

JARDEL RIBEIRO DE LIMA

Componentes de *Frameworks* de Governança de
Dados para *Big Data*: Uma Revisão Sistemática da
Literatura

São Paulo
2024

JARDEL RIBEIRO DE LIMA

Componentes de *Frameworks* de Governança de
Dados para *Big Data*: Uma Revisão Sistemática da
Literatura

Monografia apresentada ao Programa de
Educação Continuada da Escola Politécnica
da Universidade de São Paulo para obtenção
do título de Especialista, pelo Programa de
Pós-Graduação em Engenharia de Dados e
Big Data.

Área de Concentração:

Tecnologia da Informação - Gerenciamento
de Dados e *Big Data*

Orientador:

Profa. Ma. Maritza Carvalho Fran-
cisco

São Paulo
2024

A todas as pessoas que me apoiaram até aqui nessa jornada.

AGRADECIMENTOS

Agradeço aos professores do curso Especialização de Engenharia de Dados e *Big Data* da Escola Politécnica da USP (PECE), por todo o conhecimento passado em aula.

A Profa. Ma. Maritza, pela orientação do meu trabalho com sugestões que fizeram esse trabalho ser possível.

CURSO DE ESPECIALIZAÇÃO EM ENGENHARIA DE DADOS E *BIG DATA*

Coord.: Profa. Dra. Solange N. Alves de Souza

Vice-Coord.: Prof. Dr. Pedro Luiz Pizzigatti Corrêa

Perspectivas profissionais alcançadas com o curso: Apesar de já trabalhar na área, o curso me ajudou a ter maior conhecimento em diversas disciplinas relacionadas à Engenharia de Dados e *Big Data*. Desde disciplinas mais teóricas, entendendo os conceitos básicos de dado e informação, até disciplinas mais práticas, em que foi possível criar um pipeline completo de dados, desde a coleta até a utilização desses dados em um modelo de aprendizagem de máquina.

O curso me apresentou novas ferramentas e técnicas que posso aplicar no meu dia-a-dia, me qualificando para a função que já exerço e me dando uma base sólida para desafios futuros. Sem dúvidas é um curso que me ajudará a evoluir a minha carreira na área de dados.

RESUMO

A utilização de dados como suporte a tomada de decisão nas empresas já é uma prática consolidada. Nos últimos anos, o aumento exponencial na geração de dados em diferentes volumes, fontes e velocidades, impulsionou o fenômeno do *big data*. No entanto, a abundância de dados por si só é insuficiente; seu valor está diretamente ligado à confiabilidade e à qualidade que oferecem. É nesse ponto que entra a governança de dados, com uma série de artefatos que ajudam a controlar, proteger e entregar valor através dos ativos de dados. Contudo, não é algo tão simples de se implementar nas empresas e por essa razão pesquisadores e profissionais da área de dados desenvolveram *frameworks* para ajudar nesse desafio. *Frameworks* ajudam a estruturar ideias e conceitos complexos para facilitar o seu entendimento e implementação. Por definição, *big data* tem características distintas dos dados tradicionais armazenados em bancos de dados relacionais, *small data*. Por este motivo é necessário verificar se os *frameworks* construídos para os ambientes tradicionais, *small data*, podem ser utilizados em ambiente de *big data* ou se existe necessidade de adaptações. O presente trabalho realizou uma revisão sistemática da literatura para levantar componentes de *frameworks* de governança de *big data* e identificar se existe divergência em relação aos *frameworks* de governança de dados. Inicialmente, 515 artigos foram encontrados no processo de revisão sistemática da literatura e após a aplicação dos critérios de inclusão e exclusão, 52 trabalhos foram selecionados para a leitura completa, dos quais, 8 apresentaram componentes de *framework* de governança de *big data*. Os componentes levantados foram comparados com os componentes padrão de um *framework* de referência. Ao final concluímos que *frameworks* de governança de dados tradicionais podem ser utilizados em *big data*, desde que alguns componentes padrão sejam ajustados por conta das especificidades do mesmo, como resultado estendemos o *framework* proposto por Jang and Kim (2021) com componentes relacionados a *big data* e governança de dados. A estrutura final é mais completa e adequada para implementar uma governança de dados em ambiente de *big data*.

Palavras-Chave – *Big Data*, Governança de Dados, Frameworks de Governança de Dados, Componentes de Frameworks.

ABSTRACT

The use of data to support decision-making in companies is already a consolidated practice. In recent years, the exponential increase in data generation in different volumes, sources and speeds has driven the phenomenon of big data. However, the abundance of data itself is not enough; its value is directly linked to the reliability and quality it offers. This is where data governance comes in, with a series of artifacts that help control, protect and deliver value through data assets. However, it is not something that is simple to implement in companies and for this reason, researchers and professionals in the data area have developed frameworks to help with this challenge. Frameworks help structure complex ideas and concepts to facilitate their understanding and implementation. By definition, big data has distinct characteristics from traditional data stored in relational databases, small data. For this reason, it is necessary to verify whether frameworks built for traditional environments, small data, can be used in a big data environment or whether there is a need for adaptations. This work carried out a systematic literature review to identify components of big data governance frameworks and identify whether there is divergence in relation to data governance frameworks. Initially, 515 papers were found in the systematic literature review process and after applying the inclusion and exclusion criteria, 52 papers were selected for full reading, of which 8 presented big data governance framework components. The components identified were compared with the standard components of a reference framework. In the end, we concluded that traditional data governance frameworks can be used in big data, as long as some standard components are adjusted due to its specificities, as a result we extended the framework proposed by Jang and Kim (2021) with componente related to big data and data governance. The final framework is more completed and apporopriated to implement a data governance in big data environment.

Keywords – *Big Data*, Data Governance, Data Governance Framework, Frameworks Components.

LISTA DE FIGURAS

1	Metodologia utilizada	4
2	Organização dos componentes padrão	9
3	Quantidade de artigos em cada fase do protocolo de revisão	13
4	Número de artigos por ano de publicação	16
5	Quantidade de artigos por área de aplicação	16
6	Distribuição de artigos por tema	18
7	Distribuição de artigos com um FGBD por área de aplicação	18
8	Quantidade de artigos por componente padrão	21
9	Quantidade de componentes padrão por artigo	22
10	Incorporação dos componentes de um FGBD aos componentes padrão . . .	27

LISTA DE TABELAS

1	Trabalhos de referência utilizados por Jang and Kim (2021)	7
2	Componentes padrão definidos por Jang and Kim (2021)	8
3	Critérios de Inclusão e Exclusão	11
4	Artigos selecionados para leitura e extração	13
5	Quantidade de artigos por tema	17
6	Componentes padrão	20
7	Mapeamento de componentes padrão	20

LISTA DE SIGLAS

DEMATEL	<i>Decision Making Trial and Evaluation Laboratory Model</i> (Modelo de Laboratório de Avaliação e Teste de Tomada de Decisão)
FGBD	<i>Framework</i> de Governança de <i>big data</i>
FGD	<i>Framework</i> de Governança de Dados
FAIR	<i>Findable, Accessible, Interoperable and Reusable</i> (Localizável, Acessível, Interoperável e Reutilizável)
GD	Governança de Dados
GDPR	General Data Protection Regulation (Regulação Geral de Proteção de Dados)
IoT	<i>Internet of Things</i> (Internet das Coisas)
LGPD	Lei Geral de Proteção de Dados
RSL	Revisão Sistemática da Literatura

SUMÁRIO

1	Introdução	1
1.1	Objetivo	3
1.2	Contribuição	3
1.3	Metodologia	3
2	Contextualização Teórica	5
2.1	Trabalhos Relacionados	5
2.2	Componentes de um <i>Framework</i> de Governança de Dados	6
3	Revisão Sistemática da Literatura	10
3.1	Questão de Pesquisa	10
3.2	CrITÉrios de Inclusão e Exclusão	10
3.3	Protocolo de Revisão	11
3.3.1	Pesquisa na Literatura	12
3.3.2	Triagem Prática, Revisão de Qualidade e Leitura e Extração	12
3.4	Análise Quantitativa	15
3.4.1	Artigos que Não Apresentam um FGBD	17
3.4.2	Artigos que Apresentam um FGBD	17
4	Mapeamento de Componentes de Governança de Dados	20
4.1	Componentes de <i>Framework</i> de Governança de <i>Big Data</i>	23
4.1.1	Governança de Leis e Regulações	23
4.1.2	Integração	23
4.1.3	Desenvolvimento de Produtos de Dados	24
4.1.4	Pessoas	24

4.1.5	Princípios Guia	24
4.1.6	Gerenciamento de Processamento Distribuído	25
4.1.7	Comunicação	25
4.2	Discussão de Resultados	25
4.2.1	Respostas das Questões de Pesquisa	27
5	Conclusão	28
5.1	Limitações e Trabalhos futuros	29
	Referências Bibliográficas	30
	Apêndice A – <i>Strings</i> de Busca	38

1 INTRODUÇÃO

Uma grande quantidade de dados gerados em formatos e fontes distintas têm sido usada para gerar valor para as empresas, principalmente ajudando nos processos de tomada de decisão. Por esta razão, as empresas estão cada vez mais dependentes desses dados, sendo necessário garantir sua devida qualidade (Nadal et al., 2022).

Um programa de governança de dados é um dos pilares que auxilia as empresas a conseguirem sucesso na jornada da qualidade de dados. Governança de dados (GD) é uma área de conhecimento, uma função de grande importância no gerenciamento de dados, pois ela normatiza e orquestra as demais funções de dados, tais como a modelagem e projetos de dados, a gestão de metadados, a segurança de dados, dentre outras (Ekundayo et al., 2023).

Governança de dados desenvolve, executa e supervisiona planos, políticas, programas e práticas que controlam, protegem, entregam e melhoram o valor dos ativos de dados, alinhado com a estratégia da empresa. Além disso, garante que os dados estão de acordo com regulações locais e internacionais, para preservar a validade dos mesmos (Ekundayo et al., 2023). Devido a sua relevância, tem se tornado um item obrigatório em empresas que trabalham com dados (Alsaad, 2023).

Para ajudar as empresas a implementarem governança de dados, pesquisadores e profissionais de dados desenvolveram *frameworks* específicos que orientam na definição, estruturação e implantação da governança de dados (Bento et al., 2022).

Os *frameworks* de governança de dados definem, dentre outras coisas, componentes para melhor estruturação, entendimento e implementação dos conceitos, como é mostrado no trabalho de Jang and Kim (2021). Nesse trabalho, os autores listaram os componentes de alguns *frameworks* de governança de dados, sumarizando todos os componentes dos *frameworks* estudados em um conjunto final de componentes padrão utilizando uma técnica de mapeamento de relações.

Alguns trabalhos mencionam a importância de se ter uma boa governança de dados

nas empresas que pretendem ser data driven, usando os dados como ativo de valor (Alsaad, 2023; Bento et al., 2022). Todavia, não é algo fácil e simples de se fazer, o trabalho de Bassi and Souza (2023) cita os principais desafios na implementação de governança de dados, analisando vários estudos de caso sobre o tema. Zhang et al. (2022) afirmam que governança de dados é um projeto complexo e de longo prazo que precisa de várias ações para que o seus objetivos sejam alcançados.

