



Universidade de Brasília

Instituto de Ciências Exatas
Departamento de Ciência da Computação

Proposta de Framework para Avaliar a Maturidade da Governança de Dados: Uma Abordagem Empírica Combinando Revisão Sistemática da Literatura e Survey

Luis Filipe Campos Cardoso

Dissertação apresentada como requisito parcial para conclusão do
Mestrado Profissional em Computação Aplicada

Orientadora
Prof.a Dr.a Edna Dias Canedo

Brasília
2025

Ficha catalográfica elaborada automaticamente,
com os dados fornecidos pelo(a) autor(a)

CC198pp Campos Cardoso, Luis Filipe
Proposta de Framework para Avaliar a Maturidade da
Governança de Dados: Uma Abordagem Empírica Combinando
Revisão Sistemática da Literatura e Survey / Luis Filipe
Campos Cardoso; orientador Edna Dias Canedo. Brasília, 2025.
84 p.

Dissertação(Mestrado Profissional em Computação Aplicada)
Universidade de Brasília, 2025.

1. Governança de Dados. 2. Maturidade de Dados. 3.
Revisão Sistemática da Literatura. 4. Survey. 5. Qualidade
de Dados. I. Dias Canedo, Edna, orient. II. Título.



Universidade de Brasília

Instituto de Ciências Exatas
Departamento de Ciência da Computação

Proposta de Framework para Avaliar a Maturidade da Governança de Dados: Uma Abordagem Empírica Combinando Revisão Sistemática da Literatura e Survey

Luis Filipe Campos Cardoso

Dissertação apresentada como requisito parcial para conclusão do
Mestrado Profissional em Computação Aplicada

Prof.a Dr.a Edna Dias Canedo (Orientadora)
CIC/UnB

Prof.a Dr.a Rejane Maria da Costa Figueiredo
FCTE, UnB

Prof.a Dr.a Ana Paula Bernardi da Silva
Universidade Católica de Brasília (UCB)

Prof.a Dr.a Edna Dias Canedo
Coordenadora do Programa de Pós-graduação em Computação Aplicada

Brasília, 25 de Julho de 2025

Agradecimentos

A jornada para a conclusão deste trabalho foi desafiadora, mas cada passo foi viabilizado pelo apoio fundamental de pessoas e instituições a quem sou profundamente grato.

Minha gratidão à minha orientadora, que com sua paciência, dedicação e orientação rigorosa, não apenas aprimorou este estudo, mas também me ensinou a pensar de forma mais crítica e a buscar sempre a excelência.

Aos meus colegas de mestrado, meu obrigado pelas enriquecedoras trocas e pela parceria nesta jornada. Sou grato à Universidade de Brasília (UnB) e ao Programa de Pós-Graduação em Computação Aplicada (PPCA) por oferecerem um ambiente que inspira a pesquisa e por reunir profissionais e docentes que elevaram nosso conhecimento a um novo patamar.

Um agradecimento especial vai para os profissionais do Superior Tribunal de Justiça (STJ), do Governo Federal, da iniciativa privada, do Centro Universitário de Brasília (CEUB) e da Universidade de São Paulo (USP). A participação de vocês no *survey* foi a base para a validação deste trabalho.

A minha família, em particular minha mãe (Catia Campos de Miranda) e meu companheiro, Dr. Luiz Antonio Lira Junior, que me deram apoio e incentivo incondicionais. E ao meu pai *in memoriam* (Luis de Lemes Cardoso), que sempre foi uma inspiração.

A todos que, com um gesto de carinho ou uma palavra de encorajamento, tornaram esta caminhada mais leve, meu sincero reconhecimento.

Resumo

Contexto: Impulsionada pelo volume crescente de dados e por regulamentações como a Lei Geral de Proteção de Dados (LGPD) e a *General Data Protection Regulation* (GDPR), a Governança de Dados (GD) tornou-se uma área estratégica para garantir a conformidade e o uso adequado dos dados. Referenciais internacionais estabelecem diretrizes e modelos de maturidade em GD são utilizados para avaliar e otimizar as capacidades organizacionais, tratando os dados como ativos estratégicos. **Objetivo:** Este trabalho propõe e valida o DG2M, um *framework* para a avaliação da maturidade em GD. O modelo busca diagnosticar o nível de maturidade e sugerir práticas para o aprimoramento da GD. **Método:** O estudo empregou uma Revisão Sistemática da Literatura (RSL) como processo estruturado, dividida em três fases principais: Planejamento, Condução e Relatório. A RSL teve como objetivo analisar avaliações de maturidade e práticas de GD na academia e na indústria, resultando na seleção de 22 estudos. A partir dos achados da RSL, desenvolveu-se a versão inicial do DG2M, que foi posteriormente validada por meio de um *survey* com 46 profissionais da área de dados. **Resultados:** A RSL identificou e caracterizou *frameworks* de maturidade (como Stanford, Loshin's, GBDMM, MD3M) e práticas recorrentes em GD, incluindo a formalização de políticas, a definição de responsabilidades e processos e programas de treinamento. O principal resultado da pesquisa é o DG2M, um modelo de maturidade em GD estruturado em 6 níveis hierárquicos (do Nível 0 ao Nível 5), com dimensões como Estratégia e Governança de Dados; Cultura, Pessoas e Estrutura Organizacional; Formalização e Processos de Gestão de Dados; Qualidade de Dados; Infraestrutura para Dados e Dados Abertos. A validação pelo *survey* confirmou a estrutura de 6 níveis e a adequação das dimensões, destacando a utilidade do DG2M para diagnóstico e sugestões de aprimoramento. O *feedback* dos respondentes levou a aprimoramentos no *framework*. **Conclusões:** Este trabalho oferece uma contribuição relevante para a GD, apresentando o DG2M como um modelo de maturidade validado. O *framework* visa fortalecer a GD como área estratégica, aprimorando os processos e as decisões fundamentadas em dados. Com as melhorias baseadas no *feedback* do *survey*, o DG2M representa uma ferramenta eficaz para diagnosticar e evoluir a maturidade em GD nas organizações públicas e privadas, especialmente do setor público e empresas de serviço.

Palavras-chave: Governança de Dados, Maturidade de Dados, Revisão Sistemática da Literatura, Survey, Qualidade de Dados

Abstract

Context: Driven by the growing volume of data and regulations such as the Brazilian General Data Protection Law (LGPD) and the General Data Protection Regulation (GDPR), Data Governance (DG) has become a strategic area for ensuring compliance and the proper use of data. International frameworks establish guidelines, and DG maturity models are used to assess and optimize organizational capabilities, treating data as strategic assets. **Objective:** This work proposes and validates DG2M, a framework for assessing DG maturity. The model aims to assess the level of maturity and provide recommendations for enhancing DG. **Method:** The study employed a Systematic Literature Review (SLR) as a structured process, divided into three main phases: Planning, Conducting, and Reporting. The SLR aimed to analyze the maturity assessments and practices of DG in academia and industry, resulting in the selection of 22 studies. Based on the SLR findings, the initial version of DG2M was developed and subsequently validated through a survey with 46 data professionals. **Results:** The SLR identified and characterized maturity frameworks (such as Stanford, Loshin’s, GBDMM, MD3M) and recurring DG practices, including policy formalization, definition of responsibilities and processes, and training programs. The main research result is DG2M, a DG maturity model structured into 6 hierarchical levels (from Level 0 to Level 5), with dimensions such as Data Strategy and Governance; Culture, People, and Organizational Structure; Formalization and Data Management Processes; Data Quality; Data Infrastructure; and Open Data. The survey validation confirmed the 6-level structure and the suitability of the dimensions, highlighting DG2M’s usefulness for diagnosing and improving DG practices. Respondent feedback led to framework enhancements. **Conclusions:** This work offers a significant contribution to DG, presenting DG2M as a validated maturity model. The framework aims to strengthen DG as a strategic area, improving processes and data-driven decisions. With the improvements based on survey feedback, DG2M represents an effective tool for diagnosing and evolving DG maturity in public and private organizations, especially in the government sector and service companies.

Keywords: Data Governance, Data Maturity, Systematic Literature Review, Survey, Data Quality

Lista de Figuras

1.1	Mapeamento dos processos realizados neste projeto em BPMN	5
2.1	DAMA Wheel com as Áreas de Conhecimento relacionadas à Gestão de Dados [17]	8
3.1	Visão Geral do Processo de RSL usando Notação BPMN	15
3.2	Visão Detalhada do Planejamento e da Condução da RSL usando Notação BPMN	16
3.3	Distribuição dos Estudos por Ano de Publicação	25
3.4	Distribuição dos Estudos por Tipo e Método de Pesquisa	26

Sumário

1	Introdução	1
1.1	Contextualização	1
1.2	Problema de Pesquisa	2
1.3	Justificativa	3
1.4	Objetivos	3
1.5	Resultados Esperados da Pesquisa	4
1.6	Contribuição Científica	4
1.7	Publicações dessa Dissertação	4
1.8	Metodologia de Pesquisa e Processo Metodológico	5
1.9	Disponibilização dos Dados	6
1.10	Organização do Documento	6
2	Fundamentação	8
2.1	Governança de Dados (GD)	8
2.1.1	Qualidade de Dados	9
2.1.2	Modelos de Maturidade	9
2.2	Trabalhos Correlatos	10
2.3	Resumo do Capítulo	14
3	Revisão Sistemática da Literatura	15
3.1	Planejamento	17
3.2	Condução	22
3.3	Relato	23
3.3.1	RQ.1. Quais avaliações de maturidade em Governança de Dados são utilizadas na academia e na indústria e quais são suas principais características?	25
3.3.2	RQ.2. Quais ferramentas e práticas de governança de dados são implementadas na academia e na indústria para promover uma governança eficaz?	28

3.3.3	RQ.3. Quais são os principais princípios e as melhores práticas emergentes em avaliações de maturidade e governança de dados? . .	29
3.3.4	Finalização da Revisão Sistemática da Literatura	30
3.4	Ameaças à Validade e Limitações	31
3.5	Resumo do Capítulo	32
4	Proposta de <i>Framework</i> para Avaliação de Maturidade da Governança de Dados (DG2M)	33
4.1	Descrição geral do produto	33
4.1.1	Níveis de Maturidade	34
4.1.2	Dimensões do Modelo	35
4.1.3	Aspectos	37
4.1.4	Papéis	38
4.1.5	Estratégia de Uso	42
4.2	Melhorias Realizadas no DG2M após o Survey	43
4.3	Resumo do Capítulo	45
5	Desenvolvimento do DG2M	46
5.1	Primera Versão do DG2M	46
5.2	Survey	47
5.3	Construção do Questionário	50
5.4	Piloto	50
5.5	Seleção dos Participantes	51
5.6	Resultados do Survey	52
5.7	Melhoria do Framework	58
5.8	Resumo do Capítulo	58
6	Discussão	60
6.1	Discussão dos Comentários dos Profissionais de Dados	60
6.2	Ameaças à Validade do Estudo e Limitações	63
6.3	Resumo do Capítulo	65
7	Conclusão	66
7.1	Contribuições, Implicações Práticas e Pesquisas Futuras	67
	Referências	68

Lista de Tabelas

2.1	Comparação entre Trabalhos Relacionados versus Trabalho Proposto . . .	12
3.1	Palavras do PICOC	18
3.2	CrITÉrios de Inclusão e Exclusão para a Revisão Sistemática da Literatura .	19
3.3	Avaliação de Qualidade	20
3.4	Formulário de Extração de Dados	21
3.5	Estudos Seleccionados para a Extração de Dados	23
3.6	Modelos de Maturidade de Dados Identificados na Literatura	31
4.1	Papéis e Responsabilidades na Governança de Dados	38
5.1	Principais Problemas e Soluções Adotadas no Estudo	48
5.2	Perfil dos respondentes do survey (n=46).	53
5.3	Tabela Geral de Concordância, Aplicabilidade e Estatísticas por Dimensão	57
5.4	Contagem e Percentual das Principais Dificuldades na Aplicação do Modelo na Prática	57
6.1	Principais Comentários e Sugestões dos Respondentes	62
6.2	Ameaças à Validade do Estudo	64

Lista de Abreviaturas e Siglas

AI Act Regulamento Inteligência Artificial (Artificial Intelligence Act).

BPMN Notação de Modelagem de Processos de Negócio (Business Process Model and Notation).

CBDAS Sistema Consensual de Avaliação de Maturidade de Big Data (Consensual Big Data Maturity Assessment System).

CMM Modelo de Maturidade em Capacitação (Capability Maturity Model).

DAMA Associação de Gestão de Dados (Data Management Association).

DG2M Modelo de Maturidade de Governança de Dados (Data Governance Maturity Model).

DMBOK Data Management Body of Knowledge.

DQM Gerenciamento em Qualidade de Dados (Data Quality Management).

GD Governança de Dados (Data Governance).

GDPR Regulamento Geral sobre a Proteção de Dados (General Data Protection Regulation).

IA Inteligência Artificial.

IoT Internet das Coisas (Internet of Things).

ISO Organização Internacional de Normalização (International Organization for Standardization).

LGPD Lei Geral de Proteção de Dados Pessoais (Lei nº 13.853, de 2019).

PICOC População, Intervenção, Comparação, Resultado, Contexto (*Population, Intervention, Comparison, Outcomes, Context*).

PMEs Pequenas e Médias Empresas.

RSL Revisão Sistemática da Literatura (Sytematic Literature Review).

TI Tecnologia da Informação.

Capítulo 1

Introdução

Este capítulo apresenta o contexto deste trabalho, as motivações que justificam a realização desta pesquisa, uma visão geral da solução proposta e a metodologia adotada. Além disso, descreve a estrutura de organização deste documento.

1.1 Contextualização

A Governança de Dados (Data Governance) (GD) é definida por Associação de Gestão de Dados (Data Management Association) (DAMA) [17] como o exercício de autoridade, controle e tomada de decisão compartilhada sobre a gestão de ativos de dados, garantindo a conformidade no uso e administração dos dados através do estabelecimento de princípios, políticas, procedimentos e responsabilidades claros. Complementarmente, Bento et al. [6] definem a GD como uma abordagem multidimensional, que abrange processos, papéis, responsabilidades, curadoria e gestão de mudanças. Nesse contexto, a qualidade dos dados depende diretamente da eficácia de sua gestão, tornando a GD central para garantir precisão, compartilhamento adequado e segurança, especialmente para organizações que tratam dados como ativos tangíveis e estratégicos.

Ao tratar os dados como um ativo estratégico, a GD promove a colaboração entre unidades organizacionais e formaliza políticas, padrões e procedimentos, ao mesmo tempo em que assegura a conformidade [1]. Seus objetivos primários incluem maximizar o valor dos dados e minimizar custos e riscos, um aspecto crescentemente relevante devido à expansão dos volumes de dados e às exigências regulatórias, como a Regulamento Geral sobre a Proteção de Dados (General Data Protection Regulation) (GDPR)¹ e a Lei Geral de Proteção de Dados Pessoais (Lei nº 13.853, de 2019) (LGPD)². A GD também delimita direitos e responsabilidades de decisão relacionados à gestão de dados [1].

¹<https://eur-lex.europa.eu/eli/reg/2016/679/oj>

²https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm

Geralmente, a GD apoia a padronização e o controle de qualidade nos processos internos de gestão de dados [11]. O estabelecimento de estruturas e processos é fundamental para o planejamento, monitoramento e suporte contínuo à gestão de dados, reforçando a função da governança tanto como supervisora quanto como facilitadora [9].

Apesar do papel crítico da GD na imposição de políticas e padrões [9], sua implementação eficaz é desafiadora devido à complexidade inerente e ao engajamento organizacional limitado, frequentemente resultando em abordagens inadequadas [6]. Governança inadequada e falhas de segurança podem acarretar riscos financeiros e operacionais significativos, conforme destacado por Kurniawan et al. [30], que identificaram que dados são frequentemente negligenciados por sua complexidade de gerenciamento.

O crescente interesse na GD tem levado profissionais e pesquisadores a desenvolverem ferramentas que auxiliem as organizações em sua implementação [6]. Nesse sentido, a avaliação da maturidade da GD é importante para aprimorar as práticas de gestão de dados e mitigar riscos potenciais [21]. Modelos de maturidade auxiliam na identificação de pontos fortes e fracos, além de possibilitar a construção de um *benchmarking* [30], e um *framework* de governança pode trazer benefícios como estabelecer objetivos mais claros, aumentar a confiabilidade dos dados e manter o escopo e o foco no uso dos dados.

1.2 Problema de Pesquisa

Embora a GD seja reconhecida como uma iniciativa importante para assegurar a aplicação eficaz de políticas, padrões e práticas na gestão de dados organizacionais [9], sua implementação continua a enfrentar desafios significativos. Estes incluem o engajamento insuficiente das partes interessadas, a ausência de *frameworks* adaptáveis a diversos contextos organizacionais e a maturidade limitada nas práticas existentes de gestão de dados [6]. Modelos de maturidade práticos, validados e flexíveis podem ajudar as organizações a avaliarem e avançarem com suas capacidades de GD com mais ferramentas.

Por exemplo, Gunawong [18] analisaram uma força-tarefa de dados de saúde na Tailândia e demonstraram que a falta de princípios de governança resultou em falhas relacionadas à qualidade dos dados, gestão do ciclo de vida e coordenação interinstitucional. Problemas semelhantes são observados em vários setores, como visto nos trabalhos de [39] e Kurniawan et al. [30]. Essa questão ressalta a necessidade de um modelo de maturidade de GD acessível às complexidades organizacionais diversas. É importante saber se um modelo pode ser adaptável o suficiente para atender essa diversidade.

1.3 Justificativa

Embora outros modelos de maturidade em GD tenham sido propostos, como observado nos estudos de Mouhib et al. [32, 33], Malacaria et al. [31] e Al-Sai et al. [4], ainda persistem lacunas significativas em termos de clareza conceitual, adaptabilidade contextual e aplicabilidade prática. Muitos modelos existentes são excessivamente complexos ou são altamente dependentes do contexto, fatores que dificultam seu uso eficaz, particularmente em organizações em estágio inicial. Essas limitações ressaltam a necessidade de modelos acessíveis para apoiar as organizações na avaliação e no avanço de sua maturidade em GD.

Mesmo modelos amplamente reconhecidos, como o Capability Maturity Model Integration (CMMI) [13] e a família ISO/IEC TR 330xx [26], apesar de oferecerem benefícios como economia de despesas, maior envolvimento dos funcionários e maior consistência na captura e uso de processos, não conseguem, individualmente, atender plenamente aos diversos requisitos e características de cada organização, como destacado por Bernardo et al. [7]. Isso torna indispensável uma compreensão abrangente dos diferentes modelos e seus componentes essenciais por parte da comunidade, bem como o contexto organizacional.

Diferentemente dos modelos existentes, o Modelo de Maturidade de Governança de Dados (Data Governance Maturity Model) (DG2M) foi projetado para ser mais simples, incorporando práticas de GD mais recorrentes, facilitando sua adoção e aplicação em diversos contextos organizacionais. O DG2M visa preencher lacunas tanto na literatura quanto na prática, fornecendo um *framework* estruturado e multidimensional para avaliar e aprimorar a GD.

Abordando esses desafios persistentes, os achados-chave deste estudo buscam reforçar o papel da GD como um facilitador estratégico, apoiando tanto a qualidade dos dados quanto a conformidade regulatória. Com base nessa fundamentação e reconhecendo a necessidade de um *framework* acessível, este estudo propõe um modelo de maturidade em GD composto por seis dimensões-chave para apoiar a avaliação e o avanço da maturidade organizacional.

1.4 Objetivos

O objetivo geral deste estudo é desenvolver e validar o DG2M, fornecendo um *framework* abrangente de práticas e caminhos de melhoria para capacitar as organizações a avaliar e aprimorar suas práticas de GD.

Os objetivos específicos deste estudo são os seguintes:

- Sintetizar o conhecimento sobre maturidade em GD, identificando modelos, princípios e práticas recorrentes por meio de uma Revisão Sistemática da Literatura (Sytematic Literature Review) (RSL).
- Desenvolver o DG2M, um modelo de maturidade em GD mais acessível, fundamentado nas evidências da RSL.
- Aprimorar e validar a aplicabilidade do DG2M por meio de um *survey* com perguntas objetivas e abertas, coletando *feedback* de profissionais da área de dados.

1.5 Resultados Esperados da Pesquisa

Esta pesquisa visa gerar os seguintes resultados:

- O desenvolvimento e a validação de um modelo de maturidade em GD (DG2M), validado por meio de avaliações de especialistas.
- Um instrumento de avaliação projetado para permitir que as organizações identifiquem e corrijam sistematicamente as lacunas nas práticas de GD.
- Uma contribuição para a literatura acadêmica sobre maturidade em GD, facilitando a integração dos avanços teóricos com a implementação prática.

O modelo DG2M oferece um *framework* baseado em evidências. Sua validação por revisão de profissionais em dados fortalece tanto sua base teórica quanto sua relevância prática. Além disso, o DG2M serve como uma referência para *benchmarking* em futuras pesquisas e iniciativas de GD.

1.6 Contribuição Científica

Este trabalho busca avançar na área de GD ao propor um modelo de avaliação de maturidade de dados, um tema ainda com pontos a serem explorados na literatura, como a adaptabilidade a diversos contextos organizacionais e a complexidade inerente ao tema. A contribuição científica reside no desenvolvimento de uma abordagem baseada em evidências que pode ser utilizada para *benchmarking* e melhoria contínua em GD.

1.7 Publicações dessa Dissertação

Esta dissertação gerou até o momento o artigo científico:

1. Data Governance Maturity Models and Practices: A Systematic Literature Review, que foi aceito e vai ser publicado na The 29th European Conference on Advances in Databases and Information Systems, <https://adbis2025.github.io/>.

1.8 Metodologia de Pesquisa e Processo Metodológico

A condução desta pesquisa foi realizada por meio de uma RSL e um *survey*. O macro-processo metodológico empregado é detalhado na Figura 1.1, apresentando as etapas seguidas em notação Notação de Modelagem de Processos de Negócio (Business Process Model and Notation) (BPMN).

