



Universidade de Brasília

Instituto de Ciências Exatas
Departamento de Ciência da Computação

**Proposta de Governança de Dados na Agência
Nacional de Telecomunicações – Uma Análise da
Qualidade dos Dados do Programa Banda Larga nas
Escolas**

Luigy de Freitas

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

Orientador

Prof. Dr. Márcio de Carvalho Victorino

Coorientadora

Prof.a Dr.a Maristela Terto de Holanda

Brasília
2025

Dados Internacionais de Catalogação na Publicação (CIP)
(Biblioteca Central da Universidade de Brasília - BCE/UNB)

F866p Freitas, Luigy.
 Proposta de governança de dados na Agência
 Nacional de Telecomunicações [recurso
 eletrônico] : uma análise da qualidade dos dados
 do programa banda larga nas escolas / Luigy
 Freitas. -- Brasília, 2025.
 124 f. : il.

 Inclui bibliografia.
 Dissertação (Mestrado profissional em Computação
 aplicada) - Universidade de Brasília, Instituto de
 Ciências Exatas, 2025.
 Orientação: Márcio de Carvalho Victorino.
 Coorientação: Maristela Terto de Holanda.

 1. Governança de dados. 2. Qualidade de dados.
 3. Dimensões da qualidade. 4. PBLE. 5. Anatel. I.
 Victorino, Márcio de Carvalho, orient. II.
 Holanda, Maristela Terto de, coorient. III.
 Título.



Universidade de Brasília

Instituto de Ciências Exatas
Departamento de Ciência da Computação

Proposta de Governança de Dados na Agência Nacional de Telecomunicações – Uma Análise da Qualidade dos Dados do Programa Banda Larga nas Escolas

Luigy de Freitas

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

Prof. Dr. Márcio de Carvalho Victorino (Orientador)
PPCA/UnB

Prof. Dr. Marcelo Ladeira Prof.a Dr.a Ana Paula Bernardi da Silva
PPCA/UnB Universidade Católica de Brasília

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

Brasília, 30 de junho de 2025

Dedicatória

Dedico esta dissertação à minha avó materna, Carlota Lourenço de Freitas (*in memoriam*), minha fonte inesgotável de inspiração, exemplo de força, integridade e resiliência. Vó, foi com a senhora que aprendi a encarar os desafios mais difíceis com coragem, dignidade e perseverança. A sua trajetória de vida me ensinou que, mesmo diante dos maiores obstáculos, é possível seguir em frente com fé, garra e determinação. Esta conquista também é sua, vó!

Dedico, com gratidão e carinho, ao Ricardinho (Ricardo Casé Filho), que esteve ao meu lado em cada etapa desta jornada. Seu apoio constante, incentivo generoso e presença foram fundamentais para a realização deste estudo e para a superação dos momentos mais desafiadores do mestrado. Nada disso teria sido possível sem você, Ricardinho!

“Você não pode ser analítico sem dados, e você não pode ser realmente bom em análise sem dados realmente bons.”.

Agradecimentos

Em primeiro lugar, agradeço a Deus pelo dom da vida e por ter estado ao meu lado nos momentos mais difíceis dessa caminhada. Sem Sua presença constante, força e luz, não teria sido possível concluir mais esta etapa da minha trajetória.

A palavra que define essa jornada chamada Mestrado Profissional em Computação Aplicada da Universidade de Brasília é **persistência**. Foram meses de esforço, dúvidas e superações que, com fé e determinação, se transformaram em aprendizado e crescimento.

Agradeço de coração à minha família, que, mesmo à distância, sempre se fez presente em minha trajetória. Foram eles que me ajudaram a construir meus valores, sustentaram meus passos e contribuíram, de forma essencial, para tudo o que sou hoje.

Expresso minha profunda gratidão a todos os meus professores: da Escola Estadual Madre Santa Face, especialmente à Irmã Marlene (*in memoriam*), que acreditou em mim desde cedo; do Colégio de Aplicação da Universidade Federal de Viçosa (Coluni); e da Universidade Federal de Viçosa (UFV), onde me formei em Ciência da Computação. Sou fruto da educação pública de qualidade oferecida por essas instituições da minha cidade natal, Viçosa – MG, e levo comigo cada ensinamento recebido ao longo dessa caminhada. Este mestrado também é de vocês!

Dedico um agradecimento especial ao Ricardo Casé Filho, que me acompanhou em todos os momentos dessa jornada. Sua parceria, suas revisões cuidadosas e, sobretudo, sua presença constante foram essenciais para que este trabalho alcançasse o nível de qualidade que tanto almejei. Ricardinho, você foi fundamental para que eu pudesse acreditar na conclusão deste mestrado. Muito obrigado!

Ao meu amigo e irmão de caminhada, Thiago Cunha de Souza, meu muito obrigado pelas revisões – inclusive no *abstract* –, pelo apoio generoso e pelo carinho de sempre. Thi, você é top demais!

Aos meus orientadores, Prof. Dr. Márcio Victorino e Profa. Dra. Maristela Terto de Holanda, deixo minha gratidão profunda pela escuta atenta, pela orientação cuidadosa e pelos valiosos conselhos ao longo de todo o processo. Levarei seus ensinamentos para a vida.

Agradeço à Universidade de Brasília (UnB) por acolher este sonho, e a todos os professores do PPCA por sua dedicação e generosidade na partilha do conhecimento. Foi um privilégio caminhar com vocês.

À Agência Nacional de Telecomunicações (Anatel), onde trabalho com orgulho desde 2007, agradeço o apoio institucional e, especialmente, ao colega e ex-gerente José Borges Neto, que compreendeu a importância deste estudo e me incentivou a buscar novos aprendizados com dedicação e seriedade. Sua confiança foi combustível para a realização deste projeto.

Aos colegas e amigos do PPCA, que foram exemplo de superação e companheirismo, meu carinho e admiração. Enfrentamos juntos uma pandemia avassaladora e, ainda assim, nos mantivemos firmes, reinventando-nos como estudantes. Foram muitas sextas-feiras e sábados de troca, esforço e construção conjunta de saberes. Essas vivências me deixam um legado de aprendizado, companheirismo e profunda admiração por cada um de vocês.

Hoje, ao olhar para trás, vejo que esse sonho valeu a pena. E ele só foi possível porque muitas mãos me apoiaram ao longo do caminho.

Resumo

Os dados são hoje reconhecidos como ativos estratégicos para as organizações, mas a garantia de sua qualidade constitui um desafio constante. Com o avanço tecnológico e a expansão do acesso à internet, a inclusão digital emerge como componente essencial das políticas públicas. Nesse contexto, o Programa Banda Larga nas Escolas (PBLE) configura-se como uma iniciativa para promover a conectividade em escolas públicas brasileiras. Para viabilizar as atividades de fiscalização e acompanhar o cumprimento das metas de conexão, a Agência Nacional de Telecomunicações (Anatel) recebe dados das prestadoras responsáveis por conectar as escolas. No entanto, a inexistência de um mecanismo sistemático de avaliação da qualidade desses dados compromete o acompanhamento do programa. Este estudo tem como objetivo principal realizar a avaliação da qualidade dos dados do PBLE, fundamentada em dimensões da qualidade. Foram empregadas metodologias quantitativas e qualitativas para mensurar o desempenho dos dados e identificar deficiências nos processos de coleta e validação das informações. Com base nos resultados obtidos, propõe-se um conjunto de três medidas articuladas: Regras de Qualidade de Dados, Modelos de Auditoria de Dados e Relatórios de Auditoria de Dados do PBLE. Essas ações visam não apenas melhorar a qualidade dos dados, mas também reforçar a capacidade analítica da Agência na fiscalização e no controle das obrigações do programa. Como inovação, o estudo introduz, no âmbito da Anatel, uma abordagem estruturada para avaliação da qualidade de dados, oferecendo subsídios metodológicos e operacionais que podem ser replicados internamente na Agência. Adicionalmente, fortalece a governança de dados no setor de telecomunicações ao estabelecer processos contínuos de monitoramento e melhoria da qualidade dos dados.