Sivarajah et al. (2017) aponta a governança de dados como um dos desafios do *big data*, sendo considerada como um dos principais meios de assegurar qualidade de dados, mantendo o seu valor, dando suporte a decisões de negócio e operações. Al-Badi et al. (2018a) alerta para os desafios relacionados ao gerenciamento, processamento e segurança de big data. Zeng et al. (2022) comenta que o custo de governança pode ser um problema devido à característica heterogênea e complexa do big data.

O trabalho de Kim and Cho (2017) lista uma série de desafios e características que *big data* possui que devem ser refletidas na governança de *big data*. Eles ressaltam os problemas e as diferenças que os *frameworks* de governança de dados devem considerar quando aplicados em *big data*. Os autores afirmam que a governança de dados tradicional foca no processamento de dados estruturados, sendo que a governança de *big data* deveria considerar a natureza dos dados não estruturados. Dentre outros pontos, eles levantam a divergência na área de qualidade de dados, em que atributos como tempo, confiabilidade e significância devem receber uma atenção maior devido à criticidade do uso de *big data* para a tomada de decisões, em que um erro pode ter um grande impacto no negócio. Os autores também chamam a atenção para questões relacionadas a proteção de dados pessoais, devido a casos de vazamento de dados que geram um impacto negativo nas empresas.

Após uma revisão de literatura, Obatolu (2022) comenta que a pesquisa científica na área de governança de dados é limitada, fragmentada e possui lacunas de definição. O autor adiciona que poucos trabalhos abordam governança de *big data* e afirma que essa área de pesquisa está em fases iniciais. Segundo o seu levantamento, *frameworks* de governança de dados tem limitações quando consideram a complexidade de *big data*.

Para entender melhor a diferença entre *frameworks* de governança de dados tradicionais e *framework* de governança de *big data* (FGBD), o presente trabalho realizou uma revisão sistemática da literatura (RSL) para levantar os componentes dos *frameworks* de governança de *big data*, comparando-os com componentes padrão do *frameworks* de governança de dados apresentados por Jang and Kim (2021). Foi feita uma proposta

de alteração ao *framework* de referência, além de realizar uma análise quantitativa dos trabalhos encontrados.

1.1 Objetivo

O presente trabalho tem como objetivo avaliar se existe alguma diferença significativa entre os componentes de *frameworks* de governança de *big data* e *frameworks* de governança de dados tradicionais (*small data*), avaliando a possibilidade de estender um *framework* de governança de dados tradicional para suportar essas diferenças.

1.2 Contribuição

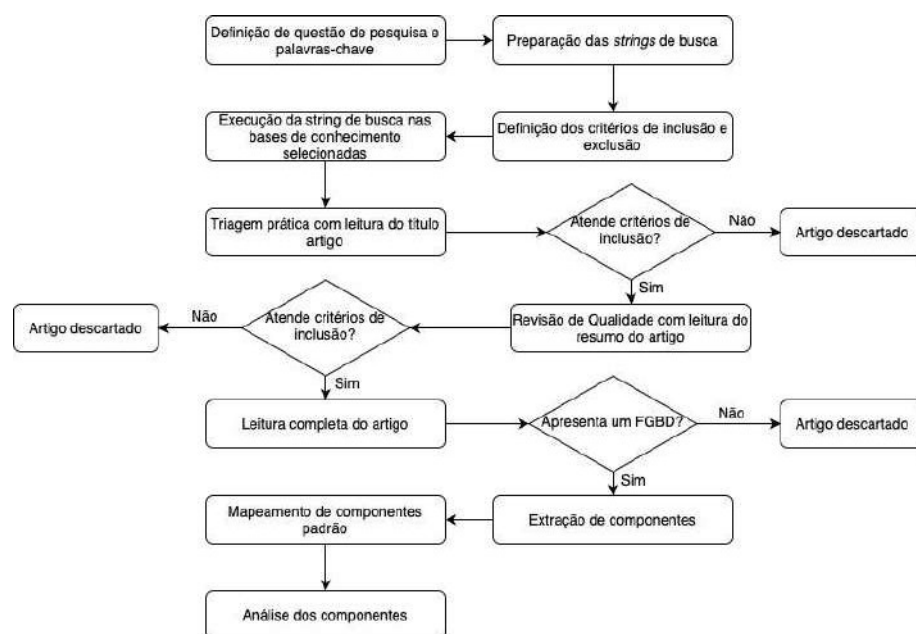
Este trabalho apresenta contribuições significativas tanto para a teoria quanto para a prática no campo da governança de dados. Por meio de uma análise comparativa entre *frameworks* tradicionais e os aplicados ao contexto de *big data*, a pesquisa destaca as adaptações necessárias para lidar com as especificidades de dados massivos. Essas adaptações foram utilizadas para realizar uma extensão do *framework* de governança de dados proposto por Jang and Kim (2021), para que ele pudesse ser utilizado com *big data*. Além disso, o mapeamento sistemático dos componentes essenciais fornece uma base estruturada para estudos futuros e orientações práticas para empresas que desejam adaptar *frameworks* existentes, reduzindo esforços no desenvolvimento de novos modelos.

1.3 Metodologia

Foi feita uma revisão sistemática da literatura (RSL) para realizar o levantamento dos componentes de *frameworks* de governança de *big data*, usando a metodologia proposta por Kitchenham (2004) que define as fases de formulação das questões de pesquisa, definição dos critérios de inclusão e exclusão de trabalhos, condução do protocolo de revisão e demonstração dos resultados, esses pontos serão detalhados no capítulo 3. A partir dos resultados da RSL foi possível analisar o *framework* proposto por Jang and Kim (2021) e estendê-lo.

A Figura 1 exemplifica a metodologia usada no presente trabalho.

Figura 1: Metodologia utilizada



Fonte: Elaborado pelo autor

2 CONTEXTUALIZAÇÃO TEÓRICA

2.1 Trabalhos Relacionados

Diversos trabalhos utilizam uma revisão sistemática da literatura para estudar algum tema de governança de dados, a exemplo de Alsaad (2023), em que os autores fizeram uma RSL para levantar as pesquisas de estado da arte em *frameworks* de governança de dados governamentais. Os autores mencionam que a diferença entre esse tipo de *framework* e os relacionados ao mundo corporativo está no objetivo final. Enquanto que os frameworks com foco em empresas do mundo corporativo foca em trazer valor para o negócio, os frameworks de governança de dados governamentais centra-se em trazer valor para as pessoas. No levantamento feito por Alsaad (2023), os autores mostram a existência de trabalhos que fazem ajustes em *frameworks* tradicionais para se adequar aos requisitos de *big data*, algo que também foi feito nesta pesquisa, mas no presente trabalho foi ajustando um *framework* de componentes padrão. O trabalho de Alsaad (2023) também ajudou a definir os critérios de inclusão e exclusão do presente trabalho.

Abraham et al. (2019) utilizaram uma revisão de literatura para criar uma visão holística sobre governança de dados, levantando quais são os blocos de construção de governança de dados e quais são as lacunas de conhecimento sobre o tema. Os autores também propuseram um *framework* conceitual de governança de dados, consideraram a distinção entre dados tradicionais e *big data* em um dos seus componentes, porém, não se aprofundam no tema. Abraham et al. (2019) recomendaram a execução de trabalhos futuros em *big data* no escopo de governança de dados como um possível tópico de pesquisa devido aos desafios que o mesmo traz, como qualidade e privacidade de dados. Diferente do trabalho de Abraham et al. (2019) o presente trabalho focou em trabalhos com *frameworks* de governança de *big data*.

Um RSL também foi realizada no trabalho de Bassi and Souza (2023) para levantar os principais desafios ocorridos na implementação de projetos de governança de dados nas empresas. Considerando trabalhos que apresentavam estudo de casos, os autores chega-

ram em 6 desafios mais importantes, que podem ser considerados pelas organizações com o objetivo de serem mais efetivas na implementação dos projetos de governança de dados, tentando mitigar o impacto desses desafios. Os autores também fazem uma análise quantitativa dos trabalhos levantados, servindo como base de comparação da análise realizada no presente trabalho.

O Trabalho de Rasheed et al. (2021) comenta da importância da governança de dados no contexto de *big data* para garantir a consistência, confiabilidade e correção dos dados. Os autores afirmam que a governança de *big data* traz benefícios como a transparência e segurança dos dados, mas também alguns desafios como a centralização dos dados, integração e conformidade com regulações. Os Autores fazem uma comparação entre *frameworks* de *big data* e o *framework* padrão da ISO 8000 de governança de dados, procurando lacunas de implementação. Por fim, propõem um *framework* de governança de big data com 8 componentes. O presente trabalho compara *frameworks* de governança de big data com um *framework* de componentes padrão proposto por Jang and Kim (2021), elaborado a partir de uma síntese dos componentes presentes em diversos *frameworks* existentes.

2.2 Componentes de um *Framework* de Governança de Dados

O estudo de Jang and Kim (2021) foi adotado como referência no presente trabalho para os componentes de *frameworks* de governança de dados. Os autores realizaram uma síntese dos componentes padrão encontrados em diversos *frameworks* de governança de dados, destacando que, embora existam várias pesquisas com soluções para governança de dados, essas abordagens frequentemente apresentam visões e escopos divergentes. Reconhecendo essa fragmentação, os autores propuseram um modelo integrado de padrões, identificando os componentes essenciais para o desenvolvimento de um *framework* de governança de dados mais unificado e consistente.

Jang and Kim (2021) utilizaram um método chamado *Decision Making Trial and Evaluation Laboratory Model* (DEMATEL) e análise de conteúdo para identificar e analisar a influência existente entre componentes de *frameworks* de governança de dados com o objetivo de definir uma integração padrão entre os componentes. O método DEMATEL é utilizado para resolver problemas estruturais que são associados a problemas do mundo real e para visualizar o valor de influência mútua por meio de um mapa de impacto-relacionamento. Os dois métodos são demonstrados no trabalho de Jang and Kim (2021).

No trabalho de Jang and Kim (2021) os autores consideram 10 *frameworks* de governança de dados, sendo 5 propostos por instituições, como a DAMA¹ e DGI², e os outros 5 propostos por pesquisadores. A Tabela 1 lista todos os 10 trabalhos usados por Jang and Kim (2021).

Tabela 1: Trabalhos de referência utilizados por Jang and Kim (2021)

Referência
Otto (2011)
DAMA-DMBOK (2009)
Thomas (2006)
IBM (2007)
Kelle (2015)
Khatri and Brown (2010)
Soup (2011)
SAS (2014)
TTA (2011)
Weber et al. (2009)

Para chegar nos componentes do *framework* de governança de dados os autores dividiram em duas etapas, cada um com uma série de tarefas. Na etapa de coleta de dados e pré-processamento, foi realizada uma análise da literatura para coletar os componentes de governança de dados. Em seguida, os dados passaram por um pré-processamento, utilizando a técnica de análise de conteúdo. Essa técnica permitiu corrigir erros e agrupar adequadamente os componentes por meio de uma padronização consensual, envolvendo pesquisadores e colaboradores. Como resultado dessa etapa, foram identificados 12 componentes padrão de governança de dados, sendo eles: Estratégia, Políticas, Metodologia, Métricas, Arquitetura de Sistema, Ciclo de Vida do Dados, Monitoramento de Dados, Gerenciamento de Qualidade, Valor de Qualidade, Segurança de Dados, Papéis e Responsabilidades e Estrutura Organizacional. Estes componentes serviram de base para a etapa seguinte. A Tabela 2 descreve cada um dos componentes identificados no trabalho de Jang and Kim (2021).