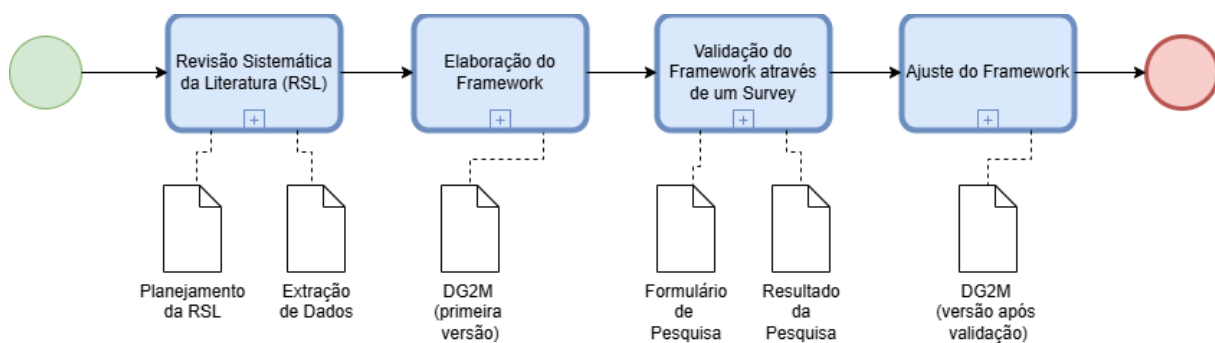


Figura 1.1: Mapeamento dos processos realizados neste projeto em BPMN

Primeiramente, foi condizida uma RSL, conforme as diretrizes de Kitchenham and Charters [29], com o objetivo de identificar os principais modelos e *frameworks* de avaliação da maturidade de dados, bem como práticas de GD, disponíveis na literatura. Este processo incluiu a definição de perguntas de pesquisa, a identificação e seleção de estudos relevantes, além da extração e síntese dos dados. A RSL permitiu mapear as práticas de GD e os modelos de avaliação de maturidade existentes.

Com base nos resultados da RSL, foi elaborada a primeira versão da proposta de modelo de maturidade em GD. Subsequentemente, realizou-se um *survey* com profissionais atuantes na área de dados, seguindo as orientações de Ghazi et al. [16]. O objetivo principal do *survey* foi validar o DG2M proposto, submetendo-o ao crivo de profissionais da área.

Ghazi et al. [16] detalham que a realização de um *survey* compreende oito etapas: definição dos objetivos, identificação e seleção do público-alvo e amostra representativa, escolha do plano de amostragem, desenho e avaliação do instrumento de coleta (questionário), coleta e análise dos dados, extração de conclusões, e documentação e relato dos resultados.

Após a conclusão do *survey*, os resultados foram analisados e utilizados para revisar e aprimorar o DG2M. Essas revisões subsidiaram o desenvolvimento do modelo final, que foi refinado e ajustado de acordo com as descobertas obtidas a partir do *survey*.

1.9 Disponibilização dos Dados

Todo o material produzido durante essa pesquisa está disponível no Zenodo. O repositório sobre o RSL pode ser acessado em: <https://zenodo.org/records/16794340>. Já o repositório com o DG2M e o *survey* pode ser encontrado em: <https://zenodo.org/records/16794340>.

1.10 Organização do Documento

Este documento está organizado em seis capítulos, além deste Capítulo de Introdução:

- **Capítulo 2 — Fundamentação:** Este capítulo explora os conceitos fundamentais relacionados ao tema do trabalho, fornecendo a base teórica necessária para a pesquisa. Além disso, apresenta e analisa estudos da literatura que abordam modelos de maturidade de dados e GD, destacando como esses temas têm sido tratados por outras pesquisas e identificando lacunas e oportunidades para contribuição no campo;
- **Capítulo 3 — Revisão Sistemática da Literatura (RSL):** Este capítulo discute os resultados da RSL e apresenta os achados que subsidiarão a construção do DG2M;
- **Capítulo 4 — Proposta de *Framework* para Avaliação de Maturidade da Governança de Dados (DG2M):** Este capítulo apresenta o DG2M que foi desenvolvido para a avaliação de maturidade em GD. Esse produto é um *framework*, que possui dimensões, aspectos, níveis e papéis que suportam esse processo de avaliação;
- **Capítulo 5 — Desenvolvimento do DG2M:** Este capítulo apresenta o desenvolvimento do DG2M, desenvolvido com base na RSL. Além disso, o capítulo descreve a condução de um *survey* com especialistas na área de dados, cujo objetivo foi validar o *framework*, bem como identificar oportunidades de melhoria. Por fim, são analisados os dados coletados no *survey*, os quais foram utilizados para refinar e aprimorar o *framework* proposto;

- **Capítulo 6 — Discussão:** Este capítulo analisa as implicações, contribuições, resultados, entregas e possíveis ameaças à validade dos resultados obtidos neste trabalho;
- **Capítulo 7 — Conclusão:** Este capítulo conclui os aspectos levantados no projeto, destacando os principais achados e contribuições encontradas e sugere direções para pesquisas futuras na área de GD, com base no *framework* elaborado.

Capítulo 2

Fundamentação

2.1 Governança de Dados (GD)

A DAMA, uma organização sem fins lucrativos dedicada à elaboração de padrões para a gestão e gerenciamento de dados, publicou a segunda edição do Data Management Body of Knowledge (DMBOK)[17]. O DMBOK oferece uma estrutura, que aborda as principais áreas de conhecimento relacionadas à Gestão de Dados, as quais podem ser visualizadas na Figura 2.1.

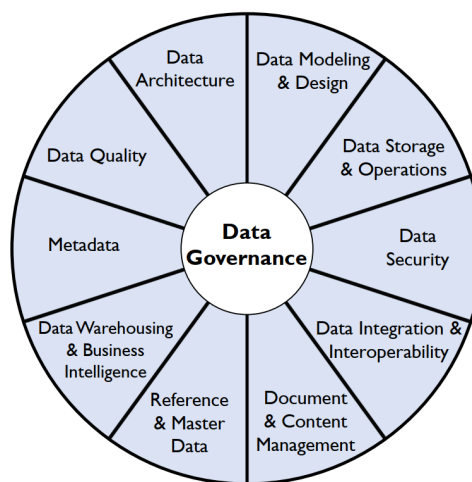


Figura 2.1: DAMA Wheel com as Áreas de Conhecimento relacionadas à Gestão de Dados [17]

A GD é definida por DAMA [17] como a execução de autoridade, controle e tomada de decisão compartilhada sobre o gerenciamento de ativos de dados. Ela fomenta a colaboração entre diferentes áreas organizacionais, permitindo que os dados sejam geridos como um ativo estratégico.

Este modelo formaliza políticas, padrões e procedimentos, além de incorporar o monitoramento da conformidade [1]. Os autores ressaltaram que a GD visa maximizar o valor dos dados, ao mesmo tempo em que minimiza os custos e riscos associados ao seu uso. Isso se torna especialmente relevante diante do aumento exponencial no volume de dados e das crescentes exigências regulatórias, como a GDPR e a LGPD. Assim, a GD especifica claramente os direitos de decisão e as responsabilidades no processo decisório relacionado aos dados da organização[1].

2.1.1 Qualidade de Dados

Kang et al. [27] destacaram que a qualidade dos dados tem recebido atenção significativa em áreas como gestão da informação, Internet das Coisas (Internet of Things) (IoT), manufatura digital, saúde e gestão de prognósticos. Diversas definições de qualidade de dados surgiram nesses contextos.

Além disso, Kang et al. [27] apontaram que, além das definições encontradas na literatura acadêmica, a norma internacional ISO 8000-2:2022[24] define a qualidade de dados como o grau em que suas características inerentes atendem aos requisitos estabelecidos. IBM [23], por sua vez, definiram a qualidade de dados com base em critérios como precisão, completude, validade, consistência, unicidade, atualidade e adequação ao propósito.

Por outro lado, Khatib et al. [28] afirmaram que o Gerenciamento em Qualidade de Dados (Data Quality Management) (DQM) envolve um conjunto de processos destinados a garantir altos padrões de qualidade da informação, desde a obtenção dos dados até a aplicação de técnicas inovadoras de processamento em *data warehouses*, assegurando a qualidade dos dados coletados. Adicionalmente, Khatib et al. [28] explicaram que, entre as políticas e técnicas utilizadas no DQM, destaca-se a análise das informações com base no valor que elas trazem para os negócios. Tais políticas são essenciais para evitar que dados incorretos comprometam o conjunto de dados como um todo.

2.1.2 Modelos de Maturidade

Aguiar et al. [2] observaram que, desde a introdução do Modelo de Maturidade em Capacitação (Capability Maturity Model) (CMM), os modelos de maturidade proliferaram em diversos domínios, impulsionados pela sua eficácia no alinhamento de objetivos organizacionais e no aprimoramento do desempenho. Esses modelos são conceituados como *frameworks* estruturados que avaliam as capacidades organizacionais e guiam a progressão através de níveis de maturidade definidos[2]. Ao apoiar a tomada de decisões estratégicas e fomentar a melhoria contínua, eles permanecem altamente relevantes, garantindo

seu desenvolvimento e adoção contínuos tanto na pesquisa acadêmica quanto na prática profissional[2].

2.2 Trabalhos Correlatos

Al-Ruithe et al. [3] realizaram uma RSL sobre GD em contextos de nuvem e não-nuvem, enfatizando seu status subdesenvolvido e a necessidade de abordagens estruturadas. Contudo, esse trabalho não propõe nem valida um modelo de maturidade específico.

Similarmente, Bližnák et al. [8] forneceram uma síntese atualizada da literatura (2017–2023), destacando ambiguidades definicionais e a evolução dos *frameworks* de maturidade, especialmente à luz de desenvolvimentos regulatórios como a GDPR e as normas Organização Internacional de Normalização (International Organization for Standardization) (ISO). Sua revisão, embora abrangente, permanece conceitual, sem instanciar ou testar empiricamente um modelo.

Ribeiro et al. [38] exploraram a GD sob a ótica da sustentabilidade, propondo um *framework* conceitual baseado em análise bibliométrica. Sua contribuição é temática e estratégica, focando na geração de valor a longo prazo em vez de níveis de maturidade organizacional.

Nesse contexto, Setiadi et al. [39] destacaram a importância da qualidade dos dados na tomada de decisões organizacionais, com foco particular na gestão de recursos humanos. Os autores argumentaram que dados de baixa qualidade podem resultar em decisões equivocadas e na perda de oportunidades de negócios. Além disso, o estudo avaliou o nível de maturidade da gestão da qualidade dos dados no departamento de RH de uma instituição de ensino superior, utilizando o *framework* de Loshin. A metodologia adotada envolveu a aplicação de um questionário dividido em duas partes: a primeira composta por 10 perguntas abertas sobre gestão de qualidade de dados e a segunda com 111 perguntas fechadas baseadas no *framework* de Loshin, cobrindo oito áreas-chave, como expectativas de qualidade, governança e tecnologia. Os autores classificaram as respostas do questionário em cinco níveis de maturidade, variando de *não executado* a *executado*. Além disso, o estudo identificou a falta de uma estrutura organizacional clara e de uma documentação adequada para a gestão da qualidade dos dados. Diante dos resultados, os autores recomendam a criação de uma equipe dedicada à gestão da qualidade dos dados e o desenvolvimento de uma documentação. Essas ações são vistas como fundamentais para aumentar a maturidade e melhorar a qualidade dos dados.

O estudo de [39] foca em uma organização específica e na avaliação de sua gestão de qualidade de dados. Embora os escopos tenham diferenças, ambos os estudos compar-

tilham o interesse pela análise da maturidade de dados e na melhoria das práticas de GD.

Kurniawan et al. [30] investigaram a importância da GD nas organizações e fez um estudo de caso no escritório de Tecnologia da Informação (TI) de uma auditoria governamental. O estudo utiliza o Stanford Data Governance Maturity Model para avaliar a maturidade da GD, medindo-a em três dimensões principais: Pessoas, Políticas e Capacidades. A pesquisa seguiu um processo estruturado em quatro etapas: entrevistas iniciais com membros do do escritório de TI, análise dos dados coletados, entrevistas de confirmação com os líderes da organização e, por fim, a elaboração de um relatório final. Os autores aplicaram questionários a 20 funcionários-chave, abrangendo administradores e administradores de banco de dados, com o objetivo de avaliar a maturidade em seis componentes da governança: conscientização, formalização, metadados, curadoria, qualidade de dados e dados mestres. O estudo recomendou melhorias contínuas para elevar o nível de maturidade da GD.

Harwanto and Hidayanto [21] investigaram a GD no setor público, com um estudo na Diretoria Geral de Correções (DGC) da Indonésia. A organização enfrentava desafios na implementação de um sistema integrado de gestão de casos criminais, no qual a qualidade dos dados desempenha um papel importante. O estudo aplicou o Stanford Data Governance Maturity Model, que classifica a maturidade em cinco níveis, e utilizou questionários e entrevistas com líderes de TI. A pesquisa avaliou duas dimensões principais: fundação, que abrange a conscientização e a formalização, e projeto, que inclui governança e qualidade de dados. Os resultados indicaram que a DGC estava em um nível de maturidade inicial, com uma média de 1,93. O estudo identificou uma lacuna significativa entre o nível atual e o desejado pela organização, que seria o nível três (definido). Os autores sugeriram 30 recomendações, como a criação de políticas de governança e o aumento da conscientização sobre a importância da qualidade de dados.

Por fim, Gunawong [18] exploraram o papel da GD em problemas complexos, utilizando um estudo de caso no sistema de saúde da Tailândia, focado no vício em ópio. Os autores abordam a GD em países de baixa e média renda, onde as infraestruturas são limitadas. A metodologia incluiu observação participante, análise de documentos governamentais e entrevistas. O estudo revelou que, apesar da ausência de uma estrutura formal de GD, estratégias como consenso e integração flexível de organizações trouxeram resultados positivos. Os autores identificaram incertezas sobre a qualidade dos dados, seu ciclo de vida e a gestão de metadados devido à falta de uma governança formal. Os autores sugerem que uma governança mais estruturada é necessária para enfrentar esses desafios.

Mouhib et al. [32] propuseram um modelo de maturidade baseado em uma revisão da

literatura e validado por meio de *surveys* com profissionais de dados. Esse estudo foca na maturidade de projetos de Big Data.

Malacaria et al. [31] apresenta a implementação do Sistema Consensual de Avaliação de Maturidade de Big Data (Consensual Big Data Maturity Assessment System) (CBDAS) em uma multinacional do setor de bens de consumo, visando entender o estágio de maturidade em Big Data e guiar iniciativas em GD. O CBDAS é organizado em oito domínios relevantes para o negócio e utiliza o Processo Analítico Hierárquico (AHP) para ponderar a relevância de cada domínio, identificando as prioridades de intervenção e gerando um roteiro de ações para preencher lacunas de maturidade e aprimorar processos.

Tabela 2.1: Comparação entre Trabalhos Relacionados versus Trabalho Proposto

Estudo	Contexto	Metodologia	Tópicos em GD
Al-Sai et al. [4]	Modelo de maturidade em GD para Big Data	Análise da Literatura	Maturidade em GD
Al-Ruithe et al. [3]	Revisão da GD em contextos de nuvem e não-nuvem	RSL	GD, desafios de implementação, fatores de sucesso e barreiras
Kurniawan et al. [30]	GD no escritório de TI de uma auditoria governamental	Estudo de caso utilizando o Stanford Data Governance Maturity Model, com questionários e entrevistas	Pessoas, políticas, capacidades, curadoria, qualidade de dados, dados mestres
Setiadi et al. [39]	Gestão da qualidade de dados no RH de uma instituição de ensino superior	Questionário baseado no <i>framework</i> de Loshin e entrevistas	Governança, dimensões de qualidade, expectativas de qualidade, tecnologia, padrões de dados, políticas, protocolos de qualidade
Ribeiro et al. [38]	GD e sustentabilidade	RSL baseada em análise bibliométrica, proposta de <i>framework</i> conceitual	GD, sustentabilidade, valor de dados a longo prazo, tecnologias de suporte
Continua na próxima página			

Tabela 2.1 (Continuação)

Estudo	Contexto	Metodologia	Tópicos em GD
Malacaria et al. [31]	Proposição de um modelo de maturidade em GD para o contexto de Big Data	Proposta de modelo de maturidade	Modelo de maturidade em GD, Big Data
Harwanto and Hidayanto [21]	GD na Diretoria Geral de Correções da Indonésia	Uso do Stanford Data Governance Maturity Model, com questionários e entrevistas com líderes de TI	Conscientização, formalização, governança, qualidade de dados, dados mestre
Gunawong [18]	GD no sistema de saúde da Tailândia, focando em problemas complexos e infraestruturas limitadas	Estudo de caso qualitativo, com observação participante e entrevistas	Qualidade de dados, ciclo de vida dos dados, metadados
Mouhib et al. [32]	Avaliação da maturidade de projetos de Big Data em empresas norte-africanas	Revisão da literatura, <i>survey</i> com especialistas e aplicação de AHP para cálculo de pesos de domínio	Alinhamento estratégico, dados, pessoas, governança, tecnologia, metodologia
Bližnák et al. [8]	Revisão da literatura recente (2017-2023) sobre GD, com foco em definições e evolução de <i>frameworks</i> frente a regulamentações como GDPR e normas ISO	RSL	Definição de GD, ambiguidade e evolução, regulamentações (GDPR, Regulamento Inteligência Artificial (Artificial Intelligence Act) (AI Act) ¹), práticas de implementação
Continua na próxima página			

¹<https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai>

Tabela 2.1 (Continuação)

Estudo	Contexto	Metodologia	Tópicos em GD
Trabalho Proposto - DG2M	Proposição de um modelo de avaliação de maturidade em GD	RSL, desenvolvimento de <i>framework</i> e <i>survey</i> com profissionais de dados	Maturidade em GD

Para conhecimento, foi elaborada a Tabela 2.1, a qual possui uma tabela comparativa entre os trabalhos correlatos e o trabalho proposto. Este estudo distingue-se pela integração de uma RSL, o desenvolvimento de um modelo de maturidade em GD (DG2M), e sua validação por meio de um *survey* com profissionais da área de dados. Comparado a trabalhos anteriores, oferece tanto fundamentação teórica quanto uma abordagem empírica. Além disso, este estudo avança no campo ao validar um modelo de maturidade em GD através de uma RSL e de pesquisa baseada em *survey*, fornecendo um *framework* acessível para avaliação de maturidade.

2.3 Resumo do Capítulo

Este Capítulo explora as definições e conceitos importantes sobre a GD. Em linhas gerais, a GD se refere à autoridade e controle sobre os ativos de dados, garantindo conformidade e uso adequado, conforme estabelecido por políticas e procedimentos estabelecidos[17].

A GD colabora com diferentes áreas organizacionais para garantir que os dados sejam tratados como ativos estratégicos. A implementação de governança eficaz reduz riscos e otimiza o valor dos dados, sendo importante em um cenário de crescente volume de dados e exigências regulatórias [1]. Além disso, o capítulo aborda a importância da qualidade dos dados, enfatizada tanto na literatura acadêmica quanto em normas internacionais[24].

A seção 2.2 apresenta dez estudos que tratam da GD em diferentes contextos, com ênfase em maturidade e qualidade de dados e no final apresenta o diferencial dos trabalhos em relação a este projeto de pesquisa. Esses estudos oferecem uma base para o presente projeto, que visa desenvolver e validar um modelo de avaliação de maturidade para a GD.

Capítulo 3

Revisão Sistemática da Literatura

A RSL é um processo estruturado para identificar, avaliar e interpretar estudos relevantes em uma área de pesquisa. Conforme Kitchenham and Charters [29], ela é dividida em três fases principais. A primeira fase, denominada Planejamento, envolve a identificação da necessidade da revisão, o estabelecimento de objetivos claros e a definição do protocolo, que inclui perguntas de pesquisa, palavras-chave (*strings* de busca), critérios de seleção e uma lista de verificação para a avaliação da qualidade dos estudos. A fase subsequente, Condução, é responsável pela implementação do protocolo, que abrange a seleção, o filtro e a extração de dados dos estudos selecionados. Por fim, a fase de Relato visa documentar os resultados obtidos, geralmente na forma de um artigo científico. No contexto desta pesquisa, os resultados estão sendo apresentados e discutidos nesta dissertação.

Para ilustrar o processo executado, foi elaborado um processo em notação BPMN com uma visão geral da RSL, que pode ser observada na Figura 3.1. Além disso, construiu-se uma visão mais detalhada do processo de planejamento e condução com os resultados de cada etapa. Essa visão pode ser observada na Figura 3.2. Para manter a reprodutibilidade e a transparência, todos os artefatos gerados estão disponibilizados no [Zenodo](#).

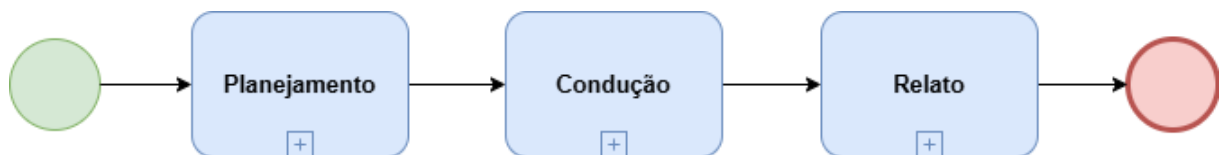


Figura 3.1: Visão Geral do Processo de RSL usando Notação BPMN

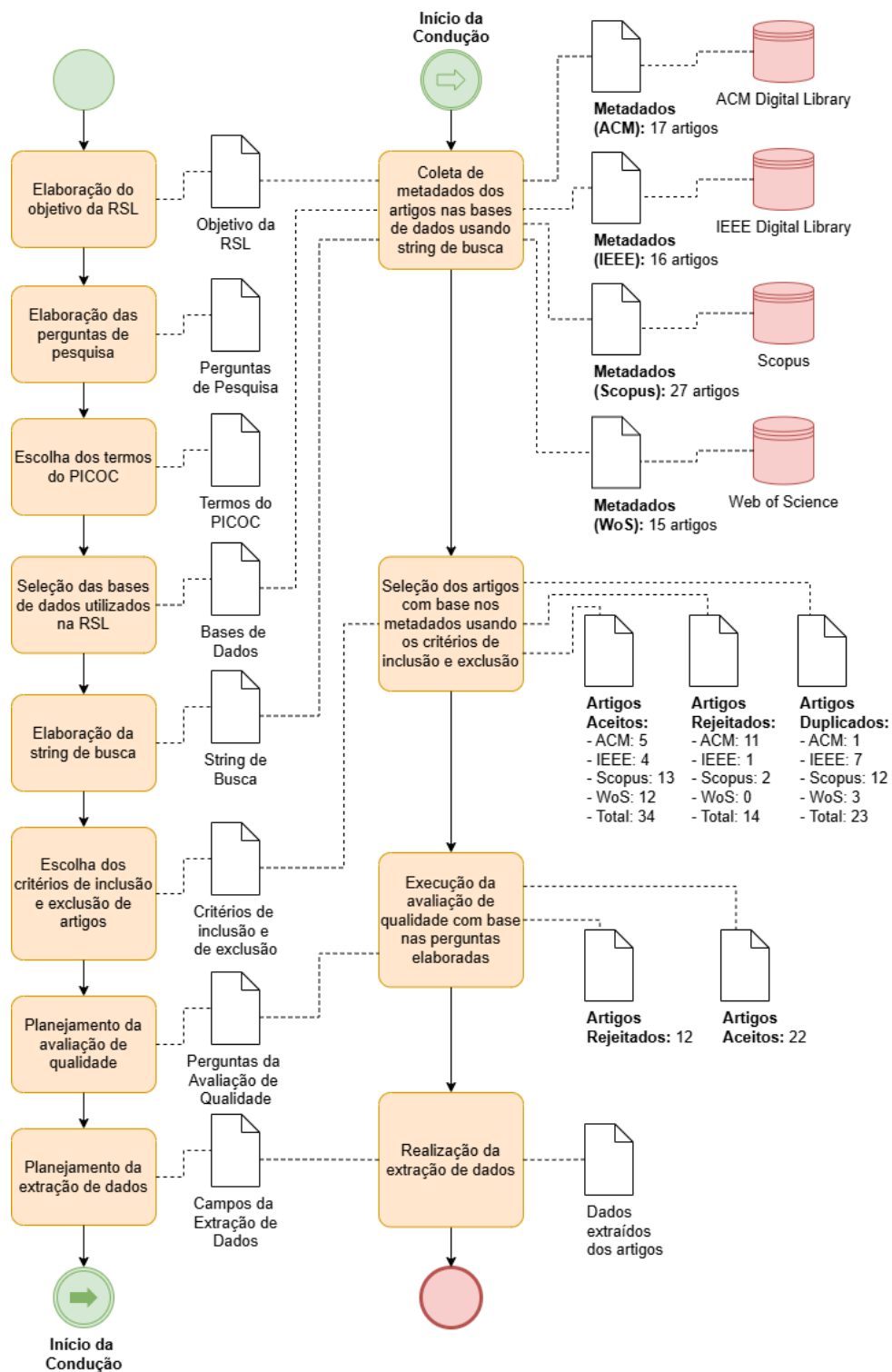


Figura 3.2: Visão Detalhada do Planejamento e da Condução da RSL usando Notação BPMN

3.1 Planejamento

A Figura 3.2 ilustra o processo de planejamento e condução da RSL. Eles foram realizados utilizando a ferramenta Parsifal¹. O processo teve início com a definição do objetivo da RSL, que foi:

Analisar criticamente as metodologias de avaliação de maturidade e as práticas de governança de dados implementadas na academia e na indústria, a fim de construir uma base sólida para o desenvolvimento de um framework de referência que apoie a melhoria dessas práticas nas organizações.