Palavras-chave: governança de dados, qualidade de dados, dimensões da qualidade, PBLE, Anatel.

Abstract

Data are increasingly recognized as strategic assets for organizations, but ensuring their quality remains a constant challenge. With technological advances and the growing expansion of internet access, digital inclusion has become an essential component of public policies. In this context, the Programa Banda Larga nas Escolas (PBLE) – Broadband in Schools Program stands out as a key initiative aimed at promoting connectivity in Brazilian public schools. To support oversight activities and monitor compliance with connection goals, the National Telecommunications Agency (Anatel) receives data from telecommunications providers responsible for connecting the schools. However, the absence of a systematic mechanism for assessing the quality of these data compromises the effective monitoring of the program. This study aims to evaluate the data quality of the PBLE based on well-established quality dimensions. Both quantitative and qualitative methodologies were employed to measure data performance and to identify deficiencies in data collection and validation processes. Based on the findings, a set of three coordinated measures is proposed: Data Quality Rules, Data Audit Models and Data Audit Reports. These actions are designed not only to improve the reliability and accuracy of the data but also to enhance Anatel’s analytical capacity in overseeing and enforcing program obligations. As an innovative contribution, this research introduces a structured and replicable approach to data quality assessment within Anatel, providing methodological and operational guidance that can be adopted across other initiatives. Additionally, it strengthens data governance in the telecommunications sector by establishing continuous processes for monitoring and improving data quality.

Keywords: data governance, data quality, dimensions of quality, PBLE, Anatel.

Sumário

1	Introdução	1
1.1	Contextualização	2
1.2	O PBLE	5
1.2.1	Números do PBLE	6
1.3	Fiscalização	7
1.3.1	Fiscalização do PBLE	9
1.4	Dados Abertos	12
1.5	Justificativa	14
1.6	Objetivos	15
1.6.1	Objetivos Específicos	16
1.7	Estrutura do Trabalho	16
2	Metodologia da Pesquisa	18
3	Mapeamento da Literatura	20
3.1	Protocolo de Pesquisa	20
3.1.1	Questões de Pesquisa	21
3.1.2	Identificação e Seleção dos Estudos	21
3.1.3	Análise de Qualidade da Pesquisa	23
3.1.4	Extração de Dados da Pesquisa	23
3.1.5	Síntese e Análise dos Dados da Pesquisa	23
4	Trabalhos Relacionados	26
4.1	Análise dos Trabalhos Relacionados	26
4.2	Resumo dos Trabalhos Relacionados	27
5	Fundamentação Teórica	31
5.1	Dados	31
5.1.1	Conceito de Dados	32
5.1.2	Classificação dos Dados	33

5.2	Governança de Dados	35
5.2.1	Definição de Governança de Dados	36
5.2.2	Princípios da Governança de Dados	38
5.2.3	Cenário Atual da Governança de Dados	39
5.2.4	Implementação da Governança de Dados	40
5.3	Qualidade de Dados	42
5.3.1	Conceito da Qualidade de Dados	42
5.3.2	Cenário da Qualidade de Dados	43
5.3.3	Áreas de Pesquisa da Qualidade de Dados	44
5.3.4	Riscos da Baixa Qualidade de Dados	46
5.3.5	Avaliação da Qualidade de Dados	47
5.4	Dados Governamentais Abertos	48
5.4.1	Qualidade de Dados Governamentais Abertos	49
5.5	Dimensões da Qualidade de Dados	49
5.6	Metodologias, Técnicas e Modelos	50
5.6.1	Modelos para Avaliação da Qualidade de Dados	51
5.6.2	Metodologias para Avaliação da Qualidade de Dados	52
6	Desenvolvimento do Trabalho	55
6.1	Arquitetura da Informação	55
6.2	Arquitetura da Informação do PBLE	56
6.3	Dados do Estudo	57
6.3.1	Coleta dos Dados	58
6.3.2	Armazenamento dos Dados	58
6.3.3	Fiscalização e Qualidade dos Dados	59
6.3.4	Avaliação da Qualidade dos Dados	60
6.4	Metodologia da Avaliação da Qualidade de Dados	61
6.5	Desenvolvimento da Pesquisa	62
6.5.1	Identificação dos Dados do PBLE	62
6.5.2	Seleção dos Dados do PBLE	63
6.5.3	Identificação dos Problemas de Qualidade dos Dados	65
6.5.4	Seleção das Dimensões de Qualidade de Dados	66
6.5.5	Avaliação da Qualidade dos Dados do PBLE	72
6.5.5.1	Análise Preliminar da Qualidade dos Dados	73
6.5.5.2	Métricas de Avaliação da Qualidade de Dados	75
6.5.5.3	Avaliação dos Dados do PBLE	77
6.6	Diagnóstico dos Dados do PBLE	78
6.6.1	Campos com Alta Qualidade de Dados	79

6.6.2	Campos com Baixa Qualidade de Dados	82
6.7	Resultado da Avaliação da Qualidade dos Dados	84
6.8	Solução Proposta	85
6.8.1	Regras de Qualidade de Dados	86
6.8.2	Modelos de Auditoria de Dados	89
6.8.2.1	Modelo de Auditoria do SICI	89
6.8.2.2	Modelo de Auditoria dos Sistemas Internos	90
6.8.2.3	Modelo de Auditoria de Fiscalização e das Escolas	91
6.8.3	Relatório de Auditoria de Dados	92
6.8.3.1	Relatório de Auditoria de Dados do SICI	93
6.8.3.2	Relatório de Auditoria de Dados dos Sistemas Internos	94
6.8.3.3	Relatório de Auditoria de Fiscalização e das Escolas	96
6.9	Análise de Riscos	97
7	Conclusão e Trabalhos Futuros	99
7.1	Conclusão	99
7.2	Contribuições	100
7.2.1	Contribuição Tecnológica	100
7.2.2	Contribuição de Inovação	101
7.2.3	Contribuição para a Fiscalização	102
7.2.4	Geração de <i>Scripts</i>	102
7.3	Impactos para a Anatel	103
7.4	Trabalhos Futuros	103
	Referências	105
	Anexo	112
I	Scripts de Tabelas de Auditoria	113
I.1	Criação da Tabela BandaLargaEscolas_Aud	113
I.2	Criação da Tabela Prestadora_Aud	114
I.3	Criação da Tabela Escola_Aud	114
I.4	Criação da Tabela Tecnologia_Aud	115
I.5	Criação da Tabela Municipio_Aud	115
I.6	Criação da Tabela Fiscalizacao_Escola_Aud	116
II	Scripts de Extração de Dados	117
II.1	Base de Dados: PBLE	
	Sistema Interno: SICI	117

II.2	Base de Dados: Prestadoras	
	Sistema Interno: STEL	118
II.3	Base de Dados: Escolas	
	Sistema Interno: INEP	119
II.4	Base de Dados: Tecnologia	
	Sistema Interno: STEL	119
II.5	Base de Dados: Municípios	
	Sistema Interno: CEP	120
III	Scripts de Geração de Relatórios	121
III.1	Relatório SICI x Prestadora_Aud	121
III.2	Relatório SICI x Escola_Aud	122
III.3	Avaliação SICI x Municipio_Aud	122
III.4	Relatório SICI x Tecnologia_Aud	123
III.5	Relatório SICI x Fiscalizacao_Escola_Aud	123