Na etapa de questionamento de dados e análise de resultados as tarefas foram questionário de especialista do DEMATEL e análise de causa efeito. Como resultado foi obtido o agrupamento dos 12 componentes padrão em 3 grupos: Conformidade de Dados, Gerenciamento de Dados e Organização de Dados.

¹<https://www.dama.org/cpages/home>

²<https://datagovernance.com>

Tabela 2: Componentes padrão definidos por Jang and Kim (2021)

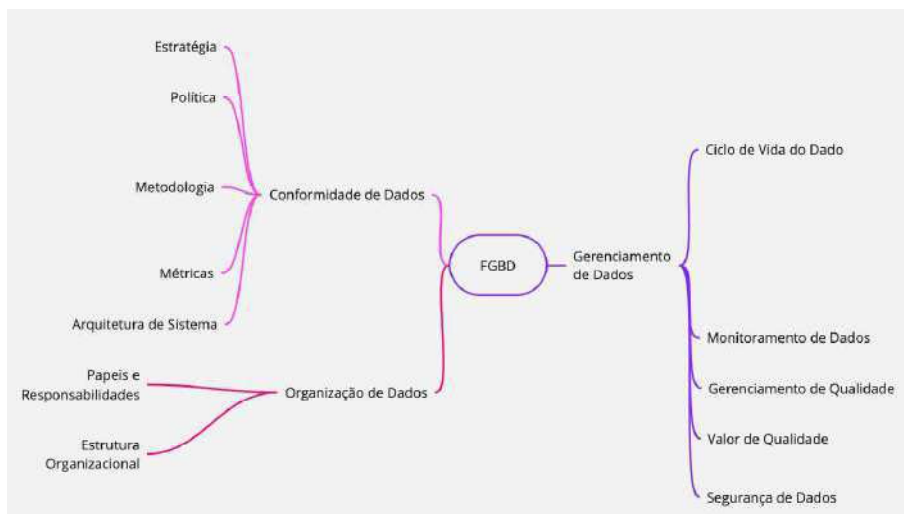
Componentes Padrão	Definição	Componentes de GD de Pesquisas Anteriores
Estratégia (C1)	Estratégia e visão para sistematização dos dados da empresa.	Guia corporativo, criação de valor, missão e visão.
Política (C2)	Políticas para alcançar a visão para o gerenciamento de dados.	Políticas, planejamento de gerenciamento de dados, controles.
Metodologia (C3)	Métodos, processos e meios de formulação de políticas.	Governança de Dados e Métodos, <i>framework</i> .
Métricas (C4)	Indicadores para avaliar se os objetivos foram alcançados.	Padrões, índice de avaliação, objetivos, medidas de métricas de governança.
Arquitetura de Sistema (C5)	Arquitetura de dados e sistema, soluções e ferramentas.	Arquitetura de sistema, soluções, classificação e metadados, ferramentas.
Ciclo de Vida do Dado (C6)	Gerenciamento de Ciclo de vida geral, incluindo coleta, armazenamento, processamento, utilização e disponibilização de dados.	Ciclo de vida dos dados, gerenciamento do ciclo de vida dos dados, gerenciamento de ciclo de vida da informação, políticas de dado e processos.
Monitoramento de dados (C7)	Registro de Dados, auditoria e relatório de atividades.	Informação de auditoria, registro e relatórios, supervisão e controle.
Gerenciamento de Qualidade (C8)	Um conjunto de atividades de gerenciamento para melhorar a qualidade dos dados.	Qualidade dos dados, gerenciamento de qualidade dos dados, regras e definições de dados de negócio.
Valor de Qualidade (C9)	Um conjunto de Atividades para melhorar a qualidade do valor dos dados.	Metadados, Disponibilidade de dados, consistência dos dados, auditabilidade dos dados, gerenciamento de metadados.
Segurança de Dados (C10)	Um conjunto de atividades para manter a segurança dos dados.	Acesso aos dados, segurança dos dados, segurança e privacidade dos da informação.
Papéis e Responsabilidades (C11)	Define papéis e responsabilidades de pessoas na organização dos dados.	Funções e tarefas de administração de dados, administração, administradores de dados, responsabilidades, usuário, participante, fornecedor.
Estrutura organizacional (C12)	Atividades Organizacionais para organização dos dados.	Organização, estrutura organizacional e conscientização, partes interessadas nos dados.

Fonte: Jang and Kim (2021) - Traduzido pelo autor

Esta técnica utiliza algumas matrizes e equações do DEMATEL para chegar em um mapa de impacto-relacionamento dos componentes, usado para agrupá-los. O grupo Conformidade de Dados é composto por componentes de causa, ou seja, que impactam diretamente outros componentes, seja dentro ou fora do grupo. O grupo Gerenciamento de Dados é constituído por componentes que sofrem influência dos componentes do primeiro grupo, Conformidade de Dados, mas só exercem influência nos componentes que estão dentro do mesmo grupo. Já o terceiro grupo, Organização de Dados, possui componentes que são influenciados pelos componentes do grupo Conformidade de Dados, mas, diferente do grupo anterior, não existe influências entre os componente do mesmo grupo.

Os componentes Estratégia, Políticas, Metodologia e Métricas formam o grupo Conformidade de Dados. Já os componentes Arquitetura de Sistema, Ciclo de Vida do Dados, Monitoramento de Dados, Gerenciamento de Qualidade, Valor de Qualidade, Segurança de Dados fazem parte do grupo Gerenciamento de Dados. Papéis e Responsabilidades junto com Estrutura Organizacional compõem o grupo Organização dos Dados. A Figura 2 apresenta um mapa mental com a organização dos componentes do *framework* proposto por Jang and Kim (2021).

Figura 2: Organização dos componentes padrão



Fonte: Elaborado pelo autor

3 REVISÃO SISTEMÁTICA DA LITERATURA

3.1 Questão de Pesquisa

A definição das questões de pesquisa é um dos pontos principais de uma revisão bibliográfica de literatura (Kitchenham, 2004). Com o objetivo de identificar se *frameworks* de governança de *big data* divergem dos *frameworks* de governança de dados para *small data*, foram elaboradas duas questões de pesquisa:

- Quais são os componentes específicos de um *framework* de governança de *big data*?
- Como um *framework* de governança de dados tradicional pode ser estendido para se adequar ao contexto de *big data*?

3.2 Critérios de Inclusão e Exclusão

Após a definição das questões de pesquisa, definiu-se os critérios de inclusão e exclusão para que os estudos primários tenham evidências diretas sobre as questões de pesquisa (Kitchenham, 2004). O trabalho de Alsaad (2023) foi usado como referência para definição destes critérios.

Para responder às questões de pesquisa, o principal critério de inclusão/exclusão foi a relação direta do trabalho com *frameworks* de governança de dados (FGDs). Apenas estudos cujo foco central fosse FGDs foram incluídos, enquanto aqueles que tratavam do tema de forma periférica ou não relacionada foram excluídos. Além disso, foram considerados critérios de credibilidade, como revisão por pares e publicação em periódicos ou conferências acadêmicas. Quanto ao acesso, os trabalhos precisavam estar disponíveis online, serem escritos em inglês e de acesso gratuito por meio do e-mail institucional da Universidade de São Paulo (USP). Por fim, optou-se por um recorte temporal de 10 anos, visando ampliar a quantidade de estudos relevantes sobre o tema.

Os critérios de inclusão e exclusão estão presentes na Tabela 3.

Tabela 3: Critérios de Inclusão e Exclusão

Critério de Inclusão	Critério de Exclusão
Trabalhos sobre FGDs	Trabalhos em que FGDs não é o tema central
Revisado por pares	Artigos pagos
Acessíveis on-line	
Escritos em Inglês	
Publicados entre 2014 e 2024	
Revistas ou periódicos de conferências científicas	

Fonte: Elaborado pelo autor

3.3 Protocolo de Revisão

Para a condução da pesquisa foi adotado um protocolo de revisão com uma sequência de etapas baseadas no trabalho de Mao et al. (2022), sendo elas:

1. Pesquisa na Literatura: Utilizando os termos de busca em bases de conhecimento e critérios de inclusão definidos;
2. Triagem Prática: Pré-seleção dos artigos encontrados fazendo uma leitura do título aplicando os critérios de exclusão;
3. Revisão de Qualidade: Leitura dos resumos dos artigos que passaram da etapa anterior, também aplicando os critérios de exclusão;
4. Leitura e Extração: Leitura completa do artigo, identificando os componentes do *framework* de governança de *big data* apresentado;
5. Análise e Síntese: Análise dos componentes encontrados e levantamento de características.

Após a execução do protocolo de pesquisa da RSL, foi feito um cruzamento dos componentes encontrados com os componentes padrão definidos no trabalho de Jang and Kim (2021). Esse processo permitiu mapear os componentes com relação direta e identificar aqueles especificamente relacionados à governança de *big data*, proporcionando uma visão mais detalhada das conexões e diferenças entre os *frameworks* analisados.

As próximas subseções detalham a execução das etapas do protocolo de revisão.

3.3.1 Pesquisa na Literatura

Antes da condução do estudo foram definidas as palavras-chaves relacionadas as perguntas de pesquisa. Para selecionar trabalhos relacionados a frameworks de governança de *big data* foram escolhidas os seguintes termos: "*Big Data*", "*Big Data Governance*", "*Big Data Governance Framework*", "*Data Governance*", "*Data Governance Framework*" e "*Framework Components*". Foram também definidos sinônimos para as palavras "*Governance*" e "*Components*", sendo eles "*Management*" e "*Elements*", respectivamente.

Com base nessas palavras-chaves a seguinte *string* de busca base foi criada: ("*Big Data*") AND ("*Governance*" OR "*Management*") AND ("*Framework*" OR ("*Framework Element*" OR "*Framework Component*")). O operador lógico *OR* foi utilizado para adicionar os termos de sinônimos e para tentar obter trabalhos que explicitamente citam elementos ou componentes de *frameworks*.

As bases de conhecimento definidas para a pesquisa foram Springer, IEEE, Science Direct e Scopus. Todas são bases confiáveis, de bom respaldo científico e que são acessíveis por meio do *e-mail* institucional da USP. Para reduzir a quantidade de material encontrado e na tentativa de aumentar a relevância dos artigos resultantes, foi feita uma restrição de busca, considerando apenas as palavras contidas no título, assunto e palavras-chaves dos artigos.

Devido às características específicas das bases de conhecimento utilizadas, foram criadas *strings* de buscas distintas para cada uma das bases. As *strings* de busca usadas em cada base de conhecimento estão presentes no Apêndice A.

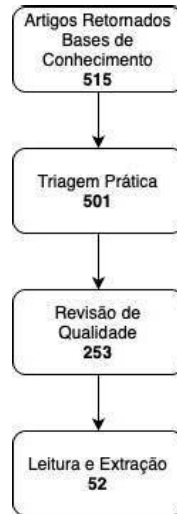
A aplicação da *string* de busca nas bases de conhecimento retornou 515 trabalhos no total, sendo 14 duplicados. Dando continuidade ao protocolo de pesquisa foram executadas as fases de triagem prática, revisão de qualidade e leitura e extração, detalhadas na próxima seção.

