era tecnológica atual, o desafio de armazenar e analisar dados científicos em larga escala alcança um novo patamar [10]. Diante de volumes exponenciais de dados, a função do *Hive* e tecnologias similares torna-se crucial para possibilitar análises eficientes e escalonáveis [11].

A integração entre bancos de dados distribuídos e aplicativos legados [12] é um ponto focal na Biologia. A busca por compatibilidade entre sistemas é um desafio premente que demanda soluções inovadoras para preservar a coesão dos dados.

No campo da saúde, a gestão eficaz de dados é imperativa [13]. A garantia de disponibilidade, relevância e qualidade dos dados exige sistemas fortes e eficazes, capazes de gerir informações sensíveis com a máxima integridade.

Este tópico delinea os desafios multidimensionais associados à incorporação da tecnologia *blockchain* na gestão de dados de pesquisa. O entendimento aprofundado desses problemas é fundamental para navegar pelo caminho rumo a soluções inovadoras que podem transformar o paradigma da gestão de dados científicos.

A perspectiva teórica

A perspectiva teórica constitui um enfoque fundamental por meio do qual a adoção e aplicação da tecnologia *blockchain* na pesquisa científica são examinadas, sobretudo no que concerne à gestão de dados. Este tópico planeja desvelar as discussões teóricas prevalentes que moldam o entendimento e a exploração da *blockchain* como uma ferramenta promissora para aprimorar a transparência, integridade e eficiência na coleta, armazenamento e compartilhamento de informações científicas.

Ao adentrar as discussões teóricas dos artigos analisados, emerge uma ênfase notável em termos-chave como “dados de pesquisa”, “gestão de dados de pesquisa”, “*blockchain*”, “banco de dados distribuído” e “dados científicos”. Estes termos, longe de serem meras nomenclaturas, refletem as aspirações centrais da comunidade científica. Destacamos que nos artigos [1, 2, 3, 4, 6, 7, 8, 9 e 13] é mencionada a tecnologia *blockchain*, enquanto nos artigos [5, 10, 11, e 12] são mencionados os bancos de dados distribuídos.

Por exemplo, o termo “banco de dados distribuído” alude a um sistema onde os dados são armazenados em múltiplos locais, proporcionando um acesso mais democratizado e seguro, qualidades intrínsecas à tecnologia *blockchain*.

No cerne destas investigações, encontra-se a busca contínua por soluções que possam garantir a confiabilidade dos resultados, a reprodutibilidade das pesquisas e o estímulo à colaboração entre cientistas. Este impulso não apenas reflete uma

conscientização crescente sobre a relevância desses elementos no âmbito acadêmico e científico, mas também destaca a *blockchain* como um veículo potencial para alcançar esses objetivos fundamentais.

Além disso, a *blockchain* oferece uma resposta potencial a dilemas éticos prementes na pesquisa científica. A capacidade de rastrear e verificar a origem dos dados pode abordar questões de plágio e autoria, aspectos cruciais para manter a integridade e a ética no ambiente acadêmico.

O processo de coleta de dados

O processo de coleta de dados é um pilar crucial em qualquer empreendimento científico, servindo como a ponte entre questões teóricas e empíricas. No contexto do uso da tecnologia *blockchain* para a gestão de dados de pesquisa, este processo assume um papel ainda mais significativo.

Exploramos, com mais profundidade, como diferentes estudos abordaram a coleta de dados, evidenciando a diversidade de métodos e o impacto potencial na gestão de dados de pesquisa.

Um artigo apresenta uma abordagem bifásica para a coleta de dados: análise de conteúdo e entrevistas. A análise de conteúdo foi realizada com dados obtidos de um *workshop* de *design thinking* que contou com a presença de 17 pesquisadores renomados nas Ciências da Vida. O foco principal era desvendar os desafios que esses pesquisadores enfrentavam no gerenciamento de seus dados de pesquisa [1]. De uma perspectiva mais humana, imagine a riqueza de percepções derivadas das experiências compartilhadas por esses profissionais.

Complementarmente, foram conduzidas entrevistas em diversas áreas, desde Biomedicina até Ciências Sociais, com o intuito de explorar as percepções dos usuários sobre o serviço oferecido, identificar possíveis falhas no design existente e descobrir os obstáculos enfrentados por eles. Esta abordagem permitiu um entendimento mais aprofundado sobre três aspectos cruciais das práticas de pesquisa cotidianas: manutenção de registros, armazenamento de dados e colaboração acadêmica.

Contrastando com o primeiro, o outro artigo adotou uma abordagem bibliográfica na coleta de dados [3]. A ênfase foi colocada na revisão metódica da literatura existente, seguida pela construção de um dispositivo eletrônico

computacional que serviu como uma plataforma experimental para testar os conceitos teóricos discutidos na literatura. Embora os detalhes específicos desta abordagem experimental não tenham sido totalmente elucidados, esta estratégia ilustra a diversidade de métodos que podem ser empregados na coleta de dados, oscilando entre análises teóricas e experimentos práticos.

No que se refere aos artigos [2, 4, 5, 6, 7, 8, 9, 10, 11, 12 e 13], a falta de informações detalhadas sobre os processos de coleta de dados é notável e destaca a necessidade de uma descrição mais transparente desses métodos. Este é um chamado para a comunidade acadêmica, ressaltando a importância de uma comunicação clara e detalhada dos métodos de coleta de dados, fundamental para a replicabilidade, validação e avanço da pesquisa científica.

A natureza dos dados

A natureza dos dados é crucial na análise das tendências do uso de *blockchain* na gestão de dados de pesquisa, conforme evidenciado pela diversidade nos artigos desta revisão. Exploramos a variedade desses dados e como diferentes áreas se beneficiam da tecnologia *blockchain*.

Inicialmente, é fascinante observar como a *blockchain* encontra relevância em uma vasta gama de áreas do conhecimento. Temos, por exemplo, a Biomedicina, Ciências Básicas, Ciências Aplicadas e Ciências Sociais [1]. A diversidade continua com a inclusão de campos de estudo especializados como Física de Partículas, Genômica e Ciências da Vida [4].

A complexidade dos dados em questão é outro aspecto que captura nossa atenção. Os dados podem variar desde serem planejados, de intervalo, incertos, temporais, espaciais, de localização até espaço-temporais [5]. Esta complexidade não só ressalta os desafios inerentes à gestão de dados de pesquisa, mas também destaca o potencial dos bancos de dados distribuídos em oferecer soluções confiáveis.

Adentrando o mundo prático, é intrigante como os dados do Sistema de Identificação Automática (AIS) de tráfego marítimo e dados de GPS podem ser gerenciados e protegidos por meio da *blockchain* [6]. Este tipo de aplicação prática ressalta a capacidade da *blockchain* de atender às necessidades reais e cotidianas de diversos setores.

A esfera acadêmica, por sua vez, não fica atrás. Dados científicos [3], manuscritos acadêmicos [7], dados acadêmicos e informativos [8] e dados digitais de pesquisa científica universitária [9] são exemplos eloquentes de como a gestão de informações críticas pode ser revolucionada pela *blockchain*. A multidimensionalidade dos dados científicos, como dados de rádio, multibanda, de espectro e polarização, sublinha a versatilidade dos bancos de dados distribuídos na manipulação de diferentes tipos de dados [10].

Ademais, é notável como os bancos de dados distribuídos é vista como uma ferramenta promissora na gestão de dados complexos e multidimensionais em campos como Astronomia, sensoriamento remoto e processamento de imagens médicas [11].

Por fim, a menção de dados relacionados à gestão de dados de pesquisa na área da saúde reforça a necessidade imperativa de tecnologias expressivas como a *blockchain*, visando a garantir a integridade, segurança e acessibilidade dos dados [13].

A ausência de informações específicas sobre a natureza dos dados nos artigos [2 e 12] sugere uma oportunidade para futuras investigações, que podem desvelar ainda mais a profunda conexão entre a *blockchain* e a gestão de dados de pesquisa.

Este exame meticuloso da natureza dos dados nos leva a apreciar a vastidão das possibilidades que a tecnologia *blockchain* pode desbloquear em diversas áreas do conhecimento.

A validação dos dados

A validação dos dados emerge como um alicerce central na gestão eficaz dos dados de pesquisa, servindo como o guardião da precisão, autenticidade e integridade das informações. No cenário de pesquisa científica, o papel da tecnologia *blockchain* na validação dos dados tem sido reconhecido por sua capacidade inovadora de oferecer um forte mecanismo de consenso, em que diversas partes interessadas alcançam um acordo sobre a veracidade das informações registradas [3].

Este processo de consenso pode ser amplamente enriquecido por revisões por pares, validações independentes ou validação cruzada, constituindo uma forte salvaguarda para garantir a qualidade dos dados. Aprofundando-se na esfera técnica,

a validação dos dados é muitas vezes garantida por meio de testes rigorosos que buscam a consistência interna e a validade discriminante [4].

Métodos como os testes *Fornell-Larcker*, empregados para assegurar a validade discriminante, e o critério *HTMT*, que avalia a magnitude da correlação estrutural entre variáveis latentes e tido como inválido se alcançar ou exceder 0,9, constituem elementos cruciais na validação de dados. Esses métodos garantem uma análise rigorosa da legitimidade dos construtos e funcionam como defensores contra a multicolinearidade, capaz de embasar as interpretações analíticas.

No coração da tecnologia *blockchain*, encontramos um aliado poderoso na validação dos dados. A utilização de assinaturas digitais para verificar a autenticidade das transações apresenta um caminho revolucionário para a segurança e rastreabilidade dos dados [6]. Este método de validação protege as informações contra adulteração ou falsificação, estabelecendo um registro imutável que se mantém íntegro ao longo do tempo.

A *blockchain* desdobra sua inovação mediante um livro-razão distribuído e assinaturas digitais, pavimentando o caminho para uma era em que os dados permanecem imutáveis e cada transação é rastreável até sua origem [9]. O carimbo de data/hora digital e outras técnicas de validação auxiliar tornam-se companheiros valiosos nesta jornada, contribuindo significativamente para a integridade dos dados.

No contexto da gestão eficaz dos dados de pesquisa, a validação dos dados é apreciada por sua capacidade de assegurar a disponibilidade, relevância e transparência dos dados [13]. A verificação contínua dos dados não apenas garante a precisão das informações utilizadas, mas também ressalta a necessidade de validação contínua, especialmente quando os dados são reutilizados em diferentes contextos, mantendo sua relevância e confiabilidade intactas.

Embora os artigos [1, 2, 5, 7, 8, 10, 11 e 12] não tenham fornecido informações específicas sobre a validação dos dados, a discussão acima destaca a evolução e o potencial da tecnologia *blockchain* como um auxílio na validação dos dados. Este tópico crucial ressoa com a necessidade incessante de garantir a precisão e a confiabilidade dos dados na pesquisa científica, lançando uma luz sobre o papel vital que a tecnologia *blockchain* pode desempenhar na conquista deste objetivo fundamental.

O processo de coleta de achados

Um achado científico é um resultado ou descoberta obtida por pesquisa científica e o tópico referente ao processo de coleta de achados científicos desdobra-se mediante uma gama diversificada de métodos elucidados nos estudos analisados, demonstrando o quão multifacetada pode ser a jornada de coleta de dados de pesquisas acadêmicas.

É destacada a utilização de análise de conteúdo para delinear as necessidades dos pesquisadores e as barreiras enfrentadas no compartilhamento de dados, indicando um caminho valioso para aprimorar a gestão de dados mediante a compreensão das perspectivas dos pesquisadores [1].

A tecnologia apresenta-se como um aliado fundamental neste processo, onde a utilização de dispositivos eletrônicos equipados com sensores, como um termômetro, facilita a coleta de dados com precisão [3]. Este ponto sublinha a preponderância da integração tecnológica na coleta de dados científicos, proporcionando meios precisos e confiáveis para a captura de informações relevantes.

A eficiência no gerenciamento dos dados coletados é enfatizada, demonstrando a importância de organizar os dados de forma sistemática em bancos de dados distribuídos [5]. Esta menção ressalta a necessidade de infraestruturas resistentes para armazenar e analisar os dados, assegurando assim a sua integridade e acessibilidade ao longo do tempo.

A diversidade metodológica na coleta de achados científicos evidencia a importância de moldar as estratégias de coleta conforme as demandas específicas de cada pesquisa, incorporando simultaneamente tecnologias e processos eficazes para assegurar a integridade e a qualidade dos dados obtidos.

A ausência de informações específicas sobre o processo de coleta de achados nos artigos [2, 4, 6, 7, 8, 9, 10, 11, 12 e 13] serve como uma chamada para a necessidade de transparência e descrição meticulosa dos métodos utilizados, facilitando assim a replicabilidade e a validação da pesquisa.

Ao desvendar o processo de coleta de achados científicos, este tópico busca não apenas elucidar as variadas estratégias empregadas, mas também enfatizar a importância de uma gestão de dados eficaz e a incorporação de tecnologia neste processo.

Os achados

Os “achados” constituem a espinha dorsal desta revisão sistemática, lançando luz sobre as potencialidades e desafios inerentes ao uso da tecnologia *blockchain* na gestão de dados de pesquisa. A primeira descoberta salienta a aspiração acadêmica de “melhorar o impacto acadêmico”, uma narrativa que ressoa profundamente com a comunidade acadêmica e enfatiza a necessidade de práticas sólidas de gestão de dados [1].

Adentrando o domínio da *blockchain*, é revelado que a mineração e a imutabilidade são pilares centrais para assegurar a autenticidade e integridade dos dados ao longo do tempo [3]. Esta descoberta nos conduz a uma visão onde os dados brutos são transformados em ativos confiáveis, pavimentando o caminho para uma expressiva pesquisa futura.

A literatura oferece uma diversidade de estratégias para a proveniência de dados no gerenciamento científico, com a *blockchain* emergindo como uma força promissora para a segurança e transparência na rastreabilidade dos dados [6]. Este achado desdobra uma gama de soluções inovadoras, como contratos inteligentes e sistemas de votação descentralizada, que podem ser integrados na estrutura de gestão de dados.

A exploração da *blockchain* em bibliotecas e centros de informação revela um cenário de desafios e oportunidades, indicando uma metamorfose potencial na maneira como as instituições acadêmicas orquestram o compartilhamento de informações e o acesso ao conhecimento [8].

O enfoque na aplicação da *blockchain* ao longo do ciclo de vida dos dados de pesquisa destaca sua instrumentalidade na eficácia e segurança dos dados, sublinhando a *blockchain* como um vetor vital para a preservação e gestão abrangente dos dados científicos [9].

Os artigos [2, 4, 5, 7, 10, 11, 12 e 13], embora não apresentem informações específicas sobre achados, contribuem para a amplitude do escopo desta revisão, refletindo a vastidão e complexidade do domínio investigado.

Este tópico instiga as descobertas críticas que ressoam com a temática central da tese, proporcionando uma compreensão aprofundada das interseções entre *blockchain* e gestão de dados de pesquisa. A narrativa tecida aqui não apenas encapsula as descobertas críticas, mas também provoca uma reflexão sobre as

implicações práticas e futuras direções no entrelaçamento da tecnologia *blockchain* com a gestão de dados de pesquisa.

O processo de interpretação dos achados

O processo de interpretação dos achados é uma etapa crucial que segue a coleta e análise de dados, servindo como uma ponte para entender o significado e a relevância das descobertas no contexto maior da gestão de dados de pesquisa com a tecnologia *blockchain*.

Inicialmente, alguns estudos se debruçam sobre a análise exploratória e teórica da literatura para embasar a interpretação dos achados, construindo uma compreensão sólida sobre temas centrais como a comunicação científica, curadoria digital e interseção com a tecnologia *blockchain* [3, 8].

Em um veio distinto, a interpretação dos achados em estudos arqueológicos é destacada, em que a análise de vestígios ósseos e faunísticos oferece visões sobre épocas pré-históricas e históricas, demonstrando a amplitude das abordagens interpretativas na pesquisa científica [5].

A tecnologia *blockchain*, sendo central para esta revisão, é discutida em diferentes contextos, focando principalmente a segurança e transparência na rastreabilidade dos dados [6]. Embora o processo exato de interpretação não seja detalhado, é evidente que a *blockchain* tem implicações significativas na gestão de dados de pesquisa.

Um destaque particular é a menção da análise *SWOT*, a qual é uma técnica usada para identificar forças, oportunidades, fraquezas e ameaças com o intuito de desenvolver um plano estratégico, instrumental na avaliação da viabilidade e do potencial da aplicação da tecnologia *blockchain* na esfera acadêmica [9]. Esta análise proporciona uma lente crítica por meio da qual os benefícios e desafios da *blockchain* podem ser avaliados.

A necessidade de um sistema eficiente para o gerenciamento e análise de dados científicos em larga escala, introduzindo o *FASTDB* como uma solução inovadora (baseado em um modelo de matriz distribuída) [10].

Nos artigos [1, 2, 4, 7, 11, 12 e 13], a falta de informações específicas sobre o processo de interpretação dos achados indica uma área que pode necessitar de

investigação adicional, ressaltando a importância de uma exploração mais profunda nesta área.

Este tópico de interpretação dos achados delinea a variedade e complexidade das abordagens interpretativas, reiterando a importância de métodos rigorosos e bem articulados para extrair significado e compreensão a partir dos dados coletados.

A interpretação dos achados

A interpretação dos achados revela-se como um ponto de vista por meio do qual as nuances do gerenciamento de dados de pesquisa são exploradas. Os pesquisadores delinham funções centrais para uma interface eficaz de gerenciamento de dados, sublinhando a necessidade de autonomia do usuário, monitoramento contínuo do progresso e uma gestão eficaz da equipe de pesquisa [1].

Essas observações sinalizam a direção para a qual as plataformas de gerenciamento de dados podem evoluir para atender melhor às necessidades dos pesquisadores.

A tecnologia *blockchain* apresenta-se como uma solução potencial para vários desafios identificados. Notavelmente, a transparência e a rastreabilidade dos registros de pesquisa são ampliadas com o uso desta tecnologia, fornecendo uma solução confiável para documentar alterações em projetos de pesquisa [3]. Isso, por sua vez, pode promover uma maior responsabilidade e controle sobre o desenvolvimento dos projetos.

A proveniência dos dados emerge como um aspecto fundamental, com a tecnologia *blockchain* oferecendo uma plataforma segura e transparente para armazenar informações de proveniência [6]. Isso é crucial para estabelecer a confiança e rastreabilidade dos dados, aspectos ressaltados como vitais na literatura.

Aventurando-se além, a exploração do potencial da *blockchain* em bibliotecas e centros de informação revela um horizonte promissor para aprimorar a gestão e o compartilhamento de informações [8]. Contudo, também são reconhecidos desafios inerentes à implementação desta tecnologia disruptiva.

Finalmente, a segurança e a autenticidade dos dados de pesquisa universitária são identificadas como áreas que podem ser significativamente aprimoradas com o uso da tecnologia *blockchain*, reforçando a disponibilidade e validade dos dados [9].

Nos artigos [2, 4, 5, 7, 10, 11, 12 e 13], a interpretação dos achados não foi discutida em detalhe, sinalizando uma possível lacuna na literatura existente.

Este exercício de interpretação não apenas esclarece os achados, mas também ilustra o papel catalisador da *blockchain* na evolução das práticas de gerenciamento de dados de pesquisa. As implicações dessas interpretações são vastas e apontam para um caminho inovador na gestão de dados de pesquisa.

Os cuidados éticos

No tocante aos cuidados éticos, ressaltamos que a integridade da pesquisa e a confiabilidade dos resultados estão intrinsecamente atreladas à ética na publicação acadêmica. A prevenção contra plágio, fraude e viés não apenas preserva a qualidade dos resultados, mas também sustenta a reputação dos pesquisadores e instituições envolvidos.

A observância de normas éticas, estabelecidas por instituições acadêmicas e periódicos, é imperativa, assim como a imparcialidade dos revisores e editores no processo de revisão por pares.

A tecnologia *blockchain*, com sua capacidade de verificar a autenticidade e a data de criação de trabalhos de pesquisa, emerge como uma ferramenta potencial para prevenir o plágio e salvaguardar a prioridade de pesquisa [7]. Este aspecto ressalta o potencial disruptivo e benéfico da *blockchain* no fortalecimento da integridade acadêmica.