Com base no objetivo da RSL, as perguntas de pesquisa foram formuladas para direcionar a pesquisa, com foco na análise de modelos de maturidade de dados e práticas de GD. A justificativa para cada uma das perguntas está detalhada a seguir:

- RQ.1: Quais avaliações de maturidade em governança de dados são utilizadas na academia e na indústria e quais são suas principais características?** - essa pergunta tem como objetivo identificar e caracterizar as avaliações de maturidade aplicadas nos contextos acadêmico e industrial. Compreender as principais características dessas avaliações é importante para mapear as ferramentas e metodologias existentes, identificar padrões, lacunas e pontos fortes e estabelecer uma base para melhorias e novas propostas.
- RQ.2: Quais ferramentas e práticas de governança de dados são implementadas na academia e na indústria para promover uma governança eficaz?** - essa pergunta busca investigar as ferramentas e práticas utilizadas para implementar e sustentar uma GD eficaz. A análise dessas práticas em diferentes contextos permite compreender como as organizações e a academia enfrentam desafios relacionados a essa área.
- RQ.3: Quais são os principais princípios e as práticas emergentes em avaliações de maturidade e governança de dados?** - dada a natureza evolutiva da GD, esta questão busca consolidar as práticas recorrentes e seus princípios e identificar tendências emergentes que possam contribuir para o avanço da maturidade em GD.

Para garantir que a RSL seja focada em seu objetivo, utilizamos o método População, Intervenção, Comparação, Resultado, Contexto (*Population, Intervention, Comparison, Outcomes, Context*) (PICOC). Este método permitiu estruturar as palavras-chave e seus sinônimos de forma sistemática. Neste contexto, a população (*population*) refere-se ao

¹<https://parsif.al/>

objeto de estudo, a intervenção (*intervention*) é o meio empregado ou efeito causado para atingir um objetivo, a comparação (*comparison*) seria um elemento comparativo, que neste caso não se aplica, o resultado (*outcome*) é o impacto da intervenção, e o contexto (*context*) define o foco do estudo, incluindo suas restrições e limitações. Foram definidas algumas palavras-chave e seus sinônimos, os quais são apresentados na Tabela 3.1.

Tabela 3.1: Palavras do PICOC

PICOC	Palavra-chave	Palavras Relacionadas
Population	"Organization"	"Company", "Academy", "Government", "Industry", "Public Administration"
Intervention	"Maturity Assessment"	Não aplicável
Comparison	Não aplicável	Não aplicável
Outcome	"Framework"	"Assessment"
Context	"Data Maturity"	"Data Governance"

Após a definição dos termos do PICOC (Tabela 3.1), foram selecionadas as bases [ACM Digital Library](#), [IEEE Xplore](#), [Web of Science \(WoS\)](#) e [Scopus](#). Essas bases foram escolhidas devido à sua relevância no campo da engenharia de software [10], por indexarem conferências e periódicos de alto impacto, e por oferecerem suporte à execução direta de *strings* de busca.

Em seguida, a *string* de busca foi elaborada com o objetivo de garantir precisão e abrangência nas bases científicas selecionadas. O refinamento da *string* foi realizado com base na seleção de palavras relacionadas ao tema, especificado no PICOC (Tabela 3.1). As palavras-chave e seus sinônimos foram conectados utilizando operadores booleanos (**AND**, **OR**), de modo a maximizar a recuperação de estudos relevantes. A *string* final foi definida da seguinte forma:

("Academy"OR "Company"OR "Government"OR "Industry"OR "Organization"OR "Public Administration") AND ("Maturity Assessment") AND ("Assessment"OR "Framework") AND ("Data Governance"OR "Data Maturity")

Kitchenham and Charters [29] recomenda a definição de critérios de inclusão e exclusão para garantir a relevância e a qualidade dos estudos selecionados para uma RSL. Esses critérios foram elaborados para que a pesquisa se concentre em estudos alinhados ao objetivo de pesquisa e que tenham a qualidade exigida. Os critérios de inclusão e exclusão podem ser observados na Tabela 3.2.

Os critérios de inclusão da Tabela 3.2 foram cuidadosamente definidos para garantir que os estudos selecionados abordem temas importantes para a pesquisa, como GD e avaliação de maturidade. Esses critérios asseguram que os resultados da RSL sejam diretamente aplicáveis e relevantes para responder às perguntas de pesquisa formuladas.

Tabela 3.2: Critérios de Inclusão e Exclusão para a Revisão Sistemática da Literatura

ID	Tipo	Descrição
I01	Inclusão	Artigos completos publicados como artigos de revista ou anais de conferências e revisados por pares.
I02	Inclusão	Artigos que descrevem ferramentas, práticas, princípios ou <i>frameworks</i> relacionados à GD ou maturidade de dados.
I03	Inclusão	Artigos escritos em inglês, português ou espanhol.
I04	Inclusão	Estudos focados em avaliações de maturidade ou GD na academia ou na indústria.
E01	Exclusão	Artigos que não sejam textos completos, como resumos expandidos, pôsteres, editoriais ou revisões.
E02	Exclusão	Artigos com menos de 5 páginas, exceto se forem explicitamente relevantes.
E03	Exclusão	Artigos escritos em idiomas diferentes de inglês, português ou espanhol.
E04	Exclusão	Estudos que não abordem diretamente aspectos de maturidade ou práticas de GD.
E05	Exclusão	Estudos que não apresentam aplicações práticas ou metodologias claras relacionadas à maturidade ou GD.

Além disso, a exigência de que os artigos sejam revisados por pares e publicados em periódicos ou conferências reconhecidas reforça a qualidade das evidências obtidas.

Por outro lado, os critérios de exclusão da Tabela 3.2 foram estabelecidos para aumentar a relevância e a precisão da análise, reduzindo o risco de inclusão de estudos irrelevantes ou de baixa qualidade. Estudos incompletos, como resumos expandidos, pôsteres ou editoriais, foram descartados por não fornecerem informações suficientes para uma análise detalhada. Trabalhos que não tratam diretamente de aspectos de maturidade ou GD foram excluídos, pois estão fora do escopo da pesquisa. A limitação linguística (inglês, português ou espanhol) foi adotada para garantir que os textos sejam acessíveis e compreensíveis pela equipe de pesquisa. Além disso, estudos com menos de 5 páginas foram excluídos, exceto se considerados explicitamente relevantes, devido à probabilidade de apresentarem análises ou resultados superficiais.

Embora os critérios de inclusão e exclusão sejam suficientes para selecionar trabalhos relevantes, é fundamental aplicar uma avaliação de qualidade para garantir que as práticas analisadas sejam válidas e que seus impactos tenham sido medidos de maneira objetiva. Para isso, foi utilizada a *checklist* de qualidade da Tabela 3.3.

A avaliação de qualidade, apresentada na Tabela 3.3, foi projetada para garantir que os estudos selecionados atendam a critérios de qualidade científica e metodológica. A questão QA1 avalia se as práticas propostas de governança ou maturidade de dados são

Tabela 3.3: Avaliação de Qualidade

QA ID	Questão de Avaliação de Qualidade
QA1	As práticas propostas de governança ou maturidade de dados são pragmáticas e replicáveis?
QA2	A metodologia está clara, adequada e bem definida?
QA3	O estudo inclui um método ou atividade para validar as práticas analisadas?
QA4	Os resultados são apresentados de forma clara e são relevantes para as questões de pesquisa?
QA5	As limitações e ameaças à validade estão explicitamente identificadas e tratadas?

pragmáticas e replicáveis, assegurando que as práticas possam ser aplicadas de fato. A questão QA2 examina a clareza, adequação e definição da metodologia empregada no estudo, garantindo a transparência e a confiabilidade das conclusões apresentadas. QA3 verifica se o estudo inclui um método ou atividade para validar as práticas mencionadas no estudo, sendo relevante para assegurar a robustez dos resultados.

A questão QA4 considera se os resultados são apresentados de forma clara e relevante para as perguntas de pesquisa, assegurando sua utilidade e aplicabilidade na RSL. Por fim, QA5 avalia se as limitações e ameaças à validade foram explicitamente identificadas e tratadas, o que é importante para avaliar a confiabilidade dos resultados e sua aplicabilidade.

Para que um artigo seja aceito na avaliação de qualidade, ele deve atender a todas as cinco questões da *checklist* (QA1 a QA5). Essa abordagem garante que apenas estudos metodologicamente sólidos, cientificamente confiáveis e alinhados com o objetivo da RSL sejam incluídos na análise final. Estudos que não atendem a qualquer uma das questões foram excluídos, reforçando a qualidade e relevância da revisão sistemática.

Para garantir que cada artigo de pesquisa fosse analisado de maneira consistente, foi desenvolvido um formulário estruturado para o processo de extração de dados. Esse formulário foi projetado para capturar informações relacionadas aos estudos analisados, assegurando a uniformidade e a qualidade dos dados coletados. As questões foram cuidadosamente elaboradas para abordar diferentes aspectos do estudo, incluindo informações gerais, objetivos, metodologia, resultados e limitações, permitindo uma análise criteriosa.

O formulário de extração de dados foi projetado com 20 questões organizadas para analisar as múltiplas perspectivas relacionadas aos estudos selecionados. Essas questões foram desenvolvidas em diferentes formatos, incluindo múltipla escolha, respostas abertas e campos numéricos, proporcionando flexibilidade e precisão na coleta de informações. Cada questão foi cuidadosamente elaborada para capturar aspectos específicos e relevantes dos artigos, assegurando uma análise consistente e bem fundamentada. A Tabela

3.4 apresenta os campos definidos para o processo de extração de dados, detalhando as informações coletadas e os tipos de entrada utilizados.

Para a coleta e organização sistemática das informações provenientes dos estudos selecionados na RSL, foi elaborado um Formulário de Extração de Dados, cujos campos estão detalhados na Tabela 3.4. Este formulário foi estruturado para assegurar a abrangência e a relevância dos dados extraídos para os objetivos da pesquisa.

Os campos F01 a F08 abrangem informações gerais e de identificação dos estudos, incluindo ID, título, autoria, ano e local de publicação, afiliação dos autores, tipo de estudo (ex.: periódico, artigo de conferência, capítulo de livro) e fonte de indexação (ex.: ACM, IEEE, Scopus). Essa seção tem a finalidade de realizar a categorização inicial e a rastreabilidade dos artigos.

Os campos F09 a F15 focam nos elementos que definem o propósito e a estrutura argumentativa dos estudos: o objetivo/motivação/meta, as palavras-chave principais, as perguntas de pesquisa formuladas e suas respectivas respostas. A extração desses dados é importante para analisar o alinhamento dos artigos com as questões e objetivos da RSL.

Os campos F16 a F18 detalham a metodologia empregada em cada estudo, especificando os métodos de pesquisa (ex.: estudo de caso, entrevista, *survey*, revisão de literatura), o número de organizações envolvidas e o tipo de análise de dados realizada (ex.: qualitativa, quantitativa, mista). Essas informações fornecem uma compreensão aprofundada do desenho da pesquisa e da natureza das evidências apresentadas.

Por fim, os campos F19 e F20 são dedicados aos resultados/conclusões dos estudos e às limitações ou ameaças à validade identificadas pelos próprios autores. A análise desses campos permite avaliar a contribuição científica de cada trabalho e identificar potenciais lacunas ou direções para futuras investigações.

Tabela 3.4: Formulário de Extração de Dados

ID	Descrição	Tipo de Entrada
F01	ID do artigo.	Numérico
F02	Título do artigo.	Questão Aberta
F03	Autores do artigo.	Questão Aberta
F04	Ano de publicação.	Numérico
F05	Local de publicação (ex. nome do jornal, conferência).	Questão Aberta
F06	Afiliação dos autores.	Questão Aberta
F07	Tipo de estudo (ex.: publicação de revista, artigo de conferência, capítulo de livro, outros).	Múltipla Escolha
F08	Fonte (ex.: ACM, IEEE, WoS, Scopus).	Múltipla Escolha

Continua na próxima página

Tabela 3.4 (Continuação)		
ID	Descrição	Tipo de Entrada
F09	Qual é o objetivo/motivação/meta do estudo?	Questão Aberta
F10	Quais são as principais palavras-chave do estudo?	Questão Aberta
F11	Resumo do estudo.	Questão Aberta
F12	Perguntas de pesquisa abordadas no estudo (Ex. RQ.1, RQ.2, RQ.3).	Múltipla Escolha
F13	Resposta à Pergunta de Pesquisa 1.	Questão Aberta
F14	Resposta à Pergunta de Pesquisa 2.	Questão Aberta
F15	Resposta à Pergunta de Pesquisa 3.	Questão Aberta
F16	Método usado no estudo (ex.: Pesquisa-Ação, Estudos de Caso, Análise de Documentação, Grupo Focal, Análise de Lacunas, Entrevistas, Revisão de Literatura, Análise Quantitativa, Pesquisas).	Múltipla Escolha
F17	Número de organizações no estudo.	Numérico
F18	Tipo de análise de dados usada (ex.: Qualitativa, Quantitativa, Mista).	Múltipla Escolha
F19	Resultados ou conclusões do estudo.	Questão Aberta
F20	Limitações e ameaças à validade identificadas no estudo.	Questão Aberta

3.2 Condução

Após a conclusão da etapa de planejamento da RSL, conforme ilustrado na Figura 3.2, foi iniciada a fase de condução da revisão, seguindo os passos definidos no protocolo estabelecido por Kitchenham and Charters [29]. O primeiro passo consistiu na coleta de estudos nas bases de dados selecionadas, utilizando a *string* de busca desenvolvida durante o planejamento. Essa *string* foi aplicada uniformemente nas quatro bases de dados escolhidas (Scopus, IEEE Xplore, ACM Digital Library e Web of Science). Nenhum filtro adicional foi aplicado, com exceção da ACM Digital Library, onde as coletâneas de artigos foram excluídas, garantindo que apenas artigos individuais, e não coleções, fossem recuperados.

Os critérios de inclusão e exclusão foram aplicados em seguida, conforme as diretrizes estabelecidas na fase de planejamento, para filtrar os estudos. A aplicação inicial dos critérios de exclusão na revisão de literatura revelou as principais razões para a rejeição de artigos. Entre os estudos analisados, 23 foram identificados como duplicados, 13 foram

excluídos por não estarem diretamente relacionados à temática de GD e Maturidade de Dados (E04), e 1 foi removido por não estar disponível em inglês, espanhol ou português (E03), sendo o idioma identificado como chinês. Como resultado, 34 estudos foram aceitos (I01, I02, I03 e I04 - Tabela 3.2), 14 rejeitados (E04 e E03 - Tabela 3.2), e 23 eliminados por duplicidade.

Para garantir a relevância e a qualidade dos estudos incluídos, foi realizada uma avaliação de qualidade dos 34 estudos que atenderam aos critérios de inclusão I01 a I04. A análise foi conduzida com base nas cinco perguntas de qualidade apresentadas na Tabela 3.3. Apenas os estudos que atenderam integralmente a todos os critérios de qualidade foram selecionados para as etapas subsequentes.

3.3 Relato

Após a avaliação de qualidade, 22 estudos foram considerados adequados e selecionados para a etapa de extração de dados, representando as melhores evidências disponíveis para responder às perguntas de pesquisa estabelecidas. A Tabela 3.5 sintetiza os principais aspectos desses estudos. A coluna *Referência* apresenta cada estudo conforme as referências utilizadas no trabalho. A coluna *Base Digital* detalha as bases de dados acadêmicas em que os estudos foram encontrados, como IEEE Xplore, ACM Digital Library, Scopus e Web of Science. Por fim, a coluna *Questões de Pesquisa Abordadas* indica as perguntas de pesquisa que os estudos ajudam a responder, seja de forma integral ou parcial, ressaltando sua relevância para os objetivos desta revisão.

Tabela 3.5: Estudos Selecionados para a Extração de Dados

Referência	Base Digital	Questões de Pesquisa Abordadas
[21]	IEEE Xplore	RQ.1, RQ.2, RQ.3
[36]	IEEE Xplore, Scopus	RQ.1, RQ.2, RQ.3
[32]	Scopus	RQ.1, RQ.2, RQ.3
[31]	Web of Science	RQ.1, RQ.2, RQ.3
[19]	Scopus, Web of Science	RQ.1, RQ.2, RQ.3
[5]	Scopus	RQ.2, RQ.3
[43]	IEEE Xplore, Scopus	RQ.1, RQ.2, RQ.3
[30]	IEEE Xplore, Scopus	RQ.1, RQ.2, RQ.3
[35]	Scopus	RQ.1, RQ.2, RQ.3
Continua na próxima página		

Tabela 3.5 (Continuação)		
Referência	Base Digital	Questões de Pesquisa Abordadas
[33]	Scopus	RQ.1, RQ.2, RQ.3
[20]	Scopus, Web of Science	RQ.1, RQ.2, RQ.3
[7]	Scopus, Web of Science	RQ.1, RQ.2, RQ.3
[41]	Scopus	RQ.1, RQ.2, RQ.3
[42]	ACM Digital Library	RQ.2, RQ.3
[22]	ACM Digital Library	RQ.1, RQ.2, RQ.3
[15]	ACM Digital Library	RQ.1, RQ.2, RQ.3
[34]	ACM Digital Library	RQ.1, RQ.2, RQ.3
[25]	IEEE Xplore, Scopus, Web of Science	RQ.1, RQ.2, RQ.3
[12]	ACM Digital Library, IEEE Xplore, Scopus, Web of Science	RQ.1, RQ.2, RQ.3
[40]	Scopus, Web of Science	RQ.1, RQ.2, RQ.3
[14]	Scopus	RQ.1, RQ.2, RQ.3
[37]	IEEE Xplore, Scopus, Web of Science	RQ.1, RQ.2, RQ.3

A Figura 3.3 apresenta a distribuição dos estudos selecionados para extração de dados de acordo com seu ano de publicação. Observa-se que há uma concentração maior de estudos publicados nos anos mais recentes, especialmente em 2021, 2022 e 2024. Enquanto períodos anteriores, como 2014 e 2016, registram uma quantidade reduzida de publicações.

Por sua vez, a Figura 3.4 ilustra a distribuição dos métodos de pesquisa utilizados nos estudos selecionados, considerando que um mesmo estudo pode empregar múltiplos métodos. Nota-se que a RSL foi o método mais frequente, com 15 ocorrências, sendo amplamente utilizada tanto em abordagens qualitativas quanto mistas (qualitativas e quantitativas). Os Estudos de Caso aparecem em seguida, com 14 ocorrências, destacando-se pela predominância em estudos de abordagem mista. Métodos como Entrevistas (11 ocorrências) e Análise de Documentação (5 ocorrências) também foram consideravelmente empregados, evidenciando sua relevância no campo. Por outro lado, métodos como Action Research e Grupos Focais foram utilizados de forma mais restrita, refletindo aplicações específicas em determinados contextos. Dessa forma, observa-se uma diversidade significativa de métodos empregados para investigar a área de maturidade e GD.