3.3.2 Triagem Prática, Revisão de Qualidade e Leitura e Extração

Na fase de triagem prática foi feita a leitura dos títulos dos artigos. Nela foram removidos 248 artigos por não estarem diretamente associados ao tema de *framework* de governança de dados. Na etapa de revisão de qualidade foi feita a leitura do resumo, 201 artigos foram removidos por não terem governança de dados como tema principal. Após essas duas fases restaram 52 artigos para a leitura completa, destes, apenas 8 apresentaram

componentes de *framework* de governança de *big data*. A Figura 3 exemplifica e detalha a quantidade de artigos em cada fase do protocolo de revisão.

Figura 3: Quantidade de artigos em cada fase do protocolo de revisão



Fonte: Elaborado pelo autor

A Tabela 4 lista os 52 trabalhos que passaram pelos filtros de título e resumo, detalhando o ano de publicação, área de aplicação e se o mesmo apresenta um FGBD.

Tabela 4: Artigos selecionados para leitura e extração

Id	Referência	Área Aplicação	FGBD
A1	Shankarnarayan and Ramakrishna (2020)	Agricultura	Não
A2	Salierno et al. (2020)	<i>Big Data</i>	Não
A3	Yakobi et al. (2020)	Governamental	Não
A4	Chauhan et al. (2021)	Governamental	Não
A5	Al-Shdifat and Emmanouilidis (2018)	Educação	Não
A6	Poger et al. (2023)	Saúde	Não
A7	Salierno et al. (2020)	Infraestrutura	Não
A8	Marelli et al. (2021)	Saúde	Não
A9	Wen et al. (2024)	Agricultura	Não
A10	Bani-Salameh et al. (2021)	Saúde	Não
A11	Janković et al. (2022)	Indústria	Sim
A12	Naghieb et al. (2023)	IoT	Não
A13	Kaufmann (2019)	IoT	Sim
Continua na próxima página.			

Tabela 4 – Continuação da página anterior

Id	Referência	Área Aplicação	FGBD
A14	Huang et al. (2018)	Agricultura	Não
A15	Bashir et al. (2020)	Agricultura	Não
A16	Wolfert et al. (2017)	Biologia	Não
A17	Raghupathi and Raghupathi (2014)	<i>e-commerce</i>	Não
A18	Kamal and Dave (2019)	Educação	Não
A18	Hussain et al. (2023)	Energia	Não
A20	Ghorbani et al. (2022)	Energia	Não
A21	Liu et al. (2022)	Meio-ambiente	Não
A22	Bhuvana et al. (2016)	Financeiro	Não
A23	Yang et al. (2022)	Financeiro	Não
A24	Irani et al. (2018)	Alimentar	Não
A25	Yukhno (2024)	Governamental	Não
A26	Evans et al. (2020)	Saúde	Não
A27	Noor et al. (2015)	Saúde	Não
A28	Campos et al. (2016)	Indústria	Não
A29	Stergiou and Psannis (2022)	IoT	Não
A30	Ilie-Zudor et al. (2014)	Logística	Não
A31	Yousfi et al. (2022)	Ciência	Não
A32	Faroukhi et al. (2020a)	<i>Big Data</i>	Não
A33	Faroukhi et al. (2020b)	Segurança	Não
A34	Al-Badi et al. (2018a)	<i>Big Data</i>	Sim
A35	Wang et al. (2018)	Energia	Sim
A36	Zeng et al. (2022)	Energia	Sim
A37	Li et al. (2019)	Saúde	Sim
A38	Serrano and Zorrilla (2019)	Indústria	Sim
A39	Yakobi et al. (2020)	IoT	Sim
A40	Chao et al. (2023)	<i>Big Data</i>	Não
A41	Zhang et al. (2024)	Governamental	Não
A42	Arundel et al. (2023)	Governamental	Não
A43	Kalkman et al. (2019)	IoT	Não
A44	Yang et al. (2014)	Ciência	Não
A45	Karras et al. (2023)	IoT	Não
A46	Shah et al. (2021)	Saúde	Não
Continua na próxima página.			

Tabela 4 – Continuação da página anterior

Id	Referência	Área Aplicação	FGBD
A47	Akter and Wamba (2016)	IoT	Não
A48	Ng et al. (2017)	IoT	Não
A49	Holom et al. (2020)	IoT	Não
A50	Wang et al. (2024)	Governamental	Não
A51	Brasil (2018)	Indústria	Não
A52	Benítez-Hidalgo et al. (2021)	Ciência	Não

Fonte: Elaborado pelo autor

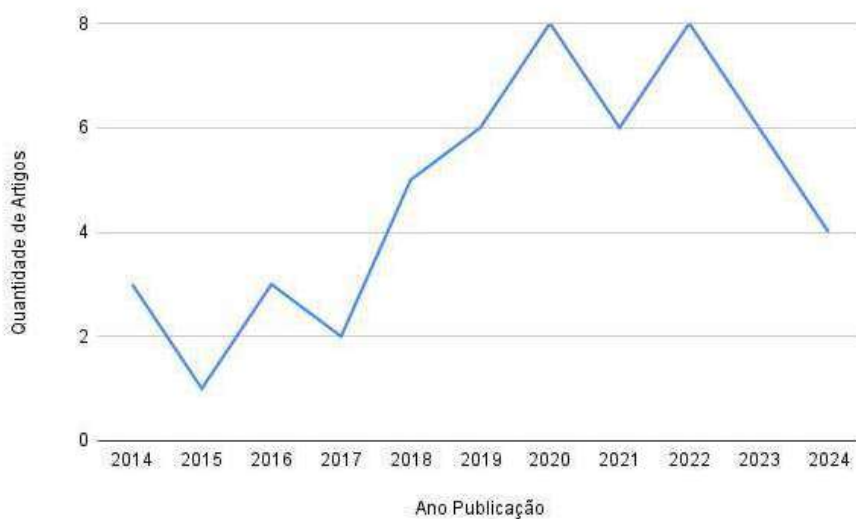
3.4 Análise Quantitativa

Foi feita uma análise quantitativa dos 52 artigos que passaram para a fase final de leitura completa do artigo. A Figura 4 apresenta um gráfico com a quantidade de artigos por ano de publicação. Observamos o aumento das publicações ao longo dos anos quase que de forma contínua desde 2017, com exceções de 2021, 2023 e 2024 que apresentaram um número de publicação menor do que o ano anterior. A queda de 2021 pode ser justificada pela pandemia de Covid 19 que impactou a publicação de trabalhos (Sohrabi et al., 2021). Tivemos um recorde de publicações em 2020 e 2022, com 8 trabalhos cada. Não foi encontrada evidências na literatura do que pode ter causado a queda do números de publicações em 2023. Quanto a quantidade de publicações em 2024, não se pode afirmar que houve uma queda no número de publicações, pois a pesquisa bibliográfica do presente trabalho foi finalizada no primeiro semestre deste ano.

A Figura 5 apresenta um gráfico com a quantidade de artigos por área de aplicação. Existe uma diversidade de áreas, como Ciência, Financeira, Segurança, Logística, entre outras, que representam 34.62% (18/52) dos artigos. Essa diversidade de áreas de aplicação pode ser vista também na Tabela 2.

Fazendo o percentual de distribuição das áreas de aplicação observamos uma concentração de trabalhos nas áreas de Saúde (15,38%, 8/52), IoT (15,38%, 8/52), Governamental (11,54%, 6/52), Indústria (7,69%, 4/52), Energia (7,69%, 4/52) e Agricultura (7,69%, 4/5). Essas concentrações mostram um forte interesse em temas de *big data* em diversas áreas, evidenciando que *big data* pode ser aplicada em vários segmentos. Vale observar o interesse das áreas Governamental, Saúde, Indústria e Agricultura pelo tema,

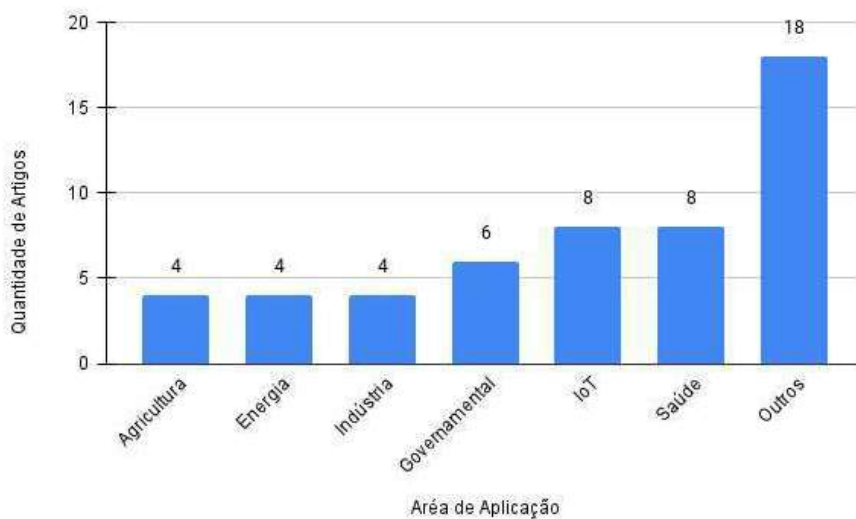
Figura 4: Número de artigos por ano de publicação



Fonte: Elaborado pelo autor

que corrobora com o trabalho de Bassi and Souza (2023), sobre desafios de governança de dados, que apontam essas áreas como áreas com mais estudos de caso na implementação de governança de dados.

Figura 5: Quantidade de artigos por área de aplicação



Fonte: Elaborado pelo autor

3.4.1 Artigos que Não Apresentam um FGBD

Foi realizada uma análise quantitativa dos 46 artigos que não apresentaram componentes de *framework* de governança de *big data*. Fazendo um levantamento dos temas de *big data* para identificar quais outros temas de *big data* estão sendo foco de pesquisa. Observa-se a presença de uma diversidade de temas listados na Tabela 5.

Tabela 5: Quantidade de artigos por tema

Tema	Quantidade de Artigos
Uso de <i>Big Data</i>	18
Análise de <i>Big Data</i>	4
Arquitetura de <i>Big Data</i>	3
<i>Framework</i> de <i>Big Data</i>	2
Ciclo de vida de <i>Big Data</i>	2
Cadeia de valor de <i>Big Data</i>	2
Plataforma de <i>software</i>	1
Movimentação de <i>Big Data</i>	1
IoT	1
Gerenciamento de qualidade	1
Gerenciamento de plataforma de dados	1
Gerenciamento de metadados	1
Gerenciamento de <i>Big Data</i>	1
GDPR	1
Desafios de <i>Big Data</i>	1
Armazenamento de dados	1
Aprendizado federado	1

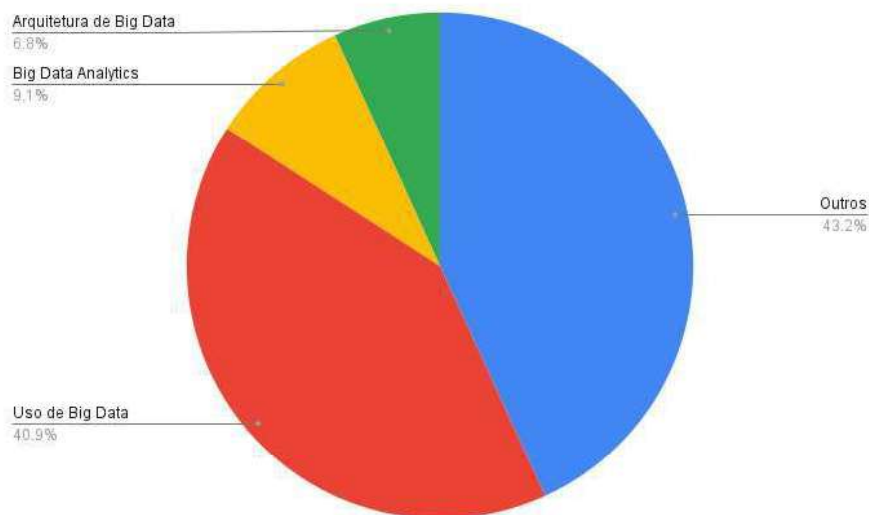
Fonte: Elaborado pelo autor

A Figura 6 mostra um gráfico com a distribuição de temas. Observa-se que os trabalhos que abordam o uso de *big data*, análise de *big data* e arquitetura de *big data* representam 40,9%, 9,1% e 6,8% respectivamente. Os outros 43,2% dos trabalhos estão distribuídos em 14 temas diferentes, como exemplo, gerenciamento de qualidade, gerenciamento de metadados, compartilhamento de dados. Artigos sobre o uso de *big data* destacam como *big data* pode ser utilizado para ajudar organizações a tomarem melhores decisões.