Profissionais de bibliotecas e da informação têm o dever de orientar suas comunidades sobre o uso ético da tecnologia *blockchain*. Ao promover uma compreensão genuína do potencial e das limitações desta tecnologia, eles podem ajudar a fomentar a confiança do público na integridade acadêmica [8].

No âmbito da saúde, a ética assume uma importância crítica no manuseio de dados sensíveis. Os desafios éticos incluem a proteção da privacidade, controle de acesso, propriedade dos dados e direitos de acesso a dados de pesquisa médica. A adoção dos princípios FAIR — que promovem acessibilidade, localização, interoperabilidade e reutilização dos dados — é vista como um passo significativo para garantir a integridade e qualidade dos resultados de pesquisa médica [13].

A análise dos artigos [1, 2, 3, 4, 5, 6, 9, 10, 11 e 12] não revelou informações adicionais sobre os cuidados éticos, apontando para uma área que pode necessitar de exploração adicional na literatura.

A implementação dos estudos

Em relação às implementações dos estudos, destaca-se a utilização de análise de conteúdo pelos pesquisadores para entender as necessidades de seus pares, culminando no design de uma interface de usuário intuitiva [1]. Este exemplo salienta a importância de métodos qualitativos e do design centrado no usuário, garantindo que as soluções propostas atendam às demandas reais da comunidade acadêmica.

Segue-se um exemplo prático com a criação de um dispositivo eletrônico computacional, ilustrando a aplicação direta de tecnologias na coleta de dados relevantes [3]. Tal iniciativa mostra como dispositivos customizados podem ser instrumentalizados para alcançar objetivos específicos de pesquisa.

A segurança e integridade dos dados de pesquisa são enfatizadas na discussão sobre sistemas de detecção de plágio avançados e a aplicação de carimbos de data/hora descentralizados via tecnologia *blockchain* [7]. Isso evidencia como a *blockchain* pode ser instrumentalizada para assegurar a autenticidade dos resultados científicos.

Adicionalmente, são propostas soluções e recomendações para integrar a tecnologia *blockchain* em bibliotecas e centros de informação [8]. Essa discussão sublinha a importância do treinamento e sensibilização dos profissionais, visando a uma adoção bem-sucedida de inovações tecnológicas.

A possibilidade de empregar a tecnologia *blockchain* para gerir todo o ciclo de vida dos dados de pesquisa universitária é explorada [9], destacando o potencial abrangente desta tecnologia em abordar desafios de segurança, autenticidade e integridade dos dados.

A proposta de um sistema de banco de dados distribuído para gerenciamento de dados científicos é apresentada, embora sem detalhamento específico sobre a implementação [10].

Nos artigos [2, 4, 5, 6, 11, 12 e 13], não foram identificadas informações adicionais sobre as implementações dos estudos.

As limitações dos estudos

Em relação às limitações dos estudos analisados, uma restrição recorrente manifesta-se nos estudos de caso, especialmente pela ausência de testes práticos [2]. Os estudos de caso, embora valiosos para desvendar concepções teóricas, carecem de fortes evidências empíricas, exigindo uma interpretação cautelosa dos achados.

É sublinhada a necessidade de uma exploração mais aprofundada, bem como a aplicação prática da tecnologia *blockchain*, indicando uma incerteza existente quanto à eficácia desta tecnologia em determinados domínios [7].

No contexto de bibliotecas e centros de informação, a implementação da *blockchain* é vista como desafiadora devido à sua arquitetura *criptográfica* e de rede, ao investimento inicial elevado, à possível perda de controle sobre recursos de informação e dados do usuário, além da associação da *blockchain* com atividades ilícitas [8]. Ressalta-se também a importância do treinamento e da sensibilização tanto para os profissionais quanto para a comunidade de usuários, visando a uma utilização ética e eficaz da tecnologia.

Outra limitação apontada é a necessidade de uma investigação mais ampla sobre a aplicação prática da *blockchain*, além de desafios em quantificar ou proteger certos tipos de dados por meio desta tecnologia, e a falta de padronização de dados e aplicativos [9].

A ausência de um benchmark (um referencial de mercado adotado para avaliar a eficácia de uma estratégia empresarial) para o processamento de dados científicos em plataformas distribuídas evidencia uma lacuna crucial na avaliação da eficiência dessas plataformas [11], sinalizando uma área de investigação futura potencial.

Para os artigos [1, 3, 4, 5, 6, 10, 12 e 13], não foram identificadas informações adicionais referentes às limitações dos estudos.

A figura 25 apresenta um esquema com os tópicos apresentados na revisão sistemática da literatura.

Figura 25 - Síntese da revisão sistemática



Fonte: Elaborado pelo autor.

Após concluir as análises pertinentes à revisão sistemática, no capítulo 8 apresentamos as razões para a escolha do Projeto Archangel e exploramos seu estudo de caso.

8 PROJETO ARCHANGEL (ARQUIVO NACIONAL DO REINO UNIDO)

Archangel é uma plataforma descentralizada desenvolvida para assegurar a integridade a longo prazo dos documentos digitais armazenados em arquivos públicos (Collomosse *et al.*, 2018). A tecnologia desenvolvida pelo Projeto Archangel e os resultados do estudo piloto se mostraram extremamente promissores. Eles demonstraram como a aplicação adequada de tecnologias emergentes pode transformar as metodologias de arquivamento digital e fomentar novas colaborações entre instituições (The ODI, 2024).

O Projeto Archangel exemplifica uma aplicação inovadora da tecnologia *blockchain* fora do âmbito das criptomoedas, abordando uma questão de grande importância: a possibilidade de que tecnologias semelhantes àquelas usadas em *blockchain*, particularmente no contexto de criptomoedas como o *bitcoin*, possam oferecer soluções eficazes para o gerenciamento de dados relacionados a uma ampla gama de informações provenientes de pesquisas científicas.

Muitas vezes considerado sinônimo de *bitcoin*, a *blockchain* é a tecnologia que sustenta uma série de moedas digitais, mas tem potencial para uma aplicação muito

mais ampla e dentro dessa perspectiva, The ODI (2024) em sua essência, explica que a *blockchain* é o equivalente digital de um livro-razão, como um banco de dados, mas com três recursos que o diferenciam dos bancos de dados padrão.

- Primeiro, uma tecnologia de registo distribuído é imutável – ou “apenas anexada” – o que significa que os dados não podem ser substituídos, alterados ou eliminados; ele só pode ser adicionado.
- Segundo, é distribuído. Nenhuma autoridade ou organização central possui a posse exclusiva dos dados. Em vez disso, cada membro da rede mantém uma cópia de todo o banco de dados e eles colaboram para validar cada nova entrada antes que ela seja gravada no livro-razão. Como resultado, não existe uma autoridade centralizada no controlo dos dados e cada participante tem um estatuto igual na rede: responsabilidade igual, direitos iguais e participação igual.
- Terceiro, é transparente. Todas as entradas no livro-razão são visíveis para todos que possuem uma cópia (The ODI, 2024, p. 8, tradução nossa).

Sendo assim, Bell *et al.* (2019) esclarecem que o projeto Archangel está investigando as oportunidades proporcionadas pela tecnologia de registo distribuído (DLT, frequentemente referida como *blockchain*) e examinando como essa tecnologia pode solucionar os desafios relacionados à confiança, integridade e autenticidade inerentes à preservação de materiais originados digitalmente. Dentro desse contexto, após essa breve explicação, apresentamos o estudo de caso do projeto Archangel.

8.1 O estudo de caso do Projeto Archangel (Arquivo Nacional do Reino Unido)

Este tópico concentra-se especificamente no estudo de caso do Projeto Archangel, uma iniciativa pioneira na aplicação da tecnologia *blockchain* para a preservação de registros digitais. Archangel é um projeto de pesquisa que, ao longo de 24 meses, explora a viabilidade sociotécnica por meio da criação e avaliação conjunta de um novo protótipo de serviço DLT (Tecnologia de *Ledger* Distribuído). Envolve usuários finais na determinação do impacto que uma tecnologia descentralizada e confiável pode ter sobre as práticas de arquivamento, desenvolvimento de modelos sustentáveis e mudanças nas percepções públicas. O objetivo é estabelecer formas de comprovar a integridade do conteúdo e assegurar o acesso livre aos arquivos públicos digitais (Archangel, 2024; UK Research and Innovation, 2024).

Na elaboração deste estudo de caso sobre o Projeto Archangel, focamos nos seguintes aspectos metodológicos: visão geral, funcionamento, parcerias, relevância, abordagem, impacto esperado e resultados.

A visão geral

Para enfrentar os desafios que as novas tecnologias impõem à preservação do nosso patrimônio digital, é essencial que os arquivos estejam mais abertos e dispostos a adotar e experimentar essas novas tecnologias (Bell *et al.*, 2019).

O Projeto Archangel surge como uma iniciativa pioneira e multidisciplinar na área de preservação digital, enfatizando o uso de tecnologias de *ledger* distribuído (DLT). Este projeto atende a uma necessidade crítica e premente em âmbito nacional: a de adaptar e proteger arquivos em um contexto digital que evolui rapidamente. Archangel visa fornecer ciência avançada, práticas inovadoras e modelos de trabalho eficientes para modernizar os arquivos, assegurando sua acessibilidade e integridade para as futuras gerações (University of Surrey, 2028).

Este projeto, desenvolvido em estreita colaboração com usuários finais, visa a criar, implementar e avaliar um protótipo que não só atenda às necessidades atuais de arquivamento digital, mas também explore as possíveis mudanças em práticas arquivísticas, modelos de sustentabilidade e percepções públicas.

Green *et al.* (2018) esclarecem que em um período em que as tecnologias capazes de modificar conteúdo digital estão se tornando cada vez mais acessíveis, não é de estranhar que a confiança do público em elementos digitais esteja diminuindo. A *blockchain* proporciona uma proteção que os arquivos podem utilizar para assegurar a autenticidade dos registros e diante desses desafios. O Projeto Archangel visa estabelecer um novo paradigma de confiança e estabilidade na preservação de arquivos digitais públicos, assegurando sua integridade no longo prazo em um cenário digital em constante evolução.

Como funciona

Segundo Collomosse *et al.* (2018) o projeto Archangel é estruturado como uma plataforma descentralizada que visa preservar a integridade de longo prazo de documentos digitais em arquivos públicos, movendo a base da confiança pública de

instituições para tecnologia de ledger distribuído (DLT), que criptografa e assegura a proveniência e imutabilidade dos documentos.

O Projeto Archangel destaca-se no campo da preservação digital pela sua aplicação pioneira da tecnologia de registro distribuído (DLT), frequentemente associada ao *blockchain*. Este projeto visa empregar a Tecnologia *Distributed Ledger* (DLT), visando validar a integridade e a origem de registros digitais. Essa verificação ocorre tanto no estágio de preservação dos registros (curadoria) quanto no momento de sua divulgação (apresentação) (Green *et al.*, 2018). Este estudo de caso ilustra que a *blockchain* vai além de sua associação inicial com criptomoedas, emergindo como um elemento crucial para a proteção da integridade de registros digitais.

Por meio da *blockchain*, Archangel não somente restaura, mas também fortalece a confiança nos registros digitais, demonstrando sua eficácia não apenas em transações financeiras, mas também como uma solução robusta para a segurança documental. O Projeto Archangel está investigando as oportunidades proporcionadas pela tecnologia de registro distribuído DLT e examinando como essa tecnologia pode solucionar os desafios relacionados à confiança, integridade e autenticidade inerentes à preservação de materiais originados digitalmente (Bell *et al.*, 2019; The ODI, 2024).

Bell *et al.* (2019) ao analisar o DLT, destacam que, apesar da complexidade tecnológica e criptográfica, pode-se compreendê-lo de forma mais clara como um banco de dados com duas características fundamentais: é um registro imutável, que permite apenas adições; distribuído, ou seja, cada participante da rede detém uma cópia. Isso implica que, uma vez que os dados de *hash* relativos à proveniência e ao conteúdo são adicionados ao *blockchain*, eles se tornam permanentes e inalteráveis.

Esta abordagem permite a verificação contínua da autenticidade de um documento, independentemente de eventuais mudanças de formato ao longo do tempo. Bell *et al.* (2019) explica que o Projeto Archangel criou um protótipo de aplicação para *desktop* que diversas instituições arquivísticas se propuseram a testar. Esse teste também revelará desafios relacionados à implementação do *software* em distintos ambientes tecnológicos e a viabilidade de integrá-lo a outros processos.

Assim, visando examinar esses aspectos e descobrir como a *blockchain* pode ser utilizada na autenticação de registros, o Archangel criou uma plataforma experimental privada fundamentada no Ethereum. Essa plataforma possibilita aos usuários a integração a uma *blockchain* autorizada, na qual é permitido o carregamento de metadados relativos aos seus registros, a validação de metadados

provenientes de outras entidades e a pesquisa e visualização dos resultados na *blockchain* (Bell *et al.*, 2019).

Os Arquivos e Instituições de Memória (AMIs) funcionam como as lentes por meio das quais as futuras gerações enxergarão o presente; eles constituem a memória oficial, econômica, social e cultural de uma nação, conforme descrito por (UK Research and Innovation, 2024) . Conforme apontado em The ODI (2024), um dos principais desafios enfrentados pelas AMIs é a transição de objetos predominantemente físicos para objetos majoritariamente digitais.

Em linguagem técnica, o Archangel gera *hashes* dos documentos e registra estas impressões em uma *blockchain* autorizada. Collomosse *et al.* (2018) informam que o Archangel utiliza um modelo de *blockchain* permitido, no qual apenas operadores ou processos automáticos autorizados podem adicionar conteúdo à AMI. Esses operadores comprometem blocos na cadeia, os quais codificam evidências de conteúdo. Caso essas evidências sejam derivadas de uma forma, não padrão, inclui-se um *hash* do código associado para a sua derivação.

Continuam informando que a segurança da *blockchain* é garantida pela imutabilidade dos dados nos blocos, alcançada pelo mecanismo de cada novo bloco ser *criptograficamente* ligado aos anteriores por meio de *hashes*. Portanto, conforme o conteúdo é adicionado a *blockchain*, a segurança desse conteúdo é progressivamente reforçada. Além disso, a *blockchain* mantém-se acessível publicamente, permitindo a verificação aberta de documentos do arquivo em qualquer momento (Collomosse *et al.*, 2018).

O registro do Archangel é mantido colaborativamente por diversos arquivos, incluindo internacionais, assegurando a verificação constante da integridade dos documentos em relação à sua assinatura original. Para Archangel (2024), do ponto de vista tecnológico, o projeto se beneficia do uso de aprendizado de máquina de última geração para coletar assinaturas digitais robustas, tanto de conteúdos físicos digitalizados quanto de digitais, num DLT autorizado. As assinaturas e o código usado para renderizar o conteúdo e verificar sua procedência e integridade são incorporados na DLT. Serão explorados novos modelos de negócios para sustentar a DLT, por exemplo, por meio de esforço contribuído (prova de trabalho), nos pontos de criação e consumo, utilizando um modelo *cross-AMI*. Neste modelo, uma única DLT é alimentada por múltiplas AMIs, abrangendo diferentes disciplinas e nações, mitigando o risco de distorção de arquivo pela AMI que o opera (UK Research and Innovation,

2024). Esta capacidade de verificação se mantém mesmo após a divulgação dos documentos, assegurando a inalterabilidade do registro.

As AMIs fundamentam-se no princípio da confiança pública, sustentando-se na ideia de serem entidades neutras e completamente confiáveis. A imutabilidade e a integridade das AMIs são cruciais para preservar a sua objetividade (UK Research and Innovation, 2024).

De acordo com Green *et al.* (2018) por meio desta tecnologia, arquivos ao redor do mundo conseguem oferecer proteção mútua, monitorando e validando a integridade dos dados para evitar corrupções, sejam elas acidentais ou intencionais.

O compromisso do Projeto Archangel com a confiabilidade de longo prazo dos arquivos públicos é crucial, atendendo não só às necessidades acadêmicas e de pesquisa, mas também a aplicações críticas como em litígios legais, inquéritos oficiais e manutenção da responsabilidade de organizações e governos.

Suas parcerias

O Projeto Archangel é uma iniciativa que se destaca por sua colaboração estratégica e multidisciplinar, crucial para alcançar seus objetivos ambiciosos e inovadores. Para empreender este projeto aventureiro e ambicioso, formamos uma parceria estratégica multidisciplinar unindo um grupo líder mundial em processamento de sinais multimodais (CVSSP)⁶, o Centro para a Economia Digital (CODE) da *Surrey Business School*⁷ e um consórcio de Partes interessadas da AMI, incluindo o Arquivo Nacional⁸ e o Instituto de Dados Abertos (ODI)⁹ (Archangel, 2024; UK Research and Innovation, 2024).

Esta parceria sinérgica congrega especialistas do Centro de Visão por Computador e Processamento de Sinais Multimodais (CVSSP), renomado globalmente por sua experiência em processamento de dados e sinais. Esta experiência é complementada pela participação do Centro para a Economia Digital (CODE) da *Surrey Business School*, que se destaca na análise das interseções entre tecnologia e negócios, crucial para entender as implicações econômicas do projeto.

⁶ <https://www.surrey.ac.uk/centre-vision-speech-signal-processing> acesso em 07/01/2024.

⁷ <https://www.surrey.ac.uk/business-school> acesso em 09/01/2024.

⁸ <https://www.nationalarchives.gov.uk/> acesso em 09/01/2024.

⁹ <https://theodi.org/> acesso em 07/01/2024.

Para UK Research and Innovation (2024) a colaboração no projeto Archangel é fortalecida pela inclusão de um consórcio de partes interessadas da *Advanced Music Information* (AMI), que inclui entidades renomadas como os Arquivos Nacionais e o *Open Data Institute* (ODI), cofundado por Tim Berners-Lee. Por meio da parceria com o ODI, o Archangel buscará o envolvimento de diversas organizações, tanto do setor público quanto do setor comercial, além de grupos de diferentes setores. O objetivo é construir a infraestrutura do Archangel para aumentar a eficiência e extrair mais valor dos dados, sejam eles abertos, compartilhados ou restritos. A descentralização é um fator crucial na criação de infraestruturas de dados compartilhadas e confiáveis, indicando que este projeto terá aplicações relevantes em várias áreas da economia.

Esses parceiros compartilham uma visão que está em harmonia com os princípios do Archangel, focados em dados abertos e transparentes. A implementação técnica do projeto fica a cargo da *Guardtime*¹⁰, uma provedora de infraestruturas de DLT, que traz consigo uma experiência tecnológica crucial para a realização do serviço. Um aspecto de potencial impacto significativo é a adaptação da plataforma global DLT da *Guardtime* para criar uma infraestrutura inovadora baseada em DLT, dedicada a arquivos digitais públicos confiáveis (UK Research and Innovation, 2024).

A Archangel colabora com o *Open Data Institute* de Sir Tim Berners-Lee, trazendo experiência líder mundial em plataformas confiáveis de dados abertos em escala *web*. O Arquivo Nacional, um parceiro de renome internacional, desempenha um papel duplo no projeto: como usuário final e também como uma organização de pesquisa independente, que co-criou o Archangel. O impacto do Archangel vai além das práticas arquivísticas, alcançando também a esfera política. O projeto planeja gerar um impacto intersetorial mais amplo por meio de um programa criativo de atividades (University of Surrey, 2018).

Estas alianças permitem que o Projeto Archangel avance na fronteira do conhecimento e na aplicação prática da tecnologia DLT, contribuindo significativamente para a preservação do patrimônio digital global.

¹⁰ <https://guardtime.com/> acesso em 09/01/2024.

Sua importância

O Projeto Archangel emerge como uma iniciativa de grande importância no contexto da preservação digital, tendo como objetivo primordial garantir a durabilidade dos registros digitais, uma questão crucial na era da informação. Para Archangel (2024), University of Surrey, 2018 e UK Research and Innovation, 2024) este projeto está focado no desenvolvimento e implementação de uma inovadora tecnologia de *ledger* distribuído (DLT) e planeja assegurar a sustentabilidade a longo prazo dos arquivos digitais, por meio do design, desenvolvimento e teste de nova tecnologia transformadora de *ledger* distribuído (DLT). A finalidade é aumentar a acessibilidade e garantir a integridade do conteúdo, ao mesmo tempo, em que maximiza o impacto do projeto por meio de novos modelos de negócios voltados para a comercialização e o acesso aberto aos dados.