Lista de Figuras

1.1	Escolas Conectadas pelo PBLE por Ano	6
1.2	Etapas da Fiscalização da Anatel	8
1.3	Etapas do Procedimento de Fiscalização do PBLE	10
5.1	Classificação dos Dados	33
5.2	Relações Entre Tipos de Governança	37
5.3	Diagrama do DAMA DMBOK V2 - Evoluído	41
5.4	Áreas de Pesquisa em Qualidade de Dados - Adaptada	45
5.5	Pesquisas em Qualidade de Dados - Adaptada	46
5.6	Fases da Metodologia de Avaliação da Qualidade de Dados - Adaptada	52
6.1	Arquitetura da Informação do PBLE	56
6.2	Fluxo dos Dados do PBLE	57
6.3	Metodologia de Avaliação da Qualidade de Dados	61
6.4	Resultado dos campos <i>CodINEP</i> e <i>IP</i>	74
6.5	Resultado dos campos <i>VelAcessoInstalada</i> e <i>IdtTecnologia</i>	74
6.6	<i>DataMovimentoAno</i> - Resultado da Avaliação da Qualidade	80
6.7	<i>DataMovimentoMes</i> - Resultado da Avaliação da Qualidade	80
6.8	<i>NumFistel</i> - Resultado da Avaliação da Qualidade	81
6.9	<i>IdtSituacao</i> - Resultado da Avaliação da Qualidade	81
6.10	<i>NumeroEscola</i> - Resultado da Avaliação da Qualidade	82
6.11	<i>DataPrevisao</i> - Resultado da Avaliação da Qualidade	83
6.12	<i>codTipoObrigacao</i> - Resultado da Avaliação da Qualidade	83
6.13	<i>DataDesativacao</i> - Resultado da Avaliação da Qualidade	84
6.14	Modelo de Auditoria do Sistema SICI	89
6.15	Modelo de Auditoria do Sistemas Internos	90
6.16	Modelo de Auditoria de Fiscalização e das Escolas	91
6.17	Relatório de Auditoria de Dados do SICI	93

Lista de Tabelas

3.1	Trabalhos Identificados sobre Governança de Dados - QP1	24
3.2	Trabalhos sobre Melhoria da Qualidade de Dados - QP2	24
3.3	Áreas Abordadas sobre Melhoria da Qualidade de Dados - QP2	25
3.4	Dimensões da Qualidade de Dados mais Exploradas na Literatura - QP3 . .	25
5.1	Classificação de Dimensão Por Critérios	48
6.1	Campos do PBLE	63
6.2	Dados do PBLE Selecionados	65
6.3	Tipos de Problemas de Dados do PBLE	66
6.4	Critérios da Dimensão Completude	67
6.5	Critérios da Dimensão Consistência	68
6.6	Critérios da Dimensão Conformidade	69
6.7	Critérios da Dimensão Integridade	70
6.8	Critérios da Dimensão Precisão	71
6.9	Critérios da Dimensão Atualidade	72
6.10	Relação entre Dimensões e Problemas de Qualidade do PBLE	73
6.11	Conjunto de Dados Históricos do PBLE	74
6.12	Métricas dos Indicadores de Avaliação de Qualidade	76
6.13	Avaliação dos Campos do PBLE	78
6.14	Regras de Qualidade - Dados de Identificação de Escolas	87
6.15	Regras de Qualidade - Dados de Controle de Instalação	88
6.16	Regras de Qualidade - Dados Técnicos de Telecomunicações	89
6.17	Relatório de Auditoria de Dados do Sistema STEL	94
6.18	Relatório de Auditoria de Dados do Sistema INEP	94
6.19	Relatório de Auditoria de Dados do Sistema CEP	95
6.20	Relatório de Auditoria de Dados de Fiscalização e Escolas	96
6.21	Matriz de Impactos de Riscos	98

Lista de Abreviaturas e Siglas

Anatel Agência Nacional de Telecomunicações.

CSV Comma-separated Values.

DAMA Data Management Association.

DGA Dados Governamentais Abertos.

DMBoK Data Management Body of Knowledge.

FNDE Fundo Nacional de Desenvolvimento da Educação.

GDPR General Data Protection Regulation.

INEP Instituto Nacional de Estudos e Pesquisas Educacionais.

MCom Ministério das Comunicações.

MEC Ministério da Educação.

OKF Open Knowledge Foundation.

PAC Processo de Acompanhamento e Controle.

PAF Plano Anual de Fiscalização.

PBLE Projeto Banda Larga nas Escolas.

SCM Serviço de Comunicação Multimídia.

SGBD Sistema Gerenciador de Banco de Dados.

SICI Sistema de Coleta de Informações.

STFC Serviço Telefônico Fixo Comutado.

TIC Tecnologia da Informação e Comunicações.

UIT União Internacional de Telecomunicações.

Capítulo 1

Introdução

Os dados têm se consolidado como um dos ativos mais valiosos para as organizações contemporâneas, uma vez que seu uso adequado gera valor significativo e informações cruciais para a tomada de decisões estratégicas.

Contudo, para que esses dados possam efetivamente contribuir para os objetivos organizacionais, é imprescindível garantir sua governança e qualidade. Isso assegura que os dados mantenham relevância nas atividades decisórias, na avaliação de cenários futuros, na mitigação de riscos de negócios, além de garantir a produção e o uso de informações confiáveis.

No cenário atual, os dados deverão, cada vez mais, ser controlados mediante uma gestão eficiente e uma governança presente, que também estenderá o seu radar para a forma automática e independente com que são tratados e os resultados produzidos [1].

É amplamente reconhecido que o dado é a base para a informação, e, por sua vez, a informação constitui o alicerce do conhecimento. Esse processo exige que as organizações adotem práticas eficazes de gestão do conhecimento, assegurando que os dados sejam não apenas coletados, mas também gerenciados de forma a gerar insights e apoiar decisões estratégicas [2].

Em um ambiente global interconectado, onde as tendências e as demandas evoluem rapidamente, a disponibilidade de informações atualizadas e confiáveis torna-se essencial para o sucesso organizacional. Nesse contexto, a qualidade dos dados é um fator determinante para garantir a entrega de informações precisas, oportunas e relevantes [3].

A transformação digital e o crescente uso de estratégias baseadas em dados inteligentes tornaram os conceitos de Governança de Dados e Qualidade de Dados componentes centrais deste novo paradigma. Esses conceitos são fundamentais para garantir que os dados sejam utilizados de forma eficiente, segura e ética, agregando valor real ao processo organizacional.

Assim como as atividades empresariais devem ser constantemente aperfeiçoadas, os dados, como ativos estratégicos, precisam ser avaliados em diversos aspectos, incluindo integridade, exatidão, disponibilidade, confiabilidade e atualidade.

Nesse sentido, a Governança de Dados – conceito em evolução que envolve diversas disciplinas com foco central em Qualidade de Dados – passa por uma busca de maturidade da empresa na gerência de seus recursos, melhoria na valoração e produção dos dados, monitoração de seu uso, além de aspectos críticos de segurança, privacidade, ética e aderência a regras de *compliance* associadas a ele [1].

Pesquisas apontam que a Qualidade de Dados desempenha um papel crucial no conhecimento organizacional e na eficácia das soluções de Tecnologia da Informação dentro das empresas, impactando diretamente as decisões baseadas em dados [4]. A falta de qualidade nos dados pode comprometer significativamente a análise e interpretação das informações, prejudicando a competitividade e a capacidade de inovação das organizações.

Esse impacto negativo da Qualidade de Dados é evidenciado por estudos, como o relatório do *Data Warehouse Institute*, que estima que os problemas relacionados à qualidade dos dados geram um custo anual superior a 600 bilhões de dólares para as empresas norte-americanas [5].

Além disso, a Qualidade de Dados tem sérias consequências para a eficiência e eficácia de organizações e seus processos de negócios. Assim, garantir a qualidade aos dados é fundamental no mundo informacional para evitar que elementos incompletos, campos sem significado, dados com erros ou com precisões duvidosas contaminem a tomada de decisão das organizações [1].

Portanto, garantir a qualidade dos dados configura-se como um elemento estratégico crucial para a competitividade e sustentabilidade das organizações no contexto atual, onde decisões baseadas em informações precisas e confiáveis são determinantes para o sucesso a longo prazo.