3.4.2 Artigos que Apresentam um FGBD

Fazendo uma análise quantitativa dos 8 artigos que apresentaram componentes de governança de *big data*, observamos na Figura 7 que as áreas de aplicação ficaram concentradas em Indústria, Internet das Coisas (IoT), Energia, Saúde e *Big Data*. Este resultado que também corrobora com o trabalho de Bassi and Souza (2023). Indústria, Energia,

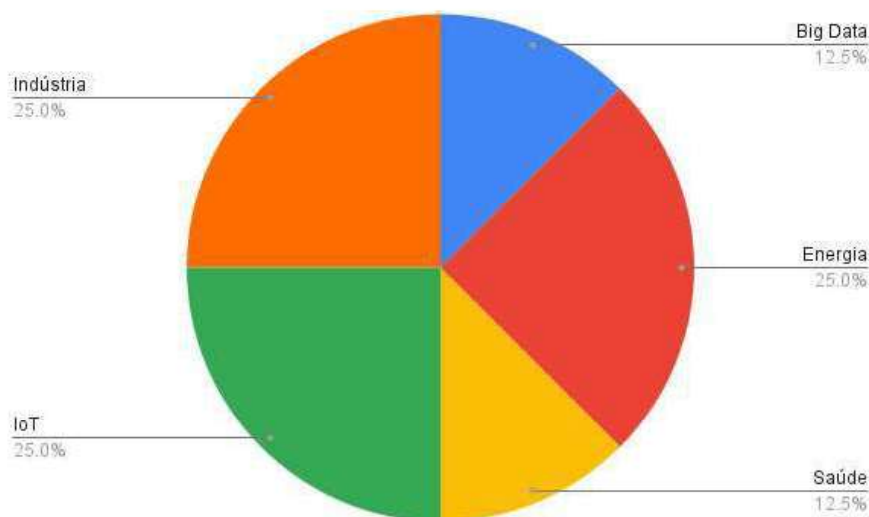
Figura 6: Distribuição de artigos por tema



Fonte: Elaborado pelo autor

Saúde são setores regulados, o que justifica o interesse dessas áreas em governança de *big data*. IoT é uma das áreas precursoras do *big data* e está associada à área da Indústria.

Figura 7: Distribuição de artigos com um FGBD por área de aplicação



Fonte: Elaborado pelo autor

Observou-se que 25% dos trabalhos realizaram uma RSL antes de propor um *framework* de governança de *big data*. O único trabalho que referenciou os 12 componentes padrão do framework proposto por Jang and Kim (2021), fez uma revisão de literatura, mostrando a importância dessa técnica para aproveitar pesquisas já realizadas. Serrano and Zorrilla (2019) realizaram uma RSL em que foram selecionados 22 artigos que aju-

daram os autores a definir os componentes do *framework* proposto.

Quando se analisa trabalhos que explicitamente comentam a diferença entre *framework* de governança de dados tradicionais (*small data*) e *big data*, obtêm-se que 62,5% dos artigos selecionados fizeram essa distinção, sendo um diferencial do presente trabalho a tentativa de comparar os componentes de *frameworks* de governança de *big data* com um *framework* governança de dados tradicionais. Esse percentual corrobora com o objetivo do deste trabalho em abordar essa distinção e responde a pergunta de pesquisa relacionada a diferenciação entre *big data* e *small data* em *frameworks* de governança de dados.

4 MAPEAMENTO DE COMPONENTES DE GOVERNANÇA DE DADOS

Foi feito o cruzamento entre os termos e definições dos componentes dos FGBDs apresentados nos 8 artigos selecionados neste estudo e os componentes padrão do trabalho de Jang and Kim (2021), com o objetivo de mapear os componentes padrão referenciados em cada um desses trabalhos. A Tabela 6 lista os componentes padrão e suas respectivas identificações utilizadas na Tabela 7 que apresenta o resultado do mapeamento realizado.

Tabela 6: Componentes padrão

Componente Padrão	Identificador
Estratégia	C1
Política	C2
Metodologia	C3
Métricas	C4
Arquitetura de Sistema	C5
Ciclo de Vida do Dado	C6
Monitoramento de Dados	C7
Gerenciamento de Qualidade	C8
Valor de qualidade	C9
Segurança de Dados	C10
Papéis e Responsabilidades	C11
Estrutura Organizacional	C12

Fonte: Elaborado pelo autor

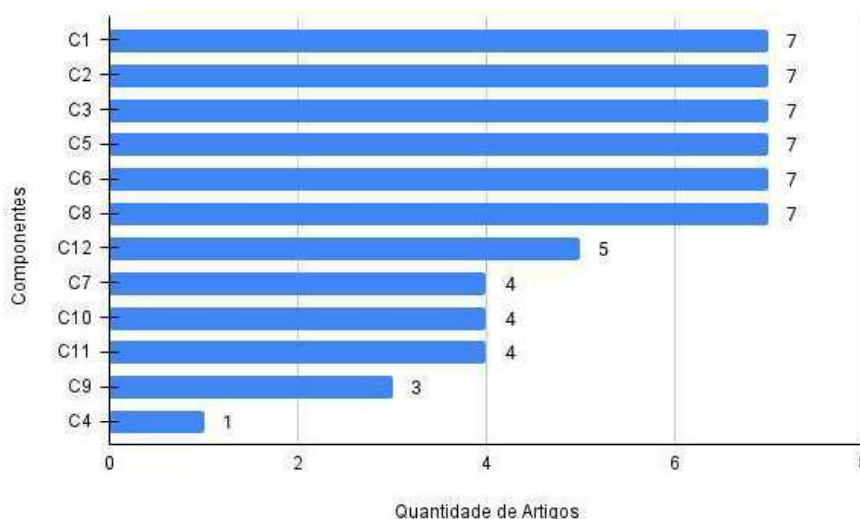
Tabela 7: Mapeamento de componentes padrão

Referência	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12
Li et al. (2019)	✓	✓	✓		✓	✓		✓		✓		✓
Kaufmann (2019)	✓	✓	✓		✓	✓						✓
Al-Badi et al. (2018a)	✓	✓	✓		✓		✓	✓		✓	✓	✓
Janković et al. (2022)	✓	✓			✓	✓	✓	✓			✓	✓
Wang et al. (2018)			✓			✓	✓	✓	✓			
Zeng et al. (2022)	✓	✓	✓		✓	✓		✓				
Serrano and Zorrilla (2019)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Yakobi et al. (2020)	✓	✓	✓		✓	✓		✓	✓	✓	✓	

Fonte: Elaborado pelo autor

Para auxiliar na análise de mapeamento de componentes foram construídos dois gráficos, um comparando as quantidade de artigos que tiveram os componentes padrão mapeados e outro com a quantidade de componentes padrão mapeados nos artigos.

Figura 8: Quantidade de artigos por componente padrão



Fonte: Elaborado pelo autor

Na Figura 8 observa-se que os componentes de Estratégia (C1), Políticas (C2), Metodologia (C3), Arquitetura de Sistemas (C5), Ciclo de Vida dos Dados (C6) e Gerenciamento de Qualidade (C8) estão presentes em 7 dos 8 artigos selecionados, evidenciando que são componentes básicos em um *framework* de governança de dados. Tais componentes estão relacionados diretamente ao objetivo da governança de dados de gerenciar dados com estratégia para trazer valor para o negócio, garantindo qualidade por meio de políticas e processos (Ekundayo et al., 2023). Isso também reforça o trabalho de Jang and Kim (2021), pois, com exceção do componente de Métricas (C4), todos os outros componentes padrão que constituem o grupo Conformidade de Dados, estão entre os componentes mais mapeados. Como foi mostrado em Jang and Kim (2021), os componentes que formam o grupo de Conformidade de Dados são aqueles que não são influenciados por outros componentes principais, mas exercem influencia em outros componentes.

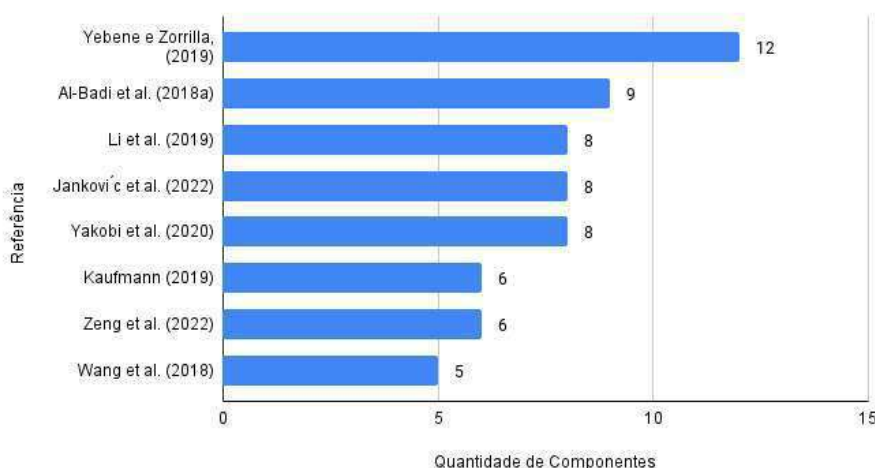
Por outro lado, os componentes de Métricas (C4) e Valor de Qualidade (C9) foram menos mapeados, cada um aparecendo em 1 e 3 artigos, respectivamente. Vale observar que ambos possuem uma relação direta com dois outros componentes, Monitoramento de Dados (C7) e Gerenciamento de Qualidade (C8), o que pode justificar a não aparição de forma direta dos mesmos. Os componentes de Monitoramento de Dados (C7), Segurança

dos Dados (C9), Papeis e Responsabilidades (11) foram mapeados em 4 artigos, enquanto que Organização Estrutural (12) foi mapeado em 5 artigos. Isso mostra uma relação intermediária entres esses componentes e *frameworks* de governança de *big data*.

Trabalhos da área da Indústria se preocuparam em considerar Monitoramento de Dados (C7) em seus componentes. Considerando que parte dos dados desse setor são gerados por dispositivos de IoT e que eles ajudam a tomar decisões que impactam o processo de produção (Janković et al., 2022), esses dados precisam ser constantemente monitorados para evitar problemas na linha produtiva, justificando a preocupação dos autores em trazer esse componente nos *frameworks* apresentados. Trabalhos que consideraram componentes relacionados a Papeis e Responsabilidades (C11), em sua maioria, trouxeram componentes relacionados à Estrutura Organizacional (C2), o que corrobora com o agrupamento feito por Jang and Kim (2021), colocando esses dois componentes no mesmo grupo, como foi monstrado na Seção 2.2.