Para concluir a análise geral dos estudos, foram avaliadas as limitações e ameaças à validade, que foram cuidadosamente consideradas no processo de análise dos artigos. Um

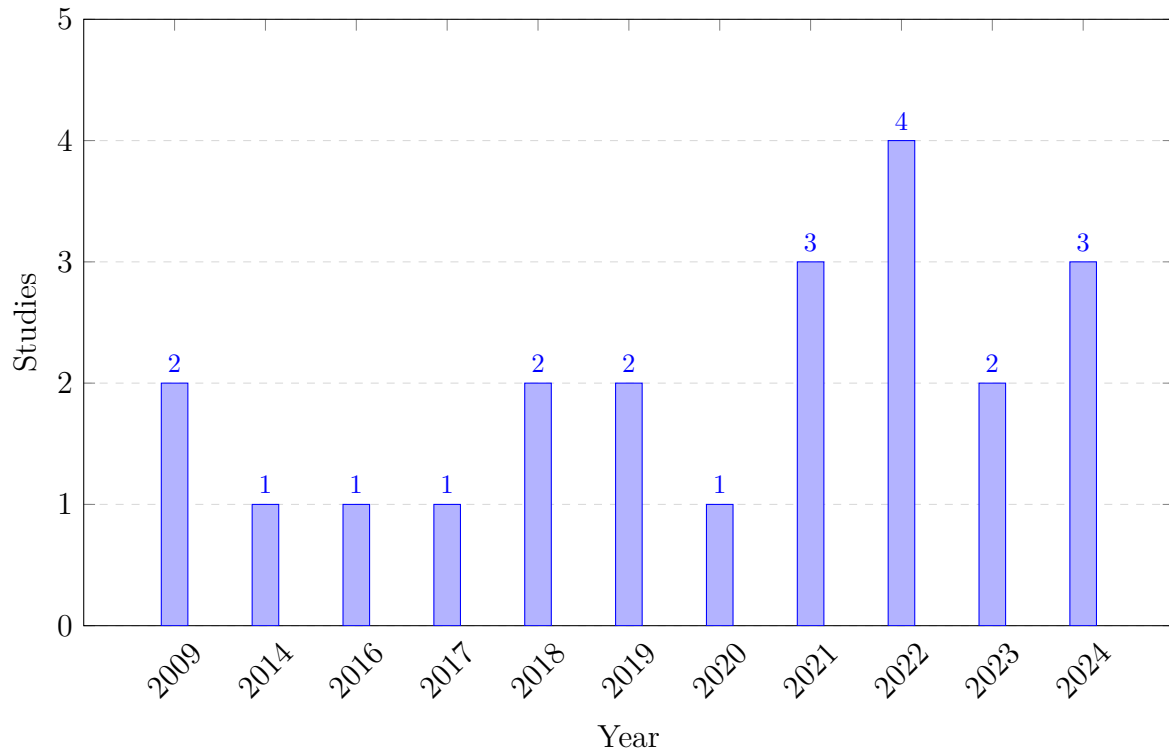


Figura 3.3: Distribuição dos Estudos por Ano de Publicação

aspecto foi o escopo limitado de alguns estudos, conduzidos em organizações ou setores específicos, restringindo a possibilidade de generalização para outros contextos organizacionais e industriais. A ausência de análises comparativas entre diferentes organizações também foi identificada como uma lacuna em alguns estudos, dificultando a identificação de variações ou padrões contextuais. Além disso, o foco predominante em grandes organizações revelou-se um desafio significativo, o que pode reduzir a aplicabilidade dos resultados às Pequenas e Médias Empresas (PMEs). Essas limitações ressaltam a importância de validações adicionais e de estudos mais abrangentes, capazes de fortalecer as evidências disponíveis e ampliar a aplicabilidade dos modelos propostos para maturidade e GD em contextos mais diversos.

3.3.1 RQ.1. Quais avaliações de maturidade em Governança de Dados são utilizadas na academia e na indústria e quais são suas principais características?

Os estudos analisados fornecem uma visão abrangente sobre modelos de maturidade utilizados para a GD. Seguem os exemplos coletados:

- **Stanford Maturity Model:** Este modelo foi amplamente adotado em estudos

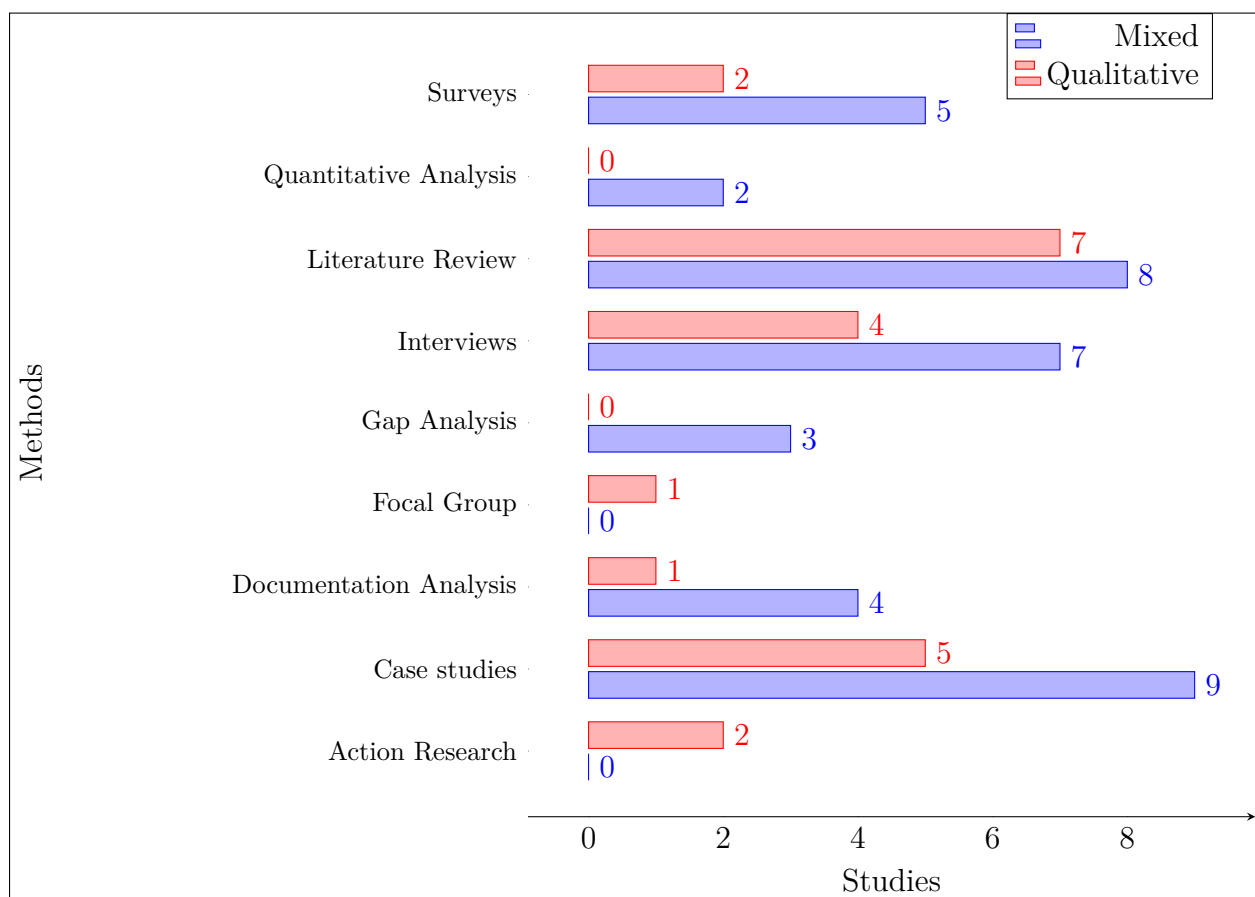


Figura 3.4: Distribuição dos Estudos por Tipo e Método de Pesquisa

como o de Harwanto and Hidayanto [21], Kurniawan et al. [30] e Permana and Su-roso [36]. Ele avalia seis componentes-chave (consciência, formalização, metadados, governança, qualidade de dados e dados mestres) em três dimensões (pessoas, políticas e capacidades). O modelo é estruturado em níveis de maturidade, indo de processos *ad hoc* (nível inicial) até otimização contínua.

- **Loshin's Data Quality Maturity Model:** foi adotado no estudo do Nugraha et al. [35]. Ele avalia oito componentes-chave (expectativas de qualidade, dimensões da qualidade dos dados, políticas, procedimentos, governança, padronização, tecnologia e gestão de desempenho) e estrutura esses componentes em cinco níveis de maturidade (Inicial, Repetível, Definido, Gerenciado e Otimizado).
- **Global Big Data Maturity Model (GBDMM):** Apresentado por Mouhib et al. [32] e Mouhib et al. [33], o GBDMM inclui domínios como estratégia, governança e metodologia. Este modelo se destaca por integrar o método AHP (Analytic Hierarchy Process) para priorizar domínios e propor um *framework* temporal evolutivo.
- **Consensual Big Data Maturity Assessment System (CBDAS):** Introduzido por Malacaria et al. [31], o CBDAS avalia a maturidade de iniciativas de *big data* em oito domínios, como estratégia de dados e interface humana. Ele é aplicado especialmente em empresas de bens de consumo.
- **Data Science Maturity Model (DSMM):** Proposto por Gökalp et al. [19] e Gökalp et al. [20], este modelo é baseado no ISO/IEC 330xx e avalia a maturidade de ciência de dados em manufatura, focando em capacidades estruturadas e alinhamento estratégico.
- **Master Data Management Maturity Model (MD3M):** Utilizado por Iqbal et al. [25], Spruit and Pietzka [40] e Qodarsih et al. [37], o MD3M avalia práticas de gerenciamento de dados mestres com base em cinco níveis de maturidade e 13 áreas de foco, como qualidade e propriedade dos dados.
- **Cloud Data Governance Maturity Model (CDGM):** Proposto por Cheng et al. [12], este modelo endereça desafios específicos de GD em ambientes de computação em nuvem, com categorias como estratégia, segurança e qualidade.
- **Big Data Maturity Model (BDMM):** Descrito por Comuzzi and Patel [14], o BDMM integra dimensões técnicas e gerenciais, cobrindo domínios como governança e alinhamento estratégico.

- **Outros Modelos:** Estudos como Xue et al. [43], Visintin et al. [41] e Bernardo et al. [7] propõem modelos adaptados para contextos específicos, como GD geográficos, dados abertos e integração de *frameworks* para auditoria e garantia.

Esses estudos evidenciam que os modelos de maturidade variam em termos de domínios, níveis e metodologias, sendo amplamente utilizados tanto na academia quanto na indústria para avaliar capacidades organizacionais e propor melhorias na GD.

3.3.2 RQ.2. Quais ferramentas e práticas de governança de dados são implementadas na academia e na indústria para promover uma governança eficaz?

Os estudos analisados apresentam uma variedade de ferramentas e práticas de GD implementadas tanto na academia quanto na indústria, com o objetivo de promover uma GD eficaz. Abaixo, destacam-se os principais exemplos:

- **Formalização de Políticas e Treinamento de Pessoal:** Estudos como Harwanto and Hidayanto [21] e Permana and Suroso [36] destacam a importância de estabelecer políticas documentadas e oferecer treinamento para gestores de dados e *data stewards*. Essas práticas foram recomendadas para lidar com inconsistências e melhorar a governança organizacional.
- **Modelos e Frameworks de Avaliação:** Modelos de maturidade como o GBDMM (Mouhib et al. [32], Mouhib et al. [33]) e CBDAS (Malacaria et al. [31]) são frequentemente utilizados para avaliar e aprimorar a GD. Estes modelos integram ferramentas como o AHP (*Analytic Hierarchy Process*) para priorização de domínios e mapeamento de processos críticos.
- **Estratégias de Gerenciamento de Qualidade de Dados:** A adoção de *frameworks* como o DMBOK (Nugraha et al. [35]) e o ISO 8000 (Bena et al. [5]) é destacada para monitoramento e validação contínuos da qualidade de dados. Essas abordagens incluem processos iterativos como o ciclo “Plan-Do-Check-Act” e o uso de tecnologias emergentes.
- **Integração de Tecnologias Avançadas:** Estudos como Bena et al. [5] e Gökalp et al. [20] enfatizam o uso de tecnologias avançadas, incluindo aprendizado de máquina, IoT e arquiteturas escaláveis para lidar com desafios de integração e heterogeneidade de dados.
- **Ferramentas de Gerenciamento de Metadados e Dados Mestres:** Práticas como a definição de metadados padronizados e o uso de modelos de dados mestres

centralizados foram amplamente abordadas em estudos como Spruit and Pietzka [40], Cheng et al. [12] e Iqbal et al. [25].

- **Governança de Dados Geográficos e Abertos:** O artigo de Xue et al. [43] sugere modelos hierárquicos para gestão de informações geoespaciais e *frameworks* legais para coordenação interdepartamental, enquanto Visintin et al. [41] foca na definição de políticas para dados abertos.
- **Práticas Corporativas e Estratégicas:** Estudos como Comuzzi and Patel [14] sugerem a implementação de estratégias para alinhar iniciativas de Big Data às metas corporativas, bem como o uso de infraestrutura para processamento em tempo real e análises avançadas.
- **Reforço da Segurança e Privacidade:** A gestão de riscos e a conformidade com normas de privacidade aparecem como elementos essenciais, destacados por Cheng et al. [12] e Bernardo et al. [7], para garantir a eficácia das práticas de GD.

Esses achados apontam para uma ampla gama de ferramentas e práticas utilizadas para sustentar a GD, destacando a integração entre estratégias organizacionais, o uso de tecnologias emergentes e a utilização de *frameworks* para a condução dos processos.

3.3.3 RQ.3. Quais são os principais princípios e as melhores práticas emergentes em avaliações de maturidade e governança de dados?

Com base nos estudos analisados, foram mapeados os principais princípios e melhores práticas emergentes em avaliações de maturidade e GD. Podemos destacar:

- **Definição de papéis e ferramentas automatizadas:** Estudos como Harwanto and Hidayanto [21] destacam a importância de definir claramente os papéis organizacionais, utilizar ferramentas automatizadas para coleta de metadados e estabelecer padrões de qualidade para garantir a interoperabilidade e transparência.
- **Formalização de políticas e processos:** A adoção de *frameworks* como o DM-BOK foi apontada em Permana and Suroso [36] como uma prática essencial para formalizar políticas de GD e melhorar a qualidade dos dados por meio de iniciativas de *stewardship*.
- **Ciclos de melhoria contínua e cultura organizacional favorável:** O CBDAS (Malacaria et al. [31]) sugere ciclos iterativos de melhoria baseados em prioridades específicas de domínio, integração de sistemas de inteligência de negócios e desenvolvimento de programas de talentos.

- **Adoção de padrões ISO e métricas quantitativas:** Estudos como Gökalp et al. [19] enfatizam o alinhamento com padrões como o ISO/IEC 330xx² e a adoção de métricas quantitativas para avaliação de qualidade.
- **Enfoque iterativo e visualização de dados:** Bena et al. [5] introduz frameworks como o ciclo *Plan-Do-Check-Act* e ferramentas analíticas, como *dashboards*, para monitoramento contínuo e validação das práticas de governança.
- **Gestão do ciclo de vida dos dados e interoperabilidade:** Xue et al. [43] destaca o gerenciamento integrado do ciclo de vida dos dados e a promoção de interoperabilidade entre plataformas.
- **Alinhamento com objetivos estratégicos e governança adaptativa:** Bernardo et al. [7] e Mouhib et al. [33] reforçam a importância de alinhar iniciativas de dados aos objetivos corporativos, além de utilizar metodologias ágeis e tecnologias avançadas na governança, como aprendizado de máquina.
- **Práticas de melhoria contínua em dados abertos:** Visintin et al. [41] sugere melhorias iterativas em políticas de dados abertos e a adoção de *frameworks* para assegurar qualidade e interoperabilidade.
- **Ciclos iterativos de avaliação de qualidade:** Spruit and Pietzka [40] defende a necessidade de avaliações regulares da qualidade dos dados e o estabelecimento de papéis claros de responsabilidade.

Esses princípios e práticas emergentes fornecem uma base sólida para a evolução contínua da maturidade e governança de dados, destacando a integração de tecnologia, políticas robustas e ciclos iterativos de melhoria como elementos-chave.

3.3.4 Finalização da Revisão Sistemática da Literatura

Para concluir a fase de RSL, a Tabela 3.6 resume os Modelos de Maturidade de Dados identificados na nossa pesquisa. Além disso, os achados deste estudo culminaram na elaboração do artigo *Data Governance Maturity Models and Practices: A Systematic Literature Review*, que foi aceito para publicação no *The 29th European Conference on Advances in Databases and Information Systems*³.

²<https://committee.iso.org/sites/jtc1sc7/home/projects/flagship-standards/isoiec-33000-family.html>

³<https://adbis2025.github.io/>

Tabela 3.6: Modelos de Maturidade de Dados Identificados na Literatura

Ano	Referência	Autores	Modelos de Maturidade de Dados
2024	[5]	Bena et al	ISO 8000-62
2024	[35]	Nugraha et al	DQM, Loshin, Ryu et al, Kirikoglu, Caballero & Piattini
2024	[7]	Bernardo et al	DGMM, McDowall, Zorrilla and Yebenes
2023	[32]	Mouhib et al	GBDMM
2023	[33]	Mouhib et al	GBDMM
2022	[21]	Harwanto et al	Stanford, CMMI, Dataflux, IBM, Kalido
2022	[31]	Malacaria et al	CBDAS
2022	[43]	Xue et al	GB/T 36073-2018
2022	[15]	Song	CMM, R&D, D&A
2021	[19]	Gökalp et al	DSMM
2021	[20]	Gökalp et al	DSCMM, CMMI, ISO/IEC TR 330xx
2021	[41]	Visintin et al	European Open Data Maturity, Dodds and Newman, Solar et al
2020	[34]	Muniz et al	DWMM, DWBP, CMM, OGMM, Open Data Maturity Model
2019	[30]	Kurniawan et al	CMMI, Stanford, IBM, Data Flux, Kalido
2019	[25]	Iqbal et al	MD3M
2018	[36]	Permana et al	CMMI, Stanford
2018	[37]	Qodarsih et al	MD3M
2017	[12]	Cheng et al	CDGM, DMM
2016	[14]	Comuzzi et al	BDMM and other Big Data Models
2014	[40]	Spruit et al	MD3M
2009	[42]	Weber et al	DQM
2009	[22]	Hüner et al	CMMI, EFQM, AIMQ, DQMMM, IQMM, GDQMM

3.4 Ameaças à Validade e Limitações

Este trabalho apresenta certas ameaças à validade comuns em RSL. A primeira ameaça está relacionada à validade de construção, uma vez que a definição do escopo e das *strings* de busca pode ter limitado a recuperação de estudos relevantes. Apesar do uso do método PICOC para estruturar as palavras-chave, publicações potencialmente relevantes podem ter sido excluídas devido à ausência de termos específicos nas *strings* de busca. Além disso, os critérios de inclusão e exclusão adotados, como a limitação de idiomas a inglês, português e espanhol, podem ter resultado na rejeição de estudos significativos oriundos de outras regiões e contextos.

No que diz respeito à validade interna, a seleção de estudos pode ter sido influenciada por vieses, como a priorização de trabalhos mais recentes ou de autores reconhecidos, em detrimento de outros igualmente relevantes. A extração manual dos dados também

representa um risco, pois pode introduzir inconsistências ou interpretações subjetivas. Para mitigar esses riscos, utilizamos um formulário de extração de dados estruturado, visando garantir maior consistência e transparência no processo.

A validade externa também enfrenta desafios, especialmente no que se refere à generalização dos resultados. Muitos dos estudos incluídos concentram-se em grandes organizações ou setores específicos, como o setor público ou indústrias de tecnologia, o que limita a aplicabilidade das conclusões a outros contextos ou organizações de outros portes. Ademais, a predominância de estudos acadêmicos pode não refletir integralmente as práticas da indústria.

3.5 Resumo do Capítulo

Este estudo empregou uma RSL como processo estruturado, dividida em três fases principais: Planejamento, Condução e Relato, conforme preconizado por Kitchenham and Charters [29]. A fase de Planejamento envolveu a definição do objetivo da RSL — analisar criticamente metodologias de avaliação de maturidade e práticas de GD na academia e na indústria para desenvolver um *framework* de referência —, a formulação de perguntas de pesquisa específicas (RQ.1, RQ.2, RQ.3) e a construção de um protocolo de pesquisa. Este protocolo utilizou o método PICOC para estruturar as *strings* de busca, que foram aplicadas em bases de dados como ACM Digital Library, IEEE Xplore, Web of Science e Scopus, e estabeleceu critérios de inclusão e exclusão, além de uma *checklist* de avaliação de qualidade.

Na fase de Condução, os estudos foram coletados, filtrados (eliminando duplicados e estudos irrelevantes ou de baixa qualidade) e 22 artigos foram selecionados para a extração de dados. A fase de Relato, por sua vez, apresenta a síntese dos achados, incluindo a distribuição dos estudos por ano de publicação e métodos de pesquisa e discute as respostas às perguntas de pesquisa, mapeando modelos de maturidade (como Stanford, Loshin's, GBDMM, CBDAS, DSMM, MD3M, CDGM, BDMM), ferramentas e práticas de governança (formalização de políticas, integração de tecnologias avançadas, gerenciamento de metadados) e princípios emergentes (papéis claros, ciclos de melhoria contínua, alinhamento estratégico).

O estudo reconhece limitações relacionadas à validade de construção (escopo da busca), validade interna (vieses na seleção e extração manual) e validade externa (generalização dos resultados de estudos focados em grandes organizações ou setores específicos). Um artigo derivado desta RSL foi aceito para publicação no *The 29th European Conference on Advances in Databases and Information Systems*.

Capítulo 4

Proposta de *Framework* para Avaliação de Maturidade da Governança de Dados (DG2M)

Este capítulo detalha o DG2M, um modelo desenvolvido a partir do conhecimento coletado na RSL e na validação por meio de um *survey*. O DG2M visa diagnosticar o nível de maturidade de organizações em GD e fornecer diretrizes para o aprimoramento contínuo.

4.1 Descrição geral do produto

O DG2M é um modelo de maturidade em GD que descreve seis níveis de maturidade organizacional, do Nível 0 (Não Iniciado) ao Nível 5 (Otimizado), com foco em cinco dimensões: Formalização e Processos de Gestão de Dados, Infraestrutura para Dados, Cultura, Pessoas e Estrutura Organizacional, Qualidade dos Dados e Estratégia e GD. Há também uma dimensão opcional para Dados Abertos, relevante para organizações que buscam transparência. A primeira versão completa do *DG2M* é apresentada no site <https://luisfccardoso.github.io/dg2m-v1/>, desenvolvido com base na RSL, e a segunda versão completa, <https://luisfccardoso.github.io/dg2m/>, após realizada a validação por meio de um *survey*.

Cada nível de maturidade é detalhado com base nas dimensões, descrevendo o estado da organização nos aspectos de pessoas, políticas e capacidades. O modelo também inclui um conjunto de melhorias para cada nível, sugerindo as organizações sobre as próximas etapas para avançar em sua maturidade. A descrição do modelo também inclui papéis e suas responsabilidades na GD, agrupados por área de atuação (como, Estratégia e Liderança, Gestão e Curadoria de Dados, entre outras).