Na era digital atual, Arquivos e Instituições de Memória (AMIs) enfrentam novos desafios sociotécnicos na preservação de dados. Registros públicos digitais, por serem intangíveis, podem ser facilmente removidos ou alterados de forma imperceptível. De fato, em certos casos, é necessário modificar esses registros para assegurar sua acessibilidade contínua, especialmente com a mudança de formatos. Além disso, a curadoria de dados envolve a manutenção do código associado para apresentar esses dados, frequentemente por décadas (Archangel, 2024; UK Research and Innovation, 2024). O foco não está apenas na proteção da integridade dos conteúdos arquivísticos, mas também em ampliar sua acessibilidade.

Além disso, conforme indicado por UK Research and Innovation (2024), o Archangel tem o potencial de transformar a sustentabilidade dos arquivos públicos digitais. Esse projeto promete ter um impacto horizontal e de longo prazo em todos os setores da Economia Digital, beneficiando-se da maior integridade e acessibilidade desses arquivos.

Em um ambiente caracterizado pela rápida mudança de informações e pela obsolescência tecnológica, soluções como as oferecidas pelo Archangel são essenciais para assegurar a preservação confiável e sustentável do patrimônio digital humano. Para Collomosse *et al.* (2018) a era digital atual traz desafios novos e urgentes relacionados à confiança e imutabilidade. Documentos digitais são efêmeros e produzidos em grande volume. Sua natureza intangível os torna suscetíveis a modificações — não apenas adulterações, mas também degradações ao longo do

tempo (como em períodos de décadas). Por exemplo, formatos de arquivo podem tornar-se obsoletos, levando à necessidade de transcodificação dos documentos.

Bell *et al.* (2019) nos dá o exemplo do *Word Star* que foi um processador de texto muito popular nos anos 1980, mas hoje não há uma versão compatível com os computadores modernos, embora existam emuladores criados por entusiastas. Um arquivo *Word Star* pode ser aberto em uma versão atual do Microsoft Word, mas somente após a instalação de um suplemento de conversão. Isso mantém o formato acessível por ora, mas não há garantias de que esses arquivos poderão ser renderizados em um computador padrão em 20 ou 50 anos.

Além disso, um processador de texto moderno pode não reproduzir o arquivo original com exatidão. Pensando na preservação a longo prazo e na conveniência dos usuários, o arquivo pode criar cópias desses arquivos *Word Star* e convertê-los para um formato aberto, que provavelmente ainda será legível em algumas décadas (Bell *et al.*, 2019).

O Archangel supera desafios técnicos ao inovar nos modelos de negócios associados aos arquivos digitais. Para UK Research and Innovation (2024) o Archangel desenvolveu um *software* baseado na tecnologia *blockchain*, visando aumentar a confiança de arquivistas e do público nos arquivos digitais. O *software*, inicialmente desenvolvido para arquivos de dados de pesquisas universitárias, foi expandido para proteger registros de vídeo digital e pode ser adaptado para qualquer tipo de registro digital.

Ele foi testado com usuários finais na gestão de dados de pesquisa em setores comerciais e acadêmicos, e em arquivos nacionais, por meio de uma implementação piloto em diversos arquivos governamentais internacionais. Essa experiência proporcionou novas percepções sobre o valor da tecnologia Archangel nesses contextos e orientará o projeto em sua próxima fase, focando em outros tipos de conteúdo digital nos Arquivos Nacionais do Reino Unido (UK Research and Innovation, 2024). Ao investigar novas formas de comercialização e acesso aberto, visa a maximizar o impacto social e econômico desses arquivos, estendendo seu valor além do convencional.

Este projeto possibilita inéditas formas de utilização e geração de renda a partir dos arquivos digitais. A relevância do Archangel estende-se para além do âmbito tecnológico, oferecendo uma abordagem holística que engloba tanto aspectos

técnicos quanto comerciais, contribuindo significativamente para a remodelação do panorama dos arquivos digitais em um contexto global.

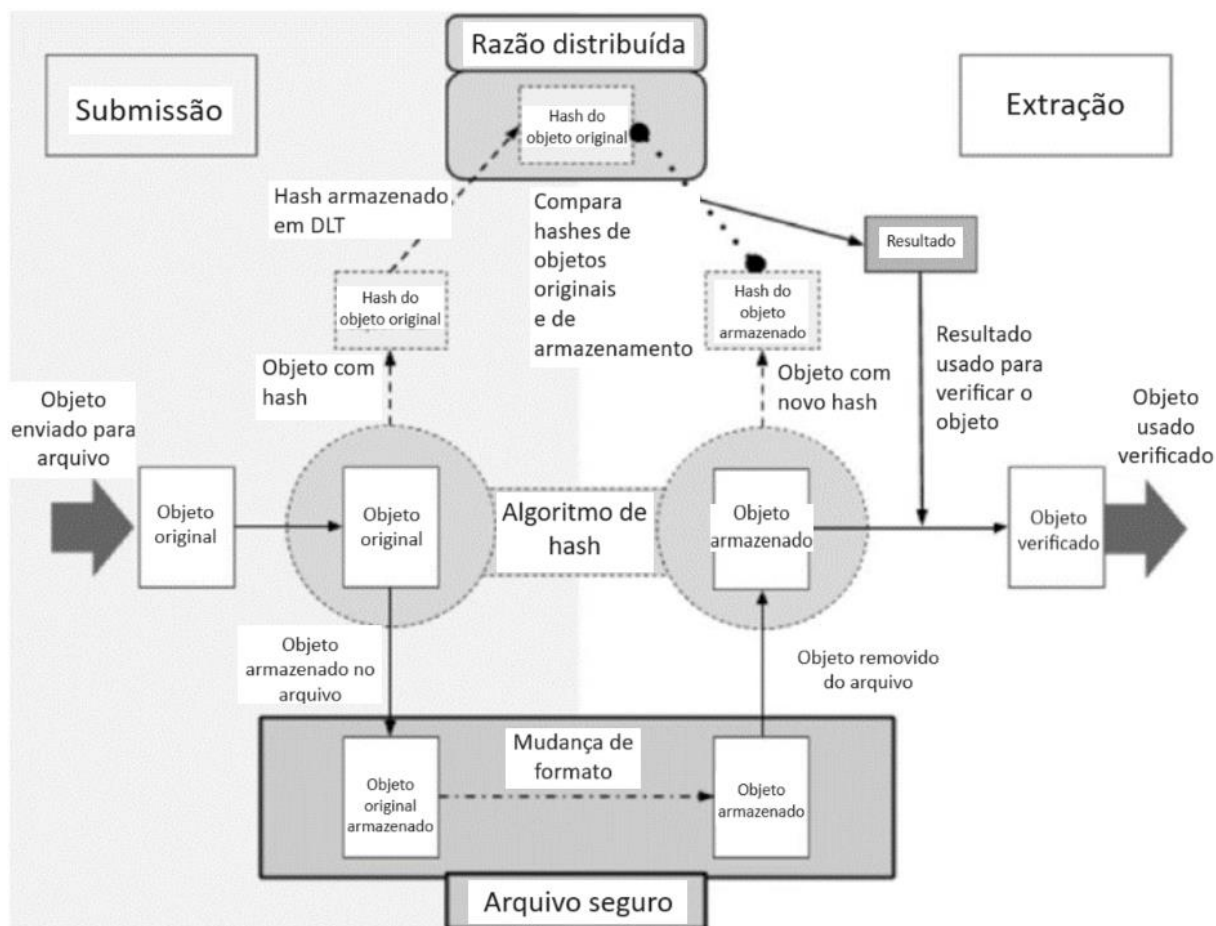
A abordagem

No Projeto Archangel, adota-se uma metodologia tecnológica avançada e multidisciplinar, integrando o que há de mais moderno em aprendizado de máquina e tecnologias de *ledger* distribuído (DLT).

Descrevemos o Archangel como uma arquitetura descentralizada projetada para assegurar a integridade de documentos digitais em arquivos públicos que utiliza *blockchains*, uma forma de tecnologia de *ledger* distribuído (DLT) segura, como alicerce para garantir a procedência e a integridade dos documentos. *Blockchains* registram dados de transações de maneira cronológica e preservam permanentemente esses dados por uma distribuição *peer-to-peer* e um mecanismo de verificação de consenso, eliminando a necessidade de uma terceira parte confiável (Collomosse *et al.*, 2018).

Esta metodologia concentra-se na coleta e no registro de assinaturas digitais, obtidas tanto de documentos físicos digitalizados quanto de registros digitais nativos, em um sistema DLT autorizado. Apresentamos na figura 26 a proposta da arquitetura da plataforma Archangel.

Figura 26 - A arquitetura da plataforma Archangel proposta



Fonte: Bell *et al.*, 2019; Collomosse *et al.*, 2018, tradução nossa.

Conforme apresentado na figura 26, os registros passam por um processo de análise para extrair evidências de seu conteúdo, as quais são armazenadas de maneira imutável em uma *blockchain*, juntamente com metadados que identificam tanto o conteúdo em si quanto o algoritmo utilizado para extrair tais evidências. Isso permite que, em qualquer momento, a integridade e a procedência de um registro sejam verificadas ao reextrair e comparar a evidência de conteúdo presente no *blockchain* com a evidência recém-extraída.

Esta abordagem é vital para garantir que, apesar das futuras mudanças tecnológicas ou de formato, a integridade e a procedência dos registros sejam irrefutavelmente estabelecidas e mantidas. O Archangel aborda o desafio multidimensional de assegurar a integridade e a sustentabilidade de longo prazo do conteúdo arquivístico digital, propondo uma plataforma dedicada à verificação da integridade e da origem dos documentos digitais (Collomosse *et al.*, 2018).

Além disso, o Archangel inova ao investigar e desenvolver novos modelos de negócios para sustentar a infraestrutura de DLT. Esses modelos serão aplicados tanto nos pontos de criação quanto de consumo, utilizando um modelo transversal às AMIs. Nesse modelo, uma única DLT recebe contribuições de diversas AMIs, abrangendo diferentes disciplinas e países. Isso permite mitigar o risco de distorção dos arquivos por parte da AMI que os opera (Archangel, 2024).

Estes modelos são explorados tanto na fase de criação quanto na de consumo dos registros digitais, examinando, mecanismos, como o esforço contribuído, análogo à prova de trabalho em criptomoedas.

Para Collomosse *et al.* (2018, p. 3) em nossa arquitetura propomos dois modos de verificação de consenso, ambos baseados em um modelo DLT permitido:

(1) O *Blockchain* é mantido através de prova de trabalho através de um conjunto privado de nós, que são mantidos coletivamente por múltiplas AMIs, cada uma com uma estrutura de governança independente, por exemplo, arquivos nacionais de diferentes estados-nação. Como tal, seria necessário um nível de conluio sem precedentes para corromper o *Blockchain*.

(2) O *Blockchain* é mantido através de prova de trabalho em um *Blockchain* público mantido globalmente. Nesse caso, um programa incorporado no *Blockchain* (um 'contrato inteligente') com permissão exclusiva para escrever no *Blockchain* é invocado para anexar dados. O acesso ao terminal do contrato inteligente é concedido por meio de uma chave secreta. Neste caso, a corrupção exigiria que mais de metade dos mineiros de infra-estruturas públicas de DLT conspirassem, o que é mais uma vez improvável, por exemplo, na rede principal Ethereum (Collomosse *et al.*, 2018, p. 3, tradução nossa).

Este aspecto é crucial para a viabilidade e sustentabilidade do sistema, representando uma exploração pioneira de estruturas econômicas que buscam estabelecer um sistema de arquivamento digital não apenas tecnologicamente avançado, mas também economicamente sustentável.

O impacto planejado

Os Arquivos e Instituições de Memória fundamentam-se nos princípios da confiança pública, pautando-se pela neutralidade e total confiabilidade. A imutabilidade e a integridade dessas instituições são cruciais para preservar sua objetividade (University of Surrey, 2018). O Projeto Archangel está posicionado para ter um impacto significativo e abrangente no âmbito da economia digital,

transformando a maneira como os arquivos públicos digitais são geridos e preservados.

Além do âmbito arquivístico, os resultados do projeto Archangel possuem o potencial de influenciar e auxiliar outros campos onde a verdade e a integridade da informação ao longo do tempo, como no jornalismo, são fundamentais para a sua sustentabilidade a longo prazo (The ODI, 2024). Esta transformação estende-se para além de um único domínio, impactando uma variedade de setores de forma vertical e horizontal, garantindo maior integridade e acessibilidade aos registros digitais.

No âmbito governamental e de serviços públicos, o Archangel apresenta um grande potencial para reforçar a gestão e a preservação de evidências políticas essenciais. Segundo Green *et al.* (2018) a tecnologia *blockchain* atua como um escudo que os arquivos podem utilizar para assegurar a autenticidade dos registros. Ao possibilitar que pesquisadores comparem a soma de verificação e a identificação dos metadados do registro com os armazenados na *blockchain*, eles podem constatar que não houve alterações (sejam elas deliberadas ou acidentais) no registro desde sua preservação no arquivo.

Este método oferece aos pesquisadores a certeza de que o registro é confiável como evidência proveniente do arquivo. O Archangel propicia uma transição da confiança institucional para a confiança tecnológica.

No contexto do cumprimento de normas regulatórias, especialmente nos setores financeiro e comercial, o Archangel surge como uma promissora solução para um repositório de dados confiável, com potencial para revolucionar o armazenamento e a verificação de informações de conformidade.

O Archangel proporcionará essa garantia por meio de *blockchains*. Os arquivos criarão *hashes* e registrarão assinaturas digitais de documentos em um *blockchain* autorizado, mantido colaborativamente por diversos arquivos participantes, inclusive em âmbito internacional. Assim, os documentos poderão ser verificados em relação à assinatura original a qualquer momento, inclusive no ato da divulgação, assegurando a integridade do registro (University of Surrey, 2018).

A indústria da propriedade intelectual, incluindo agências de registro de patentes, também se beneficiará da capacidade do projeto de autenticar de forma indelével a origem e titularidade dos registros.

O impacto do Archangel é percebido de maneira ampla na sociedade, beneficiando arquivos e outras instituições de memória. Ele se estende desde a

preservação de arquivos culturais, cruciais para a nossa identidade coletiva, até os Arquivos Nacionais, responsáveis por manter registros governamentais.

O alcance do projeto não se restringe apenas às AMIs tradicionais, mas abrange qualquer arquivo público digital. Isso inclui repositórios de dados de pesquisa universitária, além de melhorar a gestão da memória corporativa em multinacionais, como a conformidade financeira/regulatória e a gestão de registros de estado da técnica em empresas de tecnologia (Archangel, 2024).

Este amplo impacto reforça a infraestrutura da memória coletiva da sociedade, promovendo sua integridade e acessibilidade para as futuras gerações. Em resumo, o Archangel está preparado para ser um catalisador de mudança significativa, promovendo a confiabilidade e a longevidade dos arquivos digitais em um mundo crescentemente orientado para o digital.

Os resultados

O estudo de caso do Projeto Archangel demonstra um progresso notável na confiabilidade dos registros digitais, marcado pelo desenvolvimento de uma solução de software inovadora baseada na tecnologia *blockchain*. A tecnologia desenvolvida pelo projeto Archangel e os resultados do estudo piloto se mostraram extremamente promissores. Eles demonstraram como a aplicação adequada de tecnologias emergentes pode transformar as metodologias de arquivamento digital e fomentar novas colaborações entre instituições (The ODI, 2024).

Este *software* foi inicialmente concebido e testado com conjuntos de dados de pesquisa acadêmica, em que comprovou sua eficácia, e foi posteriormente adaptado para a proteção de registros de vídeo digital, destacando sua versatilidade em diversos domínios de arquivamento digital.

A aplicação prática deste *software* passou por testes rigorosos com usuários finais em contextos acadêmicos e comerciais, gerenciando dados de pesquisa. Além disso, uma implementação piloto foi realizada com arquivos nacionais, expandindo seu alcance para incluir vários arquivos governamentais internacionais. De acordo com (University of Surrey, 2018), a plataforma de teste Surrey *Blockchain* compreende:

1. Infraestrutura tecnológica que integra múltiplas plataformas DLT executando protótipos de pesquisa para nossos projetos,

principalmente em redes de teste Ethereum públicas e privadas das quais somos membros.

2. Experiência líder mundial na fusão de tecnologias de IA e *Blockchain* com foco sociotécnico em nossa pesquisa que explora não apenas as tecnologias, mas também os sistemas socioeconômicos dentro dos quais é viável para elas operarem
3. Acesso a um ecossistema diversificado de usuários finais com os quais nos envolvemos para design participativo e teste de nossos protótipos DLT (University of Surrey, 2018, tradução nossa).

Estes experimentos trouxeram visões valiosas sobre o potencial da tecnologia Archangel, revelando diferentes contextos de uso e fornecendo informações essenciais para o aprimoramento contínuo do projeto. De acordo com The ODI (2024), conforme o projeto progrediu, esses resultados iniciais proporcionaram uma base robusta para expandir a aplicação da tecnologia Archangel em uma escala mais ampla nos Arquivos Nacionais do Reino Unido, abrangendo uma variedade de conteúdos digitais. O sistema Archangel, desenvolvido por meio deste projeto, emprega tecnologia de *ledger* distribuído para assegurar a imutabilidade das impressões digitais dos documentos e utiliza aprendizado de máquina para criar impressões digitais adaptáveis às mudanças de formato.

Assim, o estudo de caso do Archangel não só ilustra o potencial transformador de um *software* baseado em *blockchain* para a gestão de arquivos digitais, mas também estabelece um marco para futuras inovações no campo da preservação digital no longo prazo. Nesse contexto, ainda, um aspecto central do trabalho dos pesquisadores em preservação digital é assegurar a integridade de objetos nativos digitais, isto é, garantir que eles permaneçam inalterados durante o período em que são armazenados no arquivo (The ODI, 2024). Na figura 27 apresentamos um esquema com os tópicos apresentados no estudo de caso do projeto Archangel.

Figura 27 - Estudo de caso do projeto Archangel



Fonte: Elaborado pelo autor.

Após concluir o estudo de caso do Projeto Archangel, no capítulo 9 apresentamos as razões para a escolha do Projeto dARK e exploramos seu estudo de caso. Além disso, discutimos sobre os identificadores persistentes e introduzimos o conceito de objeto digital FAIR.

9 PROJETO dARK (INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA E DA REDE LA REFERENCIA)

De acordo com Washington Segundo *et al.* (2022), o projeto dARK foi concebido com o objetivo fundamental de permitir que múltiplas instituições gerenciem, colaborativamente, seus sistemas de Identificadores Persistentes (PID). Isso inclui a atribuição e reutilização desses identificadores em uma infraestrutura descentralizada comum. Esta abordagem descentralizada é uma resposta direta às limitações observadas em sistemas centralizados, que dependem de uma única autoridade para executar e coordenar operações.

Em contraste, os sistemas descentralizados distribuem essas tarefas entre os “nós” (qualquer servidor computacional que pode armazenar uma cópia da *blockchain* e usá-la para verificar transações, ou seja, uma cópia do livro-razão operado pelos participantes da rede *blockchain*) da rede, promovendo uma gestão mais autônoma e resiliente (Washington Segundo *et al.*, 2022).

Os identificadores persistentes, conforme descritos por Santos e Maimone (2023), são códigos alfanuméricos com configurações específicas adaptadas ao seu contexto e finalidade. Eles desempenham um papel crucial na otimização do processo

de recuperação de informações científicas em ambientes digitais. Ao empregar esses identificadores, o projeto dARK facilita o acesso e a localização precisa de dados e recursos científicos, contribuindo significativamente para a eficiência da pesquisa científica.

Ao considerar um sistema descentralizado que emprega a tecnologia *blockchain*, identificam-se várias vantagens, incluindo:

- Segurança aprimorada: Devido às ferramentas criptográficas robustas empregadas no núcleo desses sistemas, cada usuário deve ser identificado e assinar cada transação usando suas chaves criptográficas. Essa característica adiciona uma camada de segurança aos sistemas;
- Transparência e confiabilidade: Cada transação realizada nesse sistema é verificada por todos os nós da rede. Todas as transações e dados armazenados nesses sistemas são auditáveis, tornando o sistema transparente e confiável;
- Garantias de evidência de violação: Os dados armazenados não podem ser modificados após inserção. Contrapondo um sistema centralizado, onde qualquer pessoa no topo da hierarquia com autorização pode fazer alterações;
- Procedência de dados inseridos: uma vez que os dados são imutáveis e cada transação indica quem e quando foi executada, esses sistemas têm um sistema incorporado de procedência de dados (Washington Segundo *et al.*, 2023, p. 152 e 153).