A Figura 9 mostra a quantidade de componentes padrão mapeados em cada artigo que apresentou um FGBD.

Figura 9: Quantidade de componentes padrão por artigo



Fonte: Elaborado pelo autor

Observamos que o trabalho de Serrano and Zorrilla (2019) contém todos os 12 componentes da referência utilizada. Importante notar que esse artigo é um trabalho da área da Indústria 4.0, que tem como foco a utilização de dados para melhorar processos e produtos (Serrano and Zorrilla, 2019).

No outro extremo temos o trabalho de Wang et al. (2018) que apresentou apenas 5 componentes: Metodologia (C3), Ciclo de Vida do Dados (C6), Monitoramento de

Dados (C7), Gerenciamento de Qualidade (C8) e Valor de Qualidade (C9). Vale ressaltar que esse foi um dos trabalhos que pouco apresentou descrição para os componentes do *framework*, o que pode ter impactado o mapeamento entre os componentes apresentados e os da referência utilizada.

4.1 Componentes de *Framework* de Governança de *Big Data*

Analizando os componentes de *framework* de governança de *big data* apresentados nos 8 trabalhos que não foram mapeados de forma direta para componentes padrão, temos os seguintes componentes: Governança de Leis e Regulações, Integração, Desenvolvimento de Produtos de Dados, Pessoas, Princípios Guia, Gerenciamento de Processamento Distribuído e Comunicação. Cada componente é apresentado nas subseções seguintes.

4.1.1 Governança de Leis e Regulações

Segundo Li et al. (2019), *big data* tem atributos legais que devem ser regulamentados por legislações e devem ser consideradas como um dos guias na implementação de um programa de governança de dados. As leis e regulamentos devem ser balanceadas em três aspectos: proteção efetiva de dados pessoais, promoção do compartilhamento aberto de dados governamentais e promoção do desenvolvimento da indústria de *big data* com dados de transações comerciais.

Os autores ainda afirmam que a depender da estratégia do governo, ou de mercado, a legislação relacionada a dados pode ser mais conservadora, a exemplo da União Europeia, ou uma estratégia mais liberal em que é o valor comercial dos dados tem maior ênfase, existindo uma responsabilização individual e do mercado quanto a utilização dos dados, como é o caso dos Estados Unidos.

4.1.2 Integração

No trabalho de Zeng et al. (2022), os autores trazem um componente de integração de dados devido à variedade de fontes e formatos, sendo umas das características de *big data* (Al-Mekhlal and Ali Khwaja, 2019). Esse componente tem como objetivo agregar diferentes fontes de dados formando um conjunto de dados que podem ser unificados, resolvendo diferenças de dados heterogêneos, aumentando o seu valor. O componente de

integração também é apresentado por Yakobi et al. (2020) como ferramenta para ajudar a entender e contextualizar os dados. Esse componente apresenta uma camada de fusão onde é feito um mapeamento de dados para coordenar formatos e resolver conflitos de semântica que podem existir na integração de dados, baseando-se nas definições dos componentes do *framework* relacionados a estratégia e planejamento.

Dentro desse mesmo componente de integração, os autores comentam sobre a importância de um armazenamento distribuído de dados utilizando tecnologias de armazenamento NoSQL, permitindo o armazenamento em larga escala e de dados heterogêneos, ponto que também é reforçado por Yakobi et al. (2020).

4.1.3 Desenvolvimento de Produtos de Dados

No mesmo trabalho de Zeng et al. (2022), os autores apresentam um componente de aplicação, que tem como objetivo explorar o potencial valor de dados com a construção de produtos de dados. Essas aplicações devem seguir os princípios FAIR, sigla em inglês para localizável, acessível, interoperável e reutilizável. . Os produtos primários são obtidos por meio da limpeza, agrupamento e visualização dos dados brutos. Já os produtos avançados envolvem a adoção de algum modelo de previsão que utiliza como insumo os produtos primários.

4.1.4 Pessoas

O trabalho de Serrano and Zorrilla (2019) apresenta pessoas como componente relacionado a atividades operacionais em um projeto de governança de dados. Os autores consideram importante a existência de pessoas com perfis capazes de trabalhar com novas tecnologias, e lidar com novas formas de operar dados e desenvolvimento de automações de processos, evidenciando a complexidade e o cuidado nesses projetos.

4.1.5 Princípios Guia

O *framework* proposto por Yakobi et al. (2020) apresenta um componente de princípios guia com boas práticas que podem ser seguidas quando alguma atividade não está prevista nas regras e processos previamente definidos, dando um suporte adicional ao gerenciamento dos dados. Enquanto regras e processos são obrigatórios, esses princípios podem, ou não serem seguidos a depender da situação. Os autores trazem os princípios de responsabilidade, integridade, auditabilidade e transparência.

Responsabilidade está relacionada a acessibilidade dos dados e a credibilidade dos operadores dos dados. Os autores complementam que se todos os setores de uma empresa são responsáveis na utilização dos dados, vazamento de dados seria uma preocupação a menos para a organização. Integridade é um princípio que se materializa quando os dados são claramente definidos, devidamente controlados e acessados de forma adequada, garantindo um bom alinhamento das estratégias da organização. Para isso é necessário uma avaliação contínua dos dados, detectando mudanças que impactam no seu uso. Transparência está ligada à utilização e ao tratamento dos dados sensíveis, tornando possível que *stakeholders* ou auditores entendam os processos relacionados aos dados. É importante também que decisões, processamento e ações feitas nos dados devam ser auditadas por meio de documentações apropriadas seguindo requisitos de auditoria e conformidade.

4.1.6 Gerenciamento de Processamento Distribuído

Yakobi et al. (2020) também trazem outro componente, gerenciamento de processamento distribuído de dados. Devido às características de *big data*, sistemas tradicionais de processamento não são adequados para manipular esse tipo de dado. Por isso é necessário utilizar uma tecnologia de processamento de *big data* eficiente, permitindo análise de um volume grande de dados em um tempo aceitável. Este ponto também é reforçado por Zeng et al. (2022).

4.1.7 Comunicação

O *framework* proposto por Al-Badi et al. (2018b) coloca comunicação como último componente em um *framework* de governança de *big data*, reforçando que os recursos gerados devem ser comunicados e entregues aos clientes e pessoas de interesse.

4.2 Discussão de Resultados

Ao analisar os componentes foi identificado que eles podem ser integrados aos componentes padrão, seja como subcomponentes, seja por meio de adaptações específicas para atender às particularidades do *big data*. Essas diferenciações são essenciais para lidar com as características únicas desse contexto.

Os componentes Governança de Leis e Regulações e Princípios Guia podem ser incorporados ao componente padrão de Metodologia, que define os processos e métodos da

governança de dados (Jang and Kim, 2021). Esses dois componentes podem ser utilizados na definição e características a serem consideradas nesses processos, trazendo maior visibilidade para pontos críticos como privacidade, responsabilidade, integridade, auditabilidade e transparência. Por exemplo, considerando a Lei Geral de Proteção de Dados (LGPD) (Brasil, 2018), um processo a ser implementado poderia ser a verificação do consentimento para armazenamento e processamento de dados pessoais. Outro exemplo considerando o Princípio Guia de Integridade, seria um processo para verificar a integridade do dado ao longo do tempo, analisando anomalias e inconsistências da informação.

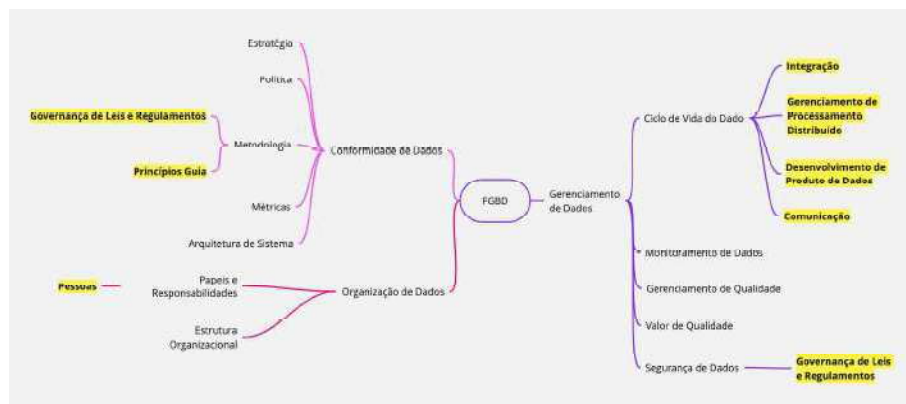
O componente padrão de Ciclo de Vida pode agregar quatro componentes: Integração, Gerenciamento de Processamento Distribuído, Desenvolvimento de Produtos de Dados e Comunicação. Neste componente padrão é considerado todas as fases do ciclo de vida do dado dentro de uma empresa, desde a coleta e geração até a sua remoção. Devido à relação com a coleta e pré-processamento dos dados, o componente de Integração pode ser colocado após as fases de coleta e armazenamento. Gerenciamento de Processamento Distribuído pode ser visto como uma característica a ser considerada no processamento de dados devido às especificidades de *big data* que exige uma capacidade maior de processamento. Dando maior visibilidade à fase de utilização do dado do componente padrão de Ciclo de Vida, o componente Desenvolvimento de Produtos de Dados pode ser associado a mesma, dando mais detalhes do que pode ser feito com o dado, por exemplo, criação de *dashboards* e modelos de aprendizagem de máquina. Por último, o componente de Comunicação pode ser incorporado ao componente de Ciclo de Vida, enfatizando a importância da divulgação dos ativos de dados que foram criados no processo.

O componente de Governança de Leis e Regulações também pode ser considerado no componente padrão de Segurança de Dados nas atividades relacionadas à segurança e privacidade de dados, ajudando a definir atividades com base em leis e regulamentos. Trazendo novamente a LGPD, uma atividade de segurança a ser aplicada, seria o controle de acesso aos dados armazenados, garantindo que apenas pessoas autorizadas tenham acesso aos mesmos.

Por último, o componente padrão Papéis e Responsabilidades pode considerar o componente Pessoas, reforçando a importância de considerar pessoas aptas a assumir papéis e a realizar as atividades definidas.

A Figura 10 estende a organização dos componentes padrão proposto por Jang and Kim (2021), incorporando os componentes discutidos nesta seção. Os componentes com destaque em amarelo são os componentes incorporados.

Figura 10: Incorporação dos componentes de um FGBD aos componentes padrão



Fonte: Elaborado pelo autor

4.2.1 Respostas das Questões de Pesquisa

Após uma análise e discussão dos resultados, as questões de pesquisa definidas para esse trabalho podem ser respondidas.

A RSL conduzida levantou os seguintes componentes específicos de *frameworks* de governança de *big data*: Governança de Leis e Regulações, Integração, Desenvolvimento de Produtos e Pessoas, Princípios Guia, Gerenciamento de Processamento Distribuído e Comunicação. Isso responde a primeira questão de pesquisa que questiona quais são os componentes específicos de um *framework* de governança de *big data*.