4.1.1 Níveis de Maturidade

O modelo DG2M define seis níveis de maturidade:

- **Nível 0: Não Iniciado:** Neste estágio, a organização não possui iniciativas nem consciência em relação à GD ou à ciência de dados. Não há reconhecimento do valor estratégico dos dados, e as decisões são tomadas sem qualquer apoio informacional estruturado. Organizações neste nível carecem completamente de práticas e estruturas para a gestão de dados. A ausência de processos e de qualquer esforço voltado à governança dificulta que se obtenham os benefícios que uma gestão eficaz de dados poderia proporcionar.
- **Nível 1: Inicial – Ad Hoc, Não Estruturado e/ou em Aprendizado:** No nível inicial, as primeiras iniciativas relacionadas à gestão de dados começam a surgir, mas de maneira desorganizada e sem alinhamento estratégico. A organização está em uma fase de aprendizado, explorando ferramentas e tecnologias sem uma visão estruturada de longo prazo. As decisões são reativas e baseadas na intuição, com projetos exploratórios ocasionais voltados à compreensão dos possíveis usos dos dados. No entanto, a falta de documentação e de processos padronizados limita a evolução dessas iniciativas.
- **Nível 2: Emergente – Iniciativas Incipientes, Transformação de Ativos e Processos Repetíveis:** Além dos avanços do estágio anterior, neste nível os dados passam a ser reconhecidos como um ativo estratégico, e a organização começa a utilizar ferramentas como *dashboards* e processos básicos para apoiar a tomada de decisões. O desenvolvimento de protótipos e os primeiros investimentos tecnológicos caracterizam esse nível. Algumas áreas começam a padronizar processos e as decisões passam a se basear em informações extraídas dos dados. Contudo, a adoção dessas práticas ainda ocorre de maneira isolada dentro da organização, sem uma governança formal consolidada.
- **Nível 3: Definido – Estruturado e Padronizado:** Com base nos avanços dos estágios anteriores, neste nível a GD torna-se parte integrante da estratégia organizacional, com processos bem definidos e documentados. A organização estabelece papéis e responsabilidades claros para a gestão de dados, promovendo uma abordagem estruturada e integrada entre diferentes áreas. A padronização dos processos permite maior eficiência e confiabilidade no uso dos dados. Além disso, ferramentas avançadas de análise e monitoramento começam a ser implementadas, aumentando a capacidade da organização de tomar decisões baseadas em dados.

- **Nível 4: Gerenciado – Medido e Preditivo:** Além dos avanços do estágio anterior, neste nível a organização atinge um controle significativo sobre seus processos de GD, utilizando métricas quantitativas e indicadores estatísticos para monitorar, gerenciar e prever resultados. As decisões tornam-se totalmente orientadas por dados, eliminando a subjetividade e reduzindo os riscos operacionais. A governança baseia-se em indicadores-chave de desempenho (KPIs), assegurando conformidade contínua e eficiência. Sistemas e processos são monitorados em tempo real para garantir que o desempenho esteja adequado e alinhado às necessidades do negócio.
- **Nível 5: Otimizado – Inovação e Melhoria Contínua:** Avançando, no nível mais avançado a GD e a ciência de dados tornam-se ativos estratégicos, impulsionando a inovação e promovendo um ciclo de melhoria contínua baseado em aprendizado e análise preditiva. A organização aprimora continuamente seus processos, métricas e análises de dados, alinhando-os ao máximo aos objetivos estratégicos, sempre mantendo conformidade com as regulamentações pertinentes e práticas de mercado adequadas à realidade da organização. Soluções inovadoras integram *stakeholders* e cadeias de valor, promovendo colaboração e sinergia em toda a estrutura da organização. Tecnologias avançadas e automação são aplicadas estrategicamente, quando necessário, para otimizar operações e gerar vantagem competitiva. Explora-se o potencial de novos produtos e serviços baseados em dados, alinhados aos objetivos do negócio. Essa abordagem consolida uma cultura de dados interna, facilitando a colaboração com o ecossistema externo, conforme a necessidade.

4.1.2 Dimensões do Modelo

O DG2M é composto por seis dimensões, cada uma com sua própria descrição e critérios de avaliação. A versão completa está no site <https://luisfccardoso.github.io/dg2m/>. As descrições de cada dimensão são apresentadas abaixo.

- **Formalização e Processos de Gestão de Dados:** Esta dimensão avalia práticas relacionadas à gestão de metadados, dados mestres, arquitetura de dados, modelos de dados e classificação de dados. Seu objetivo é avaliar o nível de maturidade organizacional no uso e na GD, promovendo eficiência operacional, alinhamento estratégico e conformidade regulatória. Além disso, esta dimensão inclui abordagens de desenvolvimento iterativas, práticas de metadados, controle de ciclo de vida, integração DevOps, padrões de documentação, licenciamento e identificação de riscos. Além disso, a dimensão destaca a participação ativa dos diversos papéis envolvidos, como gestores e consumidores de dados, e o estabelecimento de mecanismos eficazes e seguros para o compartilhamento de dados ao longo de todo o ciclo de vida do

tratamento. Também enfatiza o alinhamento de processos internos de dados com políticas internas e externas (como normas, políticas e modelos específicos de privacidade e segurança da informação), bem como a implementação de mecanismos de monitoramento de processos e acompanhamento de desempenho.

- **Infraestrutura para Dados:** Esta dimensão avalia a capacidade do ambiente tecnológica (hardware, software e arquitetura) da organização e sua adequação para armazenar, processar e integrar dados de forma eficiente. O foco está na prontidão e adequação das plataformas e ferramentas de TI ao ambiente tecnológico que a organização usa (seja *on-premises*, nuvem pública, privada ou híbrida), incluindo a consideração de arquiteturas distribuídas em borda, visando suportar efetivamente a governança e o uso estratégico dos dados dentro do contexto do negócio. A avaliação abrange aspectos relacionados à escalabilidade, segurança e conformidade da infraestrutura de TI utilizada na gestão de dados. Além disso, inclui a avaliação da prontidão da infraestrutura, agilidade, escalabilidade, consistência arquitetural, uso de plataformas em nuvem (quando necessário), adoção de padrões de metadados e mecanismos de registro e relatórios. Considerações sobre segurança, gestão de riscos e capacidades de integração também são essenciais para garantir robustez e alinhamento regulatório ao ambiente tecnológico.
- **Cultura, Pessoas e Estrutura Organizacional:** Esta dimensão avalia o papel de indivíduos e equipes na gestão, curadoria e análise de dados, com ênfase na colaboração, responsabilidade e desenvolvimento de competências. A avaliação abrange desde a conscientização inicial sobre a importância estratégica dos dados e o letramento em dados, até a consolidação de uma cultura organizacional orientada para a inovação baseada em dados. Inclui avaliação da colaboração e integração entre as áreas de Negócio e de Tecnologia da Informação e Comunicação e outras disciplinas, competências da força de trabalho, clareza de papéis, cultura de colaboração, confiança, patrocínio, capacitação e conscientização sobre dados abertos. Reflete estruturas de responsabilização e práticas de comunicação.
- **Qualidade dos Dados:** Esta dimensão avalia a capacidade organizacional de garantir alta qualidade dos dados por meio de critérios bem definidos e práticas padronizadas. Seu escopo inclui avaliação, melhoria contínua, conscientização e monitoramento da qualidade dos dados, considerando disponibilidade, acessibilidade, confiabilidade, perfilamento de dados, políticas de privacidade e segurança, mecanismos de *feedback* e integração com ferramentas de plataformas de dados. Também considera o alinhamento com estruturas de governança, expectativas de qualidade e práticas de melhoria.

- **Estratégia e GD:** Esta dimensão avalia como as organizações alinham seus objetivos estratégicos com práticas estruturadas de GD. Esse alinhamento é essencial para garantir a segurança, privacidade, qualidade e uso ético dos dados, além de estabelecer estruturas e diretrizes que possibilitem uma governança eficiente e sustentável. Para a Segurança da Informação, a organização adota um modelo apropriado a sua organização. Além disso, esta dimensão considera o desenvolvimento de metodologias para a implementação de projetos, a integração entre equipes de TI e de negócio, e a adoção de métricas que apoiam a tomada de decisões orientada por dados. Para refletir uma maturidade abrangente em governança, ela também incorpora indicadores como alocação orçamentária, patrocínio executivo, visão estratégica, planejamento de dados, alinhamento de políticas e integração na tomada de decisões. Inclui ainda a definição de papéis de curadoria, participação ativa na governança, gestão de riscos e conformidade, e aspectos de privacidade.
- **Dados Abertos (opcional):** Esta dimensão não é obrigatória e é particularmente relevante para organizações do setor público ou empresas que buscam fomentar a transparência, impulsionar a inovação e gerar impacto social positivo. Avalia a maturidade da organização na gestão, publicação e uso estratégico de dados disponibilizados publicamente (dados abertos), seguindo os princípios de serem livremente utilizáveis, reutilizáveis e redistribuíveis. A avaliação considera aspectos essenciais para garantir a reutilização, acessibilidade, interoperabilidade e a devida atribuição da fonte, ao longo de seu ciclo de vida. Abrange áreas políticas, organizacionais e técnicas, incluindo o planejamento, a priorização estratégica, o licenciamento claro, o versionamento, o formato (legível por máquina) e os mecanismos eficazes de engajamento e *feedback* com a comunidade de usuários.

4.1.3 Aspectos

O modelo avalia cada dimensão vista acima a partir de três aspectos analíticos:

- **Pessoas:** analisa o envolvimento dos profissionais na gestão de dados, incluindo conhecimento técnico, engajamento das equipes e clareza nas responsabilidades.
- **Políticas:** examina as diretrizes, normas e processos institucionais relacionados ao uso e à gestão de dados, como segurança, privacidade e conformidade com legislações.
- **Capacidades:** considera os recursos técnicos disponíveis, como ferramentas, tecnologias e infraestrutura voltadas ao processamento, análise e integração de dados.

A combinação desses três aspectos permite uma avaliação da dimensão de maturidade em GD, cobrindo desde o engajamento humano até os recursos técnicos utilizados na prática. Cada resposta da avaliação é associada a um nível de maturidade conforme critérios definidos na dimensão para cada aspecto.

4.1.4 Papéis

Uma GD eficaz requer papéis e responsabilidades definidos. Cada papel contribui de forma única para o sucesso das iniciativas de governança, garantindo que os dados sejam gerenciados como um ativo importante na organização. Os papéis foram divididos em áreas de atuação que estão dispostas abaixo:

Tabela 4.1: Papéis e Responsabilidades na Governança de Dados

Área de Atuação	Papel	Responsabilidades
Estratégia e Liderança	Patrocinadores(as) Executivos(as) / Alta Administração	Definem a visão e a estratégia geral da organização para os dados. Fornecem o patrocínio, endosso e garantem a alocação de recursos (financeiros, humanos, tecnológicos) necessários para as iniciativas de GD, reforçando a importância da cultura de dados e da responsabilidade em toda a organização.
	Chief Data Officer (CDO) / Diretor(a) de Governança de Dados	Lidera a estratégia de dados da organização, alinhando as iniciativas de dados aos objetivos de negócio. É o principal responsável pela implementação e supervisão do programa de GD. Atua como elo entre a alta administração, as áreas de negócio e a tecnologia, e frequentemente interage com órgãos reguladores.
	Comitês de Governança de Dados (ou Comitês Diretivos de Dados)	Definem políticas de alto nível, padrões, processos e diretrizes de GD. Revisam e aprovam iniciativas estratégicas de dados, resolvem conflitos de prioridade e responsabilidade e monitoram o progresso do programa de governança, representando diversas áreas da organização.

Continua na próxima página

Tabela 4.1 (Continuação)

Área de Atuação	Papel	Responsabilidades
	Escritório de Governança de Dados (EGD / DGO)	Apoia o CDO e o Comitê de Governança, sendo responsável pela operacionalização do programa de governança, incluindo a documentação de políticas, orquestração de atividades, comunicação, monitoramento de métricas de governança e mediação de conflitos. Atua como ponto central para o programa de GD.
Gestão e Curadoria de Dados	Curadores(as) de Dados (Data Stewards)	Gerenciam a qualidade, as definições (metadados de negócio) e o ciclo de vida dos dados dentro de um domínio de negócio específico. Traduzem requisitos de negócio em regras de dados, monitoram a conformidade com essas regras e trabalham para resolver problemas de qualidade de dados. São responsáveis por gerenciar o catálogo de dados, glossário de negócios e outras documentações relacionadas aos metadados, assegurando que as definições de dados sejam claras, consistentes e acessíveis. Atuam como guardiões(ãs) dos dados e ponte entre as áreas de negócio e a tecnologia, facilitando a descoberta e o entendimento dos ativos de dados pela organização.
	Gerentes de Equipes de Dados / Gestores(as) de Dados	Coordenam equipes e projetos de dados, gerenciando recursos, cronogramas e entregas. Asseguram que as práticas de GD sejam integradas aos processos de desenvolvimento e operação de dados.
Operações de Dados	Usuários(as) / Operadores(as) de Dados (Data Creators / Operators)	São responsáveis pela criação, inserção, atualização e manutenção diária de dados em sistemas operacionais e ferramentas de negócio. Devem seguir estritamente as políticas e diretrizes de qualidade de dados, contribuindo fundamentalmente para a integridade dos dados na origem. Inclui colaboradores(as) que geram ou inserem dados no seu dia a dia.
Continua na próxima página		

Tabela 4.1 (Continuação)

Área de Atuação	Papel	Responsabilidades
Engenharia e Infraestrutura de Dados	Engenheiro(a) de Dados (Data Engineer)	Projeta, constrói, otimiza e mantém a infraestrutura de dados (como, <i>data lakes</i> , <i>data warehouses</i> , bancos de dados, etc.) e os pipelines de dados (ETL/ELT). Garante a ingestão, transformação, armazenamento e disponibilidade dos dados em escala, colaborando com cientistas e analistas para atender às suas necessidades de acesso e processamento de dados.
	Especialistas em TI e Bancos de Dados (DBAs e Administradores(as) de Sistemas)	Garantem o desempenho, a disponibilidade, a segurança física e lógica da infraestrutura de dados e dos sistemas de banco de dados. Realizam backups, otimizam consultas e gerenciam acessos. Oferecem suporte essencial aos(as) Engenheiros(as) e Curadores(as) de Dados.
	Especialistas em Cybersecurity (Segurança Cibernética)	Focam na proteção dos ativos de dados contra acessos não autorizados, perdas ou ataques cibernéticos. Desenvolvem e implementam políticas de segurança da informação, realizam auditorias de segurança e respondem a incidentes, garantindo a confidencialidade, integridade e disponibilidade dos dados.
Análise e Consumo de Dados	Consumidores(as) de Dados	Utilizam os produtos de dados (como, relatórios, <i>dashboards</i> , modelos e aplicações) para análise, tomada de decisão e operações de negócio. Fornecem <i>feedback</i> valioso sobre a usabilidade, relevância e qualidade dos dados, orientando as prioridades da governança e melhorias nos sistemas.

Continua na próxima página

Tabela 4.1 (Continuação)

Área de Atuação	Papel	Responsabilidades
	Analistas de Dados e Cientistas de Dados	Analisam, interpretam, modelam e visualizam dados para gerar análises, relatórios e documentos que informam e fortalecem as estratégias de negócio. Criam e validam modelos preditivos e descritivos. São consumidores(as) avançados(as) de dados e frequentemente criam novos produtos de dados para outros(as) consumidores(as).
Conformidade e Governança Legal	Encarregado(a) de Proteção de Dados (DPO)	Foca no estrito cumprimento da LGPD (e outras leis de proteção de dados relevantes). Atua como ponto de contato para autoridades reguladoras e titulares dos dados, monitora a conformidade interna e externa e aconselha a organização sobre as obrigações de proteção de dados.
	Especialistas em Compliance e Auditoria	Garantem que a organização esteja em conformidade com as leis, regulamentos e padrões internos e externos aplicáveis ao uso e gestão de dados. Realizam auditorias regulares para identificar riscos e lacunas de conformidade, e ajudam a desenvolver planos de ação corretivos.
Habilitação e Transformação Cultural	Data Product Managers / Product Owners	Lideram o ciclo de vida de produtos de dados (como, <i>dashboards</i> , pipelines de dados específicos, modelos de Inteligência Artificial (IA)), desde a concepção até a entrega e manutenção. Atuam como ponte entre as necessidades do negócio e as equipes técnicas, garantindo que os produtos de dados agreguem valor e atendam aos padrões de governança. Reconhece a tendência de produtos de dados.
Continua na próxima página		

Tabela 4.1 (Continuação)

Área de Atuação	Papel	Responsabilidades
	Agentes de Cultura de Dados / Change Leaders	Focados em disseminar a cultura de dados na empresa, promovendo a conscientização sobre a importância dos dados, boas práticas de governança, ética no uso de dados e a responsabilidade compartilhada. Desenvolvem programas de treinamento e comunicação para fomentar uma mentalidade orientada a dados em todos os níveis da organização.

4.1.5 Estratégia de Uso

Para utilizar o DG2M, é importante que a organização adote uma abordagem estruturada para diagnosticar seu nível de maturidade em GD e planejar sua evolução contínua. A estratégia de uso pode ser dividida nas seguintes etapas principais:

1. **Adaptação do Modelo à Realidade da Organização:** A equipe de avaliação deve analisar as dimensões, níveis, papéis e aspectos do modelo DG2M e realizar as adaptações necessárias para que ele se ajuste à realidade da organização. É fundamental ajustar a linguagem e os termos do modelo para que se alinhem à realidade e ao contexto institucional, garantindo sua aplicabilidade, sem alterar o que é exigido em cada nível ou dimensão. Além disso, é importante determinar se a dimensão de Dados Abertos, que é opcional e relevante para organizações que buscam transparência ou são do setor público, deve ser incluída na avaliação.
2. **Diagnóstico da Maturidade Organizacional:** Essa etapa consiste em selecionar profissionais-chave da organização, em particular aqueles que exercem os papéis ligados à GD, como visto no modelo. As dimensões, os níveis e os aspectos do modelo devem ser apresentados a esses indivíduos, por meio de um formulário ou ferramenta de avaliação, para que possam diagnosticar o nível atual para cada dimensão e seus aspectos.
3. **Identificação do Nível Atual:** As respostas dos avaliadores devem ser analisadas e consolidadas para determinar o nível de maturidade da organização em cada dimensão. Dentro de cada dimensão, a avaliação se baseia em três aspectos: Pessoas, Políticas e Capacidades. O nível de maturidade de uma dimensão pode ser considerado o do aspecto com o menor nível, pois isso reflete o ponto de maior limitação da organização naquela dimensão. Com o valor de nível da dimensão definido para

cada avaliador, faça a média para cada dimensão e encontre o nível atual de cada uma das dimensões.

4. **Planejamento e Execução das Melhorias:** Com os níveis diagnosticados para cada dimensão, a equipe deve analisar o conjunto de práticas sugeridas oferecidas pelo modelo para cada nível, indicando as próximas etapas para avançar sua GD. O planejamento e a execução das melhorias devem ser priorizados com base nos objetivos de cada organização, bem como em sua realidade e benefícios que podem ser alcançados.
5. **Definição de Papéis e Responsabilidades:** Para garantir a implementação eficaz das iniciativas de governança mapeadas, a organização deve estabelecer de forma clara os papéis e suas responsabilidades. A organização pode usar, como base, os papéis e responsabilidades presentes no modelo, fazendo adaptações, se necessário.
6. **Acompanhamento e Melhoria Contínua:** O processo de maturidade é cíclico. Após a implementação das melhorias ou em momento oportuno, a organização pode refazer a avaliação para medir o progresso e planejar os próximos ciclos de aprimoramento, buscando alcançar níveis de maturidade mais elevados e a inovação contínua.

4.2 Melhorias Realizadas no DG2M após o Survey

Durante o processo de validação, a versão 1 do DG2M foi adequada de acordo com as sugestões dos respondentes do *survey*. Uma das principais alterações reside na "Dimensão Estratégia e GD". O Nível 5, "Otimizado – Inovação e Melhoria Contínua", agora reflete uma perspectiva em que a GD transcende a mera otimização interna para se posicionar como um motor de inovação para o negócio. Essa visão expandida visa a extração de valor dos dados, a consecução de objetivos estratégicos e o fomento à colaboração com ecossistemas externos.

Adicionalmente, aspectos como a cultura de dados e as considerações legais foram amplamente integrados. O modelo agora enfatiza que a GD deve estar plenamente incorporada aos processos e à cultura organizacional e que suas políticas devem ser dinâmicas, assegurando a conformidade com legislações de privacidade e antecipando tendências, como a governança de inteligência artificial responsável.

O suporte e o compromisso da alta gestão também foram explicitamente reforçados, com a inclusão de ações para sua conscientização e engajamento contínuo em diferentes níveis de maturidade. A prática de *benchmarking* com líderes em privacidade e estratégia de dados foi antecipada para o Nível 2 desta dimensão, visando uma comparação mais antecipada. Outra mudança significativa foi a substituição do termo "Repetível" por

"Emergente" em diversas instâncias para aprimorar a clareza e o alinhamento com a progressão da maturidade do modelo, resolvendo ambiguidades hierárquicas percebidas.

A distinção entre as dimensões de "Infraestrutura para Dados" e "Formalização e Processos de Gestão de Dados" foi mantida e reforçada. A primeira concentra-se no ambiente tecnológico (o "o quê": ferramentas, plataformas, arquitetura, escalabilidade, segurança, conformidade e gestão de riscos), enquanto a segunda aborda os fluxos de trabalho e procedimentos (o "como": classificação, gestão de metadados, integração, interoperabilidade e gestão de riscos dos dados). O texto descritivo da dimensão "Infraestrutura para Dados" foi expandido para cobrir ambientes *on-premises*, nuvem (pública, privada ou híbrida), arquiteturas distribuídas, e a capacidade técnica das equipes, garantindo que o foco permaneça no suporte tecnológico essencial à GD, sem se aprofundar em modelos específicos de Segurança da Informação, que faz parte de outra área.

Houve a mudança da dimensão "Equipes, Pessoas e Estrutura Organizacional" para "Cultura, Pessoas e Estrutura Organizacional". Nessa dimensão, o foco no letramento em dados foi evidenciado, com ajustes nos níveis que preveem desde treinamentos básicos até o aprendizado contínuo da força de trabalho. A colaboração sistemática entre Negócio e TI foi explicitada como um indicativo de maturidade e a comunicação transparente, inclusive através de plataformas de capacitação, foi incorporada. A exploração do potencial da IA e Aprendizado de Máquina, quando necessário, para otimizar a curadoria e análise de dados também foi incluída no Nível 5.

A dimensão "Formalização e Processos de Gestão de Dados" foi aprimorada para destacar o compartilhamento de dados e a participação ativa de todos os papéis envolvidos no ciclo de vida dos dados. A realocação da prática de *benchmarking* para o Nível 2 e a antecipação da resolução de problemas para níveis anteriores na dimensão "Qualidade de Dados" também são exemplos de otimizações sequenciais.

Finalmente, a estrutura de papéis foi significativamente revisada para melhor refletir a granularidade e a dinâmica da GD. Foram criadas "Áreas de Atuação" distintas, como "Engenharia e Infraestrutura de Dados" (que agora inclui Engenheiros de Dados e Especialistas em Cybersecurity) e "Habilitação e Transformação Cultural" (com Agentes de Cultura de Dados / Change Leaders). O papel do "Escritório de GD" (EGD / DGO) foi formalmente introduzido para apoiar a operacionalização do programa de governança e a inclusão de Usuários(as)/Operadores(as) de Dados formaliza a responsabilidade de quem cria e insere dados diariamente. O Encarregado de Proteção de Dados (DPO) foi explicitamente incluído, assegurando o alinhamento com a LGPD e outras legislações de proteção de dados. Essas mudanças, em conjunto, visam tornar o modelo mais adaptável, oferecendo papéis definidos que as organizações podem customizar conforme suas necessidades específicas.