1.1 Contextualização

Nas últimas décadas, a Tecnologia da Informação e Comunicações (TIC) promoveram uma revolução na forma como nos comunicamos e interagimos. O uso da internet, e-mail, redes sociais e aplicativos de mensagens transformou a maneira como as pessoas se conectam, ampliando as possibilidades de interação de maneira sem precedentes.

Com os avanços tecnológicos impulsionados pelas TIC e a rápida expansão do acesso à internet, a inclusão digital tornou-se uma questão de grande relevância. Garantir a inclusão digital de todas as camadas da população é essencial, uma vez que o acesso à rede mundial de computadores possibilita o aprendizado e o desenvolvimento econômico,

contribuindo para a redução das desigualdades sociais e ampliando as oportunidades em diversos setores da sociedade [6].

A inclusão digital visa assegurar que todas as pessoas tenham a oportunidade de utilizar e se beneficiar das tecnologias digitais. Ao promover a inclusão digital, um país pode reduzir a lacuna existente entre os cidadãos que têm acesso e habilidades para utilizar as tecnologias e aqueles que não dispõem desses recursos.

Neste contexto, o Estado brasileiro tem a necessidade urgente de adotar políticas públicas voltadas para a inclusão digital, especialmente no âmbito educacional, com o objetivo de democratizar o acesso à internet nas escolas públicas. A implementação dessas políticas pode reduzir desigualdades sociais e expandir as oportunidades educacionais para estudantes de diferentes regiões do país [6].

A universalização do acesso à banda larga nas escolas públicas brasileiras tem sido um tema central nas discussões sobre a educação digital, especialmente após a pandemia de covid-19. Esse debate ressaltou a importância da conectividade escolar, não apenas para a criação de novas oportunidades pedagógicas, mas também para ampliar o uso dos recursos educacionais já existentes.

Estudos internacionais, como o realizado na Nova Zelândia, evidenciam o impacto positivo da instalação de banda larga de alta velocidade no desempenho acadêmico. A pesquisa realizada em mais de 1.800 escolas daquele país revelou que a implementação da fibra ótica contribuiu para um aumento significativo nas taxas de aprovação dos alunos nas disciplinas de Matemática, Leitura e Escrita, com crescimentos de 31,3%, 4,8% e 26,2%, respectivamente, após três anos de adoção da tecnologia [7].

Este estudo também demonstrou que o impacto foi mais pronunciado em escolas com maior concentração de estudantes de origens socioeconômicas mais baixas, reforçando a importância da conectividade para a inclusão educacional [7].

Outro estudo realizado na Noruega, em 2023, reforça a integração crescente das TICs no cotidiano escolar. O estudo destacou a importância da cobertura de banda larga nas escolas norueguesas, considerando-a essencial para o desenvolvimento de oportunidades de aprendizagem e para o aprimoramento das competências dos estudantes. Além disso, evidenciou que a qualidade da conectividade tem um impacto direto no desempenho acadêmico e influencia as oportunidades profissionais futuras no mercado de trabalho [8].

Ademais, o setor de telecomunicações da União Internacional de Telecomunicações (UIT), da qual o Brasil é membro desde 1947, destaca o crescimento expressivo do número de usuários de internet, o que tem impulsionado uma transformação digital global. Esse processo tem gerado impactos significativos em todos os países, incluindo o Brasil, promovendo uma sociedade mais conectada e com maior acesso à informação [9].

Embora a internet tenha um potencial significativo para promover a educação e o bem-estar social e econômico, esse potencial permanece em grande parte inexplorado. Atualmente, cerca de um terço da população mundial, aproximadamente 3 bilhões de pessoas, ainda não tem acesso à internet, e muitos dos usuários conectados fazem uso da tecnologia de forma limitada, utilizando apenas serviços básicos [9].

Nesse contexto, a conectividade significativa, definida como a capacidade de proporcionar uma experiência digital segura, satisfatória, produtiva e acessível para todos os cidadãos, tornou-se essencial para a concretização dos Objetivos de Desenvolvimento Sustentável (ODS) da União Internacional de Telecomunicações UIT na próxima década [9].

Essa conectividade desempenha um papel fundamental na garantia de uma educação de qualidade, inclusiva e igualitária, capaz de promover oportunidades de aprendizagem para todos os estudantes, independentemente de sua localização ou condição socioeconômica [9].

No Brasil, um estudo realizado pela empresa MegaEdu [10] revela que cerca de 90% dos estudantes mundialmente interromperam suas atividades escolares durante a pandemia de covid-19. De acordo com o Censo Escolar 2021, divulgado pelo INEP, quase todas as escolas públicas brasileiras (99,3%) suspenderam as atividades presenciais durante o período de pandemia, com uma média nacional de 279 dias de suspensão das aulas presenciais [11].

Esse cenário evidenciou a necessidade urgente de políticas públicas eficazes para reduzir as lacunas de conectividade nas escolas públicas e expandir o uso da internet no país. O estudo da MegaEdu [10] aponta que uma ação estratégica por parte do governo é essencial para conectar as escolas públicas e garantir o acesso contínuo à educação digital.

Nesse contexto, em 2008, o governo brasileiro criou o Programa Banda Larga nas Escolas (PBLE), por meio do Decreto nº 6.424 [12], com o objetivo de assegurar o acesso gratuito à tecnologia de banda larga em 100% das escolas públicas urbanas do Brasil até 2025. A iniciativa visa promover a inclusão digital dos estudantes e modernizar a educação brasileira, integrando as tecnologias ao processo de ensino-aprendizagem.

O PBLE destaca-se como uma estratégia essencial para o fortalecimento das políticas públicas de inclusão digital nas escolas públicas brasileiras, sendo parte de um esforço contínuo para modernizar o sistema educacional.

A Agência Nacional de Telecomunicações (Anatel), órgão regulador responsável pela organização e fiscalização dos serviços de telecomunicações no Brasil, desempenha um papel crucial nesse processo. A atuação da Anatel é fundamental para garantir que as empresas prestadoras de serviços de telecomunicações cumpram suas obrigações relativas à instalação de conexões de banda larga nas escolas públicas participantes, assegurando que o programa atenda aos seus objetivos de forma eficaz.

1.2 O PBLE

O PBLE foi instituído em abril de 2008, por meio da assinatura de Termos Aditivos aos respectivos Termos de Autorização para Exploração do Serviço de Comunicação Multimídia (SCM) – também conhecido como Serviço de Banda Larga – entre Anatel e as então concessionárias do Serviço Telefônico Fixo Comutado (STFC ou Telefonia Fixa): Companhia de Telecomunicações do Brasil Central (**Algar**), Telemar Norte Leste S.A. e Brasil Telecom S.A. (**Oi**), Telecomunicações de São Paulo S/A (**Telefônica**) e Sercomtel S/A Telecomunicações (**Sercomtel**) [13].

O principal objetivo do PBLE é garantir o acesso gratuito à internet em banda larga para todas as escolas públicas urbanas do país até dezembro de 2025. Para alcançar essa meta, as instituições de ensino elegíveis devem ser conectadas pelas prestadoras de serviços de telecomunicações, seguindo as diretrizes definidas pelo programa e utilizando tecnologias de conexão terrestre (como fibra ótica ou cabo) ou via satélite.

O PBLE representa uma importante ampliação da política de tecnologia educacional do Estado brasileiro, consolidando-se como uma iniciativa decorrente dos avanços tecnológicos e da necessidade de desenvolvimento da infraestrutura de conectividade no ambiente escolar. Trata-se de uma ação colaborativa entre o Estado e as concessionárias do setor de telecomunicações, fundamentada em compromissos regulatórios e estratégicos [14].

A gestão do programa é conduzida de forma conjunta pela Anatel e pelo Ministério da Educação (MEC), por meio do Fundo Nacional de Desenvolvimento da Educação (FNDE), em cooperação com o Ministério das Comunicações (MCom) e com as Secretarias Estaduais e Municipais de Educação.