A tecnologia *blockchain* é fundamental para o dARK, oferecendo um armazenamento imutável que serve como prova contra violações. Neste sistema, as transações são registradas em uma cadeia de blocos, com os dados armazenados em blocos interconectados por meio de ferramentas *criptográficas* (Washington Segundo *et al.*, 2023). Esta característica assegura a integridade e a segurança dos registros, fundamentais para a gestão de dados de pesquisa.

Em resumo, a escolha pelo projeto dARK reflete um compromisso com a inovação na gestão de dados de pesquisa, adotando uma infraestrutura descentralizada que não apenas supera as limitações dos sistemas centralizados, mas também promove a segurança, a integridade e a acessibilidade dos dados científicos.

Verificamos que por meio da implementação de tecnologia *blockchain* e do uso de identificadores persistentes, o dARK está estabelecendo um novo paradigma para a recuperação e gestão de informações científicas, marcando um avanço significativo

na promoção da ciência aberta e colaborativa. Dentro desse contexto, após essa breve explicação, apresentamos o estudo de caso do projeto dARK.

9.1 O estudo de caso do Projeto dARK (Instituto Brasileiro de Informação em Ciência e Tecnologia e da Rede LA Referencia)

Este tópico concentra-se especificamente no estudo de caso do Projeto dARK, projeto dARK, uma iniciativa que visa desenvolver uma ferramenta de atribuição de identificadores persistentes (PIDs) de forma aberta e descentralizada. A ferramenta combina o identificador *Archival Resource Key* (ARK) com a tecnologia *Blockchain*, permitindo que as instituições parceiras possam gerenciar e preservar seus objetos digitais de forma segura e confiável (Washington Segundo *et al.*, 2022).

Antes de mergulharmos no estudo de caso do projeto dARK, é essencial contextualizar o papel dos identificadores persistentes. Eles são cruciais na gestão de informações digitais. Atuando como uma solução eficaz para garantir um acesso seguro e de longo prazo a recursos digitais num ambiente online em constante mudança, os identificadores persistentes visam facilitar o acesso a informações científicas e elucidar diversos métodos de organização (Santos; Maimone, 2023).

A digitalização e a expansão rápida da internet facilitaram a distribuição de informações como nunca, mas também introduziram desafios significativos para encontrar e recuperar dados e documentos digitais com o passar do tempo. Santos e Maimone (2023) destacam como a volatilidade das URLs e a mudança contínua nos repositórios digitais podem comprometer o acesso ininterrupto a recursos específicos, criando barreiras significativas para a pesquisa, educação e comunicação acadêmica.

Sayão (2007) esclarece que a denominação atribuída a um identificador persistente é a de um recurso digital que se mantém inalterado permanentemente, independentemente de onde o recurso esteja localizado. A implementação de um identificador persistente garante que, mesmo se um documento for deslocado ou a sua posse mudar, os links apontando para ele continuem funcionais.

Nesse cenário, os indicadores persistentes emergem como ferramentas indispensáveis, assegurando que os recursos digitais possam ser consistentemente localizados, acessados e referenciados, independente de alterações em suas localizações ou nas estruturas de armazenamento. Gabriel Junior *et al.* (2020) ressaltam a importância dos indicadores persistentes não apenas como mecanismos

tecnológicos, mas também como elementos estratégicos vitais para a gestão do conhecimento digital projetados para garantir uma identificação exclusiva e duradoura de objetos digitais na *web*, desvinculada do local físico de armazenamento.

Além de facilitarem o acesso a recursos valiosos, os indicadores persistentes desempenham um papel crucial na manutenção e perpetuação do conhecimento digital. Santos e Maimone (2023) observam que, devido ao seu caráter dinâmico, esses identificadores estão constantemente sendo atualizados e aprimorados. Este fator sublinha a importância de contextualizar as análises e referências aos indicadores persistentes, considerando o momento específico em que foram utilizados. Tal consideração é essencial para a interpretação adequada dos dados e para a aplicabilidade das informações em pesquisas futuras.

De acordo com Gabriel Junior *et al.* (2020) apresentamos no quadro 8 detalhes sobre os identificadores persistentes URN, ARK, PURL, Handle System e DOI.

Quadro 8 - Identificadores persistentes URN, ARK, PURL, Handle System e DOI

Identificadores persistentes	
<i>Uniform Resource Name (URN)</i>	URN, ou Nome Uniforme de Recurso, é uma categoria específica dentro do Identificador Uniforme de Recurso (URI) que exclui as URLs. Trata-se de um identificador estável e desvinculado de localizações específicas, que facilita a organização de diversos <i>namespaces</i> em um único <i>namespace</i> URN. Como ilustração, os <i>namespaces</i> isbn e issn são exemplos registrados para a referência de livros e publicações periódicas, respectivamente.
<i>Archival Resource Key (ARK)</i>	O ARK, sigla para Chave de Recurso de Arquivamento, é um sistema de identificação desenvolvido pela Biblioteca Digital da Califórnia (CDL), destinado à identificação persistente de objetos. Este sistema foi criado sob a premissa de que a persistência é uma característica relacionada exclusivamente aos serviços providos e não está intrinsecamente ligada ao objeto nem é determinada por uma sintaxe de nomeação específica.
<i>Persistent Uniform Resource Locator (PURL)</i>	PURL, ou Localizador Uniforme de Recursos Persistente, representa uma URL duradoura que oferece um meio para acessar um recurso na Web por meio de códigos de status HTTP convencionais. Assim, um PURL funciona como um endereço web fixo que possui instruções para redirecionar a outro local, que pode variar ao longo do tempo. Isso permite que, caso um autor necessite modificar o endereço onde um

	recurso está disponível, ele possa simplesmente atualizar o PURL para direcionar à nova localização.
Handle System	Os <i>handles</i> representam identificadores únicos e duradouros para recursos na Internet, operando com um registro central encarregado de atualizar as URLs para seus locais atuais. Embora sejam permanentes, os <i>handles</i> oferecem flexibilidade para modificar as informações de localização e acesso aos objetos. Cada <i>handle</i> é vinculado exclusivamente a um recurso e à entidade responsável por sua criação ou manutenção. Adicionalmente, o Sistema <i>Handle</i> fornece a base técnica para o DOI, que constitui uma forma específica de <i>handle</i> .
<i>Digital Object Identifier (DOI)</i>	O DOI (Identificador Digital de Objetos) é um sistema de identificação para objetos digitais (sejam eles originariamente digitais, digitalizados ou conceituais) que pode ser concedido por entidades filiadas a uma das agências de registro do DOI. Este sistema conta com uma infraestrutura projetada para assegurar que uma URL direcione corretamente para o local do objeto digital. A identidade DOI de um objeto é constante ao longo de sua existência, embora sua localização e outros metadados associados possam ser modificados conforme necessário.

Fonte: Elaborado pelo autor com base em Gabriel Junior *et al.*, 2020.

Santos e Maimone (2023) salientam que, além das várias alternativas de identificadores persistentes, apresentam no quadro 9 o ORCID, *ResearcherID* e *ScopusID*, voltados especificamente para a identificação de pesquisadores.

Quadro 9 - Identificadores persistentes ORCID, ResearcherID e ScopusID

Identificadores de pessoas (pesquisadores)	
ORCID	Visando a interoperabilidade e o estabelecimento como um identificador autoral persistente, o ORCID tem conquistado adoção global desde sua introdução, oferecendo a possibilidade de criação de um perfil gratuitamente. Neste perfil, o pesquisador tem a oportunidade de adicionar informações sobre sua carreira acadêmica e profissional. Este identificador tornou-se um requisito por várias revistas, tanto nacionais quanto internacionais, para a submissão de trabalhos, bem como por agências de financiamento, para a apresentação de projetos.

<i>ResearcherID</i>	O <i>ResearcherID</i> é um identificador que faz parte de um conglomerado empresarial e está associado à base de dados Web of Science, de certa forma. No entanto, oferece ao pesquisador autonomia para administrar seu perfil, permitindo-lhe adicionar publicações independentemente da fonte, detalhar sua trajetória acadêmica e profissional e, crucialmente, dar destaque às suas contribuições como revisor, um aspecto menos visível, mas fundamental para o avanço científico. Esta funcionalidade enfatiza a participação ativa do pesquisador em processos essenciais para o desenvolvimento da ciência.
<i>ScopusID</i>	O <i>ScopusID</i> proporciona ao pesquisador um perfil rico em visualizações e funcionalidades úteis para os usuários que acessam a base de dados. No entanto, a limitação na autonomia do pesquisador para efetuar modificações no seu próprio perfil e a restrição que permite apenas a exibição de publicações indexadas pela própria base de dados, podem desencorajar os pesquisadores a promover e manter seus perfis atualizados, a menos que entrem em contato com o suporte para realizar essas alterações.

Fonte: Elaborado pelo autor com base em Santos e Maimone, 2023.

Os identificadores persistentes não apenas superam os desafios impostos pela volatilidade da *web*, garantindo acesso confiável e duradouro a recursos digitais, mas também fortalecem a estrutura de organização e disseminação do conhecimento científico e acadêmico. As contribuições de Santos e Maimone (2023), bem como as de Gabriel Junior *et al.* (2020), ilustram a diversidade e a especificidade dos identificadores persistentes, desde URN, ARK, PURL, *Handle System*, até DOI, além de destacar a importância dos identificadores de pesquisadores como ORCID, *ResearcherID* e *ScopusID*.

Esse entendimento prepara o terreno para nosso estudo de caso do projeto dARK, permitindo-nos apreciar plenamente como essas ferramentas podem ser aplicadas para resolver questões práticas no gerenciamento de dados e documentos digitais. O projeto dARK, com sua ênfase em identificação, preservação e acessibilidade de informações digitais, serve como um exemplo contemporâneo da aplicação dos princípios e tecnologias discutidos.

Ao avançarmos, manteremos um foco especial na maneira como os identificadores persistentes podem facilitar não apenas a localização e o acesso a recursos digitais, mas também como podem promover a integridade, autenticidade e persistência desses recursos ao longo do tempo. Este estudo de caso nos permitirá explorar as implicações práticas dessas tecnologias e como elas estão sendo implementadas para enfrentar os desafios contemporâneos na gestão de informações digitais.

Na elaboração deste estudo de caso sobre o Projeto dARK, focamos nos seguintes aspectos metodológicos: visão geral, funcionamento, parcerias, relevância, abordagem, impacto esperado e resultados.

Visão geral

Washington Segundo *et al.* (2022) ressaltam que o projeto dARK, uma iniciativa do Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT) e da Rede *LA Referencia*, constitui um avanço importante nos desafios de preservação digital e no acesso livre a dados de pesquisa científica no Brasil. O projeto aplica a tecnologia *blockchain* para estabelecer uma rede de nós institucionais, promovendo uma gestão de dados distribuída, na qual cada entidade participante adiciona à rede, reduzindo assim a necessidade de infraestrutura computacional centralizada.

Para Gabriel Junior *et al.* (2020) a relevância do dARK no cenário científico brasileiro é destacada pela adoção do *Archival Resource Key* (ARK), um esquema de identificação desenvolvido pela *California Digital Library*¹¹ (CDL) para identificar persistentemente objetos digitais e físicos. Esta abordagem não só promove a ciência aberta, mas também facilita a colaboração científica por uma infraestrutura que suporta esses ideais.

Este projeto visa criar uma infraestrutura descentralizada para o armazenamento, preservação e acesso contínuo a dados científicos, com foco especial nos dados em risco de desaparecimento ou obsolescência digital. A imutabilidade garantida pela *blockchain* aumenta a transparência e confiança nos Identificadores Persistentes (PID), assegurando que os dados, uma vez registrados,

¹¹ <http://www.cdlib.org/inside/diglib/ark/> acessado em 20/01/2024.

permaneçam inalterados, promovendo assim a ciência aberta e a transparência na pesquisa científica brasileira (Washington Segundo *et al.*, 2022).

Um dos aspectos inovadores do dARK é a adoção de indicadores persistentes, que asseguram a preservação eficaz e a acessibilidade contínua dos dados ao longo do tempo, aumentando sua visibilidade e potencial de uso futuro. Esse identificador permite que o recurso seja corretamente citado ao longo do tempo e possibilita que múltiplas cópias estejam disponíveis em diferentes locais (Gabriel Junior *et al.*, 2020).

De forma contrária, o *Uniform Resource Locator* (URL) é um identificador volátil, o que significa que seu uso para descrever uma fonte de informação na *web* pode se tornar problemático ao longo do tempo, devido a uma série de fatores (Gabriel Junior *et al.*, 2020). As razões mais comuns para a inacessibilidade dos recursos do URL incluem: a) Reorganização do sistema de arquivos em um servidor *web*, já que o URL associa diretamente o endereço do servidor ao nome e caminho do recurso na mesma cadeia de caracteres; b) Modificação na infraestrutura da instituição que hospeda a fonte, incluindo mudanças no nome de domínio ou transferência do recurso para um novo URL; c) Atualizações tecnológicas, como em sistemas de banco de dados e servidores de aplicações; d) Obsolescência do recurso ou perda de seu valor comercial, intelectual ou histórico; e) Ausência de garantia para a preservação de dados (Sayão, 2007).

O dARK aborda proativamente os desafios contemporâneos na gestão de dados científicos, documentando todas as operações realizadas em cada PID. Isso permite um rastreamento detalhado de modificações, quem as realizou e quando, mitigando os desafios associados à proveniência dos dados. Esta funcionalidade eleva os sistemas de PID a um novo nível de transparência e confiança (Washington Segundo *et al.*, 2022).

Assim, compreende-se que a implementação do projeto dARK no Brasil representa um passo significativo para situar o país em consonância com as práticas internacionais de excelência em ciência aberta. Essa iniciativa sublinha a capacidade inovadora do dARK na gestão de dados de pesquisa brasileiros, impulsionando uma mudança de paradigma rumo à sustentabilidade, interoperabilidade e acessibilidade dos dados científicos.

Como funciona

O projeto dARK, descrito por Washington Segundo *et al.* (2023), introduz uma arquitetura cliente-servidor inovadora. Nessa estrutura, a rede de consórcio *blockchain* do dARK desempenha o papel de servidor, possibilitando a criação de clientes com tecnologias diversas, que vão desde páginas, *web* em *HTML* até repositórios acadêmicos e plataformas de revistas eletrônicas. Essa estrutura promove uma gestão colaborativa e descentralizada dos Identificadores Persistentes (PID), facilitando a contribuição e o acesso seguro e regulado por diversas instituições à rede.

Santos e Maimone (2023) ressaltam a crescente complexidade na localização de informações específicas no ambiente digital, devido à proliferação de conteúdos. O dARK, ao integrar identificadores persistentes, simplifica significativamente a localização e o acesso a essas informações, promovendo práticas de ciência aberta e a sustentabilidade dos dados de pesquisa.

A rede de consórcio *blockchain*, conforme explicado por Washington Segundo *et al.* (2022), representa uma abordagem parcialmente descentralizada e autorizada, onde a integração dos participantes é mediada por uma entidade reguladora. Tal configuração assegura a replicação segura, confiável e auditável dos dados do PID em todos os nós da rede, garantindo a preservação dos dados mesmo na eventualidade do desaparecimento de uma instituição participante.

Neste cenário, Henning *et al.* (2019) exploram a adoção dos princípios FAIR (Localizáveis, Acessíveis, Interoperáveis, Reutilizáveis), focados no desenvolvimento de metadados que sejam compreensíveis tanto por indivíduos quanto por máquinas, promovendo a interoperabilidade entre diversos conjuntos de dados. Esse enfoque incentiva a criação e implementação de sistemas projetados para a coleta e preservação de dados a longo prazo (Rekunen *et al.*, 2022). A adoção desses princípios pelo dARK reforça seu compromisso com a gestão eficiente e acessível de dados de pesquisa.

Adicionalmente, Sales e Sayão (2024) introduzem o conceito de Objeto Digital FAIR (FDO), que incorpora a ideia de virtualização e encapsulamento de dados e diversos tipos de metadados em uma unidade de conhecimento. Isso aumenta consideravelmente o nível de confiança com o qual o objeto pode ser analisado transversalmente por diferentes disciplinas e modos de compartilhamento.

Na visão de um cientista de dados, um FDO é uma unidade acionável estável que reúne informações suficientes para permitir a interpretação e o processamento confiáveis dos dados nele contidos (De Smedt; Koureas; Wittenburg, 2020).

Os autores Sales e Sayão, (2024) complementam enfatizando que os objetos digitais FAIR têm um papel fundamental na ciência aberta, pesquisa colaborativa e na preservação de recursos digitais. Essencial a este conceito é a ideia de tornar o conhecimento e os dados científicos acessíveis e utilizáveis por um público vasto, fomentando, assim, progressos na pesquisa e no desenvolvimento de soluções inovadoras.

No quadro 10, De Smedt, Koureas e Wittenburg, (2020) apresentam as vantagens dos FDOs referente aos desafios para a ciência de dados.

Quadro 10 - Vantagens dos FDOs

Capacidades interdisciplinares dimensionadas	Os FDOs possibilitam construir um ecossistema integrado envolvendo diversos participantes e múltiplas camadas, por meio do emprego de um protocolo específico (DOIP), que estabelece uma comunicação padronizada para todos os envolvidos neste ambiente global de objetos digitais interconectados. Isso nos permitirá alocar recursos em um novo conjunto de ferramentas de suporte à pesquisa interdisciplinar, de maneira mais eficaz do que os métodos de gestão de dados atualmente em voga.
Dados tornados acessíveis	O descompasso entre a enorme quantidade de dados gerados e nossa habilidade para aproveitá-los adequadamente se deve a vários fatores. Entre eles, a escassez de habilidades especializadas e a pouca valorização do trabalho interdisciplinar são aspectos que os Objetos de Dados Funcionais (FDOs) não conseguem resolver diretamente. No entanto, um dos problemas - a ausência de informações contextuais - pode ser atenuado pelo conceito de FDO, que facilita a associação de dados com seu contexto de forma consistente e duradoura.
Interpretando evidências científicas em um contexto confiável	Nos Objetos de Dados Funcionais (FDOs), é possível vincular dados contextuais e identificadores únicos aos objetos digitais em vários momentos do seu ciclo de vida. As informações relacionadas à privacidade também podem e devem ser anexadas a cada objeto digital de maneira segura e inalterável. De fato, todos os metadados permanecem constantemente associados aos dados, garantindo que os pesquisadores tenham acesso contínuo à origem e a outras informações vitais para determinar sua adequação ao propósito — fortalecendo, assim, a confiança.
Domínio do raciocínio	A complexidade e a constante evolução do conhecimento em todas as áreas científicas nos impulsionam em direção

	ao processamento automatizado na nossa procura por percepções orientadas por dados. Os Objetos de Dados Funcionais (FDOs), como unidades acionáveis, documentam e estabelecem ligações complexas ao longo do tempo. Desse modo, eles atuam como peças fundamentais na construção das estruturas de conhecimento que compõem a nossa memória científica digital em desenvolvimento.
Avançando dados para unidades de conhecimento acionáveis	Conforme os Objetos de Dados Funcionais (FDOs) se movimentam pelo ciberespaço e atravessam o tempo, o seu mecanismo de encapsulamento assegura que, mesmo após décadas e diante das transformações tecnológicas e da alternância de participantes, os dados e seus contextos se mantenham acessíveis como entidades íntegras. Eles não sofrerão perdas informativas, mas poderão enriquecer seu contexto de reutilização ao longo dos anos.
Proliferação de ferramentas e decisões fundamentais	O conceito de Objetos de Dados Funcionais (FDO) proporciona uma abstração que oculta complexidades tecnológicas dos pesquisadores, prevenindo o aprisionamento por tecnologias específicas e fomentando a inovação tecnológica sem comprometer o contínuo desenvolvimento do Domínio do Conhecimento Digital. Os serviços de registro e armazenamento, virtualizados, podem ser integrados a um sistema federado por meio de protocolos de Objetos Digitais unificados, facilitando uma interação clara e direta com os usuários na interface dos serviços.