Foi feita uma análise considerando os componentes listados acima e o *framework* proposto por Jang and Kim (2021) para responder à segunda pergunta de pesquisa: Como um *framework* de governança de dados tradicional pode ser estendido para se adequar ao contexto de *big data*? Esse trabalho mostrou que os componentes padrão definidos no trabalho de Jang and Kim (2021) podem ser utilizados, considerando as extensões propostas em alguns componentes como discutido na Seção 4.2. Como apresentado, os componentes Metodologia, Ciclo de Vida, Segurança de Dados e Estrutura Organizacional devem ser estendidos para agregar componentes específicos levantados na RSL.

5 CONCLUSÃO

O presente trabalho realizou uma revisão sistemática da literatura com o objetivo de levantar as diferenças entre *frameworks* de governança de dados para dados tradicionais e para *big data*. Seguindo um protocolo de pesquisa, foram extraídos os componentes de *framework* de governança de *big data* de 8 trabalhos, fez-se um cruzamento com os componentes padrão de governança levantados por Jang and Kim (2021) e identificou-se diferenças entre os componentes apresentados. Sete componentes de *framework* de governança de *big data* não foram mapeados de forma direta em componentes padrão do trabalho de Jang and Kim (2021), sendo eles: Governança de Leis e Regulações, Princípios Guia, Integração, Gerenciamento de Processamento Distribuído, Desenvolvimento de Produtos de Dados, Pessoas e Comunicação.

A análise dos componentes permitiu associar os componentes encontrados a componentes padrão, trazendo maior detalhamento e completude ao framework proposto por Jang and Kim (2021). O componente padrão Metodologia incorporou os componentes Governança de Leis e Regulações e Princípios Guia; os componentes de Integração, Gerenciamento de Processamento Distribuído, Desenvolvimentos de Produtos de Dados e Comunicação foram adicionados ao componente padrão de Ciclo de Vida do Dado; Segurança de Dados também incorporou Governança de Leis e Regulações e o componente de Pessoas foi adicionado ao componente padrão de Estrutura Organizacional.

A principal contribuição do presente trabalho consistiu em levantar componentes de *frameworks* de governança de *big data*, comparando-os com componentes padrão, estendendo o *framework* proposto por Jang and Kim (2021) para que os componentes mapeados a partir da RSL fossem incorporados. Isso permitiu adequar um *framework* com componentes padrão as particularidades de *big data*. Além disso, o estudo apresenta um mapeamento abrangente das áreas de atuação e dos temas relacionados a *big data* identificados nos trabalhos analisados na RSL, ampliando a compreensão do campo.

Por fim conclui-se que é possível utilizar *frameworks* de governança de dados tradicionais em *big data*, desde que as características desse tipo de dados sejam consideradas

nos componentes e atividades dos *frameworks* utilizados.

5.1 Limitações e Trabalhos futuros

Como fator limitante para o presente trabalho pode ser citada a utilização de um único trabalho de referência que apresentou componentes de *framework* de governança de dados. Um trabalho futuro poderia considerar outros trabalhos de referência e a comparação entre eles.

Outra limitação está na estratégia de mapeamento entre os componentes de *framework* de governança de *big data* e os componentes padrão. A aplicação da técnica de análise de conteúdo usada por Jang and Kim (2021), que utilizou especialista para realizar o agrupamento dos componentes padrão, ajudaria a garantir uma maior assertividade nos mapeamentos feitos.

Como sugestão de trabalhos futuros, podem ser considerados a atualização de mais bases de conhecimento com o objetivo de encontrar mais componentes de *frameworks* de governança de *big data* e a execução de uma análise de conteúdo dos componentes de um FGBD utilizando especialistas.

REFERÊNCIAS BIBLIOGRÁFICAS

- Abraham, R., Schneider, J., vom Brocke, J., 2019. Data governance: A conceptual framework, structured review, and research agenda. doi:10.1016/j.ijinfomgt.2019.07.008.
- Akter, S., Wamba, S.F., 2016. Big data analytics in e-commerce: a systematic review and agenda for future research. *Electronic Markets* 26. doi:10.1007/s12525-016-0219-0.
- Al-Badi, A., Tarhini, A., Khan, A.I., 2018a. Exploring big data governance frameworks, in: *Procedia Computer Science*, Elsevier B.V.. pp. 271–277. doi:10.1016/j.procs.2018.10.181.
- Al-Badi, A., Tarhini, A., Khan, A.I., 2018b. Exploring big data governance frameworks, Elsevier B.V.. pp. 271–277. doi:10.1016/j.procs.2018.10.181.
- Al-Mekhlal, M., Ali Khwaja, A., 2019. A synthesis of big data definition and characteristics, in: 2019 IEEE International Conference on Computational Science and Engineering (CSE) and IEEE International Conference on Embedded and Ubiquitous Computing (EUC), pp. 314–322. doi:10.1109/CSE/EUC.2019.00067.
- Al-Shdifat, A., Emmanouilidis, C., 2018. Development of a context-aware framework for the integration of internet of things and cloud computing for remote monitoring services, in: *Procedia Manufacturing*. doi:10.1016/j.promfg.2018.10.155.
- Alsaad, A., 2023. Governmental data governance frameworks: A systematic literature review, *Institute of Electrical and Electronics Engineers (IEEE)*. pp. 150–156. doi:10.1109/iccece59400.2023.10238504.
- Arundel, S.T., McKeehan, K.G., Campbell, B.B., Bulen, A.N., Thiem, P.T., 2023. A guide to creating an effective big data management framework. *Journal of Big Data* 10. doi:10.1186/s40537-023-00801-9.
- Bani-Salameh, H., Al-Qawaqneh, M., Taamneh, S., 2021. Investigating the adoption of big data management in healthcare in jordan. *Data* 6. doi:10.3390/data6020016.
- Bashir, M.R., Gill, A.Q., Beydoun, G., McCusker, B., 2020. Big data management and analytics metamodel for iot-enabled smart buildings. *IEEE Access* 8. doi:10.1109/ACCESS.2020.3024066.

- Bassi, C.A., Souza, S.N.A.d., 2023. Challenges to implementing effective data governance: A literature review, in: Proceedings of the 15th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management - KMIS, INSTICC. SciTePress. pp. 17–28. doi:10.5220/0012185900003598.
- Bento, P., Neto, M., Corte-Real, N., 2022. How data governance frameworks can leverage data-driven decision making: A sustainable approach for data governance in organizations. doi:10.23919/CISTI54924.2022.9866895.
- Benítez-Hidalgo, A., Barba-González, C., García-Nieto, J., Gutiérrez-Moncayo, P., Paneque, M., Nebro, A.J., del Mar Roldán-García, M., Aldana-Montes, J.F., Navas-Delgado, I., 2021. Titan: A knowledge-based platform for big data workflow management. Knowledge-Based Systems 232. doi:10.1016/j.knosys.2021.107489.
- Bhuvana, M., Thirumagal, P.G., Vasantha, S., 2016. Big data analytics - a leveraging technology for indian commercial banks. Indian Journal of Science and Technology 9. doi:10.17485/ijst/2016/v9i32/98643.
- Brasil, 2018. Lei nº 13.709, de 14 de agosto de 2018. Diário Oficial [da] República Federativa do Brasil URL: https://www.planalto.gov.br/ccivil_03/ato2015-2018/2018/lei/l13709.htm. Campos, J., Sharma, P., Jantunen, E., Baglee, D., Fumagalli, L., 2016 *Procedia CIRP*. doi : 10.1016/j.procir.2016.03.059.
- Cassoli, B.B., Jourdan, N., Nguyen, P.H., Sen, S., Garcia-Ceja, E., Metternich, J., 2022. Frameworks for data-driven quality management in cyber-physical systems for manufacturing: A systematic review, in: Procedia CIRP. doi:10.1016/j.procir.2022.09.062.
- Chao, K., Sarker, M.N.I., Ali, I., Firdaus, R.B., Azman, A., Shaed, M.M., 2023. Big data-driven public health policy making: Potential for the healthcare industry. Heliyon 9. doi:10.1016/j.heliyon.2023.e19681.
- Chauhan, A.S., Cuzzocrea, A., Fan, L., Harvey, J.D., Leung, C.K., Pazdor, A.G., Wang, T., 2021. Predictive big data analytics for service requests: A framework, in: Procedia Computer Science. doi:10.1016/j.procs.2021.12.216.
- DAMA-DMBOK, 2009. The DAMA Guide to The Data Management Body of Knowledge. doi:10.1161/CIRCULATIONAHA.108.834176.
- Ekundayo, T., Bhaumik, A., Chinoperekweyi, J., 2023. Identifying the core data governance framework principle: A framework comparative analysis. Organization Leadership and Development Quarterly 5.

- Evans, E.A., Delorme, E., Cyr, K., Goldstein, D.M., 2020. A qualitative study of big data and the opioid epidemic: recommendations for data governance. *BMC Medical Ethics* 21. doi:10.1186/s12910-020-00544-9.
- Faroukhi, A.Z., Alaoui, I.E., Gahi, Y., Amine, A., 2020a. An adaptable big data value chain framework for end-to-end big data monetization. *Big Data and Cognitive Computing* 4. doi:10.3390/bdcc4040034.
- Faroukhi, A.Z., Alaoui, I.E., Gahi, Y., Amine, A., 2020b. A multi-layer big data value chain approach for security issues, in: *Procedia Computer Science*. doi:10.1016/j.procs.2020.07.109.
- Ghorbani, Y., Zhang, S.E., Nwaila, G.T., Bourdeau, J.E., 2022. Framework components for data-centric dry laboratories in the minerals industry: A path to science-and-technology-led innovation. doi:10.1016/j.exis.2022.101089.
- Holom, R.M., Rafetseder, K., Kritzing, S., Sehrschön, H., 2020. Metadata management in a big data infrastructure, in: *Procedia Manufacturing*. doi:10.1016/j.promfg.2020.02.060.
- Huang, Y., xin CHEN, Z., YU, T., zhi HUANG, X., fa GU, X., 2018. Agricultural remote sensing big data: Management and applications. doi:10.1016/S2095-3119(17)61859-8.
- Hussain, M., Zhang, T., Seema, M., 2023. Adoption of big data analytics for energy pipeline condition assessment - a systematic review. doi:10.1016/j.ijpvp.2023.105061.
- IBM, 2007. The ibm data governance council maturity model : Building a roadmap for effective data governance. *Governance An International Journal Of Policy And Administration* .
- Ilie-Zudor, E., Kemény, Z., Ekárt, A., Buckingham, C.D., Monostori, L., 2014. A solution for information management in logistics operations of modern manufacturing chains, in: *Procedia CIRP*. doi:10.1016/j.procir.2014.10.047.
- Irani, Z., Sharif, A.M., Lee, H., Aktas, E., Topaloglu, Z., van't Wout, T., Huda, S., 2018. Managing food security through food waste and loss: Small data to big data. *Computers and Operations Research* 98. doi:10.1016/j.cor.2017.10.007.
- Jang, K.A., Kim, W.J., 2021. Development of data governance components using dematel and content analysis. *Journal of Supercomputing* 77, 3695–3709. doi:10.1007/s11227-020-03405-9.