No âmbito do PBLE, a Anatel é a entidade responsável por fiscalizar, acompanhar, controlar e verificar o cumprimento das obrigações assumidas pelas prestadoras. Já o FNDE é incumbido de elaborar, com base no Censo Escolar Anual, conduzido pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP), a lista oficial das escolas públicas urbanas de ensino fundamental e médio elegíveis para atendimento pelo programa.

Dessa forma, todas as escolas públicas urbanas – federais, estaduais ou municipais – identificadas no Censo Escolar do INEP podem ser consideradas elegíveis a integrar o PBLE, desde que atendam aos critérios técnicos estabelecidos. A instalação da infraestrutura de conexão é de responsabilidade das prestadoras de telecomunicações autorizadas, conforme cronograma previamente definido.

Esse cronograma é elaborado anualmente pelas prestadoras, em articulação com o FNDE e com o acompanhamento da Anatel. Ele estabelece prazos e metas para assegurar que a implementação da infraestrutura ocorra de forma ordenada e dentro do período previsto.

Além disso, como mecanismo de controle e transparência, as prestadoras devem encaminhar mensalmente à Anatel informações detalhadas e atualizadas sobre o progresso das conexões realizadas. Esses dados são enviados por meio de um sistema eletrônico próprio e servem de base para as atividades de acompanhamento, controle e fiscalização conduzidas pela Agência.

1.2.1 Números do PBLE

Até junho de 2024, o PBLE atendia a um total de 62.551 escolas públicas urbanas, distribuídas em 5.513 municípios das 27 unidades da Federação, conforme dados disponibilizados no portal institucional da Anatel [15]. Esses números representam uma cobertura de aproximadamente 99% dos municípios brasileiros, evidenciando a capilaridade e a abrangência nacional do programa.

A expressiva cobertura do PBLE reforça sua relevância como política pública estratégica para a inclusão digital e o fortalecimento da infraestrutura educacional. O programa tem contribuído significativamente para a democratização do acesso à internet nas escolas públicas, desempenhando um papel essencial na promoção da equidade no sistema educacional.

A série histórica apresentada na Figura 1.1 permite acompanhar a evolução do PBLE ao longo dos últimos seis anos [15].

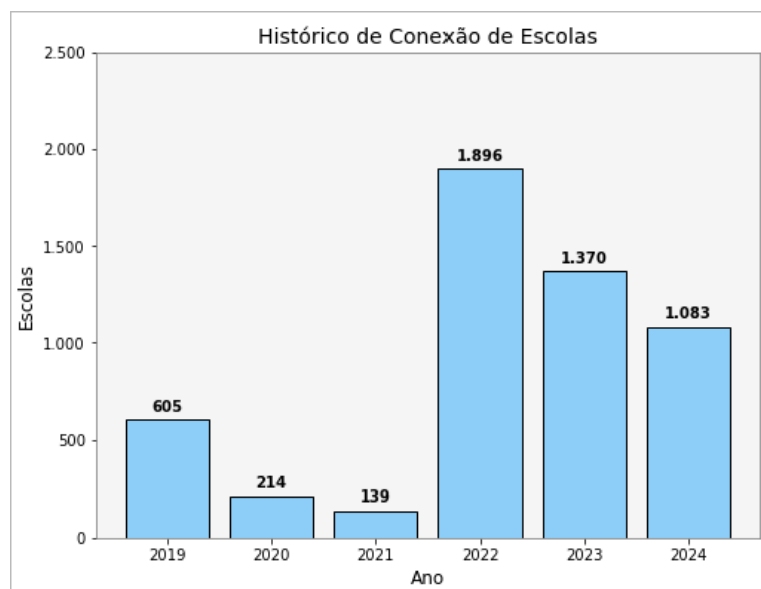


Figura 1.1: Escolas Conectadas pelo PBLE por Ano
[15]

A análise da série histórica apresentada na Figura 1.1 indica uma desaceleração no ritmo de instalação de conexões em escolas públicas durante os anos de 2020 e 2021. Esse período coincide com os efeitos mais severos da pandemia de covid-19, que ocasionou o fechamento temporário das instituições de ensino. Como consequência, o acesso físico das prestadoras às unidades escolares foi substancialmente limitado, impactando diretamente a execução das instalações previstas pelo PBLE.

Adicionalmente, observa-se que os dados referentes ao número de escolas conectadas em 2024 correspondem ao acumulado até o mês de junho, conforme os registros mais recentes disponibilizados no portal da Anatel [15].

1.3 Fiscalização

A atividade de fiscalização configura-se como uma das funções institucionais mais relevantes atribuídas à Anatel, desempenhando um papel estratégico e essencial na promoção da qualidade, continuidade e operação dos serviços de telecomunicações ofertados em território nacional.

Sua principal finalidade é assegurar a prestação adequada desses serviços, em conformidade com os dispositivos legais e normativos vigentes, promovendo a proteção dos direitos dos consumidores, o interesse público, o desenvolvimento do setor e a conformidade com o arcabouço regulatório das telecomunicações.

As atividades fiscalizatórias conduzidas pela Anatel envolvem a supervisão contínua das prestadoras de serviços de telecomunicações, de modo a verificar sua aderência às normas técnicas, operacionais e de qualidade.

Essa atuação abrange aspectos como a entrega de serviços com níveis satisfatórios de qualidade, a prática de tarifas justas e a inexistência de interferências indevidas no espectro radioelétrico, fator essencial para assegurar a eficiência, a segurança e a integridade das telecomunicações em todo o país.

O procedimento de fiscalização conduzido da Anatel fundamenta-se em um conjunto normativo composto por regulamentos específicos aplicáveis a cada modalidade de serviço – incluindo telefonia celular, telefonia fixa, internet e TV por assinatura –, além de atos normativos internos que complementam e detalham os requisitos técnicos e operacionais a serem observados pelas prestadoras.

Esses instrumentos normativos conferem maior robustez, transparência e coerência às ações fiscalizatórias conduzidas pela Anatel, ao estabelecerem parâmetros técnicos objetivos e procedimentos operacionais padronizados que orientam a atuação dos agentes de fiscalização da Agência.

Nesse cenário, o processo fiscalizatório é iniciado, em geral, a partir de uma solicitação da área demandante – responsável pelo acompanhamento e controle da prestação dos serviços –, da identificação de não conformidade ou, ainda, em decorrência do recebimento de denúncias provenientes de usuários, entidades públicas ou outros canais institucionais.

A partir da formalização da demanda, cabe à área técnica de fiscalização estruturar sua atuação, por meio da definição de diretrizes operacionais específicas que orientarão todo o processo. Essas diretrizes estão consolidadas no Procedimento de Fiscalização, o qual integra o Plano Anual de Fiscalização (PAF) – instrumento estratégico da Anatel que organiza e prioriza as ações fiscalizatórias no âmbito da Agência.

Concluídas todas as etapas do Procedimento de Fiscalização, é elaborado o Relatório de Fiscalização, documento técnico que consolida os resultados obtidos. O relatório reúne as evidências coletadas, seja em campo ou por meio de sistemas, identificando possíveis inconformidades, descumprimentos ou violações às normas regulatórias por parte das prestadoras.

Posteriormente, o Relatório de Fiscalização é encaminhado à área demandante, que passa a assumir a responsabilidade de prosseguir com a apuração das irregularidades identificadas. Essa continuidade se concretiza por meio da instauração do Processo de Acompanhamento e Controle (PAC).

A partir da abertura do PAC, são realizadas análises para aplicação de medidas corretivas adequadas, assegurando que as inconformidades sejam devidamente corrigidas. Além disso, é realizada a avaliação quanto à aplicação de sanções administrativas, quando for o caso, observando os critérios técnicos e jurídicos estabelecidos nas normas vigentes.

Esse processo assegura a efetividade das ações fiscalizatórias e reforça o compromisso da Anatel com a conformidade e a qualidade dos serviços de telecomunicações ofertados pelas prestadoras.

A Figura 1.2, a seguir, apresenta as principais etapas que compõem o ciclo da atividade de fiscalização conduzida pela Anatel.