Fonte: Elaborado pelo autor com base em De Smedt; Koureas; Wittenburg, 2020.

Para De Smedt; Koureas; Wittenburg, (2020) a implementação de FDOs estabilizados por identificadores persistentes não alterará o volume predominante de informação não estruturada na *web*, mas proporcionará uma alternativa mais segura e perene para o armazenamento de conjuntos de dados em um domínio registrado, necessitando de preservação a longo prazo. Essa estabilidade contribuirá para elevar a confiança de pesquisadores e demais interessados que dedicam esforços significativos na conservação do conhecimento científico.

Em síntese, o projeto dARK representa uma iniciativa abrangente que combina tecnologia de ponta, políticas de preservação rigorosas e uma adesão estrita aos princípios FAIR para fomentar o acesso aberto e a reutilização de dados de pesquisa.

Esta abordagem integrada não apenas facilita a gestão eficaz de dados científicos, mas também promove a colaboração e a inovação em pesquisa globalmente, estabelecendo um novo paradigma para a ciência aberta e a gestão de dados de pesquisa.

Suas parcerias

O projeto dARK, apoiado pela tecnologia *blockchain*, facilita a governança compartilhada e a redução de custos entre redes nacionais e regionais de pesquisa e educação. Esta colaboração, conforme descrito por Washington Segundo *et al.* (2023), envolve institutos e universidades de pesquisa que contribuem com recursos computacionais limitados, demonstrando a eficácia da descentralização na gestão de dados científicos.

Essas colaborações são fundamentais para ampliar o alcance do projeto, permitindo a preservação e a disponibilização de uma gama mais ampla de dados de pesquisa. Santos e Maimone (2023) destacam a importância dos identificadores persistentes, tanto para a comunidade científica quanto para as organizações que os gerenciam, sublinhando a necessidade de uma infraestrutura confiável para a gestão de dados.

Além disso, a parceria com instituições acadêmicas e de pesquisa reforça a criação de uma rede de conhecimento robusta, com foco nas melhores práticas de gestão e preservação de dados científicos. Washington Segundo *et al.* (2022) apresentam o dARK como um projeto aberto e comunitário, promovido pelo Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT-Brasil) e pela Rede *LA Referencia*¹² / *RedCLARA*¹³, com apoio da *SCOSS*¹⁴. O objetivo é fornecer um modelo de serviço de resolução e uma fábrica de identificadores persistentes abertos e descentralizados para o ecossistema de ciência aberta.

As parcerias estabelecidas pelo dARK são cruciais para o sucesso do projeto, ressaltando a importância de uma abordagem colaborativa para superar os desafios associados à gestão de dados de pesquisa.

O modelo que denominamos dARK, para a atribuição de identificadores ARK descentralizados, implementa uma abordagem que permite a várias entidades colaborarem na gestão de um sistema de identificadores persistentes ARK próprio. Este método se apoia em uma infraestrutura descentralizada e compartilhada, baseada em nós que fazem parte de uma *blockchain* de consórcio. Com isso, a posse, armazenamento ou gestão dos dados não fica restrita a uma única instituição, mas

¹² <https://www.lareferencia.info/pt/> acessado em 02/02/2024.

¹³ <https://www.redclara.net/index.php/pt/> acessado em 02/02/2024.

¹⁴ <https://scoss.org/> acessado em 02/02/2024.

sim distribuída entre todos os participantes da rede (Washington Segundo *et al.*, 2023).

A integração de indicadores persistentes nos esforços colaborativos do projeto destaca o compromisso com a criação de um ecossistema de dados de pesquisa coeso, não só facilitando o acesso e a reutilização de dados, mas também garantindo sua preservação a longo prazo.

Sua importância

O projeto dARK surge como uma resposta inovadora aos desafios enfrentados na preservação digital e no acesso aberto a dados de pesquisa científica no Brasil. Segundo Santos e Maimone (2023), a eficiência na recuperação de conteúdos científicos é crucial, e o dARK contribui significativamente para otimizar este processo, garantindo que informações relevantes sejam acessíveis de maneira assertiva para os pesquisadores.

Washington Segundo *et al.* (2022) destacam a importância do dARK na oferta de uma solução descentralizada e inclusiva para a gestão de identificadores persistentes, abrangendo uma ampla gama de objetos digitais e físicos. Esta abordagem não apenas facilita a preservação e acessibilidade dos dados de pesquisa, mas também promove a replicabilidade dos estudos, a transparência científica e, conseqüentemente, o avanço do conhecimento.

A adoção do identificador persistente ARK, conforme discutido por Gabriel Junior *et al.* (2020), é particularmente adequada para a gestão de recursos de longa duração, tanto em arquivos quanto em repositórios digitais. Além disso, o compromisso do dARK com o acesso aberto aos dados de pesquisa representa um passo significativo na democratização do conhecimento científico, permitindo que pesquisadores, estudantes e o público tenham acesso a informações valiosas que podem impulsionar novas descobertas e inovações.

A relevância do dARK vai além, estendendo-se à promoção da ciência aberta e à melhoria da acessibilidade e reutilização de dados de pesquisa. Santos e Maimone (2023) salientam que a estrutura dos identificadores persistentes é projetada para favorecer a interoperabilidade e a conexão entre diferentes fontes de informação, permitindo a construção de perfis detalhados de autores e instituições.

A implementação de indicadores persistentes reforça a importância do projeto, assegurando que os dados sejam não apenas acessíveis, mas também rastreáveis e reconhecidos como contribuições científicas válidas.

Em resumo, o projeto dARK é uma iniciativa essencial para o avanço da ciência aberta no Brasil, enfrentando desafios críticos na gestão de dados de pesquisa. Suas ações promovem a transparência, acessibilidade dos dados científicos, colaboração, inovação e avanço do conhecimento em várias áreas da pesquisa, estabelecendo um marco importante na ciência brasileira e contribuindo para o desenvolvimento sustentável da pesquisa científica no país.

A abordagem

O projeto dARK se destaca no cenário científico brasileiro por sua abordagem inovadora focada na sustentabilidade, interoperabilidade e acessibilidade dos dados de pesquisa. Implementando uma versão descentralizada dos identificadores ARK, o dARK adota uma metodologia colaborativa que permite a múltiplas instituições gerenciar conjuntamente seus sistemas de identificadores persistentes. Esta gestão é realizada por meio de uma infraestrutura comum baseada em *blockchain*, distribuindo os dados equitativamente entre todos os participantes da rede, conforme elucidado por Washington Segundo *et al.* (2023).

A utilização de tecnologias avançadas e padrões abertos pelo projeto garante a preservação eficaz dos dados, além de facilitar o acesso e a reutilização por pesquisadores de diferentes áreas. Santos e Maimone (2023) ressaltam as características dos identificadores persistentes, fundamentais para a análise e aplicabilidade na organização e no acesso à informação científica, evidenciando a importância de uma estrutura de dados bem definida e acessível.

Ressaltando os identificadores persistentes, o quadro 11, fundamentado nas observações de Santos e Maimone (2023), expõe as propriedades desses identificadores para estudo e uso na organização e acesso à informação científica, fornecendo detalhes sobre diversos esquemas de identificadores persistentes.

Quadro 11 - Características dos identificadores persistentes para análise e aplicabilidade na organização e acesso à informação científica

Característica	Descrição
Univocidade	Identificação apenas com uma única <i>Uniform Resource Identifier</i> (URI), que se trata de um tipo de protocolo de transferência de hipertexto com o qual o usuário consegue colar o endereço na barra de seu navegador de internet para então acessar o recurso que o identificador representa. De forma direta, é possível afirmar que a <i>Uniform Resource Locator</i> (URL) é o endereço de um recurso online e a URI identifica esse recurso. Como a característica aqui mencionada é a univocidade aplicada aos identificadores persistentes, a URI é a característica a ser analisada em detrimento da URL.
<i>Web-resolvable</i>	Direcionamento para o local exato em que as informações sobre o recurso são acessadas.
Atribuição livre	Atribuição sem nenhum custo às entidades que o aderem.
Acesso e uso abertos	Acesso realizado livremente por qualquer pessoa de qualquer local.
Interface amigável	Possibilitar a fácil identificação do pesquisador por meio de nome, área de pesquisa e/instituição afiliada e do conjunto de itens de pesquisa atrelados ao perfil.
Rede acadêmica	Oferecer ferramentas que favorecem o relacionamento e troca de experiências ou estimula o contato entre pares.
Inserção de outros identificadores	Possibilitar que o pesquisador insira em seu perfil o ID de outro identificador.
Integração de dados entre identificadores	Possibilitar o intercâmbio de registros entre identificadores, otimizando o tempo do pesquisador no preenchimento e manutenção das plataformas.

Fonte: Santos; Maimone, 2023, p. 523 e 524.

Além de sua infraestrutura tecnológica, o dARK é comprometido em fomentar a conscientização sobre a gestão de dados de pesquisa. Esta iniciativa visa atender a requisitos críticos como a tolerância a falhas, a preservação de longo prazo, a rastreabilidade da proveniência e a interoperabilidade dos sistemas de Identificadores Persistentes (PID), conforme apontado por Washington Segundo *et al.* (2022).

Para Sayão (2007) há um leque limitado de abordagens para a implementação de identificadores persistentes, cuja escolha depende das condições técnicas, administrativas e políticas de cada organização, em particular, da sua perspectiva futura e do nível de interoperabilidade almejado com outros sistemas.

As abordagens incluem: a) Redirecionamento - é uma estratégia mínima, posto que utiliza os recursos padronizados do servidor *web* para redirecionar as solicitações para a posição corrente do recurso; b) Instalação de um resolvedor apoiado em banco de dados - pressupõe um *software* servidor de links, rodando sobre um banco de dados e tendo como finalidade, mapear a localização corrente do recurso, ou seja, o URL corrente; e c) Contratação de sistema de identificação persistente, oferecido por

outra organização - existem vários sistemas de identificação persistente projetados para uso na Internet, baseados em padrões abertos, com objetivos e enfoques distintos (Sayão, 2007, p. 68).

A estratégia diversificada do dARK abarca tanto o desenvolvimento técnico de infraestruturas quanto a criação de políticas e diretrizes específicas, além da capacitação de pesquisadores e gestores de dados. Esse esforço reflete um comprometimento com a adaptabilidade e a capacidade de expansão das soluções oferecidas, as quais são cuidadosamente planejadas para atender às variadas exigências da comunidade científica no Brasil e superar os desafios atuais na gestão de dados.

Assim, o dARK estabelece um novo paradigma na ciência brasileira, contribuindo significativamente para o avanço da pesquisa científica e tecnológica no país.

O impacto planejado

O projeto dARK estabelece metas ambiciosas com o intuito de revolucionar a gestão de dados de pesquisa no Brasil, visando muito além da simples preservação de dados. Para Washington Segundo *et al.* (2023), o projeto almeja implementar uma solução de Identificador Persistente (PID) acessível globalmente, especialmente focada no Sul Global, onde recursos para atribuir PIDs são escassos. Esta iniciativa é um passo crucial para democratizar o acesso à infraestrutura de pesquisa e promover a inclusão científica.

Como o projeto dARK está em andamento, apresentamos a proposta de trabalho no quadro 12, iniciada em 2022.

Quadro 12 - Proposta de trabalho do Projeto dARK

Roteiro 2022 - 2023	
Fase 1	Desenvolvimento do documento conceitual: Desenvolver um documento conceitual abrangente descrevendo os objetivos, o escopo e os detalhes técnicos do projeto dARK.
Fase 2	Desenvolvimento de Prova de Conceito (PoC): Validar a viabilidade técnica do projeto dARK implementando uma versão simplificada de contratos inteligentes e demonstrando os principais recursos em uma rede de pequena escala.
Fase 3	Integração com sistemas externos: Permitir referência cruzada e resolução de identificadores entre o dARK e sistemas externos para garantir a interoperabilidade.
Fase 4	Implementação Piloto no Brasil: Instalar e configurar a infraestrutura instalando nós <i>Hyperledger Besu</i> compatíveis com a configuração da rede alvo para uma implementação piloto no Brasil.
Roteiro 2024 - 2025	
Fase 5	Teste, melhoria e lançamento de código aberto: Conduzir testes abrangentes, incluindo testes unitários, de integração e de estresse, para melhorar a estabilidade do sistema e lançar uma primeira versão de código aberto.
Fase 6	Documentação e Treinamento: Criar documentação abrangente, incluindo guias de instalação e APIs de contratos inteligentes, e conduza workshops de treinamento para educar as partes interessadas sobre o uso e os benefícios do identificador dARK.
Fase 7	Governança e Atualizações: Estabelecer um modelo de governança para tomada de decisões, atualizações de consenso e resolução de disputas, e planejar melhorias regulares com base no feedback e na evolução dos requisitos.
Fase 8	Construção e Adoção da Comunidade: Promover uma comunidade em torno do projeto dARK, envolvendo-se com pesquisadores, instituições e desenvolvedores por meio de eventos comunitários, fóruns e canais de comunicação para colaboração e compartilhamento de conhecimento.

Fonte: Washington Segundo *et al.*, 2023.

A implementação bem-sucedida das estratégias do dARK promete não somente ampliar a visibilidade e o impacto da pesquisa brasileira, mas também contribuir significativamente para o avanço do conhecimento científico e tecnológico global, reforçando a integridade e a confiabilidade da pesquisa científica.

Para Washington Segundo *et al.* (2023) o dARK se diferencia de outros identificadores persistentes devido à sua descentralização, custo reduzido, governança colaborativa e natureza aberta, desempenhando um papel crucial no progresso da ciência aberta e na promoção do acesso global à informação científica.

Em suma, o projeto dARK representa uma iniciativa pioneira com o potencial de transformar a paisagem da ciência brasileira, promovendo uma transição significativa em direção a práticas de ciência aberta. Ao facilitar o acesso e a gestão eficiente de dados de pesquisa, o dARK visa estabelecer um novo paradigma para a colaboração científica e a inovação, assegurando que a pesquisa brasileira seja reconhecida e valorizada em uma escala global.

Os resultados

Washington Segundo *et al.* (2022) caracterizam o dARK como uma iniciativa em progresso, destacando o desenvolvimento contínuo e os planos que incluem aprimoramentos nas funcionalidades de pesquisa e na recuperação avançada de registros por meio dos metadados dos Identificadores Persistentes (PID).

Os autores apresentam a primeira versão de prova de conceito do ARK, um serviço de Identificador Persistente Descentralizado (PID) criado para ser um recurso de acesso aberto e direcionado ao benefício público no contexto do ecossistema de Ciência Aberta desde o início. Utilizando a *blockchain* como sua base tecnológica, esse serviço permite uma governança compartilhada e a diminuição de custos para as redes de pesquisa e educação nacionais e regionais. Este modelo de colaboração também se aplica a instituições e universidades de pesquisa individuais, que contribuem com recursos computacionais limitados para sustentar o sistema (Washington Segundo *et al.*, 2023).

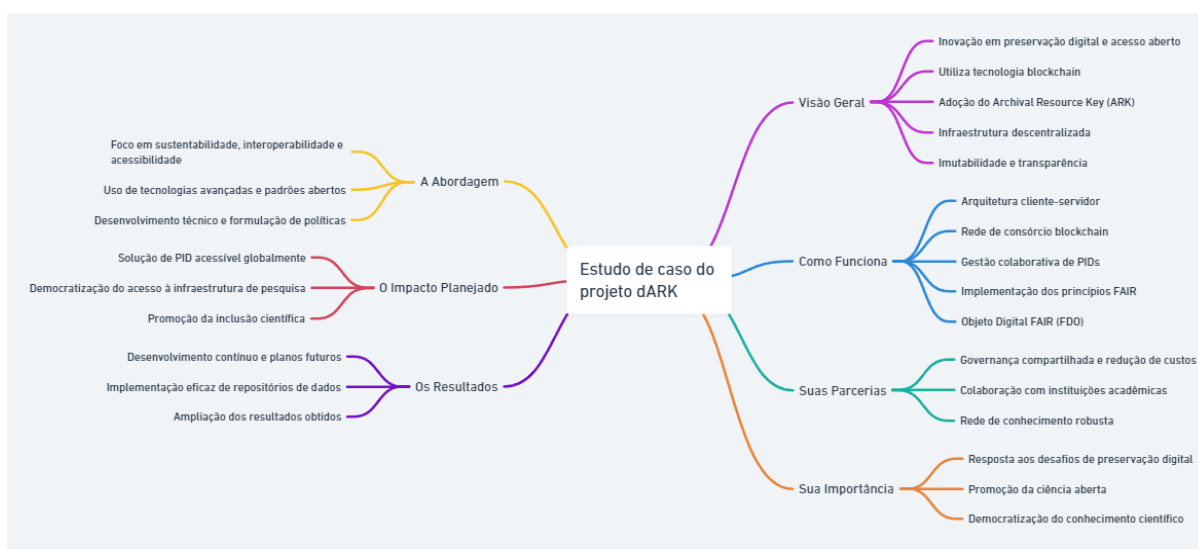
Sales e Sayão (2016) ressaltam a importância da implementação eficaz de repositórios de dados de pesquisa para estimular a colaboração científica e elevar a qualidade da pesquisa. A integração de indicadores persistentes a essas métricas é crucial para avaliar o impacto do projeto, garantindo que os dados sejam não apenas acessíveis, mas também rastreáveis e reconhecidos como contribuições científicas significativas.

À medida que o projeto avança, espera-se uma ampliação dos resultados obtidos, consolidando a posição do dARK como um componente essencial na infraestrutura de pesquisa científica brasileira. Os desenvolvimentos atuais, ainda em fase de protótipo, apontam para o lançamento da primeira versão de produção do consórcio dARK até o final de 2023. Este lançamento, inicialmente focado no Brasil, tem potencial para expandir-se para a América Latina, integrando outros países membros da LA Referencia, conforme indicado por Washington Segundo *et al.* (2022).

Em resumo, os indicadores persistentes são elementos-chave na infraestrutura da gestão de informações digitais, promovendo não apenas a acessibilidade e a recuperação eficiente de dados, mas também a integridade e a preservação do conhecimento no ambiente digital. A compreensão e a aplicação adequadas desses mecanismos são imperativas para superar os desafios impostos pela natureza fluida da internet e para sustentar o avanço da comunicação e da pesquisa acadêmica.

Com uma abordagem pioneira e colaborativa, o projeto está lançando as bases para uma infraestrutura de pesquisa mais aberta, acessível e sustentável. Este avanço representa um marco importante para a ciência aberta e para a elevação da qualidade da pesquisa científica no país, estabelecendo um novo paradigma na gestão de dados de pesquisa. Na figura 28 apresentamos um esquema com os tópicos apresentados no estudo de caso do projeto dARK.

Figura 28 - Estudo de caso do projeto dARK



Fonte: Elaborado pelo autor.

Após a apresentação da figura 28, passamos para o capítulo 10, que contém as considerações finais. Nesse capítulo, sintetizamos os achados da pesquisa, refletimos sobre as implicações teóricas e práticas do estudo e apontamos direções futuras para a pesquisa nessa área promissora.

10 CONSIDERAÇÕES FINAIS

O presente estudo investiga a aplicabilidade e as vantagens da tecnologia *blockchain* na preservação e gestão de dados em ambientes de pesquisa científica, com foco especialmente, nos projetos Archangel e dARK. Por meio de uma revisão sistemática da literatura e análise detalhada de casos específicos, constatou-se que a *blockchain* oferece soluções inovadoras para enfrentar desafios comuns na Ciência da Informação, como a proveniência, rastreabilidade, confiabilidade e autenticidade dos dados.

Esta pesquisa marca um avanço significativo na integração da tecnologia *blockchain* ao campo da Ciência da Informação, especialmente na gestão de dados de pesquisa. Evoluindo além de sua função inicial como base para criptomoedas, a *blockchain* emerge como uma ferramenta revolucionária para questões de segurança, confiabilidade e rastreabilidade em contextos científicos.

Os resultados obtidos e os modelos semânticos propostos delineiam um caminho promissor para pesquisas futuras e aplicações práticas. A adoção da *blockchain* na gestão de dados de pesquisa demonstra não apenas viabilidade teórica, mas também prática, com potencial para redefinir métodos de preservação, validação e compartilhamento de dados científicos. Destaca-se a necessidade de explorar e adaptar essas tecnologias inovadoras para superar os desafios emergentes na gestão de dados científicos, garantindo sua integridade e acessibilidade para futuras gerações.