4.3 Resumo do Capítulo

Esse capítulo apresentou o DG2M, um modelo de maturidade em GD desenvolvido a partir do conhecimento coletado em uma RSL e validado por meio de um *survey*. O modelo visa diagnosticar o nível de maturidade das organizações em GD e fornecer sugestões para o aprimoramento contínuo.

A estrutura do DG2M é composta por seis níveis de maturidade organizacional, que vão do Nível 0 (Não Iniciado) ao Nível 5 (Otimizado). Ele se concentra em cinco dimensões principais — Formalização e Processos de Gestão de Dados, Infraestrutura para Dados, Cultura, Pessoas e Estrutura Organizacional, Qualidade dos Dados e Estratégia e GD — e uma dimensão opcional para Dados Abertos, relevante para organizações que buscam transparência e que são do setor público. Cada dimensão é avaliada com base em três aspectos analíticos: Pessoas, Políticas e Capacidades. O modelo também inclui um conjunto de melhorias para cada nível, orientando as organizações sobre as próximas etapas para avançar em sua maturidade e uma tabela de papéis e responsabilidades agrupados por área de atuação, que facilitam as atividades da área de dados dentro da organização.

Após a validação por meio do *survey*, a versão 1 do modelo foi aprimorada, resultando na versão 2. O Nível 5 da dimensão sobre Estratégia foi expandido para refletir a GD como um motor de inovação e um meio para a extração de valor e a consecução de objetivos estratégicos. Aspectos de cultura de dados e conformidade legal foram mais integrados ao modelo, com ênfase no letramento em dados e na conformidade com legislações como a LGPD e a governança de IA responsável. Inclusive, a dimensões de Pessoas teve seu nome aprimorado para "Cultura, Pessoas e Estrutura Organizacional". O compromisso da alta gestão foi explicitamente reforçado em diferentes níveis.

Além disso, a prática de *benchmarking* com líderes em privacidade e estratégia de dados foi antecipada para o Nível 2 desta dimensão. O termo "Repetível" foi substituído por "Emergente" para aprimorar a clareza e o alinhamento com a progressão da maturidade do modelo. A distinção entre as dimensões de "Infraestrutura para Dados" e "Formalização e Processos de Gestão de Dados" foi mantida e reforçada, com a primeira focando no ambiente tecnológico e a segunda, nos fluxos de trabalho e procedimentos. A estrutura de papéis foi significativamente revisada para melhor refletir a granularidade e a dinâmica da GD, formalizando a inclusão do Escritório de GD (EGD / DGO) e outros papéis.

Capítulo 5

Desenvolvimento do DG2M

Este capítulo detalha o desenvolvimento do DG2M, um modelo desenvolvido a partir do conhecimento coletado na RSL. Adicionalmente, este capítulo descreve o *survey* conduzido para validar a versão inicial do DG2M. A partir dessa validação, uma versão aprimorada do modelo será então apresentada.

5.1 Primera Versão do DG2M

Com base na RSL, foi desenvolvido o DG2M. Este *framework* é estruturado em seis níveis hierárquicos, variando do Nível 0 (Não Iniciado) ao Nível 5 (Otimizado). Esses níveis indicam o desenvolvimento progressivo das capacidades de uma organização em GD e no uso estratégico de dados. A estrutura fundamental do DG2M deriva principalmente do MD3M [25, 37, 40] e do Stanford Maturity Model [21, 30, 36]. A inclusão do Nível 0 é embasada pelo BDMM [14], DSMM [19, 20] e pelo Modelo de Visintin [41]. Seu objetivo foi representar a ausência de práticas formalizadas de GD, conforme identificado nesses estudos.

O processo de desenvolvimento integrou tanto elementos metodológicos quanto componentes estruturais, como dimensões, aspectos, práticas sugeridas e níveis de maturidade, extraídos de uma gama diversa de modelos de maturidade existentes para estabelecer uma base abrangente e teoricamente fundamentada. Isso inclui modelos propostos em estudos anteriores, como [12], [32, 33], [35], [19], [20], [21], [36] e [30].

A RQ.3 da RSL (*Quais são os principais princípios e práticas emergentes em GD e avaliações de maturidade?*) mostrou-se fundamental neste estudo. Ela permitiu a identificação de princípios orientadores amplamente reconhecidos, como transparência, responsabilidade, qualidade de dados e conformidade regulatória, bem como práticas organizacionais recorrentes, incluindo o estabelecimento de comitês de dados, a formalização de políticas de GD e a implementação de indicadores de qualidade de dados.

Esse processo de identificação foi importante para definir cada nível de maturidade dentro do *framework* proposto e detalhar suas correspondentes práticas de melhoria recorrentes. Além disso, essa abordagem focou na identificação das práticas mais recorrentes e comumente compartilhadas entre os modelos de maturidade analisados, capturando assim os elementos mais substanciais e contribuindo para uma compreensão mais abrangente da progressão da maturidade em GD.

O processo de construção está documentado em um documento "DG2M_Development-en_US.xlsx" no [Zenodo](#). O DG2M, em sua versão inicial, está disponível para acesso e interação por meio de uma aplicação web. Você pode conferir o modelo no link: <https://luisfccardoso.github.io/dg2m-v1/>.

5.2 Survey

O estudo desenvolvido por Ghazi et al. [16] apresentou uma abordagem sobre a realização de *surveys* na Engenharia de Software, destacando desafios comuns e estratégias para mitigação. O processo estruturado sugerido pelo estudo Ghazi et al. [16] para a condução de *surveys* é dividido em oito etapas principais:

1. Definição dos objetivos da pesquisa;
2. Identificação do público-alvo e da amostragem;
3. Planejamento da amostragem;
4. Construção do instrumento do *survey*;
5. Avaliação do instrumento;
6. Coleta de dados;
7. Análise dos resultados;
8. Extração de conclusões com a devida documentação.

Um dos principais desafios apontados por Ghazi et al. [16] é a complexidade na seleção do público-alvo e na definição do tamanho da amostra, especialmente na Engenharia de Software, onde a dificuldade de acesso à população-alvo pode comprometer tanto a representatividade quanto a validade dos resultados.

Para garantir a transparência do trabalho realizado, todos os documentos referentes ao *survey*, incluindo o questionário, as respostas coletadas e as análises efetuadas, estão disponíveis para consulta no [Zenodo](#).

Neste estudo, incorporaram-se estratégias de Ghazi et al. [16] para mitigar os problemas mais comuns desse tipo de pesquisa. Para isso, seguiu-se principalmente as recomendações apresentadas na Tabela 5.1, que reúne as soluções identificadas no estudo de Ghazi et al. [16].

Tabela 5.1: Principais Problemas e Soluções Adotadas no Estudo

Problema	Solução Adotada
Tamanho da amostra insuficiente (LP01)	<ul style="list-style-type: none"> • Utilizar rede de contatos pessoais (LS01). • Consciência cultural (LS02). • Amostragem por conveniência (LS04). • Autoridade e credibilidade (LS08) - Líderes de dados compartilharam nosso formulário de pesquisa com suas redes. • Contatos de pessoas conhecidas (LS09). • Definir critérios claros para seleção da amostra (LS14) • Publicidade de terceiros - Utilizou-se o LinkedIn (LS15)
Nenhuma utilidade prática (LP04)	<ul style="list-style-type: none"> • Motivar explicitamente o benefício prático da pesquisa (LS21).
Falhas na formulação das perguntas (LP05)	<ul style="list-style-type: none"> • Pré-teste da pesquisa (LS22).
Problemas na escala Likert (LP08)	<ul style="list-style-type: none"> • Evitar escalas de dois pontos (LS28).
Apreensão de avaliação (LP13)	<ul style="list-style-type: none"> • Garantir anonimato (LS30).
Duplicação de respostas (LP18)	<ul style="list-style-type: none"> • Verificação de duplicações.

Conforme o primeiro passo indicado por Ghazi et al. [16], o principal objetivo do *survey* é validar o DG2M para organizações públicas e privadas. Para isso, busca-se coletar percepções de especialistas e profissionais da área de dados sobre a adequação dos níveis de maturidade, das dimensões avaliadas e das práticas associadas ao *framework* proposto. Além disso, pretende-se analisar a aplicabilidade prática do modelo e identificar possíveis ajustes para torná-lo mais preciso e relevante às necessidades organizacionais.

Seguindo o segundo passo recomendado por Ghazi et al. [16], o público-alvo do *survey* foi definido com base em critérios que garantem a participação de especialistas e profissionais diretamente envolvidos com governança e/ou gestão de dados, abrangendo diferentes setores e perspectivas. Assim, os participantes devem ter alguma experiência na área, de modo a contribuir com a avaliação do *framework*. O grupo inclui alguns perfis, como:

- **Especialista em governança de dados:** Profissionais com experiência na definição, implementação e monitoramento de estratégias voltadas à GD;

- **Gestores(as) de TI e/ou de dados:** Líderes responsáveis pela administração de ativos de dados e/ou infraestrutura de TI, com foco na segurança, qualidade e disponibilidade dos dados organizacionais;
- **Cientistas de dados:** Profissionais especializados(as) em inteligência de negócio a partir de dados, utilizando técnicas de estatística, aprendizado de máquina e inteligência artificial;
- **Analistas de dados:** Profissionais responsáveis pela coleta, processamento e análise de dados para apoiar a tomada de decisão organizacional;
- **Analistas de qualidade de dados:** profissionais responsáveis por monitorar e garantir a integridade, consistência e confiabilidade dos dados utilizados pela organização;
- **Pessoas consultoras, pesquisadoras, docentes e acadêmicas da área de dados e de tecnologia da informação e comunicação:** profissionais que estudam, ensinam e/ou aplicam práticas da área de dados em contextos acadêmicos ou organizacionais.

Já o terceiro passo informado por Ghazi et al. [16], a estratégia de amostragem adotada neste estudo foi planejada para garantir que o conjunto de respondentes represente uma diversidade de experiências e perspectivas sobre GD. Para mitigar riscos associados ao viés de amostragem e maximizar a validade dos resultados do *survey*, foram adotadas medidas que asseguram a inclusão de profissionais que atuaram na área. Diante dos desafios apontados por Ghazi et al. [16] em relação ao acesso à população-alvo, ações foram implementadas para fortalecer a representatividade da amostra, como:

- **Amostragem não probabilística por julgamento:** Escolha intencional dos participantes com base na relevância de sua experiência para a pesquisa;
- **Amostragem por conveniência:** A divulgação do *survey* foi ampliada por meio de diversas estratégias, incluindo a disseminação em fóruns e grupos especializados da área de dados em redes sociais (como, o WhatsApp), postagens e anúncios no LinkedIn e abordagem presencial a possíveis respondentes em empresas e universidades (com o uso de placas com QR-Code), visando maximizar a participação;
- **Tamanho da amostra:** Considerando a natureza exploratória da pesquisa e a necessidade de um grupo representativo de especialistas, estima-se a participação mínima de 30 a 50 respondentes qualificados para garantir uma análise dos resultados.

5.3 Construção do Questionário

Na condução do quarto passo, foi realizada a construção do instrumento do *survey*, que seguiu as diretrizes metodológicas propostas por Ghazi et al. [16], garantindo sua adequação aos objetivos da pesquisa. Para assegurar uma estrutura clara e objetiva, o questionário foi organizado em sete seções, conforme observado no documento "survey_questionnaire-pt_BR-Original.xlsx" no [Zenodo](#). As seções são:

1. **Seção 0: Boas-Vindas:** Mensagem inicial e Termo de Consentimento Livre e Esclarecido (TCLE);
2. **Seção 1: Perfil da(o) Respondente:** Coleta de informações sobre cargo, tempo na área de dados e setor da organização;
3. **Seção 2: Modelo e Níveis de Maturidade:** Avaliação da adequação da divisão em seis níveis (0 a 5) e suas descrições (Não Iniciado, Inicial, Repetível, Definido, Gerenciado e Otimizado), com espaço para comentários;
4. **Seção 3: Avaliação das Dimensões e Práticas:** Análise da pertinência e aplicabilidade das práticas sugeridas para seis dimensões do DG2M, bem como identificação de possíveis dimensões ausentes no *framework*;
5. **Seção 4: Papéis e Responsabilidades na Estrutura Organizacional de Governança de Dados:** Validação das descrições dos papéis propostos e sugestão de novos;
6. **Seção 5: Aplicabilidade do Modelo:** Avaliação da utilidade do DG2M para diagnóstico e contribuição para estratégias de aprimoramento, incluindo desafios na aplicação;
7. **Seção 6: Agradecimento e Contato:** Campo para *feedback* adicional e opção de recebimento do estudo finalizado.

A formulação das perguntas seguiu as práticas recomendadas para a construção de *surveys* descritas em Ghazi et al. [16], com o objetivo de minimizar vieses, assegurar clareza e utilizar uma terminologia precisa e acessível aos profissionais da área.

5.4 Piloto

Antes da aplicação definitiva do questionário, um teste piloto foi realizado com cinco participantes selecionados por conveniência, abrangendo 5 profissionais da área de dados

e gestores de TI de instituições públicas e privadas. O questionário foi disponibilizado por meio da plataforma Google Forms. Cabe ressaltar que os dados coletados nessa fase foram descartados após a validação do instrumento.

O principal objetivo do piloto foi avaliar o tempo necessário para o preenchimento do questionário, a clareza das questões e possíveis dificuldades na compreensão dos participantes. O tempo médio de resposta foi de aproximadamente 25 a 60 minutos, incluindo a leitura das questões com o modelo de maturidade de governança de dados. Dessa forma, iniciou-se a etapa 5 do processo descrito por Ghazi et al. [16], que consiste na análise da coerência e validade das perguntas, garantindo que elas fossem interpretadas corretamente pelos participantes.

A partir do *feedback* recebido, certos ajustes na redação de algumas perguntas foram feitos para melhorar sua clareza e garantir uma melhor experiência à pessoa participante. Além disso, foi incluída a opção "Não Sei" nas perguntas objetivas, permitindo que os respondentes as ignorem caso não possuam conhecimento sobre o tema, contribuindo para a precisão dos dados.

5.5 Seleção dos Participantes

Avançou-se para a etapa 6 do processo descrito por Ghazi et al. [16], direcionando o *survey* ao público detalhado na Seção 5.2. A participação na pesquisa foi voluntária e anônima, tendo como único propósito contribuir para o sucesso do estudo. Cabe mencionar que, antes de iniciar o questionário, os participantes precisaram consentir com os termos de participação, incluindo a aceitação de que suas respostas podem ser utilizadas em publicações científicas, conferências e postagens profissionais e acadêmicas.

O recrutamento dos participantes iniciou-se por meio de contatos pessoais com profissionais da área de dados, sendo subsequentemente expandido para redes sociais. A divulgação ocorreu em grupos de profissionais de Dados no WhatsApp. No LinkedIn, foi utilizada uma estratégia multifacetada que incluiu postagens orgânicas do autor, postagens por outros profissionais referência da área e a veiculação de anúncios pagos. Complementarmente, foram empregadas abordagens presenciais em empresas e universidades, usando *QR codes* para facilitar o acesso ao questionário.

Para otimizar a participação e esclarecer dúvidas, foram desenvolvidas estratégias adicionais, como a criação de uma [página sobre o survey](#), que incluía o link direto para o questionário, informações de contato do autor e uma funcionalidade para agendamento de reuniões individuais com o autor. Além de divulgar reuniões para preenchimento coletivo.

Após a fase de coleta, identificou-se que a veiculação de anúncios pagos demonstrou baixa efetividade, com apenas um clique registrado para o formulário. Similarmente, as

reuniões em grupo apresentaram pouca adesão, revelando-se menos eficientes. Apenas 3 respondentes de 46 responderam dessa forma. Em contraste, o agendamento de reuniões individuais, através do contato direto ou via página de marcação, mostrou-se mais eficaz, sendo o método de contato preferido por aproximadamente 32% dos respondentes.

Essas abordagens, alinhadas às recomendações de Ghazi et al. [16], foram importantes para garantir que a pesquisa fosse direcionada exclusivamente a especialistas em dados, assegurando a qualificação e relevância das respostas obtidas.

5.6 Resultados do Survey

Para dar prosseguimento às etapas 7 e 8 da metodologia proposta por Ghazi et al. [16], os resultados do *survey* são apresentados neste documento. Com isso, as respostas do *survey* foram analisadas e encontram-se disponíveis no [Zenodo](#), divididas em três arquivos: um com todas as respostas ("answers-pt_BR.csv"), outro com a análise dos resultados ("survey_analysis-pt_BR.ipynb") e outro com comentários dos profissionais ("professionals_comments-pt_BR.xlsx").

Os dados apresentados delineiam o perfil dos participantes do *survey*, fornecendo informações sobre seus cargos atuais, experiência profissional e os setores de suas organizações. Um total de 46 respostas foi coletado. Cabe destacar que um profissional pode ocupar um ou mais cargos. Observa-se uma concentração significativa de Gestores(as) de TI e/ou dados (21,74%), seguidos por analistas de dados (19,57%) e cientistas de dados (15,22%). Isso indica a participação de profissionais diretamente envolvidos com a gestão e análise de dados em diversos níveis hierárquicos. Além disso, a presença de acadêmicos/pesquisadores (23,91% combinando as áreas de TI/dados) e professores (19,56% combinando as áreas de TI/dados) destaca a participação de especialistas do ambiente acadêmico, o que pode enriquecer as perspectivas teóricas e práticas das respostas.

A maioria dos respondentes possui entre 2 e 5 anos de experiência (39,13%) na área de dados, sugerindo um perfil com conhecimento relevante no campo. Um grupo considerável de 11 participantes (23,91%) possui mais de 10 anos de experiência, indicando a presença de profissionais seniores com vasta expertise. O setor público, particularmente o Poder Executivo Federal, representa a maior parcela (30,43%) dos respondentes. Setores como o Poder Judiciário Federal (19,57%) e o setor privado de serviços (19,57%) também são significativos. A presença de diferentes esferas governamentais e do setor privado sublinha a amplitude do *survey*, capturando perspectivas tanto de entidades públicas quanto de empresas privadas. Essa diversidade é importante para a avaliação do *framework* proposto. A Tabela 5.2 apresenta uma visão geral do perfil dos 46 participantes do *survey*.

Perguntas e Alternativas	#	%
1. Qual(is) o(s) seu(s) cargo(s) atual(is)? (uma ou mais opções podem ser escolhidas)		
Gestor(a) de TI e/ou Dados	10	21.74
Analista de Dados	9	19.57
Cientista de Dados	7	15.22
Acadêmico/Pesquisador(a) em Tecnologia da Informação e Comunicação	6	13.04
Professor(a) em Tecnologia da Informação e Comunicação	6	13.04
Acadêmico(a)/Pesquisador(a) na Área de Dados	5	10.87
Especialista em Governança e/ou Gestão de Dados	5	10.87
Professor(a) na Área de Dados	3	6.52
Consultor(a) em Governança e Gestão de Dados	3	6.52
Analista de Qualidade de Dados	2	4.35
Outros Cargos no Serviço Público	2	4.35
Especialista em Governança de TI	1	2.17
Estatístico(a)	1	2.17
Analista de Sistemas	1	2.17
Especialista em Infraestrutura de TI	1	2.17
Especialista em UX (Dados e UI)	1	2.17
Encarregado(a) de Proteção de Dados (DPO)	1	2.17
2. Há quanto tempo você atua na área de dados?		
2 a 5 anos	18	39.13
Mais de 10 anos	11	23.91
6 a 10 anos	9	19.57
Menos de 2 anos	8	17.39
3. A qual setor a sua organização pertence?		
Público - Poder Executivo Federal	14	30.43
Público - Poder Judiciário Federal	9	19.57
Privado - Serviço	9	19.57
Consultoria	3	6.52
Empresa Pública ou Terceiro Setor	2	4.35
Público - Poder Executivo Estadual ou Distrital	2	4.35
Pesquisa em Instituição Pública de Pesquisa	2	4.35
Privado - Indústria	2	4.35
Público - Ministério Público Federal	1	2.17
Público - Poder Legislativo Federal	1	2.17
Pesquisa em Instituição Privada de Pesquisa	1	2.17

Tabela 5.2: Perfil dos respondentes do survey (n=46).

A análise concentra-se na adequação da divisão em seis níveis, na relevância das características dos níveis e na adequação e aplicabilidade das práticas dentro de cada dimensão. O *survey* coletou 46 respostas e o teste Qui-Quadrado foi utilizado para verificar se a distribuição das respostas para cada pergunta diferia significativamente de uma distribuição uniforme, indicando assim uma clara preferência ou tendência de opção de resposta entre os respondentes. As respostas "Não sei" não foram incluídas na análise.

A pergunta Q4.1 teve como objetivo verificar a aceitação da divisão de níveis proposta. Observou-se um nível de concordância significativo de 93,48%, resultante da soma das opções *Concordo* (50,00%) e *Concordo Fortemente* (43,48%). Isso sugere uma forte preferência entre os respondentes pela divisão dos níveis de maturidade em seis categorias. O teste Qui-Quadrado, com uma estatística de 55,52 e um p -valor de $2,53 \times 10^{-11}$, indica que a distribuição das respostas difere significativamente de uma distribuição uniforme.

A outra pergunta, Q5.1, buscou avaliar a pertinência do conteúdo em cada nível. Neste caso, a concordância total (soma de "Concordo" e "Concordo Fortemente") atingiu 84,78%. Esse resultado sugere que a maioria dos respondentes percebe os nomes, descrições e características dos níveis como apropriados para representar a realidade da maturidade em GD. O teste Qui-Quadrado resultou em uma estatística de 39,87 e um p -valor de $4,61 \times 10^{-8}$, reforçando a rejeição de uma distribuição uniforme entre as opções de resposta.

A adequação de cada dimensão e a aplicabilidade de suas práticas sugeridas foram avaliadas individualmente. A adequação da dimensão *Estratégia e Governança de Dados* foi avaliada pela pergunta Q6.1.1. Observou-se um nível de concordância de 84,78%, com 47,83% dos respondentes concordando fortemente e 36,96% concordando. O teste Qui-Quadrado ($\chi^2 = 42,04$, $p = 1,63 \times 10^{-8}$) indica que a distribuição das respostas difere significativamente de uma distribuição uniforme. A aplicabilidade das práticas para essa dimensão foi investigada pela pergunta Q7.1.1. Verificou-se um alto nível de concordância de 93,48%, com 58,70% dos respondentes na categoria "Concordo" e 34,78% em "Concordo Fortemente". O teste Qui-Quadrado ($\chi^2 = 61,61$, $p = 1,33 \times 10^{-12}$) demonstra que a distribuição das respostas é significativamente diferente da uniforme.