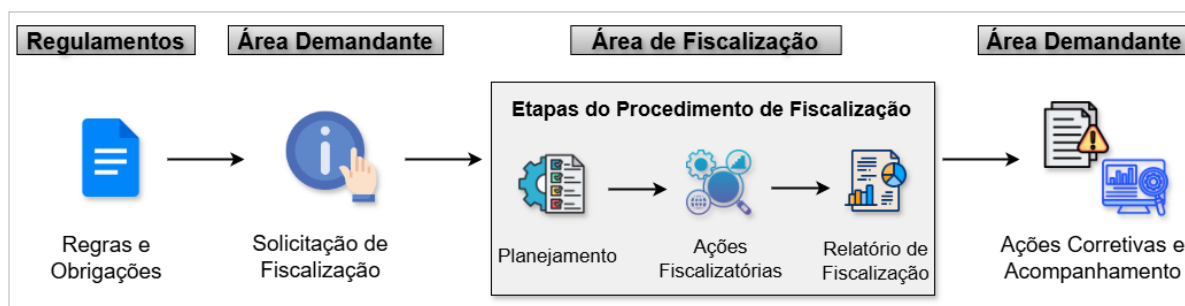


Figura 1.2: Etapas da Fiscalização da Anatel

1.3.1 Fiscalização do PBLE

A fiscalização do PBLE é regulamentada pela Portaria Anatel nº 2740, publicada em 28 de novembro de 2023, que estabelece o procedimento de fiscalização específico para a verificação do cumprimento das obrigações assumidas pelas prestadoras de serviços de telecomunicações responsáveis pela oferta gratuita de conexão à internet em escolas públicas [16].

Esse procedimento normativo estabelece um conjunto de regras, critérios técnicos e diretrizes operacionais que estruturam e orientam a atuação fiscalizatória da Anatel, promovendo maior padronização dos processos, precisão e transparência às ações executadas no âmbito do PBLE.

O principal objetivo é assegurar que as prestadoras atendam integralmente às exigências previstas no escopo do PBLE, promovendo a efetividade da conectividade nas unidades escolares elegíveis e contribuindo para a ampliação do acesso à informação e ao uso pedagógico das tecnologias digitais no ambiente educacional.

Os objetivos específicos da fiscalização do PBLE consistem em assegurar a conformidade das obrigações assumidas pelas prestadoras, conforme os parâmetros técnicos e normativos estabelecidos no Termo Aditivo [13] que rege o programa. Entre esses objetivos, destacam-se:

- Verificar se a velocidade de conexão instalada nas escolas públicas atendidas pelo PBLE corresponde à melhor oferta comercialmente disponibilizada pelas prestadoras aos seus clientes, conforme estabelecido no Termo Aditivo [13] que estabelece as regras do PBLE ;
- Verificar se as escolas atendidas pelo PBLE estão conectadas por meio de endereço IP (*Internet Protocol*) fixo e público/global, em conformidade com o que estabelece o Termo Aditivo [13];
- Verificar se a tecnologia utilizada para conectar as escolas atendidas pelo PBLE está devidamente regulamentada e se os equipamentos utilizados estão homologados pela Anatel, garantindo a segurança e a compatibilidade com os padrões técnicos vigentes; e
- Verificar eventuais casos em que escolas públicas atendidas pelo PBLE estejam operando com velocidade de conexão inferior ao mínimo obrigatório definido no Termo Aditivo [13].

A fiscalização do PBLE é conduzida de forma estruturada e sistematizada, sendo organizada em três etapas principais: Planejamento, Ações Fiscalizatórias e Relatório

de Fiscalização. Cada uma dessas etapas desempenha um papel essencial na condução das atividades fiscalizatórias e na garantia do cumprimento das obrigações contratuais assumidas pelas prestadoras.

A Figura 1.3 ilustra de forma esquemática essas etapas e seus respectivos desdobramentos operacionais.

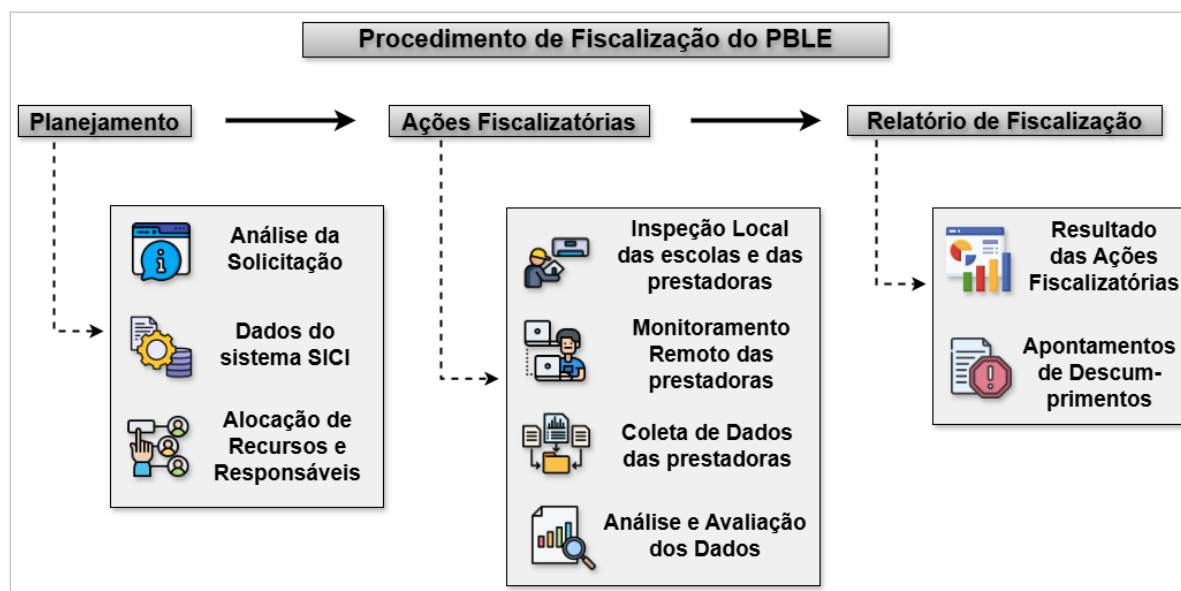


Figura 1.3: Etapas do Procedimento de Fiscalização do PBLE

A primeira etapa do processo, denominada Planejamento, é conduzida pela área técnica responsável pela fiscalização e consiste na análise da solicitação encaminhada pela área demandante. Nessa fase, são avaliados os dados disponíveis no sistema interno de coleta de informações da Anatel, com o objetivo de subsidiar a definição do escopo da fiscalização.

Além disso, são realizadas a alocação de recursos e a designação das equipes técnicas envolvidas na execução das atividades, para garantir um processo eficiente, coordenado e em conformidade com as diretrizes institucionais e normativas da Agência.

A etapa subsequente, correspondente às Ações Fiscalizatórias, compreende a execução de um conjunto estruturado de procedimentos para verificar o cumprimento integral das obrigações atribuídas às prestadoras no âmbito do PBLE.

O foco principal dessa fase é assegurar a conformidade dos serviços prestados com os parâmetros técnicos e regulatórios definidos, bem como com os marcos normativos vigentes, garantindo que as escolas públicas atendidas recebam conexões de qualidade, conforme estabelecido pelo programa.