Nossa análise evidencia a crescente adoção da *blockchain* na gestão de dados de pesquisa, refletindo um campo em expansão. Identifica-se um interesse crescente por essa tecnologia devido ao seu potencial disruptivo e capacidade de resolver problemas essenciais como proveniência, segurança, autenticidade e transparência. A *blockchain* facilita a rastreabilidade e a preservação da integridade dos dados, promovendo a criação de ecossistemas descentralizados que fomentam a colaboração e acessibilidade na pesquisa.

A integração da *blockchain* na gestão de dados científicos indica uma transformação significativa nas práticas de gerenciamento e disseminação de informações acadêmicas, com diversas implementações inovadoras. A sua utilização em bibliotecas e centros de informação leva a uma gestão mais eficaz dos acervos, impulsionando avanços na comunicação acadêmica. Sistemas de carimbo de

data/hora e revisões por pares baseadas em *blockchain* reforçam a segurança e integridade dos dados de pesquisa, contribuindo para a prevenção de fraudes e manipulações.

Os projetos Archangel e dARK ilustram a aplicabilidade e o potencial da *blockchain*, oferecendo percepções valiosas para futuras implementações. Esses projetos destacam-se pela adaptação da *blockchain* para atender às necessidades específicas de preservação digital e gestão de dados de pesquisa.

A implementação dos modelos semânticos, integrando conceitos de *blockchain* aos princípios de gestão de dados de pesquisa, representa um avanço significativo. Esses modelos não apenas reforçam a integridade dos dados, mas também promovem maior transparência e acessibilidade, cruciais para o avanço da ciência aberta. Contudo, há desafios a serem superados, como escalabilidade, custos e complexidade técnica, além da necessidade de compreensão das normativas éticas e regulatórias.

Este estudo sublinha a *blockchain* como um recurso promissor para aprimorar a gestão de dados de pesquisa, ressaltando a importância de uma avaliação cuidadosa de suas implicações e desafios. As tendências atuais e as implementações bem-sucedidas da *blockchain* fornecem uma base sólida para futuras investigações e debates, permitindo escolhas informadas e estratégicas na comunidade acadêmica, otimizando o uso da *blockchain* em pesquisa, colaboração e desenvolvimento científico.

A *blockchain* se destaca por sua descentralização, imutabilidade e robustez em termos de segurança, diferenciando-se dos bancos de dados convencionais. Empregando criptografia avançada, previne modificações indevidas e eleva a confiabilidade dos dados armazenados. A introdução desta tecnologia no domínio da Ciência da Informação desvela novas possibilidades para o gerenciamento de informações, especialmente em relação à preservação digital de dados de pesquisa. Ela oferece novos métodos para superar desafios antigos relacionados à segurança, autenticidade e rastreabilidade dos dados, que são fundamentais para a integridade da pesquisa científica. Assegurando a imutabilidade e a segurança dos dados, a *blockchain* estabelece uma base sólida para a gestão de informações científicas.

Este trabalho evidencia a *blockchain* como uma ferramenta essencial para a Ciência da Informação, pavimentando o caminho para futuras investigações sobre sua integração com outras inovações tecnológicas. Apesar dos desafios de

desenvolvimento, o potencial da *blockchain* para transformar a gestão e preservação de dados digitais de pesquisa é notável, tornando-se um componente chave na inovação da gestão de informações científicas.

A integração da *blockchain* na Ciência da Informação sugere uma mudança paradigmática na maneira como os dados de pesquisa são geridos e conservados. Esta tecnologia não só promove uma gestão de dados mais segura e eficiente, mas também fomenta a transparência e a colaboração entre pesquisadores, ampliando o acesso à informação científica. A sua adoção em contextos acadêmicos e de pesquisa reflete um compromisso com a inovação e a melhoria contínua dos processos de gestão de dados.

Portanto, à medida que avançamos, é crucial que a comunidade científica continue investigando e desenvolvendo esta tecnologia, superando os desafios existentes e explorando todo o seu potencial para revolucionar a gestão de dados de pesquisa. A *blockchain* promete não apenas maior segurança e integridade dos dados, mas também inaugura uma nova era de colaboração e acesso aberto à informação científica, contribuindo significativamente para o avanço do conhecimento e da inovação.

A proposta de um modelo inovador, aplicando a tecnologia *blockchain* na gestão de dados de pesquisa, simboliza uma mudança significativa, incorporando avanços tecnológicos à Ciência da Informação. Este modelo enfrenta desafios contemporâneos na gestão de dados, ressaltando a segurança, interoperabilidade e acessibilidade como aspectos cruciais.

Esta abordagem multidimensional combina a robustez da tecnologia *blockchain* com a flexibilidade da modelagem semântica, sugerindo um caminho promissor na superação das limitações atuais na gestão de dados de pesquisa. A sinergia entre a segurança e a fidelidade dos dados oferecida pela *blockchain*, juntamente com a otimização e o reaproveitamento possibilitados pela modelagem semântica, cria um novo paradigma para a colaboração no campo científico.

O futuro da Ciência da Informação, portanto, demanda não apenas a adaptação às evoluções tecnológicas, mas também a incorporação ativa dessas inovações para enfrentar desafios atuais e futuros na gestão de dados de pesquisa. A colaboração contínua entre tecnólogos, pesquisadores e instituições acadêmicas é vital para explorar plenamente o potencial dessa sinergia inovadora.

Por fim, este estudo enfatiza a *blockchain* como um recurso valioso para aprimorar a gestão de dados de pesquisa, instigando a continuação do diálogo e da pesquisa sobre como essa tecnologia pode ser explorada e integrada nas práticas de gestão de informação científica. Por meio da colaboração interdisciplinar e do desenvolvimento contínuo de soluções baseadas em *blockchain*, podemos aspirar a uma Ciência da Informação não só mais segura e confiável, mas também mais aberta e acessível a todos, abrindo novos caminhos para enfrentar desafios antigos e impulsionar a inovação na gestão de dados de pesquisa.

REFERÊNCIAS

ACADEMY BIT2ME. **O que é um ledger distribuído?** 2019. Disponível em: <https://academy.bit2me.com/pt/qual-%C3%A9-o-livro-distribu%C3%ADdo-do-livro/> Acesso em: 27 out. 2023.

AKORAMURTHY B.; DHIVYA, K., VENNIRA SELVI G.; PRASAD, M. **The Data-Driven Blockchain Ecosystem**. (1ª ed.). In: *A Blockchain Approach to Improving Digital Linked Management Information Systems (MIS)*. CRC Press eBooks, p. 61–82, 15 dez. 2022. Disponível em: <https://doi.org/10.1201/9781003269281>. Acesso em: 20 fev. 2023.

ALTAMIRA. **Blockchain vs Cloud Computing - which is better?** Disponível em: <www.altamira.ai>. Acesso em: 4 jan. 2024.

ANDROULAKI, E.; BARGER, A.; BORTNIKOV, V.; CACHIN, C.; CHRISTIDIS, K.; DE CARO, A.; ENYEART, D.; FERRIS, C.; LAVENTMAN, G.; MANEVICH, Y.; MURALIDHARAN, S.; MURTHY, C.; NGUYEN, B.; SETHI, M.; SINGH, G.; SMITH, K.; SORNIOTTI, A.; STATHAKOPOULOU, C.; VUKOLIĆ, M.; COCCO, S. W.; YELICK, J. Hyperledger fabric: um sistema operacional distribuído para blockchains permissionados. In: EUROSYS CONFERENCE, 13, 23-26 abr. 2018, Porto. **Anais da Décima Terceira Conferência EuroSys (EuroSys '18)**. Nova York: Association for Computing Machinery, 2018. Artigo 30, p. 1–15. Disponível em: <https://doi.org/10.1145/3190508.3190538>. Acesso em: 27 nov. 2023.

ARCHANGEL. **Securing our National Archives for Future Generations**. Disponível em: <https://www.archangel.ac.uk/>. Acesso em: 05 jan. 2024.

BECK, R.; AVITAL, M.; ROSSI, M.; THATCHER, J. B. Blockchain technology in business and information systems research. **Business & Information Systems Engineering: The International Journal of WIRTSCHAFTSINFORMATIK**, v. 59, n. 6, p. 381-384, dez. 2017. Disponível em: <https://doi.org/10.1007/s12599-017-0505-1>. Acesso em: 18 nov. 2023.

BELL, M.; GREEN, A.; SHERIDAN, J.; COLLOMOSSE, J.; COOPER, D.; BUI, T.; THEREAUX, O.; HIGGINS, J. Underscoring archival authenticity with blockchain technology. **Insights: The UKSG Journal**, v. 32, 2019. Disponível em: <https://doi.org/10.1629/uksg.470>. Acesso em: 15 set. 2023.

BELLEGGARDA, J. **Exploiting latent semantic information in statistical language modeling**. Proceedings of the IEEE, Nova York, v. 88, p. 1279-1296, 2000, doi: 10.1109/5.880084.

BHUTTA, M. N. M. *et al.* A Survey on *Blockchain* Technology: Evolution, Architecture and Security. **IEEE Access**, New York, v. 9, p. 61048-61073, 2021.

BRALIĆ, V.; STANČIĆ, H.; STENGÅRD, M. A *blockchain* approach to digital archiving: digital signature certification chain preservation. **Records Management Journal**, Bingley, v. 30, n. 3, p. 345-362, 2020.

CAPURRO, R.; HJORLAND, B. O conceito de informação. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 12, n. 1, p. 148-207, jan./abr. 2007. Disponível em: <https://periodicos.ufmg.br/index.php/pci/article/view/22360>. Acesso em: 11 ago. 2020.

CASEY, M. J.; VIGNA, P. **The truth machine**: the *blockchain* and the future of everything. Nova York: St. Martin's Press, 2018.

CHEN, G.; XU, B.; LU, M.; CHEN, N. Exploring blockchain technology and its potential applications for education. **Smart Learning Environments**, v. 5, p. 1-10, 2018. Disponível em: <https://doi.org/10.1186/s40561-017-0050-x>. Acesso em: 05 jul. 2023.

DE SMEDT, K.; KOUREAS, D.; WITTENBURG, P. **FAIR Digital Objects for Science**: From Data Pieces to Actionable Knowledge Units. *Publications*, v. 8, n. 2, p. 21, 11 abr. 2020. Disponível em: <https://www.mdpi.com/2304-6775/8/2/21>. Acesso em: 08 ago. 2023.

DELOITTE INSIGHTS. **Tech Trends 2019**. [S.l.]: Deloitte, 2019. Disponível em: <https://www2.deloitte.com/us/en/insights/focus/tech-trends/2019.html>. Acesso em: 15 nov 2023.

DENG, Z. *et al.* *Blockchain-based trusted electronic records preservation in cloud storage*. **Computers, Materials & Continua**, v. 58, n. 1, p. 135-151, Nanjing, 2019.

DENNEULIN, Y. *et al.* Merging file systems and data bases to fit the grid. *In*: HAMEURLAIN, A.; MORVAN, F.; TJOA, A. M. (Eds.). **Data management in grid and peer-to-peer systems**. Globe 2010. Berlin; Heidelberg: Springer, 2010. p. 13-25. *Lecture Notes in Computer Science*, v. 6265.

DENYER, D.; TRANFIELD, D. Using qualitative research synthesis to build an actionable knowledge base. **Management Decision**, v. 44, n. 2, p. 213-227, Bradford: Emerald Group Publishing Limited. 2006.

EPPLER, M. J. A. Comparison between Concept Maps, Mind Maps, Conceptual Diagrams, and Visual Metaphors as Complementary Tools for Knowledge Construction and Sharing. **Information Visualization**, v. 5, n. 3, p. 202-210, Londres, 22 jun. 2006. Disponível em: <https://doi.org/10.1057/palgrave.ivs.9500131>. Acesso em: 14 jun. 2023.

FENG, Q. *et al.* A survey on privacy protection in *blockchain* system. **Journal of Network and Computer Applications**, v. 126, p. 45-58, Amsterdã, jan. 2019.

FUNDAÇÃO BIBLIOTECA NACIONAL (Brasil). **Política de preservação digital da Biblioteca Nacional**. Rio de Janeiro: FBN, 2020. 36 p. ISBN: 978-65-5940-008-9. Disponível em: <https://bit.ly/3tx29FS>. Acesso em: 13 jan. 2024.

GABRIEL JUNIOR, R. F.; AZAMBUJA, L. A. B.; MOURA, A. M. M.; CAREGNATO, S. E.; PASSOS, P. C. S. J.; PAVÃO, C. G.; BORGES, E. N.; ROCHA, R. P. da; VANZ, S. A. de S. **Acesso aberto a dados de pesquisa no Brasil**: identificadores persistentes para dados de pesquisa. 2020. Disponível em: <http://hdl.handle.net/20.500.11959/1258>. Acesso em: 17 abr. 2023.

GANGADEVI, K.; RENUGA DEVI, R. A survey on data integrity verification schemes using blockchain technology in cloud computing environment. **IOP Conference Series: Materials Science and Engineering**, Mysuru, v. 1110, n. 1, p. 1-16, 1 mar. 2021.

GIL, A. C. **Métodos e técnicas de pesquisa social**. 5. ed. São Paulo: Atlas, 1999.

_____. **Como elaborar métodos de pesquisa**. 4. ed. São Paulo: Atlas, 2002.

GIPP, B. *et al.* CryptSubmit: introducing securely timestamped manuscript submission and peer review feedback using the *blockchain*. In: IEEE/ACM JOINT CONFERENCE ON DIGITAL LIBRARIES (JCDL), 19-23 jun. 2017, Toronto. **Proceedings of the IEEE/ACM Joint Conference on Digital Libraries (JCDL)**. Toronto, ON, Canadá, 2017, p. 1-4, doi: 10.1109/JCDL.2017.7991588.

GOLOSOVA, J.; ROMĀNOVS, A. **The Advantages and Disadvantages of the Blockchain Technology**. In: 2018 IEEE 6th Workshop on Advances in Information, Electronic and Electrical Engineering (AIEEE), Vilnius, 2018, p. 1-6, doi: 10.1109/AIEEE.2018.8592253.

GOMES, A. N.; COUTINHO, E. F. Um Estudo Inicial sobre a Importância de Simular Contratos Inteligentes em Blockchain. *In: Workshop em Modelagem e simulação de Sistemas Intensivos em Software (MSSIS)*, 4, 2022, Uberlândia/MG. **Anais [...]**. Porto Alegre: Sociedade Brasileira de Computação, 2022. p. 1-10. DOI: <https://doi.org/10.5753/mssis.2022.225407>.

GONÇALVES, N. C. da S.; RODRIGUES, F. de A. Arquivologia e *blockchain*: discussão teórica sobre oportunidades e barreiras. **Ciência da Informação em Revista**, Maceió, v. 7, n. 3, p. 21-38, 31 dez. 2020.

GREEN, A. *et al.* Using blockchain to engender trust in public digital archives. In: INTERNATIONAL CONFERENCE ON DIGITAL PRESERVATION (iPres 2018), 28 set. 2018, Boston. **Proceedings of the 15th International Conference on Digital Preservation (iPRES 2018)**. 2018. Disponível em: <https://puma.ub.uni-stuttgart.de/bibtex/65ded8bbb68d2bb1e307364bb794665f>. Acesso em: 27 nov. 2023.

HENNING, P. C.; RIBEIRO, C. J. S.; DA SILVA SANTOS, L. O. B.; DOS SANTOS, P. X. GO FAIR e os princípios FAIR: o que representam para a expansão dos dados de pesquisa no âmbito da Ciência Aberta. **Em Questão**, Porto Alegre, v. 25, n. 2, p. 389–412, 2019. DOI: 10.19132/1808-5245252.389-412. Disponível em: <https://seer.ufrgs.br/index.php/EmQuestao/article/view/84753>. Acesso em: 8 ago. 2024.

HWANG, H. C.; SHON, J. G.; PARK, J. S. Design of an enhanced web archiving system for preserving content integrity with *blockchain*. **Electronics**, Korea, v. 9, n. 8, p. 1-13, 5 ago. 2020.

IBM. **Blockchain**: Como funciona? 2017. Disponível em: <https://www.ibm.com/ibm/br/pt/blockchain/what-is-blockchain.html>. Acesso em: 27 nov. 2023.

IHODL. **Guia de Blockchain para principiantes**. 2017. Disponível em: <https://pt-br.ihodl.com/tutorials/2017-06-29/guia-de-Blockchain-para-principiantes/>

KIM, M. et al. Design of secure protocol for cloud-assisted electronic health record system using blockchain. **Sensors**, v. 20, n. 10, p. 2913, 2020. Disponível em: <https://www.mdpi.com/1424-8220/20/10/2913>. Acesso em: 21 mai. 2020.

KRIEGEL, H.-P. et al. Towards archaeo-informatics: scientific data management for archaeobiology. In: GERTZ, M.; LUDÄSCHER, B. (Eds.). **Scientific and statistical database management**. SSDBM 2010. Berlin; Heidelberg; Springer, 2010. p. 169-177. Lecture Notes in Computer Science, v. 6187.

LANDAUER, T.; FOLTZ, P.; LAHAM, D. **An introduction to latent semantic analysis**. Discourse Processes, Londres, v. 25, p. 259-284, 1998. Disponível em: <https://doi.org/10.1080/01638539809545028>. Acesso em: 25 mai. 2023

LEMIEUX, V. Blockchain and Distributed Ledgers as Trusted Record Keeping Systems: an archival theoretic evaluation framework. In: **FUTURE TECHNOLOGIES CONFERENCE (FTC 2017)**, 2017, Vancouver. *Proceedings of the Future Technologies Conference*. Vancouver: SAI Conference, 2017. p. 41-48. Disponível em: https://saiconference.com/Downloads/FTC2017/Proceedings/4_Paper_279-Blockchain_and_Distributed_Ledgers_as_Trusted.pdf. Acesso em: 08 mar. 2023.

LEMIEUX, V. Trusting records: is *blockchain* technology the answer? **Records Management Journal**, Bingley, v. 26, n. 2, p. 110-139, 18 jul. 2016.

LI, H. et al. FASTDB: an array database system for efficient storing and analyzing massive scientific data. In: WANG, G.; ZOMAYA, A.; MARTINEZ, G.; LI, K. (eds.). **Algorithms and Architectures for Parallel Processing. ICA3PP 2015. Lecture Notes in Computer Science**, v. 9532. Cham: Springer, 2015. Disponível em: https://doi.org/10.1007/978-3-319-27161-3_55. Acesso em: 10 abr. 2023.

LIU, G.; HE, J.; XUAN, X. A data preservation method based on blockchain and multidimensional hash for digital forensics. **Complexity**, v. 2021, p. 1-12, 2021. Disponível em: <https://www.hindawi.com/journals/complexity/2021/6695239/>. Acesso em: 20 abr. 2021.

LIU, T. et al. A performance evaluation of Hive for scientific data management. In: IEEE INTERNATIONAL CONFERENCE ON BIG DATA, 6-9 out. 2013. **2013 IEEE International Conference on Big Data**. 2013. Disponível em: <https://ieeexplore.ieee.org/xpl/conhome/6679357/proceeding>. Acesso em: 27 nov. 2023.

LIVTI. **Blockchain**: tecnologia por trás da revolução das moedas digitais. 2017. Disponível em: <https://www.livti.com.br/blog/blockchain-tecnologia-por-tras-da-revolucao-das-moedas-digitais/>. Acesso em: 05 mar. 2020.

LOC.GOV. **Understanding PREMIS**. 2021. Disponível em: https://loc.gov/standards/premis/understandingPREMIS_portuguese_2021.pdf. Acesso em: 29 set. 2022.

MARIOTTO, M.; CARNEIRO, R.; ZANCAN, F.; SAWICKI S. Um estudo de mapeamento sistemático da mineração de dados para cenários de big data. **Revista GEDECON - Gestão e Desenvolvimento em Contexto**, [S. l.], v. 9, n. 1, p. 1–23, 2021. DOI: 10.33053/gedecon.v9i1.506. Disponível em: <https://revistaeletronica.unicruz.edu.br/index.php/gedecon/article/view/506>. Acesso em: 5 jan. 2024.