A adequação da dimensão *Equipes, Pessoas e Estrutura Organizacional* foi avaliada pela pergunta Q6.2.1. O nível de concordância para essa dimensão foi de 86,96%, com 43,48% em "Concordo" e 43,48% em "Concordo Fortemente". A análise do Qui-Quadrado ($\chi^2 = 42,48$, $p = 1,33 \times 10^{-8}$) indica uma divergência da distribuição uniforme para as opções de resposta. Para a aplicabilidade das práticas na dimensão "Equipes, Pessoas e Estrutura Organizacional", a pergunta Q7.2.1 revelou um nível de concordância de 84,78%. O teste Qui-Quadrado ($\chi^2 = 43,35$, $p = 8,76 \times 10^{-9}$) reitera que a distribuição observada das opções de resposta é significativamente diferente da uniforme.

A adequação da dimensão *Formalização e Processos de Gestão de Dados* foi avaliada pela pergunta Q6.3.1. Os resultados indicam um nível de concordância de 89,13% ("Concordo Fortemente": 47,83%, "Concordo": 41,30%). O teste Qui-Quadrado ($\chi^2 = 47,26$, $p = 1,35 \times 10^{-9}$) demonstra que a distribuição observada das opções de resposta é significativamente diferente da uniforme. Para a aplicabilidade das práticas na dimensão "Formalização e Processos de Gestão de Dados", a pergunta Q7.3.1 mostrou um nível de concordância de 89,13%. O teste Qui-Quadrado ($\chi^2 = 49,87$, $p = 3,84 \times 10^{-10}$) sugere

uma distribuição das opções de resposta diferente da uniforme. A adequação da dimensão "Qualidade de Dados"(Q6.4.1) obteve um nível de concordância de 91,30% ("Concordo Fortemente": 56,52%, "Concordo": 34,78%). O teste Qui-Quadrado ($\chi^2 = 56,17$, $p = 1,84 \times 10^{-11}$) demonstra uma distribuição das opções de resposta divergente da uniforme.

Para a aplicabilidade das práticas na dimensão "Qualidade de Dados"(Q7.4.1), o nível de concordância foi de 84,78%. O teste Qui-Quadrado ($\chi^2 = 42,89$, $p = 1,09 \times 10^{-8}$) sugere uma distribuição das opções de resposta divergente da uniforme. A adequação da dimensão "Infraestrutura para Dados"(Q6.5.1) apresentou um nível de concordância de 82,61% ("Concordo Fortemente": 54,35%, "Concordo": 28,26%). O teste Qui-Quadrado ($\chi^2 = 48,27$, $p = 8,28 \times 10^{-10}$) indica uma divergência da distribuição das opções de resposta da uniforme.

Para a aplicabilidade das práticas na dimensão "Infraestrutura para Dados"(Q7.5.1), o nível de concordância foi de 84,78%. O teste Qui-Quadrado ($\chi^2 = 45,77$, $p = 2,75 \times 10^{-9}$) sugere uma divergência da distribuição das opções de resposta da uniforme. A dimensão "Dados Abertos"(Q6.6.1) obteve um nível de concordância de 80,43% ("Concordo": 41,30%, "Concordo Fortemente": 39,13%). O teste Qui-Quadrado ($\chi^2 = 34,44$, $p = 6,04 \times 10^{-7}$) indica uma distribuição das respostas diferente da uniforme. Para a aplicabilidade das práticas na dimensão "Dados Abertos"(Q7.6.1), o nível de concordância foi de 78,26%. O teste Qui-Quadrado ($\chi^2 = 34,18$, $p = 6,84 \times 10^{-7}$) sugere uma distribuição das respostas diferente da uniforme.

Para a pergunta Q8.1 "Você acredita que alguma dimensão essencial está faltando no *framework* de maturidade em governança de dados?", os respondentes mostraram uma opinião mais variada. Embora alguns concordassem, uma parte notável — mais de 50% combinados — discordou (37,21%) ou discordou fortemente (16,28%) que uma dimensão estivesse faltando. Isso indica uma satisfação geral com as dimensões propostas, sugerindo que o *framework* atual é amplamente considerado abrangente. No entanto, essa porcentagem não é conclusiva o suficiente para afirmar definitivamente que nenhuma dimensão está faltando, deixando espaço para explorações futuras. O teste Qui-Quadrado ($\chi^2 = 6,20$, $p = 0,1850$) para esta pergunta não rejeitou a hipótese de uma distribuição uniforme.

A pergunta Q9.1.1 teve como objetivo validar a clareza e relevância dos papéis no modelo. Com um nível de concordância de 82,61% (soma de "Concordo" e "Concordo Fortemente"), a maioria dos 46 respondentes considera as descrições dos papéis adequadas. O teste Qui-Quadrado ($\chi^2 = 35,52$, $p = 3,63 \times 10^{-7}$) confirma que essa concordância não é aleatória, rejeitando a hipótese de uma distribuição uniforme das respostas. Em relação à inclusão de novos papéis, a pergunta Q9.2.1 resultou em achados inconclusivos. O teste

Qui-Quadrado, com uma estatística de 6,20 e um p-valor de 0,185, não permitiu rejeitar a hipótese nula. Isso significa que não há uma diferença estatisticamente significativa entre a distribuição de respostas observada e uma distribuição uniforme. Em termos mais simples, os participantes não mostraram uma tendência forte e unânime em sugerir papéis adicionais.

A utilidade do modelo para o diagnóstico de maturidade foi avaliada pela pergunta Q10.1. Os resultados demonstraram um alto nível de concordância de 97,83%, com 50,00% em "Concordo Fortemente" e 47,83% em "Concordo". O teste Qui-Quadrado ($\chi^2 = 64,22$, $p = 3,76 \times 10^{-13}$) sugere fortemente que não há uma distribuição uniforme das respostas. Por fim, a capacidade do *framework* de contribuir para a definição de estratégias de melhoria foi investigada pela pergunta Q11.1. A resposta foi unânime, com 100,00% de concordância (58,70% em "Concordo Fortemente" e 41,30% em "Concordo"). O teste Qui-Quadrado ($\chi^2 = 72,48$, $p = 6,80 \times 10^{-15}$) reforça a não-uniformidade da distribuição.

A Tabela 5.3 apresenta um resumo consolidado das avaliações de concordância e aplicabilidade para todas as dimensões do modelo em análise. Para a construção desta tabela, as opções da escala Likert foram convertidas em valores numéricos: 1 para "Discordo Fortemente", 2 para "Discordo", 3 para "Neutro", 4 para "Concordo" e 5 para "Concordo Fortemente". Essa conversão foi importante para o cálculo das médias e modas de concordância e aplicabilidade para cada dimensão.

Na análise das dimensões (Tabela 5.3), "Qualidade de Dados" desponta como um ponto forte, evidenciado por 91,30% de concordância e a maior média de concordância (4,43). Paralelamente, a dimensão "Estratégia e Governança de Dados" lidera em aplicabilidade, atingindo 93,48%. Enquanto isso, "Infraestrutura para Dados" se destaca com a maior média de aplicabilidade (4,36). As modas de concordância e aplicabilidade, que frequentemente apontam para 4,0 ou 5,0, reforçam a forte tendência de avaliação positiva por parte dos participantes, sugerindo um consenso para "Concordo" ou "Concordo Fortemente".

A pergunta "Quais seriam as principais dificuldades na aplicação do modelo na prática?" permitiu identificar obstáculos percebidos pelos respondentes. A Tabela 5.4 lista as dificuldades mais frequentemente citadas. Os respondentes puderam escolher uma ou mais opções.

As três principais dificuldades mencionadas são "Resistência organizacional à adoção de novas práticas" (65,22%), "Falta de apoio da alta gestão" (54,35%) e "Falta de alinhamento entre as áreas/setores da organização" (54,35%). Essas respostas evidenciam que os desafios na implementação do modelo são predominantemente culturais e gerenciais, superando questões de capacitação técnica ou recursos financeiros.

Tabela 5.3: Tabela Geral de Concordância, Aplicabilidade e Estatísticas por Dimensão

Dimensão	Concor- dância (%)	Aplica- bili- dade (%)	Média Con- cor- dância	Média Apli- cabili- dade	Moda Con- cor- dância	Moda Apli- cabili- dade
Estratégia e Gover- nança de Dados	84.78	93.48	4.20	4.26	5.0	4.0
Equipes, Pessoas e Estrutura Organi- zacional	86.96	84.78	4.20	4.15	4.0	4.0
Formalização e Processos de Ges- tão de Dados	89.13	89.13	4.33	4.24	5.0	4.0
Qualidade de Da- dos	91.30	84.78	4.43	4.31	5.0	5.0
Infraestrutura para Dados	82.61	84.78	4.36	4.36	5.0	5.0
Dados Abertos	80.43	78.26	4.13	4.14	4.0	4.0
Média Geral	85.87	85.87	4.27	4.24	4.67	4.33
Máximo Geral	91.30	93.48	4.43	4.36	5.0	5.0
Mínimo Geral	80.43	78.26	4.13	4.14	4.0	4.0
Variância Geral	16.54	25.99	0.01	0.01	0.27	0.27

Tabela 5.4: Contagem e Percentual das Principais Dificuldades na Aplicação do Modelo na Prática

Dificuldade	Contagem	%
Resistência organizacional à adoção de novas práticas	30	65.22
Falta de apoio da alta gestão	25	54.35
Falta de alinhamento entre as áreas/setores da organização	25	54.35
Falta de treinamento ou conhecimento técnico na área	20	43.48
Dificuldade em mensurar indicadores de maturidade	18	39.13
Falta de dados/informações sobre a maturidade organizacio- nal	14	30.43
Recursos financeiros ou tecnológicos insuficientes	13	28.26
Complexidade na adaptação do modelo à realidade organiza- cional	13	28.26
Dimensionamento insatisfatório da equipe	1	2.17
Avaliar grande volume de aspectos em instituições	1	2.17
Restrições na operação da equipe de dados devido a modelos estabelecidos	1	2.17

5.7 Melhoria do Framework

Com base no *feedback* dos respondentes, o texto do *framework* foi aprimorado. Além disso, foram levantados alguns pontos para discussão, que estão detalhados no documento hospedado no [Zenodo](#) em uma planilha com a nova versão do DG2M ("DG2M-1st_Version_and_Validated_Version-pt_BR.xlsx") e em uma planilha com a análise dos comentários ("professionals_comments-pt_BR.xlsx"). Os comentários que possuem como ação "Colocar nos questionamentos sobre o modelo" serão discutidos na próxima seção. A consolidação das melhorias pode ser observada no *website* <https://luisfccardoso.github.io/dg2m/> e no Capítulo 4.

5.8 Resumo do Capítulo

Este capítulo apresenta o processo de desenvolvimento do DG2M, um modelo de maturidade em GD desenvolvido com base na RSL.

O DG2M é estruturado em seis níveis hierárquicos (do Nível 0 ao Nível 5), inspirada principalmente no MD3M e no Stanford Maturity Model, com a inclusão do Nível 0 baseada em modelos como BDMM e DSMM para representar a ausência de práticas formalizadas de GD. A RSL foi importante para identificar princípios orientadores e práticas recorrentes que definem cada nível de maturidade e suas respectivas melhorias. O processo de construção do modelo e sua versão inicial estão documentados e acessíveis publicamente no Zenodo.

Para validar o DG2M, foi conduzido um *survey* seguindo a metodologia de Ghazi et al. [16], que incluiu a definição de objetivos, público-alvo, construção e avaliação do questionário, coleta e análise de dados. O principal objetivo foi validar o modelo e coletar percepções de profissionais de dados sobre sua adequação e aplicabilidade. Um piloto precedeu a aplicação final, ajustando a clareza das questões.

Os resultados do *survey*, com 46 respondentes, revelaram um perfil predominante de gestores e analistas de dados, com experiência na área e representatividade de setores público e privado. A estrutura de seis níveis do modelo e a adequação de seus nomes e descrições foram validadas pelos participantes. As dimensões do modelo (Estratégia e Governança de Dados; Cultura, Pessoas e Estrutura Organizacional; Formalização e Processos de Gestão de Dados; Qualidade de Dados; Infraestrutura para Dados e Dados Abertos) também obtiveram aceitação quanto à sua adequação e aplicabilidade.

A "Qualidade de Dados" se destacou pela concordância, e "Estratégia e Governança de Dados" pela aplicabilidade. A utilidade do DG2M para diagnóstico e para a definição de estratégias de aprimoramento foi reconhecida pelos respondentes. As principais dificuldades

para a aplicação prática do modelo foram identificadas como resistência organizacional, falta de suporte da alta gestão e desalinhamento entre áreas.

Capítulo 6

Discussão

6.1 Discussão dos Comentários dos Profissionais de Dados

No contexto da validação do DG2M, os comentários dos respondentes forneceram informações importantes, abordando tanto a estrutura geral do modelo quanto aspectos específicos de suas dimensões e aplicabilidade. Os comentários completos podem ser lidos no arquivo disponibilizado no [Zenodo](#). A análise qualitativa dessas contribuições permitiu identificar pontos de destaque e sugestões de aprimoramento, que são importantes para o refinamento da ferramenta apresentada na seção 4. Certos aspectos são discutidos abaixo.

Em relação à estrutura de níveis de maturidade, a divisão em seis níveis (0 a 5) foi bem recebida pelos respondentes, que consideraram a gradação razoável para diagnóstico e planejamento. Um dos comentários expressa essa percepção: "Considero 6 níveis uma gradação razoável." O Nível 0 (Não Iniciado) foi particularmente elogiado por preencher uma lacuna comum em outros modelos, representando a ausência total de práticas. "Achei o Nível 0: Não Iniciado adequado para o tema. É um nível que não costumo ver em outros modelos de maturidade," afirmou uma pessoa respondente.

A similaridade do modelo com outros *frameworks*, como o CMMI e MPS.Br, também foi notada e bem-vista, com comentários como: "Achei interessante a abordagem semelhante ao CMMI" e "Possui alguma semelhança com outros níveis de maturidade, como o MPS.Br e CMMi". No entanto, alguns participantes sugeriram que a distinção entre níveis próximos pode ser sutil na prática. Outra crítica apontou que o uso de números em vez de nomes para os níveis poderia ser mais claro para o público, direcionando o foco para os requisitos objetivos de cada etapa.

No que tange às dimensões do *framework*, a dimensão "Infraestrutura para Dados" foi considerada importante, pois aborda aspectos além de pessoas e processos. Um respon-

dente destacou: "Bem interessante ter uma para tratar só de infra, pois a governança precisa ir além de pessoas e processos." A dimensão "Dados Abertos" gerou debates sobre sua aplicabilidade, com sugestões para torná-la opcional em alguns contextos e para que o termo "Transparência" fosse usado para ampliar seu escopo. Um comentário que ilustra essa questão foi: "Sugiro Transparência, pois uma empresa pode ser transparente sem dados abertos por meio de *blockchain*, por exemplo". A conformidade com a LGPD e outras legislações de privacidade foi mencionada como um ponto a ser reforçado, especialmente nas políticas de "Estratégia e Governança de Dados".

Em "Formalização e Processos de Gestão de Dados", um comentário levantou uma preocupação sobre o risco de burocratização: "Há o perigo de uma formalização excessiva e isso burocratizar novas ideias e soluções." Essa questão foi mitigada no modelo proposto, por meio de adequação do texto e possibilidade de adaptação do modelo para a realidade organizacional. O que foi informado na estratégia do DG2M. Já na dimensão "Qualidade de Dados", a importância da colaboração interdepartamental foi enfatizada, pois a manutenção da qualidade dos dados exige o envolvimento de áreas técnicas, de negócio, jurídico e TI, sendo um desafio em "ambientes compartimentalizados".

A aplicabilidade das práticas sugeridas foi majoritariamente confirmada pelos respondentes. Contudo, foi destacado que organizações com limitações de porte, orçamento ou maturidade tecnológica podem ter dificuldades, especialmente nos níveis mais avançados. Um dos comentários aponta: "É possível que algumas organizações encontrem dificuldades em aplicar certos elementos devido a limitações de porte, orçamento ou maturidade tecnológica. Mais especificamente nos últimos níveis." A importância de uma abordagem adaptável e incremental também foi mencionada, com a recomendação de que o modelo seja visto como um "menu evolutivo, não como um *checklist* obrigatório".

A percepção equivocada da governança como custo, e não como investimento, e a falta de uma cultura de dados consolidada, foram identificadas como uma das barreiras. Um respondente resumiu o problema: "Minha experiência diz que a governança ainda é vista como custo, não como investimento. Não tem cultura de dados consolidada."

Uma sugestão foi a inclusão de uma dimensão que demonstre o valor e o impacto da GD no negócio, o que pode ser discutida em trabalhos futuros. O alinhamento com o DMBOK Pyramid (Aiken) também foi sugerido como uma melhoria. Houve um questionamento sobre a repetição de temas relacionados a "pessoas" nas diferentes dimensões, com a sugestão de deixar a dimensão de "Cultura, Pessoas e Estrutura Organizacional" mais abrangente. A necessidade de um glossário para termos técnicos e a inclusão de exemplos práticos para cada nível foram apontadas como formas de aumentar a compreensão do DG2M.

No geral, a avaliação foi muito positiva. Os respondentes elogiaram a estrutura e

a profundidade do modelo, considerando-o bem-estruturado e com potencial para gerar impactos positivos. "Parabéns pelo seu trabalho, está muito bem estruturado e tenho certeza que foi feito com muita dedicação,"concluiu um dos participantes.

Tabela 6.1: Principais Comentários e Sugestões dos Respondentes

Categoria do Comentário	Síntese dos Comentários Relevantes
Estrutura de Níveis de Maturidade	A divisão em 6 níveis é considerada apropriada, facilitando diagnóstico. Nível 0 é diferencial positivo. Abordagem semelhante ao CMMI notada. Sugestão de usar apenas números para os níveis.
Dimensão "Infraestrutura para Dados"	Validade da dimensão reconhecida. Observação de que infraestrutura "up-to-date" não garante melhores práticas de governança.
Dimensão "Dados Abertos"	Questionamento sobre aplicabilidade universal e sugestão de renomear para "Transparência" para ampliar escopo.
Conformidade Legal	Necessidade de reforçar alinhamento com legislações de privacidade (como, LGPD) em "Dados Abertos" e "Estratégia e Governança de Dados".
Dimensão "Formalização e Processos de Gestão de Dados"	Essencial e observável na prática. Preocupação com formalização excessiva que pode burocratizar novas ideias.
Dimensão "Qualidade de Dados"	Fundamental. Monitoramento e manutenção da qualidade exigem colaboração entre áreas técnicas, de negócio, jurídico e TI. Sugestão de sessão "como fazer" as melhorias com entradas e saídas de processos.
Aplicabilidade das Práticas	Práticas consideradas aplicáveis. Limitações de porte/orçamento podem dificultar aplicação nos níveis avançados. Níveis 0-2 são vistos como de baixo custo e alta visibilidade para convencer lideranças. Importância das políticas e objetivos da empresa.
Barreiras de Implementação	Governança vista como custo; ausência de cultura de dados consolidada.
Dimensões Ausentes / Repetição	Sugestão de dimensão sobre valor/impacto da GD no negócio. Questionamento sobre repetição de aspectos de "pessoas" entre dimensões e sugestão de segmentar recomendações por processos e indicadores. Sugestão de alinhamento com DMBOK Pyramid (Aiken).
Clareza e Comunicação do Modelo	Necessidade de glossário e exemplos práticos para cada nível.
Percepção Geral do Trabalho	Modelo elogiado por sua estrutura, profundidade e potencial de impacto. Considerado bem estruturado e com potencial para gerar impactos positivos.

6.2 Ameaças à Validade do Estudo e Limitações

Apesar do planejamento e execução cuidadosos, o estudo apresenta ameaças à validade e limitações inerentes a sua metodologia. A primeira versão do DG2M foi fundamentada em uma RSL, cujas ameaças e limitações foram discutidas na seção 3. A validação do DG2M ocorreu por meio de um *survey* que, embora alinhado às diretrizes de Ghazi et al. [16] para mitigar problemas comuns, possui certas restrições. Essas ameaças são apresentadas abaixo e sua síntese está na tabela 6.2.

A estratégia de amostragem não probabilística, utilizando julgamento e conveniência, pode não assegurar a representatividade completa da população de profissionais de dados. Embora esforços tenham sido feitos para diversificar as experiências e os setores (público e privado), a dificuldade de acesso à população-alvo, conforme salientado por Ghazi et al. [16], pode ter influenciado o tamanho final da amostra (46 respondentes). Este tamanho, embora considerado adequado para a pesquisa, pode restringir generalizações estatísticas mais amplas.

A elaboração do questionário, apesar do teste piloto e dos ajustes para melhorar a clareza, ainda pode conter ambiguidades ou levar a diferentes interpretações por parte dos respondentes, mesmo com a inclusão da opção "Não Sei". Para minimizar esse problema, as reuniões com os participantes foram fundamentais para esclarecer dúvidas e acompanhar o preenchimento. Uma outra ação foi a aplicação do piloto, antes de iniciar a coleta das respostas.

A confiança na autodeclaração dos participantes sobre sua experiência e cargo pode introduzir um viés de resposta. Adicionalmente, as respostas sobre a aplicabilidade das práticas são baseadas na percepção dos respondentes, que, apesar de serem profissionais da área com experiência, podem não refletir totalmente a complexidade da implementação em contextos organizacionais distintos de suas vivências. Por isso, foi importante ter representante de diversos setores.

Outra limitação reside no foco principal do *framework* na avaliação da GD, com menção à relação com Segurança da Informação e Privacidade, mas sem aprofundar a avaliação de modelos específicos dessas áreas. Essa delimitação é intencional, dado que o objetivo central do modelo é a GD e outros temas devem ser abordados por meio de modelos específicos. Contudo, essa escolha pode ser vista como uma lacuna por organizações que buscam uma avaliação integrada e detalhada de Segurança da Informação e Privacidade.

As preocupações levantadas por alguns respondentes sobre a possível burocratização devido à formalização excessiva e a necessidade de uma dimensão que evidencie o valor e o impacto no negócio, embora consideradas, representam aspectos que demandam investigação e refinamento para melhor compreender essa necessidade, por isso a estratégia do modelo informa a necessidade de adaptá-lo, dentro do possível, à realidade organizacional.

Além disso, é importante que a organização não siga o modelo como um *check-list* e sim como um menu evolutivo para se adaptar às necessidade e objetivos organizacionais.