Entre as principais atividades executadas nessa etapa, destacam-se:

- **Inspecção Local das Escolas e Prestadoras:** Consiste na realização de visitas técnicas *in loco* pelos agentes de fiscalização da Anatel, como o objetivo de avaliar a infraestrutura das prestadoras, examinar a operação dos serviços, verificar o funcionamento da conexão nas escolas e confirmar o cumprimento dos requisitos técnicos estabelecidos pelo PBLE;
- **Monitoramento Remoto das Prestadoras:** Refere-se à utilização de sistemas e ferramentas de supervisão à distância, que possibilitam o acompanhamento contínuo e em tempo real das operações das prestadoras, sem a necessidade de deslocamento físico. Essa modalidade permite a análise do cumprimento das exigências regulatórias, otimizando recursos e aumentando a eficiência do processo fiscalizatório;
- **Coleta de Dados das Prestadoras:** Envolve a obtenção sistemática de informações a partir das bases de dados técnicos e operacionais disponibilizados pelas próprias prestadoras. A coleta dessas informações é essencial para a construção de um diagnóstico preciso acerca da conformidade e cumprimento das obrigações assumidas no âmbito do PBLE; e
- **Análise e Avaliação dos Dados:** Etapa destinada ao exame crítico e detalhado das informações coletadas, com vistas à identificação de eventuais inconformidades, verificação do atendimento às normas regulatórias e subsidiação de medidas corretivas. Quando pertinente, essa fase também embasa a aplicação de sanções administrativas às prestadoras que não estiverem em conformidade com os dispositivos estabelecidos no PBLE.

Por fim, a terceira e última etapa do processo corresponde à elaboração do Relatório de Fiscalização, documento técnico que consolida os achados e resultados obtidos ao longo de toda a atividade fiscalizatória da Anatel.

Nessa fase, os dados coletados durante as etapas anteriores são sistematizados e analisados de forma criteriosa, com a identificação de eventuais inconformidades, inconsistências ou descumprimentos das obrigações regulatórias por parte das prestadoras que possuem obrigação de atendimento às escolas.

O Relatório de Fiscalização constitui um instrumento fundamental para subsidiar a tomada de decisões estratégicas, orientando a adoção de medidas corretivas, a aplicação de sanções, quando cabíveis, e o aprimoramento contínuo das atividades de acompanhamento e controle do PBLE.

1.4 Dados Abertos

Com o avanço da Tecnologia da Informação e Comunicação, diversos setores, incluindo o público, têm experimentado progressos significativos. Essa evolução tem atraído a atenção de pesquisadores e profissionais, estimulando a realização de estudos para compreender suas particularidades. Considera-se que o cenário dinâmico da gestão pública, com ênfase na “inteligência” do setor, visa resolver problemas sociais por meio da interação e colaboração com os próprios cidadãos [17].

Nas últimas décadas, houve o surgimento de práticas de dados governamentais abertos alinhadas à filosofia do movimento software livre e apoiadas pela Tecnologia da Informação. Neste novo paradigma, os dados governamentais abertos visam reforçar a transparência, responsabilização, participação e colaboração dos cidadãos no acesso, criação e inovação de serviços públicos [18].

Na década de 1970, surgiu o termo “Governo Aberto”, que significa livre acesso às informações públicas, conhecimento das atividades planejadas pelo governo e participação dos cidadãos na construção e execução de serviços públicos [18].

Ainda, Governo Aberto é a forma de relacionamento entre a administração pública e os cidadãos, caracterizando-se por estabelecer canais de comunicação e contato direto entre eles. Esse movimento foi visto como uma contribuição para tornar os processos governamentais mais eficientes, eficazes e transparentes [18].

Em 2009, os Estados Unidos publicaram a Diretiva Governo Aberto para aumentar a capacidade do público de encontrar e acessar facilmente um conjunto de dados gerados pelo Governo Federal. Essa diretiva baseou-se em três princípios fundamentais para o governo dos Estados Unidos: transparência, participação e colaboração [18].

A *Open Knowledge Foundation* (OKF) estabelece que os dados são considerados abertos quando qualquer pessoa pode acessar, usar, modificar e compartilhar livremente para qualquer finalidade, salvo requisitos que preservem a sua origem e sua divulgação. Esse fato geralmente é satisfeito por meio da publicação dos dados em formato aberto e sob uma licença aberta [19].

Recentemente, um significativo desdobramento percebido é a criação de portais de dados abertos com o objetivo de fornecer acesso aos dados públicos, que até então eram restritos à esfera governamental, por meio de sites governamentais em formatos de fácil utilização (CSV, ODS, XLS etc.) [17].

Esses sites criam valor e inovação para diversas partes interessadas, como cidadãos, academia, setores sem fins lucrativos e empresas. Ainda, os dados que estão disponíveis gratuitamente são conhecidos como Dados Públicos Abertos [17].

Dados abertos e reutilização de informações do setor público visam garantir o acesso proativo aos dados governamentais e abarca o conceito da responsabilização do governo e

da transparência para a participação dos cidadãos. Essa diretiva abrange a disponibilização e acesso aos dados e informações não sensíveis do setor público e que são consideradas cruciais para o mercado interno e para a economia de um país [20].

O uso de tecnologias, especialmente a divulgação de dados abertos, é considerada um instrumento fundamental para ajudar os governos a atingirem seus objetivos. Nesse contexto, políticas de dados abertos vêm sendo monitoradas de perto por uma variedade de atores, incluindo organizações não-governamentais, sociedade, empresas, órgãos de governo e acadêmicos [21].

Ainda, os dados governamentais abertos são vistos como ferramentas para criar novas oportunidades, eliminar ou pelo menos reduzir as desigualdades de informação e melhorar os serviços públicos de um país [22].

Adicionalmente, a garantia da qualidade dos dados abertos é um aspecto extremamente importante. Na prática, os governos não devem apenas melhorar a qualidade dos dados abertos, mas também melhorar a governança de dados para gerar valor aos dados abertos em seus países [23].

Embora os dados governamentais abertos tenham grande potencial para produzir valores políticos, sociais e econômicos, sua utilização está longe das expectativas. Estudos indicam que a utilização de dados governamentais abertos ainda está em um nível baixo, uma vez que diversos fatores influenciam na sua utilização. O principal fator é a qualidade dos dados [23].

Assim, para promover maior e melhor utilização de dados governamentais abertos e gerar valor político, social e econômico aos países, é necessário investir na melhoria da qualidade desses dados [23].

No Brasil, a política de Dados Abertos do Poder Executivo Federal foi instituída pelo Decreto nº 8.777, de 11 de maio de 2016 [24] e o governo brasileiro criou o Portal Brasileiro de Dados Abertos – ferramenta disponibilizada para que todos possam encontrar e utilizar os dados e as informações públicas [25].

O Portal de Dados Abertos do Governo Federal define dados abertos como uma metodologia para a publicação de dados do governo em formatos reutilizáveis, visando o aumento da transparência e maior participação política por parte do cidadão, além de gerar diversas aplicações desenvolvidas colaborativamente pela sociedade [25].

Ainda, o Portal de Dados Abertos preza pela simplicidade e organização para que o cidadão encontre facilmente os dados e informações de que precisa. Seu objetivo é disponibilizar dados relativos sobre diversas áreas da administração pública, além de promover a interlocução entre sociedade e governo para pensar a melhor utilização dos dados, promovendo impactos positivos sob os pontos de vista social e econômico do país [25].

Nesse cenário, diversos órgãos da Administração Pública, incluindo a Anatel, disponibilizam seus dados nesse Portal em formatos de arquivos, relatórios, gráficos e *dashboards*. A disponibilização dessas informações desempenha um papel fundamental na promoção da transparência na gestão pública, permitindo que cidadãos, pesquisadores, instituições acadêmicas e demais interessados acompanhem de forma mais eficiente os resultados das ações governamentais.

A Agência, em alinhamento com essa iniciativa, publica diversas informações de interesse público, incluindo os dados do PBLE, por meio do Portal de Dados Abertos. Essa ação visa não apenas garantir maior transparência na gestão do programa, mas também facilitar o acesso às informações sobre a conectividade das escolas públicas do país.

1.5 Justificativa

O PBLE representa uma política pública estratégica voltada à inclusão digital no Brasil. Segundo dados do Censo Escolar de 2024, mais de 41 milhões de estudantes estão matriculados em escolas públicas INEP [26], o que evidencia a relevância da conectividade escolar como ferramenta para ampliar as oportunidades educacionais e reduzir desigualdades sociais. Alinhado a esse propósito, o PBLE tem como meta conectar 100% das escolas públicas urbanas brasileiras até o ano de 2025.