METH, M. *Blockchain* in libraries. **Library Technology Reports**, Chicago, v. 55, n. 8, p. 1-29, 1 dez. 2019.

MÖLLER, J.; FRÖSCHLE, S.; HAHN, A. Blockchain com permissão para proveniência de dados em gerenciamento de dados científicos. In: AHLEMANN, F.; SCHÜTTE, R.; STIEGLITZ, S. (eds.). **Inovação por meio de sistemas de informação. WI 2021. Lecture Notes in Information Systems and Organisation**, Cham: Springer, 2021. v. 48. p. 22-38. Disponível em: https://doi.org/10.1007/978-3-030-86800-0_2. Acesso em: 26 mar. 2021.

MORI, A. *et al.* Análise comparativa entre os requisitos da RDC-Arq e a tecnologia *blockchain*: uma perspectiva de profissionais arquivistas. **Ibersid**, Brasília, v. 14, n. 1, p. 29-40, 2020.

MOUGAYAR, W. **The business blockchain**: promise, practice, and application of the next internet technology. [S.l.]: Wiley, 2016.

NAKAMOTO, S. **Bitcoin**: a peer-to-peer electronic cash system. 2008. Disponível em: <https://bitcoin.org/bitcoin.pdf>. Acesso em: 27 nov. 2023.

NARAYANAN, A. *et al.* **Bitcoin and cryptocurrency technologies**: a comprehensive introduction. Princeton: Princeton University Press, 2016.

NAWAZ, F.; KOÇ, Ç. K. Towards a systematic review of hash functions and digital signatures in *blockchain* applications. In: INTERNATIONAL CONFERENCE ON INFORMATION NETWORKING (ICOIN), 2018, Selangor. **Anais [...]**. Selangor: IEEE, 2018. p. 946-951.

OLIVEIRA, N. P. de. Tecnologia Blockchain: uma nova relação de confiança. **Revista Eixos Tech**, v. 8, n. 1, 2023. Disponível em: <https://eixostech.pas.ifsuldeminas.edu.br/index.php/eixostech/article/download/281/pdf>. Acesso em: 30 mar. 2021.

OMAME, I. M.; ALEX-NMECHA, J. C. **Application of blockchain in libraries and information centers**. 2021. Disponível em: <https://www.igi-global.com/chapter/application-of-blockchain-in-libraries-and-information-centers/285505>. Acesso em: 5 mai. 2022.

OUCHI, M. T.; ARAKAKI, A. C. S. Um estudo da blockchain aplicado ao contexto dos dados de pesquisa. **Em Questão**, v. 26, n. 3, p. 70-93, 2020. Disponível em: <https://www.redalyc.org/journal/4656/465664724004/465664724004.pdf>. Acesso em: 16 set. 2021.

PADÓ, S.; LAPATA, M. Dependency-Based Construction of Semantic Space Models. **Computational Linguistics**, Massachusetts, v. 33, p. 161-199, 2007.

PALETTA, F. C.; SILVA, A. M. D. (eds.). **Série Tecnologia e Organização da Informação: Contribuições Para a Ciência da Informação**. 1. ed. [S.l.] Blucher Open Access, 2020.

PINHEIRO, L. V. R. Ciência da Informação: desdobramentos disciplinares, interdisciplinaridade e transdisciplinaridade. *In*: GONZÁLEZ DE GÓMEZ, M. N.; ORRICO, E. G. D. (Orgs.). **Políticas de memória e informação**: reflexos na organização do conhecimento. Natal: EDUFRN, 2006. p. 111-141.

PINHEIRO, L. V. R. Mutações na Ciência da Informação e reflexos nas mandalas interdisciplinares. **Informação & Sociedade: Estudos**, João Pessoa, v. 28, n. 3, p. 115-134, 2018. Disponível em: <https://periodicos.ufpb.br/ojs2/index.php/>. Acesso em: 10 out. 2020.

POMBO, O. **Interdisciplinaridade**: ambições e limites. Lisboa: Relógio D'Água, 2004.

REKUNENKO, I. I. *et al.* Data management in healthcare research as a guarantee of its quality. **Health Economics and Management Review**, Sumy, v. 3, n. 2, 36-43, 2022.

REN, Y. *et al.* Long-term preservation of electronic record based on digital continuity in smart cities. **Computers, Materials & Continua**, Henderson, v. 66, n. 3, p. 3271-3287, 2021.

SALES, L. F.; SAYÃO, L. F. Research data: contribution to establishment of a digital curation model for Brazil. **Instituto de Engenharia Nuclear: Progress Report**, [S. l.], n. 2, p. 33, 2015. Disponível em: <https://revistas.ien.gov.br/index.php/ienprogressreport/article/view/132>. Acesso em: 11 set. 2023.

SALES, L. F.; SAYÃO, L. F. Algumas considerações sobre os repositórios digitais de dados de pesquisa. **Informação & Informação**, [S. l.], v. 21, n. 2, p. 90–115, 2016. DOI: 10.5433/1981-8920.2016v21n2p90. Disponível em: <https://ojs.uel.br/revistas/uel/index.php/informacao/article/view/27939>. Acesso em: 20 ago. 2023.

SALES, L. F.; SAYÃO, L. F. Afinal, o que é dado de pesquisa? **BIBLOS - Revista do Instituto de Ciências Humanas e da Informação**, [S. l.], v. 34, n. 2, 2020. DOI: 10.14295/biblos.v34i2.11875. Disponível em: <https://periodicos.furg.br/biblos/article/view/11875>. Acesso em: 16 ago. 2023.

SALIM, F.; HAQUE, U.; SRINIVASAN, S. A systematic review of the use of *blockchain* in healthcare. **Social Network Analysis and Mining**, Viena, v. 9, n. 1, p. 1-25, 2019.

SANTOS, H. M. dos; FLORES, D. Os impactos da obsolescência tecnológica frente à preservação de documentos digitais. **Brazilian Journal of Information Studies: Research Trends**, v. 11, n. 2, p. 28-37, 2017. ISSN 1981-1640.

SANTOS, R.; ROCHA, A. Blockchain in records and information management: opportunities and challenges. *In*: EUROPEAN CONFERENCE ON DIGITAL

GOVERNMENT (ECDG 2020), 20., 2020. **Proceedings of the 20th European Conference on Digital Government (ECDG 2020)**. [S. l.]: Academic Conferences International (ACI), 2020. p. 416-424.

SANTOS, T. V. DOS; MAIMONE, G. D. **Recuperação da informação e visibilidade científica sob a perspectiva de identificadores persistentes**. *Revista Ibero-Americana de Ciência da Informação*, v. 16, n. 3, p. 518–542. Disponível em: <https://doi.org/10.26512/rici.v16.n3.2023.44053>. Acesso em: 08 nov. 2023.

SARACEVIC, T. Ciência da Informação: origem, evolução e relações. **Perspectivas da Ciência da Informação**, Belo Horizonte, v. 1, n. 1, p. 41-62, jan./jun. 1996.

SAYÃO, L. F. Interoperabilidade das bibliotecas digitais: o papel dos sistemas de identificadores persistentes - URN, PURL, DOI, Handle System, CrossRef e OpenURL. **Transinformação**, Campinas, v. 19, n. 1, p. 65-82, abr. 2007. Disponível em: <https://doi.org/10.1590/S0103-37862007000100006>. Acesso em: 20 abr. 2023.

SAYÃO, L. F. Uma outra face dos metadados: informações para a gestão da preservação digital. **Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação**, Florianópolis, v. 15, n. 30, p. 1–31, 2010. DOI: 10.5007/1518-2924.2010v15n30p1. Disponível em: <https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2010v15n30p1>. Acesso em: 09 ago. 2023

SAYÃO, L. F.; SALES, L. **FAIR Digital Object**: expanding the boundaries of data stewardship. *Informação & Sociedade*, 2024. No prelo.

SCHWARTZ, D.; YOUNGS, N.; BRITTO, A. **The ripple protocol consensus algorithm**. 2020. Disponível em: <https://xrpl.org/consensus.html>. Acesso em: 27 nov. 2023.

THE ODI. **Archangel**: guaranteeing the integrity of digital archives. Disponível em: <https://theodi.org/insights/reports/archangel-guaranteeing-the-integrity-of-digital-archives/>. Acesso em: 05 jan. 2024.

THOMAZ, K. P.; SOARES, A. J. A preservação digital e o modelo de referência Open Archival Information System (OAIS). **DataGramaZero**, v. 5, n. 1, fev. 2004. Disponível em: http://www.dgz.org.br/fev04/F_I_art.htm. Acesso em 22 fev. 2023.

WASHINGTON LUÍS RIBEIRO DE CARVALHO SEGUNDO. *et al.* dARK: uma implementação descentralizada de identificadores persistentes ARK baseada em *blockchain*. **BiblioCanto**, Natal, v. 9, n. 2, p. 147–158, 4 dez. 2023.

WASHINGTON LUÍS RIBEIRO DE CARVALHO SEGUNDO. *et al.* **EasyChair Preprint dARK**: a Decentralized Blockchain Implementation of ARK Persistent Identifiers, 2022. Disponível em: https://easychair.org/publications/preprint_open/GpWB. Acesso em: 19 jan. 2024.

SFETCU, N. **Blockchain Design and Modelling**. Disponível em: <https://www.internetmobile.ro/blockchain-design-and-modelling/>. Acesso em: 10 fev. 2024.

SHAMSHAD, S. *et al.* A secure *blockchain*-based e-health records storage and sharing scheme. **Journal of Information Security and Applications**, Amsterdã, v. 55, p. 01-17, dez. 2020.

SHETH, A.; RANABAHU, A. Semantic Modeling for Cloud Computing, Part 2. **IEEE Internet Computing**, Los Alamitos, v. 14, p. 81-84, 2010.

SINGH, A. *et al.* Sidechain technologies in blockchain networks: An examination and state-of-the-art review. **Journal of Network and Computer Applications**, v. 149, 2020. Disponível em: <https://doi.org/10.1016/j.jnca.2019.102471>. Acesso em: 09 ago. 2022.

SKALA, K. *et al.* Prospects of digital scientific publishing on *blockchain*: The concept of DAP. **Open Research Europe**, Londres, v. 3, p. 1-27, 2023. Disponível em: <https://doi.org/10.12688/openreseurope.15771.1>. Acesso em 03 abr. 2023.

SWAN, M. **Blockchain**: blueprint for a new economy. Sebastopol: O'Reilly Media, 2015.

TAPSCOTT, D.; TAPSCOTT, A. **Blockchain revolution**: how the technology behind Bitcoin is changing money, business, and the world. Londres: Penguin, 2016.

TRANFIELD, D.; DENYER, D.; SMART, P. Towards a methodology for developing evidence-informed management knowledge by means of systematic review. **British Journal of Management**, Hoboken, v. 14, n. 3, p. 207-222, 2003.

UK RESEARCH AND INNOVATION. **Archangel**. Disponível em: <https://gtr.ukri.org/projects?ref=EP%2FP03151X%2F1>. Acesso em: 05 jan. 2024.

UNIVERSITY OF SURREY. **Archangel - Trusted Archives of Digital Public Records**. 2018. Disponível em: <https://blockchain.surrey.ac.uk/projects/archangel.html>. Acesso em: 05 jan. 2024.

VAVILENKOV, A. Basic Principles of the Synthesis of Logical–Linguistic Models. **Cybernetics and Systems Analysis**, Kyiv, v. 51, p. 826-834, 2015. Disponível em: <https://doi.org/10.1007/s10559-015-9776-z>. Acesso 08 ago. 2023.

VOSHMIGIR, S. **Token economy**: how the Web3 reinvents the internet. [S.l.]: Token Kitchen, 2020.

WANG, K.; WANG, Y.-Y.; ACERO, A. **Use and Acquisition of Semantic Language Model**. In: Demonstration Papers At Hlt-Naacl, 2004, Boston. Proceedings [...]. Boston: Association for Computational Linguistics, 2004.

WILLIAMS, K. **Disruptive Technology of the 21st Century: How Blockchain Can Improve Electoral Voting**. 2023. Disponível em: <https://www.igi-global.com/chapter/disruptive-technology-of-the-21st-century/323503>>. Acesso em: 12 jan. 2024.

WOO, C.; YOO, J. Exploring the determinants of *blockchain* acceptance for research data management. **Journal of Computer Information Systems**, Philadelphia. v. 63, n. 1, p. 1-12, 5 abr. 2022.

WOODALL, A.; RINGEL, S. *Blockchain* archival discourse: trust and the imaginaries of digital preservation. **New Media & Society**, California. v. 22, n. 12, p. 2200-2217, 22 nov. 2019.

YUN-CHI, C. *et al.* "Prove it!" A user-centered design for the *blockchain*-based research lifecycle transparency framework. **Proceedings of the Association for Information Science and Technology**, Silver Spring. v. 59, n. 1, p. 636-638, 1 out. 2022.

ZAGHLOUL, E.; LI, T.; MUTKA, M.; REN, J. Bitcoin and *Blockchain*: Security and Privacy. **IEEE Internet of Things Journal**, Michigan State University. v. 7, p. 10288-10313, 2019.

ZHANG, X. Application of *blockchain* technology in data management of university scientific research. *In*: BAROLLI, L.; PONISZEWSKA-MARANDA, A.; PARK, H. (Eds.). **Innovative mobile and internet services in ubiquitous computing**. IMIS 2020. Cham: Springer, 2020. p. 606-613. *Advances in Intelligent Systems and Computing*, v. 1195.

APÊNDICE – Memorial

Minha jornada acadêmica sempre foi moldada por uma busca constante pela formação interdisciplinar. Inicialmente, escolhi adentrar o campo tecnológico, guiado por um profundo interesse em compreender como a tecnologia pode ser empregada para resolver problemas e aprimorar a qualidade de vida das pessoas. À medida que avançava no curso, reconheci que para aplicar os conhecimentos absorvidos nesse domínio, era fundamental compreender a interação entre as pessoas e a tecnologia, assim como seus processos de aprendizado.

Esta constatação levou-me à decisão de me formar em Pedagogia, com o objetivo de aprofundar os meus conhecimentos na área da educação e melhorar a minha compreensão de como a tecnologia pode ajudar a otimizar os processos de ensino e aprendizagem. Durante esse percurso, minha crescente fascinação pelo entrelaçamento entre tecnologia e educação inspirou-me a buscar especializações nas esferas da educação e da tecnologia, ampliando assim a minha compreensão desses temas essenciais.

Após completar minhas especializações, emergiu em mim a convicção de que a Ciência da Informação poderia servir como um terreno fértil para a minha pesquisa, unindo harmoniosamente os domínios da tecnologia, educação e informação. Essa convicção levou-me a empreender meu mestrado em Ciência da Informação, no qual tive a oportunidade de investigar o valor informativo das histórias em quadrinhos como veículo de divulgação científica. Essa pesquisa, repleta de desafios, permitiu-me aprimorar minha perspectiva interdisciplinar e explorar novos domínios do conhecimento.

A transição da minha formação interdisciplinar para o foco específico desta pesquisa de doutorado ocorreu de maneira natural, mas reveladora. Ao adentrar o doutorado, também em Ciência da Informação, minha trajetória me conduziu a um tema ainda mais instigante e desafiador: a aplicação da tecnologia *blockchain* no contexto dos dados de pesquisa e preservação digital. Estou firmemente convencido de que essa tecnologia detém um potencial transformador capaz de redefinir a maneira pela qual as informações são geridas e compartilhadas.

Minha trajetória acadêmica reforçou a importância de abordagens interdisciplinares e integradas. Acredito que essa perspectiva é vital para superar os desafios enfrentados pela sociedade contemporânea. Minha pesquisa busca

combinar a tecnologia, a educação e a informação de maneira significativa e inovadora, visando a gerar conhecimentos que revolucionem a gestão e o compartilhamento de informações.

ANEXO

Dados coletados no processo de análise da revisão sistemática da literatura:

	O fenômeno investigado
[1]	Pesquisadores podem gerenciar seus dados de pesquisa de forma mais eficiente e transparente, por meio de uma interface centrada no usuário e uma estrutura de transparência de pesquisa baseada em <i>blockchain</i> .
[2]	Aplicação da tecnologia <i>blockchain</i> em bibliotecas, com o objetivo de melhorar a gestão de coleções e a comunicação acadêmica.
[3]	Aplicação de <i>blockchain</i> em dados de pesquisa.
[4]	A aceitação do <i>blockchain</i> para gerenciamento de dados de pesquisa.
[5]	Apresentar a Arqueobiologia como um domínio de aplicação altamente interessante e desafiador para a Ciência da Computação, e discute as principais questões e tarefas de gerenciamento e análise de dados neste campo de pesquisa.
[6]	Apresenta um estudo sobre o uso de <i>blockchain</i> para gerenciamento de dados científicos em geral.
[7]	O envio seguro de manuscritos com carimbo de data/hora e feedback de revisão por pares utilizando o <i>blockchain</i> , a fim de garantir a integridade e a aceitação dos manuscritos protegidos.
[8]	Aplicação da tecnologia <i>blockchain</i> em bibliotecas e centros de informação, incluindo suas características e como ela pode ser aplicada para gerenciar dados de pesquisa, entre outras possibilidades.
[9]	Aplicação da tecnologia <i>blockchain</i> na gestão de dados de pesquisa científica universitária.
[10]	Gerenciamento e análise de grandes volumes de dados científicos, mais especificamente dados gerados por levantamentos astronômicos como o Sloan Digital Sky Survey.
[11]	Gerenciamento de dados científicos no Hive, um data warehouse baseado em MapReduce, e avalia o desempenho do Hive para algumas tarefas de análise de dados científicos.
[12]	Propõe uma abordagem híbrida para gerenciamento de dados científicos em infraestruturas de grid.
[13]	Eficiência da gestão de dados no sistema de saúde. <i>blockchain</i>
	Os problemas
[1]	As motivações dos pesquisadores para registrar suas pesquisas com <i>blockchain</i> e suas expectativas sobre uma ferramenta de gerenciamento de registros de pesquisa baseada em <i>blockchain</i> .
[2]	A aplicação da tecnologia <i>blockchain</i> em bibliotecas para melhorar a gestão de coleções e a comunicação acadêmica.
[3]	A contribuição da tecnologia <i>blockchain</i> para a falta de confiabilidade e reprodutibilidade dos resultados de pesquisas científicas.
[4]	Os fatores que influenciam a intenção dos investigadores em usar a tecnologia <i>blockchain</i> para gerenciar seus dados de pesquisa, bem como identificar os riscos percebidos pelos usuários em relação ao uso dessa tecnologia?
[5]	A falta de uso de tecnologia da informação para gerenciar e analisar dados científicos na área da Arqueobiologia, levando a uma grande quantidade de dados que não são experimentados e são registrados de maneiras diferentes por cada pesquisador individual.
[6]	A falta de rastreabilidade e confiança nas informações de proveniência de dados em fluxos de trabalho científicos, especialmente em ecossistemas de dados complexos com várias fontes de dados descentralizadas.
[7]	A falta de um mecanismo confiável e verificável para os autores provarem que um trabalho de pesquisa.
[8]	Questões relacionadas à privacidade, acesso não autorizado, roubo de dados e altos custos de aquisição e manutenção de servidores.
[9]	A segurança dos dados, segurança de armazenamento e segurança de utilização de dados de pesquisa científica, que são facilmente modificáveis eletronicamente.
[10]	O desafio de armazenar e analisar dados científicos em larga escala, que aumentaram rapidamente em tamanho e complexidade com o desenvolvimento da ciência e da tecnologia.