- Janković, A., Adrodegari, F., Saccani, N., Simeunović, N., 2022. Improving service business of industrial companies through data: conceptualization and application. *International Journal of Industrial Engineering and Management* 13. doi:10.24867/IJIEM-2022-2-302.
- Kalkman, S., Mostert, M., Udo-Beauvisage, N., Delden, J.J.V., Thiel, G.J.V., 2019. Responsible data sharing in a big data-driven translational research platform: Lessons learned. *BMC Medical Informatics and Decision Making* 19. doi:10.1186/s12911-019-1001-y.
- Kamal, J., Dave, M., 2019. A framework for managing and analyzing big data in indian school education system with reference to jammu kashmir. *International Journal of Innovative Technology and Exploring Engineering* 8. doi:10.35940/ijitee.I1076.0789S19.
- Karras, A., Giannaros, A., Theodorakopoulos, L., Krimpas, G.A., Kalogeratos, G., Karras, C., Sioutas, S., 2023. Flibd: A federated learning-based iot big data management approach for privacy-preserving over apache spark with fate. *Electronics (Switzerland)* 12. doi:10.3390/electronics12224633.
- Kaufmann, M., 2019. Big data management canvas: A reference model for value creation from data. *Big Data and Cognitive Computing* 3. doi:10.3390/bdcc3010019.
- Kelle, O., 2015. Top 10 artifacts needed for data governance. First San Francisco Partners .
- Khatri, V., Brown, C.V., 2010. Designing data governance. *Communications of the ACM* 53, 148–152.
- Kim, H.Y., Cho, J.S., 2017. Data governance framework for big data implementation with a case of korea. doi:10.1109/BigDataCongress.2017.56.
- Kitchenham, B., 2004. Procedures for performing systematic reviews. Keele University,UK and National ICT Australia 33. doi:10.1.1.122.3308.
- Li, Q., Lan, L., Zeng, N., You, L., Yin, J., Zhou, X., Meng, Q., 2019. A framework for big data governance to advance rhins: A case study of china. *IEEE Access* 7, 50330–50338. doi:10.1109/ACCESS.2019.2910838.
- Liu, M., Zhang, B., Bi, J., 2022. Appreciating the role of big data in the modernization of environmental governance. doi:10.1007/s42524-021-0185-x.

- Mao, Z., Wu, J., Qiao, Y., Yao, H., 2022. Government data governance framework based on a data middle platform. *Aslib Journal of Information Management* 74. doi:10.1108/AJIM-03-2021-0068.
- Marelli, L., Testa, G., van Hoyweghen, I., 2021. Big tech platforms in health research: Re-purposing big data governance in light of the general data protection regulation's research exemption. *Big Data and Society* 8. doi:10.1177/20539517211018783.
- Nadal, S., Jovanovic, P., Bilalli, B., Romero, O., 2022. Operationalizing and automating data governance. *Journal of Big Data* 9. doi:10.1186/s40537-022-00673-5.
- Naghieb, A., Navimipour, N.J., Hosseinzadeh, M., Sharifi, A., 2023. A comprehensive and systematic literature review on the big data management techniques in the internet of things. *Wireless Networks* 29. doi:10.1007/s11276-022-03177-5.
- Ng, S.T., Xu, F.J., Yang, Y., Lu, M., 2017. A master data management solution to unlock the value of big infrastructure data for smart, sustainable and resilient city planning, in: *Procedia Engineering*. doi:10.1016/j.proeng.2017.08.034.
- Noor, A.M., Holmberg, L., Gillett, C., Grigoriadis, A., 2015. Big data: The challenge for small research groups in the era of cancer genomics. *British Journal of Cancer* 113. doi:10.1038/bjc.2015.341.
- Obatolu, R., 2022. A holistic framework for complex big data governance.
- Otto, B., 2011. One size does not fit all: Best practices for data governance .
- Poger, D., Yen, L., Braet, F., 2023. Big data in contemporary electron microscopy: challenges and opportunities in data transfer, compute and management. doi:10.1007/s00418-023-02191-8.
- Raghupathi, W., Raghupathi, V., 2014. Big data analytics in healthcare: Promise and potential. doi:10.1186/2047-2501-2-3.
- Rasheed, M.A., Ahmad, H.Z., Tanweer, H.A., Murtaza, H., Rasheed, M.A., Ahmed, M., 2021. Use of big data governance in several corporate sectors. *VFAST Transactions on Software Engineering* 9, 92–101.
- Salierno, G., Morvillo, S., Leonardi, L., Cabri, G., 2020. An architecture for predictive maintenance of railway points based on big data analytics, in: *Lecture Notes in Business Information Processing*.

doi:10.1007/978-3-030-49165-9₃.SAS, 2014. *Thesadatagovernanceframework : blueprint for success.SASInstituteInc.*

- Schaefer, C., Makatsaria, A., 2021. Framework of data analytics and integrating knowledge management. International Journal of Intelligent Networks 2. doi:10.1016/j.ijin.2021.09.004.
- Serrano, J.Y., Zorrilla, M., 2019. Towards a data governance framework for third generation platforms, in: Procedia Computer Science. doi:10.1016/j.procs.2019.04.082.
- Shah, S.I.H., Peristeras, V., Magnisalis, I., 2021. Dalif: a data lifecycle framework for data-driven governments. Journal of Big Data 8. doi:10.1186/s40537-021-00481-3.
- Shankarnarayan, V.K., Ramakrishna, H., 2020. Paradigm change in indian agricultural practices using big data: Challenges and opportunities from field to plate. doi:10.1016/j.inpa.2020.01.001.
- Sivarajah, U., Kamal, M.M., Irani, Z., Weerakkody, V., 2017. Critical analysis of big data challenges and analytical methods. Journal of Business Research 70, 263–286. URL: <https://www.sciencedirect.com/science/article/pii/S014829631630488X>, doi:<https://doi.org/10.1016/j.jbusres.2016.08.001>.
- Sohrabi, C., Mathew, G., Franchi, T., Kerwan, A., Griffin, M., Del Mundo, J.S.C., Ali, S.A., Agha, M., Agha, R., 2021. Impact of the coronavirus (covid-19) pandemic on scientific research and implications for clinical academic training—a review. International Journal of Surgery 86, 57–63.
- Soup, S., 2011. Data governance. In: IS6120 enterprise business intelligence .
- Stergiou, C.L., Psannis, K.E., 2022. Digital twin intelligent system for industrial internet of things-based big data management and analysis in cloud environments. Virtual Reality Intelligent Hardware 4, 279–291. URL: <https://www.sciencedirect.com/science/article/pii/S2096579622000444>, doi:<https://doi.org/10.1016/j.vrih.2022.05.003>.
- Thomas, G., 2006. The dgi data governance framework. The Data Governance Institute, Orlando, FL (USA) .
- TTA, 2011. Data governance part 5: framework. Telecommunications Technology Association .
- Wang, J., Li, Y.S., Song, W., Li, A.H., 2018. Research on the theory and method of grid data asset management, in: Procedia Computer Science. doi:10.1016/j.procs.2018.10.258.

- Wang, Z., Guan, X., Zeng, Y., Liang, X., Dong, S., 2024. Utilizing data platform management to implement “5w” analysis framework for preventing and controlling corruption in grassroots government. *Heliyon* 10. doi:10.1016/j.heliyon.2024.e28601.
- Weber, K., Otto, B., Österle, H., 2009. One size does not fit all -a contingency approach to data governance. *Journal of Data and Information Quality* 1. doi:10.1145/1515693.1515696.
- Wen, W., Gu, S., Zhang, Y., Yang, W., Guo, X., 2024. Standard framework construction of technology and equipment for big data in crop phenomics. *Engineering* URL: <https://www.sciencedirect.com/science/article/pii/S2095809924003187>, doi:<https://doi.org/10.1016/j.eng.2024.06.001>.
- Wolfert, S., Ge, L., Verdouw, C., Bogaardt, M.J., 2017. Big data in smart farming – a review. doi:10.1016/j.agry.2017.01.023.
- Yakobi, K., Scholtz, B., vom Berg, B., 2020. A conceptual model of the challenges of social media big data for citizen e-participation: A systematic review, in: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. doi:10.1007/978-3-030-45002-1_21.
- Yang, H., Guohua, W., Wanlong, G., Qinghai, H., Yixian, Y., 2022. *Design and implementation of* *Procedia Computer Science*. doi : 10.1016/j.procs.2022.04.016.
- Yang, L., Li, J., Elisa, N., Prickett, T., Chao, F., 2019. Towards big data governance in cybersecurity. *Data-Enabled Discovery and Applications* 3. doi:10.1007/s41688-019-0034-9.
- Yang, S., Li, J., Cai, J., Guo, K., Gao, X., Meng, F., 2014. Data-oriented method to big data standard system creation: A case of chinese financial industry. *Annals of Data Science* 1. doi:10.1007/s40745-014-0024-6.
- Yousfi, S., Chiadmi, D., Rhanoui, M., 2022. Smart big data framework for insight discovery. *Journal of King Saud University - Computer and Information Sciences* 34. doi:10.1016/j.jksuci.2021.12.009.
- Yukhno, A., 2024. Digital transformation: Exploring big data governance in public administration. *Public Organization Review* 24. doi:10.1007/s11115-022-00694-x.
- Zeng, M., Xu, Y., Wu, H., Ma, J., Gao, J., 2022. Sustainable insights for energy big data governance in china: Full life cycle curation from the ecosystem perspective. *Sustainability (Switzerland)* 14. doi:10.3390/su14106013.

- Zhang, Q., Sun, X., Zhang, M., 2022. Data matters: A strategic action framework for data governance. *Information Management* 59, 103642. URL: <https://www.sciencedirect.com/science/article/pii/S0378720622000544>, doi:<https://doi.org/10.1016/j.im.2022.103642>.
- Zhang, T., Gupta, A., Rodríguez, M.A.F., Spjuth, O., Hellander, A., Toor, S., 2024. Data management of scientific applications in a reinforcement learning-based hierarchical storage system[formula presented]. *Expert Systems with Applications* 237. doi:[10.1016/j.eswa.2023.121443](https://doi.org/10.1016/j.eswa.2023.121443).

APÊNDICE A – *STRINGS* DE BUSCA

Strings de busca por base de conhecimento:

- IEEE :(*("Document Title": "Big Data"AND ("Document Title": "Governance"OR "Document Title": "Management") AND ("Document Title": "Framework"OR "Document Title": "Framework Element"OR "Document Title": "Framework Component")) OR ("Abstract": "Big Data"AND ("Abstract": "Governance"OR "Abstract": "Management") AND ("Abstract": "Framework"OR "Abstract": "Framework Element"OR "Abstract": "Framework Component")) OR ("Author Keywords": "Big Data"AND ("Author Keywords": "Governance"OR "Author Keywords": "Management") AND ("Author Keywords": "Framework"OR "Author Keywords": "Framework Element"OR "Author Keywords": "Framework Component"))*)
- Science Direct: *TITLE-ABS-KEY ("Big Data") AND ("Governance"OR "Management") AND ("Framework"OR ("Framework Element"OR "Framework Component"))*
- Scopus: *TITLE-ABS-KEY (("Big Data Governance"OR "Big Data Management") AND ("Framework"OR ("Framework Element"OR "Framework Component"))) AND PUBYEAR > 2013 AND (LIMIT-TO (LANGUAGE , "English")) AND (LIMIT-TO (SRCTYPE , "p") OR LIMIT-TO (SRCTYPE , "j")) AND (LIMIT-TO (OA , "all")) AND (LIMIT-TO (DOCTYPE , "ar") OR LIMIT-TO (DOCTYPE , "cp"))*
- Spring: *("Big Data") AND ("Governance"OR "Management") AND ("Framework"OR ("Framework Element"OR "Framework Component"))*