Tabela 6.2: Ameaças à Validade do Estudo

Fonte da Ameaça	Descrição da Ameaça e Mitigação
Ameaças Decorrentes da RSL	<p>Risco de Viés: A versão inicial do DG2M foi construída a partir de uma RSL, que pode apresentar Viés de seleção e erros na extração manual de dados.</p> <p>Mitigação: Processo de triagem em múltiplas etapas com critérios documentados e extração estruturada.</p>
Amostragem Não Probabilística	<p>Risco: A amostragem por julgamento e conveniência pode não garantir representatividade total da população de profissionais de dados. O tamanho da amostra (46 respondentes) pode ser limitado para generalizações estatísticas mais amplas.</p> <p>Mitigação: Adoção de estratégias para maximizar a diversidade de experiências e setores (público e privado).</p>
Formulações do Questionário	<p>Risco: Apesar do pré-teste e ajustes de clareza, podem existir ambiguidades ou interpretações diversas das perguntas.</p> <p>Mitigação: Realização de reuniões com os participantes para esclarecer dúvidas e aplicação de um piloto para aprimorar a redação do questionário.</p>
Dependência da Autodeclaração	<p>Risco: A dependência da autodeclaração dos participantes sobre experiência e cargo, potencialmente introduzindo viés de resposta.</p> <p>Mitigação: Busca por respondentes de diversos setores e contextos para buscar convergências.</p>
Percepção dos Respondentes	<p>Risco: As respostas sobre aplicabilidade baseiam-se na percepção dos respondentes, podendo não refletir a complexidade da implementação em diferentes cenários organizacionais.</p> <p>Mitigação: Busca por respondentes de diversos setores e contextos para buscar convergências.</p>
Foco Específico da GD	<p>Risco: O <i>framework</i> avalia a GD, mas não aprofunda modelos específicos de Segurança da Informação e Privacidade. Essa delimitação, embora intencional, pode ser vista como uma lacuna.</p> <p>Mitigação: Delimitação clara do escopo do DG2M para manter sua objetividade. Caso seja necessário, a organização pode usar modelos específicos para Segurança da Informação e Privacidade.</p>
Preocupações dos Respondentes com Burocratização	<p>Preocupações sobre possível burocratização por formalização excessiva e a necessidade de uma dimensão sobre valor e impacto no negócio.</p> <p>Mitigação: O modelo deve ser adaptado para a realidade organizacional e deve ser seguido de forma evolutiva e não como um <i>check-list</i> obrigatório.</p>

6.3 Resumo do Capítulo

Este capítulo descreve uma discussão acerca do DG2M e seu processo de validação, abordando principalmente os comentários realizados pelos respondentes do *survey* aplicado. Os comentários dos respondentes sobre o DG2M fornecem informações importantes para a validação do modelo, cobrindo sua estrutura, dimensões e aplicabilidade.

A estrutura de seis níveis de maturidade (0 a 5) foi bem recebida, especialmente o Nível 0 ("Não Iniciado"). No entanto, houve observações sobre a distinção tênue entre níveis adjacentes e a sugestão de usar apenas números, ao invés de nomes, para mencionar os níveis. Para "Formalização e Processos de Gestão de Dados", expressou-se preocupação com formalização excessiva. Na "Qualidade de Dados", a necessidade de colaboração interdisciplinar foi apontada.

A aplicabilidade das práticas foi confirmada, mas limitações organizacionais (porte, orçamento, maturidade) foram citadas como desafios, especialmente em níveis avançados. Respondente notou que os níveis iniciais (0 a 2) são de baixo custo e alta visibilidade, úteis para convencer lideranças. Barreiras como a visão da GD como custo e a ausência de cultura de dados foram destacadas. Sugeriu-se também a adição de uma dimensão sobre valor/impacto da GD no negócio, alinhamento com o DMBOK Pyramid (Aiken) e a necessidade de um glossário técnico e exemplos práticos para maior clareza.

Capítulo 7

Conclusão

Este trabalho teve como objetivo principal desenvolver e validar o DG2M, um *framework* abrangente para auxiliar organizações a avaliar e aprimorar suas práticas de GD. Para alcançar este objetivo, foi conduzida uma RSL, que permitiu sintetizar o conhecimento existente sobre modelos e práticas de maturidade em GD. Essa etapa foi fundamental para a construção da primeira versão do DG2M, que incorpora as principais características e princípios identificados na literatura.

A validação do DG2M foi realizada por meio de um *survey* com 46 profissionais da área de dados, abrangendo diversos setores e níveis de experiência. Os resultados do *survey* demonstraram uma concordância dos participantes em relação à estrutura de seis níveis de maturidade do DG2M, bem como à adequação e aplicabilidade de suas dimensões e práticas. As dimensões "Qualidade de Dados" e "Estratégia e Governança de Dados" foram particularmente bem avaliadas, e a utilidade do DG2M para diagnosticar o estágio atual de maturidade e para guiar a definição de estratégias de aprimoramento foi amplamente reconhecida.

As principais dificuldades identificadas para a aplicação prática do modelo, como a resistência organizacional, a falta de suporte da alta gestão e o desalinhamento entre áreas, reforçam que os desafios da GD são frequentemente de natureza cultural e gerencial, e não apenas técnica. Com base no *feedback* dos respondentes, o DG2M foi aprimorado para se tornar mais adaptável às necessidades diversas das organizações. As melhorias incluíram a ênfase na GD como motor de inovação, a integração de aspectos de cultura de dados e conformidade legal (incluindo IA responsável), a substituição do termo "Repetível" por "Emergente" para o nível 2, a reestruturação dos papéis na governança e o foco no letramento em dados.

7.1 Contribuições, Implicações Práticas e Pesquisas Futuras

Este estudo oferece uma contribuição significativa para a área de GD ao propor e validar um modelo de maturidade que pode ser aplicado em diferentes tipos de organizações, como órgãos públicos, Pequenas e Médias Empresas (PMEs) ou instituições do setor educacional. O DG2M serve como uma ferramenta para organizações que buscam evoluir suas capacidades de governança, fornecendo um roteiro para a identificação de lacunas e a implementação de melhorias. A validação empírica por meio de um *survey* com profissionais da área fortalece a relevância e a aplicabilidade do modelo na prática, embora o estudo reconheça as ameaças à validade e as limitações inerentes à sua metodologia, como a amostragem não probabilística.

Para pesquisas futuras, sugere-se a realização de estudos de caso para avaliar a aplicação do DG2M em cenários reais e coletar dados qualitativos adicionais sobre sua efetividade e os desafios enfrentados durante a implementação. Explorar a integração do DG2M com modelos de maturidade de segurança da informação e privacidade de dados é um caminho promissor para oferecer uma avaliação ainda mais completa.

A sugestão sobre a inclusão de uma dimensão que demonstre explicitamente o valor e o impacto da GD no negócio é uma área de pesquisa futura relevante, que poderia fortalecer a justificativa para investimentos em GD. O desenvolvimento de um glossário técnico e a inclusão de exemplos práticos para cada nível e dimensão do DG2M também são oportunidades para aprimorar a usabilidade e a compreensão do *framework*.

Referências

- [1] Rene Abraham, Johannes Schneider, and Jan vom Brocke. Data governance: A conceptual framework, structured review, and research agenda. *International Journal of Information Management*, 49:424–438, 2019. ISSN 0268-4012. doi: <https://doi.org/10.1016/j.ijinfomgt.2019.07.008>. URL <https://www.sciencedirect.com/science/article/pii/S0268401219300787>. 1, 9, 14
- [2] Tomas Aguiar, Silvia Boguea Gomes, Paulo Rupino da Cunha, and Miguel Mira da Silva. Digital transformation capability maturity model framework. In *2019 IEEE 23rd International Enterprise Distributed Object Computing Conference (EDOC)*, pages 51–57, Paris, France, October 2019. IEEE. doi: 10.1109/EDOC.2019.00016. 9, 10
- [3] Majid Al-Ruithe, Elhadj Benkhelifa, and Khawar Hameed. A systematic literature review of data governance and cloud data governance. *Personal and Ubiquitous Computing*, 23(5):839–859, 2019. ISSN 1617-4917. doi: 10.1007/s00779-017-1104-3. URL <https://doi.org/10.1007/s00779-017-1104-3>. 10, 12
- [4] Zaher Ali Al-Sai, Rosni Abdullah, and Mohd heikal husin. A review on big data maturity models. In *2019 IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology (JEEIT)*, pages 156–161, Amman, Jordan, April 2019. IEEE. doi: 10.1109/JEEIT.2019.8717398. 3, 12
- [5] Yunusa Adamu Bena, Roliana Ibrahim, and Jamilah Mahmood. Current challenges of big data quality management in big data governance: A literature review. *Lecture Notes on Data Engineering and Communications Technologies*, 210:160–172, 2024. doi: 10.1007/978-3-031-59711-4_15. 23, 28, 30, 31
- [6] Patrícia Bento, Miguel Neto, and Nadine Côte-Real. How data governance frameworks can leverage data-driven decision making: A sustainable approach for data governance in organizations. In *2022 17th Iberian Conference on Information Systems and Technologies (CISTI)*, pages 1–5, 10.23919/CISTI54924.2022.9866895, June 2022. IEEE Computer Society. 1, 2
- [7] Bruno Miguel Vital Bernardo, Henrique São Mamede, João Manuel Pereira Barroso, and Vítor Manuel Pereira Duarte dos Santos. Data governance & quality management—innovation and breakthroughs across different fields. *Journal of Innovation and Knowledge*, 9(4):1–35, 2024. doi: 10.1016/j.jik.2024.100598. URL <https://doi.org/10.1016/j.jik.2024.100598>. Cited by: 0. 3, 24, 28, 29, 30, 31

- [8] Karol Bližňák, Michal Munk, and Anna Pilková. A systematic review of recent literature on data governance (2017–2023). *IEEE Access*, 12:149875–149888, 2024. ISSN 2169-3536. doi: 10.1109/ACCESS.2024.3476373. 10, 13
- [9] Asmae Boufassil, Fadwa Bouhafer, Mohamed Cherradi, and Anass El Haddadi. Data catalog: Approaches, trends, and future directions. In *2023 17th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS)*, pages 369–376, <https://ieeexplore.ieee.org/abstract/document/10472799>, Nov 2023. IEEE. doi: 10.1109/SITIS61268.2023.00067. 2
- [10] Pearl Brereton, Barbara A. Kitchenham, David Budgen, Mark Turner, and Mohamed Khalil. Lessons from applying the systematic literature review process within the software engineering domain. *Journal of Systems and Software*, 80(4):571–583, 2007. ISSN 0164-1212. doi: 10.1016/j.jss.2006.07.009. URL <https://doi.org/10.1016/j.jss.2006.07.009>. 18
- [11] Artur Potiguara Carvalho, Edna Dias Canedo, Fernanda Potiguara Carvalho, and Pedro Henrique Potiguara Carvalho. Anonymisation and compliance to protection data: Impacts and challenges into big data. In Joaquim Filipe, Michal Smialek, Alexander Brodsky, and Slimane Hammoudi, editors, *Proceedings of the 22nd International Conference on Enterprise Information Systems, ICEIS 2020, Prague, Czech Republic, May 5-7, 2020, Volume 1*, pages 31–41. SCITEPRESS, 2020. doi: 10.5220/0009411100310041. URL <https://doi.org/10.5220/0009411100310041>. 2
- [12] Guangming Cheng, Yao Li, Zhiwei Gao, and Xiaoyin Liu. Cloud data governance maturity model. In *2017 8th IEEE International Conference on Software Engineering and Service Science (ICSESS)*, pages 517–520, <https://doi.org/10.1109/ICSESS.2017.8342968>, 2017. IEEE. doi: 10.1109/ICSESS.2017.8342968. 24, 27, 29, 31, 46
- [13] CMMI Institute. CMMI Levels of Capability and Maturity Levels. <https://cmmiinstitute.com/learning/appraisals/levels>, -. Accessed: 26 June 2025. 3
- [14] Marco Comuzzi and Anit Patel. How organisations leverage big data: A maturity model. *Industrial Management and Data Systems*, 116(8):1468–1492, 2016. doi: 10.1108/IMDS-12-2015-0495. URL <https://doi.org/10.1108/IMDS-12-2015-0495>. 24, 27, 29, 31, 46
- [15] Ziqi Gao. Establishment and evaluation of measurement and control system model based on data analysis optimization. In *Proceedings of the 5th International Conference on Information Technologies and Electrical Engineering*, pages 343–348, New York, NY, USA, 2023. Association for Computing Machinery. doi: 10.1145/3582935.3582991. URL <https://doi.org/10.1145/3582935.3582991>. 24, 31
- [16] Ahmad Nauman Ghazi, Kai Petersen, Sri Sai Vijay Raj Reddy, and Harini Nekkanti. Survey research in software engineering: Problems and mitigation strategies. *IEEE*

- Access*, 7:24703–24718, 2019. doi: 10.1109/ACCESS.2018.2881041. URL <https://doi.org/10.1109/ACCESS.2018.2881041>. 5, 47, 48, 49, 50, 51, 52, 58, 63
- [17] DAMA. *DAMA-DMBOK: Data Management Body of Knowledge (2nd Edition)*. Technics Publications, LLC, Denville, NJ, USA, 2017. ISBN 1634622340. 1, 8, 14
- [18] Panom Gunawong. Data governance for wicked problems: A case from the thai health system. *THE ELECTRONIC JOURNAL OF INFORMATION SYSTEMS IN DEVELOPING COUNTRIES*, 90, 08 2023. doi: 10.1002/isd2.12296. 2, 11, 13
- [19] Mert Onuralp Gökalp, Ebru Gökalp, Kerem Kayabay, Altan Koçyiğit, and P. Erhan Eren. Data-driven manufacturing: An assessment model for data science maturity. *Journal of Manufacturing Systems*, 60:527–546, 2021. doi: 10.1016/j.jmsy.2021.07.011. 23, 27, 30, 31, 46
- [20] Mert Onuralp Gökalp, Ebru Gökalp, Kerem Kayabay, Altan Koçyiğit, and P. Erhan Eren. The development of the data science capability maturity model: a survey-based research. *Online Information Review*, 46(3):547–567, 2022. doi: 10.1108/OIR-10-2020-0469. URL <https://doi.org/10.1108/OIR-10-2020-0469>. Cited by: 24, 24, 27, 28, 31, 46
- [21] Ikhsan Mirza Harwanto and Achmad Nizar Hidayanto. Data governance maturity assessment: A case study directorate general of corrections. In *2022 International Conference on ICT for Smart Society (ICISS)*, pages 01–06, 10.1109/ICISS55894.2022.9915243, Aug 2022. IEEE. 2, 11, 13, 23, 27, 28, 29, 31, 46
- [22] Kai M. Hüner, Martin Ofner, and Boris Otto. Towards a maturity model for corporate data quality management. In *Proceedings of the 2009 ACM Symposium on Applied Computing*, pages 231–238, New York, NY, USA, 2009. Association for Computing Machinery. doi: 10.1145/1529282.1529334. URL <https://doi.org/10.1145/1529282.1529334>. SAC ’09. 24, 31
- [23] IBM. What is data quality?, 2024. URL <https://www.ibm.com/topics/data-quality>. Accessed: 2024-10-06. 9
- [24] International Organization for Standardization. Iso 8000-1:2022 - data quality – part 1: Overview, principles and terminology. <https://www.iso.org/standard/80000-1.html>, 2022. Accessed: 2024-10-19. 9, 14
- [25] Regin Iqbal, Pramana Yuda, Wahyu Aditya, Achmad Nizar Hidayanto, Putu Wuri Handayani, and Nabila Clyde Harahap. Master data management maturity assessment: Case study of xyz company. In *2019 2nd International Conference on Applied Information Technology and Innovation (ICAITI)*, pages 133–139, Denpasar, Indonesia, 2019. IEEE. doi: 10.1109/ICAITI48442.2019.8982123. URL <https://doi.org/10.1109/ICAITI48442.2019.8982123>. 24, 27, 29, 31, 46
- [26] ISO/IEC JTC 1/SC 7. ISO/IEC 33000 Family of Standards. <https://committee.iso.org/sites/jtc1sc7/home/projects/flagship-standards/isoiec-33000-family.html>, –. Accessed: 26 June 2025. 3

- [27] Jinlong Kang, Zeina Al Masry, Christophe Varnier, Ahmed Mosallam, and Noureddine Zerhouni. A data quality management framework for equipment failure risk estimation: Application to the oil and gas industry. *Engineering Applications of Artificial Intelligence*, 136:108834, 2024. ISSN 0952-1976. doi: <https://doi.org/10.1016/j.engappai.2024.108834>. URL <https://www.sciencedirect.com/science/article/pii/S0952197624009928>. 9
- [28] Mounir El Khatib, Aya Yaish, and Hessa Al Ali. Implementation challenges of data quality management: Cases from uae public sector. *iBusiness*, 13(3):144–153, 2021. doi: 10.4236/ib.2021.133009. URL <https://doi.org/10.4236/ib.2021.133009>. 9
- [29] Barbara Kitchenham and Stuart Charters. Guidelines for performing systematic literature reviews in software engineering. Technical Report EBSE-2007-01, Technical report, EBSE Technical Report. EBSE-2007-01, UK, 01 2007. 5, 15, 18, 22, 32
- [30] Dwitama Heryadi Kurniawan, Yova Ruldeviyani, Mohammad Rizky Adrian, Sutia Handayani, M. Rizki Pohan, and Rani Khairunnisa T. Data governance maturity assessment: A case study in it bureau of audit board. In *2019 International Conference on Information Management and Technology (ICIMTech)*, volume 1, pages 629–634, 10.1109/ICIMTech.2019.8843742, Aug 2019. IEEE. 2, 11, 12, 23, 27, 31, 46
- [31] Simone Malacaria, Andrea De Mauro, Marco Greco, Michele Grimaldi, and Benito Mignacca. Toward the implementation of a consensual maturity model for big data in consumer goods companies. In *Knowledge Drivers for Resilience and Transformation, IFKAD 2022*, pages 2380–2401, Lugano, Switzerland, 2022. IFKAD. 3, 12, 13, 23, 27, 28, 29, 31
- [32] Soukaina Mouhib, Houda Anoun, Larbi Hassouni, and Mohammed Ridouani. Analyzing the global big data maturity model domains for better adoption of big data projects. *International Journal of Information Science and Management*, 21(4):83–102, 2023. doi: 10.22034/ijism.2023.1977940.0. 3, 11, 13, 23, 27, 28, 31, 46
- [33] Soukaina Mouhib, Houda Anoun, Mohammed Ridouani, and Larbi Hassouni. Global big data maturity model and its corresponding assessment framework results. *IAENG International Journal of Applied Mathematics*, 53(1):12, 2023. URL <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85149629399>. Cited by: 6. 3, 24, 27, 28, 30, 31, 46
- [34] Rayelle Ingrid Vera Cruz Muniz, Karina Moura Silva, Marcelo Iury S. Oliveira, and Bernadette Farias Loscio. Towards a maturity model for data on the web initiatives. In *Proceedings of the XVI Brazilian Symposium on Information Systems*, page 44, New York, NY, USA, 2020. Association for Computing Machinery. doi: 10.1145/3411564.3411655. URL <https://doi.org/10.1145/3411564.3411655>. 24, 31
- [35] Tito Febrian Nugraha, Wahyu Setiawan Wibowo, Venera Genia, Ahmad Fadhil, and Yova Ruldeviyani. A practical approach to enhance data quality management in government: Case study of indonesian customs and excise office. *Journal of Information Systems Engineering and Business Intelligence*, 10(1):51–69, 2024. doi: 10.20473/jisebi.10.1.51-69. 23, 27, 28, 31, 46

- [36] Remi Indra Permana and Jarot S. Suroso. Data governance maturity assessment at pt. xyz: Case study of the data management division. In *2018 International Conference on Information Management and Technology (ICIMTech)*, pages 15–20, 10.1109/ICIMTech.2018.8528142, September 2018. IEEE. 23, 27, 28, 29, 31, 46
- [37] Nanik Qodarsih, Satrio Baskoro Yudhoatmojo, and Achmad Nizar Hidayanto. Master data management maturity assessment: A case study in the supreme court of the republic of indonesia. In *2018 6th International Conference on Cyber and IT Service Management (CITSM)*, pages 1–7, Parapat, Indonesia, 2018. IEEE. doi: 10.1109/CITSM.2018.8674373. 24, 27, 31, 46
- [38] Vítor Hugo Machado Ribeiro, João Barata, and Paulo Rupino da Cunha. Sustainable data governance: A systematic review and a conceptual framework. In *Proceedings of the 30th International Conference on Information Systems Development (ISD2022)*, pages 1–12, Cluj-Napoca, Romania, 2022. ISD, University of Coimbra. 10, 12
- [39] Yusuf Setiadi, Achmad Nizar Hidayanto, Fitri Rachmawati, and Adhi Yuniarto Laurentius Yohannes. Data quality management maturity model: A case study in higher education’s human resource department. In *2021 IEEE 7th International Conference on Computing, Engineering and Design (ICCED)*, pages 1–5, 10.1109/ICCED53389.2021.9664881, August 2021. 2, 10, 12
- [40] Marco Spruit and Katharina Pietzka. Md3m: The master data management maturity model. *Computers in Human Behavior*, 51:1068–1076, 2015. doi: 10.1016/j.chb.2014.09.030. URL <https://doi.org/10.1016/j.chb.2014.09.030>. 24, 27, 29, 30, 31, 46
- [41] Lidiane Visintin, José Leomar Todesco, and Fernando Álvaro Ostuni Gauthier. Open data maturity model: A reference matrix for organizations. In *Proceedings of the 14th Seminar on Ontology Research in Brazil (ONTOBRAS 2021) and 5th Doctoral and Masters Consortium on Ontologies (WTDO 2021)*, volume 3050, pages 291–296, 2021. URL <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85123288169>. Conference: ONTOBRAS 2021; Florianópolis, Brazil. 24, 28, 29, 30, 31, 46
- [42] Kristin Weber, Boris Otto, and Hubert Österle. One size does not fit all—a contingency approach to data governance. *J. Data and Information Quality*, 1(1):Article 4, 2009. doi: 10.1145/1515693.1515696. URL <https://doi.org/10.1145/1515693.1515696>. June 2009. 24, 31
- [43] Chao Xue, Yue Zhang, and Dan Jia. History, current situation and future challenges of china’s geographic information data governance. In *29th International Conference on Geoinformatics, Beijing, China, August 15-18, 2022*, pages 1–5, Beijing, China, 2022. IEEE. doi: 10.1109/GEOINFORMATICS57846.2022.9963876. URL <https://doi.org/10.1109/Geoinformatics57846.2022.9963876>. 23, 28, 29, 30, 31