[11]	O desempenho do Hive, um data warehouse baseado em MapReduce, para aplicações de gerenciamento de dados científicos.
[12]	À integração entre bancos de dados e aplicativos legados na área de biologia.
[13]	A necessidade de uma gestão efetiva de dados na área da saúde, garantindo sua disponibilidade, relevância e qualidade.
	A perspectiva teórica
[1]	Dados de pesquisa; <i>Blockchain</i>
[2]	Gerenciamento de dados de pesquisa; <i>Blockchain</i>
[3]	Dados de pesquisa; <i>Blockchain</i>
[4]	Gerenciamento de dados de pesquisa; <i>Blockchain</i>
[5]	Gerenciamento de dados científicos; Banco de dados distribuído
[6]	Dados em fluxos de trabalho científicos; <i>Blockchain</i>
[7]	Dados de pesquisa; <i>Blockchain</i>
[8]	Gerenciamento de dados de pesquisa; <i>Blockchain</i>
[9]	<i>Blockchain</i> ; gestão de dados de pesquisa científica
[10]	Dados científicos; Banco de dados distribuído
[11]	Dados científicos; Banco de dados distribuído
[12]	Gerenciamento de dados; Banco de dados distribuído
[13]	Gerenciamento de dados de pesquisa; <i>Blockchain</i>
	O processo de coleta de dados
[1]	O processo de coleta de dados envolveu duas etapas principais: análise de conteúdo e entrevistas. Na análise de conteúdo, foram utilizados dados de um workshop de design thinking com 17 investigadores principais em Taiwan no campo das ciências da vida, com o objetivo de entender os obstáculos que eles enfrentaram no gerenciamento de dados de pesquisa. Já nas entrevistas, foram realizadas seis entrevistas em áreas como biomedicina, ciências básicas, ciências aplicadas e ciências sociais, com o objetivo de entender melhor os sentimentos dos usuários sobre o serviço geral, bem como as deficiências do design existente e as barreiras encontradas pelos usuários. As entrevistas de acompanhamento cobriram três aspectos das práticas de pesquisa do dia-a-dia dos entrevistados: manutenção de registros, armazenamento de registros de pesquisa e colaboração acadêmica. Além disso, foram feitas perguntas sobre ética acadêmica e UX do protótipo.
[2]	Não há informações específicas sobre o processo de coleta de dados.
[3]	Bibliográfico com levantamento da literatura e experimento aplicando os dados encontrados na literatura com a construção de um dispositivo eletrônico computacional.
[4]	Não há informações específicas sobre o processo de coleta de dados.
[5]	Não há informações específicas sobre o processo de coleta de dados.
[6]	Não há informações específicas sobre o processo de coleta de dados.
[7]	Não há informações específicas sobre o processo de coleta de dados.
[8]	Não há informações específicas sobre o processo de coleta de dados.
[9]	Não há informações específicas sobre o processo de coleta de dados.
[10]	Não há informações específicas sobre o processo de coleta de dados.
[11]	Não há informações específicas sobre o processo de coleta de dados.
[12]	Não há informações específicas sobre o processo de coleta de dados.
[13]	Não há informações específicas sobre o processo de coleta de dados.
	A natureza dos dados
[1]	O processo de pesquisa coletou dados por meio de um workshop de design thinking por meio de entrevistas em áreas como biomedicina, ciências básicas, ciências aplicadas e ciências sociais.
[2]	Não é possível determinar a natureza dos dados.

[3]	Dados científicos.
[4]	Os dados mencionados no contexto são principalmente relacionados a pesquisas e estudos em diferentes áreas, incluindo física de partículas, genômica e ciências da vida.
[5]	Os arqueobiólogos processam dados de tipos bastante diferentes, incluindo dados planejados, dados de intervalo e dados incertos, dados temporais, dados espaciais, dados de localização e dados espaço-temporais.
[6]	Os dados mencionados incluem dados AIS (Automatic Identification System) de tráfego marítimo e dados de GPS.
[7]	O texto discute a natureza dos dados relacionados à submissão de manuscritos acadêmicos.
[8]	A natureza dos dados é principalmente acadêmica e informativa.
[9]	Dados digitais de pesquisa científica universitária, que precisam estar disponíveis, autênticos e válidos.
[10]	Os dados em questão são dados científicos multidimensionais, como dados de rádio multibanda, dados de espectro e polarização.
[11]	Os dados mencionados no texto são principalmente dados científicos, incluindo dados de astronomia, sensoriamento remoto e processamento de imagens médicas.
[12]	Não é possível determinar a natureza dos dados.
[13]	Os dados mencionados no texto são principalmente relacionados à gestão de dados de pesquisa na área da saúde.
A validação dos dados	
[1]	Não há informações específicas sobre como validar os dados.
[2]	Não há informações específicas sobre como validar os dados.
[3]	Mecanismo de consenso.
[4]	É possível observar que foram realizados testes de consistência interna entre os itens de medição de cada variável latente, bem como testes de validade discriminante por meio do teste Fornell-Larcker, valor de carga cruzada e valor Heterotrait-Monotrait Ratio of Correlations (HTMT).
[5]	Não há informações específicas sobre como validar os dados.
[6]	Na utilização de <i>blockchain</i> para garantir a segurança e a transparência na rastreabilidade dos dados, as transações são validadas por meio de assinaturas digitais.
[7]	Não há informações específicas sobre como validar os dados.
[8]	Não há informações específicas sobre como validar os dados.
[9]	A tecnologia <i>blockchain</i> utiliza livro-razão distribuído, assinatura digital, carimbo de data/hora digital e outras técnicas para garantir que os dados não sejam adulterados ou excluídos e possam ser rastreados.
[10]	Não há informações específicas sobre como validar os dados.
[11]	Não há informações específicas sobre como validar os dados.
[12]	Não há informações específicas sobre como validar os dados.
[13]	Ele aborda a importância da gestão efetiva dos dados e a necessidade de garantir sua disponibilidade e relevância, bem como a importância da transparência, reutilização e verificação de dados de pesquisa.
O processo de coleta de achados	
[1]	Envolveu uma análise de conteúdo para entender as necessidades dos pesquisadores e as barreiras de compartilhamento de dados de pesquisa. Posteriormente, foram realizadas entrevistas para entender melhor os sentimentos dos usuários sobre o serviço geral, bem como as deficiências do design existente e as barreiras encontradas pelos usuários. As entrevistas cobriram três aspectos das práticas de pesquisa do dia-a-dia dos entrevistados: manutenção de registros, armazenamento de registros de pesquisa e colaboração acadêmica.
[2]	Não há informações específicas sobre o processo de coleta de achados.
[3]	Esta coleta seria feita por um dispositivo eletrônico computacional (um hardware) que poderia conter diversos sensores, exemplarmente neste caso, um termômetro.
[4]	Não há informações específicas sobre o processo de coleta de achados.

[5]	Gerenciar e armazenar os dados heterogêneos em um sistema de banco de dados central para análise e recuperação eficiente.
[6]	Não há informações específicas sobre o processo de coleta de achados.
[7]	Não há informações específicas sobre o processo de coleta de achados.
[8]	Não há informações específicas sobre o processo de coleta de achados.
[9]	Não há informações específicas sobre o processo de coleta de achados.
[10]	Não há informações específicas sobre o processo de coleta de achados.
[11]	Não há informações específicas sobre o processo de coleta de achados.
[12]	Não há informações específicas sobre o processo de coleta de achados.
[13]	Não há informações específicas sobre o processo de coleta de achados.
Os achados	
[1]	Os autores descobriram que a maior consideração entre os acadêmicos ao conduzir atividades de gerenciamento de dados de pesquisa é "melhorar o impacto acadêmico".
[2]	Não há informações específicas sobre os achados.
[3]	Ao serem minerados, esses dados fariam parte da <i>blockchain</i> e tornando-se imutáveis ao longo do tempo, garantindo que os dados brutos para a análise sejam comprovadamente autênticos e teoricamente à prova de manipulações. As consequências de um sistema como esse, permitiria o desenvolvimento de uma nova pesquisa, e assim, é possível vislumbrar impactos positivos relacionados à reprodutibilidade da pesquisa e ao reuso dos dados dessas pesquisas.
[4]	Não há informações específicas sobre os achados.
[5]	Não há informações específicas sobre os achados.
[6]	A revisão da literatura foi realizada para analisar trabalhos na área de proveniência de dados em gerenciamento de dados científicos com especial atenção à segurança, arquitetura e modelos de fluxo de trabalho. Além disso, foram discutidos trabalhos que usam a tecnologia <i>blockchain</i> no contexto do armazenamento de informações de proveniência de dados. Foram analisados trabalhos que propõem diferentes abordagens para garantir a segurança e transparência na rastreabilidade dos dados, como a utilização de contratos inteligentes, votação com todos os nós ou por votação de limite aleatório, entre outras.
[7]	Não há informações específicas sobre os achados.
[8]	Os achados destacam os desafios e oportunidades da aplicação de <i>blockchain</i> em bibliotecas e centros de informação.
[9]	A aplicação da tecnologia <i>blockchain</i> pode contribuir para a gestão de todo o ciclo de vida dos dados de pesquisa científica e garantir o uso eficaz e a segurança dos mesmos.
[10]	Não há informações específicas sobre os achados.
[11]	Não há informações específicas sobre os achados.
[12]	Não há informações específicas sobre os achados.
[13]	Não há informações específicas sobre os achados.
O processo de interpretação dos achados	
[1]	Não há informações específicas sobre o processo de interpretação dos achados.
[2]	Não há informações específicas sobre o processo de interpretação dos achados.
[3]	Princípio da reprodutibilidade; o documento Replication Studies: Improving reproducibility in the empirical sciences, da Royal Netherlands Academy of Arts and Sciences (2018).
[4]	Não há informações específicas sobre o processo de interpretação dos achados.
[5]	Descreve que a arqueozologia avalia achados ósseos de animais domésticos e silvestres e outros vestígios faunísticos de épocas pré-históricas e históricas.
[6]	Discute a utilização da tecnologia <i>blockchain</i> para garantir a segurança e a transparência na rastreabilidade dos dados, cujo foco principal é na descrição de conceitos e tecnologias relacionadas ao uso de <i>blockchain</i> para a proveniência de dados de pesquisa científica.
[7]	Não há informações específicas sobre o processo de interpretação dos achados.

[8]	Os achados relacionados à aplicação de <i>blockchain</i> em bibliotecas e centros de informação foram interpretados como tendo um grande potencial disruptivo para alterar o modo de operação e prestação de serviços nessas instituições.
[9]	Por meio da aplicação da análise SWOT na gestão de dados de pesquisa científica universitária para avaliar a viabilidade e potencial da utilização da tecnologia <i>blockchain</i> .
[10]	O artigo descreve principalmente a necessidade de um sistema eficiente para gerenciamento e análise de dados científicos em larga escala e apresenta o FASTDB como uma solução baseada em modelo de matriz distribuída otimizada para gerenciamento de dados científicos massivos.
[11]	Não há informações específicas sobre o processo de interpretação dos achados.
[12]	Não há informações específicas sobre o processo de interpretação dos achados.
[13]	Não há informações específicas sobre o processo de interpretação dos achados.
A interpretação dos achados	
[1]	Por meio da análise de conteúdo, os pesquisadores identificaram três funções centrais da interface, incluindo autonomia decidida pelo usuário, um painel de progresso e gerenciamento da equipe de pesquisa. Os pesquisadores também identificaram os desafios e complexidades do gerenciamento de dados de pesquisa, incluindo a necessidade de um padrão unificado e sistemático de gerenciamento de dados.
[2]	Não há informações específicas sobre a interpretação dos achados.
[3]	Os dados coletados e gravados a partir de um dispositivo “lacrado” e cujos parâmetros foram configurados a partir de um Smart Contract que representa o projeto de pesquisa, ficará muito difícil que a pesquisa tome rumos que não os acordados no início do projeto, sem que uma nova concordância entre o pesquisador e o grupo de pesquisa e a agência financiadora. Além disso, cada alteração ficaria registrada, gerando assim um registro transparente de toda e qualquer alteração necessária ao projeto e que pode ser considerado em pesquisas de reprodutibilidade posteriores.
[4]	Não há informações específicas sobre a interpretação dos achados.
[5]	Não há informações específicas sobre a interpretação dos achados.
[6]	A importância da proveniência dos dados para o processamento de dados científicos e como a tecnologia <i>blockchain</i> pode ser usada para armazenar informações de proveniência de dados de forma segura e transparente. Além disso, o texto discute a arquitetura <i>blockchain</i> e a importância da identificação das entidades processadoras de dados para garantir a rastreabilidade e a confiança entre as diferentes partes envolvidas.
[7]	Não há informações específicas sobre a interpretação dos achados.
[8]	O artigo discute as possibilidades de uso do <i>blockchain</i> em bibliotecas e centros de informação para diversos fins, como arquivamento e rastreamento de registros bibliográficos essenciais, verificação da origem e integridade das informações da biblioteca e facilitação das atividades de circulação da biblioteca. No entanto, o artigo também destaca vários desafios na implementação da tecnologia <i>blockchain</i> em bibliotecas, como o alto custo inicial, complexidade da arquitetura, perda de controle sobre informações e dados e potencial uso indevido de criptomoedas para crimes cibernéticos.
[9]	A aplicação da tecnologia <i>blockchain</i> tem potencial para resolver os problemas de segurança de dados, segurança de armazenamento e segurança de utilização de dados de pesquisa científica na gestão de dados de pesquisa científica universitária, garantindo a disponibilidade, autenticidade e validade dos dados da pesquisa científica.
[10]	Não há informações específicas sobre a interpretação dos achados.
[11]	Não há informações específicas sobre a interpretação dos achados.
[12]	Não há informações específicas sobre a interpretação dos achados.
[13]	Não há informações específicas sobre a interpretação dos achados.
Os cuidados éticos	
[1]	Não há informações específicas sobre os cuidados éticos.
[2]	Não há informações específicas sobre os cuidados éticos.
[3]	Não há informações específicas sobre os cuidados éticos.
[4]	Não há informações específicas sobre os cuidados éticos.

[5]	Não há informações específicas sobre os cuidados éticos.
[6]	Não há informações específicas sobre os cuidados éticos.
[7]	A importância dos cuidados éticos na publicação acadêmica é fundamental para garantir a integridade e a confiabilidade da pesquisa. O plágio, a fraude e o viés podem comprometer a qualidade e a credibilidade dos resultados de pesquisa, além de prejudicar a reputação dos pesquisadores e das instituições envolvidas. Para garantir a ética na publicação acadêmica, é importante que os pesquisadores sigam as normas e diretrizes estabelecidas pelas instituições e pelos periódicos científicos. Além disso, é fundamental que os revisores e editores sejam imparciais e atuem com integridade e transparência no processo de revisão por pares. Outra forma de garantir a ética na publicação acadêmica é por meio do uso de tecnologias, como o <i>blockchain</i> , que permitem a verificação da autenticidade e da data de criação de um trabalho de pesquisa. Isso pode ajudar a prevenir o plágio e a garantir a prioridade de uma pesquisa.
[8]	Os profissionais das bibliotecas e da informação devem sensibilizar a sua comunidade de usuários e o público em geral sobre o uso ético da <i>blockchain</i> como um meio genuíno de armazenar, verificar e gerenciar informações para aumentar a confiança do público na tecnologia nas bibliotecas.
[9]	Não há informações específicas sobre os cuidados éticos.
[10]	Não há informações específicas sobre os cuidados éticos.
[11]	Não há informações específicas sobre os cuidados éticos.
[12]	Não há informações específicas sobre os cuidados éticos.
[13]	O texto aborda os cuidados éticos relacionados à gestão de dados de pesquisas na área da saúde. Ele menciona questões como privacidade, acesso restrito, propriedade dos dados e quem pode acessar os dados de pesquisa médica, além de destacar a importância dos princípios FAIR (localizáveis, acessíveis, interoperáveis e reutilizáveis) para garantir a qualidade e a integridade dos dados de pesquisa médica.
A implementação dos estudos	
[1]	Os pesquisadores usaram a análise de conteúdo para entender as necessidades dos pesquisadores e as barreiras ao compartilhamento de dados de pesquisa e projetaram a interface do usuário de acordo.
[2]	Não há informações específicas sobre as implementações dos estudos.
[3]	Construção de um dispositivo eletrônico computacional.
[4]	Não há informações específicas sobre as implementações dos estudos.
[5]	Não há informações específicas sobre as implementações dos estudos.
[6]	Não há informações específicas sobre as implementações dos estudos.
[7]	Existem várias maneiras de garantir a implementação ética e confiável de estudos na publicação acadêmica. Uma abordagem é o uso de sistemas de submissão de manuscritos que coordenam todos os estágios do processo de publicação, desde o envio de resumos e manuscritos até a organização da revisão por pares e o recebimento de manuscritos prontos para câmara. Além disso, o uso de sistemas de detecção de plágio mais sofisticados pode ajudar os examinadores a descobrir mais rapidamente a extensão do plágio, bem como aumentar o esforço que os plagiadores devem fazer para evitar a detecção. Outra abordagem é o uso de carimbos de data/hora confiáveis descentralizados, como o <i>blockchain</i> do Bitcoin, para garantir a autenticidade e a integridade dos dados e resultados de pesquisa. Isso permite que os autores e o público verifiquem de forma independente se um manuscrito, um conjunto de dados ou outros resultados de pesquisa já existiam em um formato preciso no momento da submissão a uma conferência ou periódico.
[8]	Algumas das soluções e recomendações para a implementação de <i>blockchain</i> em bibliotecas e centros de informação incluem: 1. Disponibilização de fundos adequados para a implementação da tecnologia <i>blockchain</i> e treinamento dos profissionais da biblioteca e informação para lidar com as mudanças que a tecnologia traria. 2. Capacitação dos profissionais de biblioteca e informação com habilidades e conhecimentos técnicos para trabalhar com a tecnologia <i>blockchain</i> . 3. Transferência de alguns dos privilégios e controle para os usuários em conjunto com o modelo Library 2.0, onde o foco está na mudança centrada no usuário e na colaboração na criação de conteúdo, comunidade e serviços. 4. Sensibilização adequada da comunidade de usuários e do público em geral sobre o uso ético da <i>blockchain</i> como um meio genuíno de

	armazenar, verificar e gerenciar informações para aumentar a confiança do público na tecnologia nas bibliotecas.
[9]	Descreve de forma geral uma proposta de aplicação da tecnologia <i>blockchain</i> para a gestão de todo o ciclo de vida dos dados de pesquisa científica universitária.
[10]	O artigo menciona que o sistema de banco de dados de matriz otimizada distribuída para gerenciamento de dados científicos massivos, chamado FASTDB, foi aplicado na análise interativa de dados de levantamentos astronômicos e foram projetados uma série de experimentos com tarefas de análise científica para demonstrar as features de desempenho intrínsecas do sistema.
[11]	Não há informações específicas sobre as implementações dos estudos.
[12]	Não há informações específicas sobre as implementações dos estudos.
[13]	Não há informações específicas sobre as implementações dos estudos.
As limitações dos estudos	
[1]	Não foi detalhado.
[2]	As limitações dos estudos de caso são que eles são apenas teóricos e não foram testados na prática. Eles podem ser úteis para educar e explorar a viabilidade dessas ideias, mas devem ser considerados com cautela.
[3]	Não foi detalhado.
[4]	Não foi detalhado.
[5]	Não foi detalhado.
[6]	Não foi detalhado.
[7]	A abordagem proposta não pode impedir o plágio acadêmico, mas pode ajudar os autores a provar sua propriedade intelectual e apoiar os examinadores na confirmação mais fácil da existência de plágio.
[8]	Os estudos apresentados destacam algumas restrições na aplicação de <i>blockchain</i> em bibliotecas e centros de informação, como a complexidade da arquitetura criptográfica e de rede, o custo inicial de implementação, a perda de controle de recursos de informação e dados do usuário por parte das bibliotecas, a falta de dependência do recurso de manutenção de registros de 'imutabilidade' do <i>blockchain</i> para a dinâmica evolutiva das pesquisas acadêmicas, e associação do uso de criptomoedas com atividades ilegais. Além disso, os estudos apontam a necessidade de treinamento e capacitação dos profissionais de biblioteca e informação para trabalhar com a tecnologia <i>blockchain</i> e a importância de sensibilizar a comunidade de usuários e o público em geral sobre o uso ético da tecnologia.
[9]	A necessidade de maior exploração e prática da aplicação da tecnologia <i>blockchain</i> neste contexto, bem como o fato de algumas atividades de pesquisa científica universitária dependerem de dados que são difíceis de quantificar ou proteger usando tecnologia <i>blockchain</i> . Outras limitações descritas no artigo incluem a necessidade de maior padronização de dados e aplicativos e a falta de consenso sobre quais métodos são mais adequados para a aplicação da tecnologia <i>blockchain</i> na gestão de dados de pesquisa científica universitária.
[10]	Não foi detalhado
[11]	É mencionado que atualmente existem alguns esforços no processamento de dados científicos na plataforma Hadoop, mas ainda falta um benchmark de processamento de dados científicos em plataformas distribuídas.
[12]	Não foi detalhado.
[13]	Não foi detalhado