Para viabilizar o acompanhamento e o controle das obrigações previstas pelo programa, a Anatel recebe periodicamente, por meio do Sistema de Coleta de Informações (SICI), dados enviados pelas prestadoras de serviços de telecomunicações. Esses dados contêm informações detalhadas sobre a instalação de conexões de banda larga nas instituições de ensino elegíveis, permitindo o monitoramento do avanço do PBLE em nível nacional.

Além dessas atividades, os dados armazenados no SICI são essenciais para subsidiar a atuação fiscalizatória da Anatel. A análise desses registros possibilita o planejamento e a execução de ações com base em evidências, contribuindo para a identificação de inconformidades e a adoção de medidas corretivas que assegurem o cumprimento das obrigações assumidas pelas prestadoras.

Nesse contexto, a consistência e a precisão dos dados utilizados tornam-se fatores críticos. Informações confiáveis não apenas minimizam incertezas, como também fundamentam decisões mais assertivas e eficientes por parte dos gestores públicos [27].

Contudo, apesar da importância estratégica dos dados no contexto do PBLE, a Anatel atualmente não dispõe de mecanismos sistemáticos para avaliação da qualidade dos dados recebidos das prestadoras e armazenados em seu ambiente computacional. Também há uma lacuna quanto à análise das regras de negócio atualmente implementadas no sistema

SICI, o que pode comprometer a integridade das informações utilizadas nas ações do programa.

Diante desse cenário, torna-se evidente a importância de estudos aprofundados que abordem aspectos relacionados à governança e à qualidade dos dados no âmbito do PBLE. Essas pesquisas têm o potencial de aprimorar os processos de tomada de decisão, bem como de fortalecer a eficácia das ações de fiscalização, controle e acompanhamento conduzidas pela Agência.

A relevância e a complexidade do estudo sobre o tema da qualidade de dados se evidenciam sob duas dimensões. A primeira refere-se à lacuna existente entre as metodologias desenvolvidas na literatura acadêmica e sua aplicação prática, especialmente no contexto de governança e gestão de dados [27]. A segunda diz respeito à escassez de estudos voltados à qualidade de dados em políticas públicas de conectividade educacional no Brasil. Essa carência foi identificada a partir de pesquisa realizada nas principais bases científicas internacionais, como Web of Science, Scopus e IEEE Xplore.

1.6 Objetivos

O objetivo geral desta dissertação é realizar a avaliação da qualidade dos dados de campos selecionados do PBLE, utilizando as dimensões da qualidade mais apropriadas ao escopo do estudo. Com base nessa análise, será proposta a implementação de melhorias nos procedimentos de coleta de dados adotados pelo programa.

A partir dessa abordagem, este trabalho visa apoiar a Anatel na identificação de inconsistências e falhas nos dados fornecidos pelas prestadoras de serviços de telecomunicações. A identificação dessas inconsistências será crucial para a adoção de medidas corretivas e evolutivas no sistema SICI, resultando em um impacto significativo na melhoria contínua da qualidade dos dados gerenciados pelo programa.

Ao aprimorar a qualidade dos dados do PBLE, espera-se contribuir diretamente para o fortalecimento das atividades de fiscalização e de acompanhamento e controle das obrigações das prestadoras, as quais são de responsabilidade da Anatel até a conclusão do programa. Com informações mais consistentes e confiáveis, a Agência será capaz de atuar de forma mais eficiente na supervisão do cumprimento das metas e exigências do PBLE.

A principal contribuição deste estudo será a elaboração de soluções para aprimorar a coleta de dados do PBLE, incluindo Regras de Qualidade de Dados, Modelos de Auditoria de Dados e Relatórios de Auditoria. Essas soluções impactarão diretamente as ações de fiscalização e o controle das obrigações do PBLE pela Anatel, além de aprimorar a qualidade das informações publicadas no Portal de Dados Abertos do governo federal.

1.6.1 Objetivos Específicos

Com o intuito de alcançar o objetivo geral estabelecido para esta dissertação, foram definidos os seguintes objetivos específicos:

1. Identificar e selecionar os campos de dados mais relevantes para as atividades de fiscalização e de acompanhamento e controle relacionadas ao PBLE;
2. Aplicar as metodologias e técnicas mais adequadas para avaliar os dados do PBLE, utilizando as dimensões de qualidade de dados mais apropriadas ao contexto do PBLE;
3. Elaborar soluções que incluam ações corretivas e evolutivas voltadas para o aprimoramento do sistema de coleta de dados do PBLE.

1.7 Estrutura do Trabalho

Este trabalho está organizado em capítulos, conforme descrito a seguir:

- Capítulo 2: Apresenta a metodologia de pesquisa adotada, estruturada em etapas principais, proporcionando uma abordagem sistemática e coerente para o desenvolvimento do estudo;
- Capítulo 3: Descreve o mapeamento da literatura, apresentando os principais estudos e pesquisas relevantes nos campos de governança de dados e qualidade de dados, fornecendo uma base teórica sólida para a pesquisa;
- Capítulo 4: Examina os trabalhos relacionados identificados na literatura, abordando o PBLE e as soluções implementadas em governança e qualidade de dados que se mostram pertinentes para o contexto deste estudo;
- Capítulo 5: Apresenta os conceitos teóricos fundamentais para o desenvolvimento da pesquisa, incluindo definições de dados, governança de dados, qualidade de dados, dados governamentais abertos, além das dimensões e metodologias associadas à qualidade de dados;
- Capítulo 6: Detalha o desenvolvimento da pesquisa, estruturado em seções que abrangem desde a metodologia aplicada até os resultados finais e as soluções propostas para melhorar a qualidade dos dados do PBLE;

- Capítulo 7: Conclui o trabalho com as principais observações sobre a avaliação da qualidade dos dados do PBLE, destacando as contribuições do estudo e os impactos para a Anatel, além de sugerir aprimoramentos nas práticas de fiscalização e de acompanhamento e controle.

Capítulo 2

Metodologia da Pesquisa

A metodologia adotada neste estudo foi estruturada em sete etapas principais, que guiaram o desenvolvimento da pesquisa de forma sistemática e coerente. As etapas são as seguintes: Planejamento, Mapeamento de Literatura, Análise de Metodologias, Exploração dos Dados, Construção da Proposta de Solução e Diagnóstico dos Dados e Solução Proposta, conforme descrito a seguir:

1. **Planejamento:** Nesta etapa inicial, foram definidos os objetivos da pesquisa, as questões a serem respondidas e o escopo do estudo. Foi elaborado um protocolo de pesquisa para orientar as etapas subsequentes, incluindo a definição de critérios para a seleção de trabalhos acadêmicos relevantes;
2. **Mapeamento de Literatura:** Realizou-se um mapeamento da literatura em três bases científicas – Web of Science, Scopus e IEEE Xplore – com o objetivo de identificar e analisar os estudos mais relevantes para as questões de pesquisa. Os detalhes estão no Capítulo 3;
3. **Análise de Metodologias:** Foram avaliadas as metodologias e técnicas existentes na literatura para a avaliação da qualidade de dados. Essa análise buscou identificar as abordagens mais adequadas para os objetivos deste estudo e os resultados fundamentaram as etapas seguintes da pesquisa;
4. **Exploração dos Dados:** Nessa etapa, foi realizada uma análise detalhada da base de dados do PBLE, incluindo o mapeamento da estrutura das tabelas e o tipo de campos armazenados. A exploração permitiu identificar os campos utilizados pelo programa e compreender as regras de negócio associadas a esses dados;
5. **Construção da Proposta de Solução:** Com base no estudo “*Quality Assessment for Open Government Data in China*”, foi desenvolvida uma metodologia adaptada

para a avaliação dos dados do PBLE. A metodologia proposta é composta por cinco etapas principais: Identificação dos Dados, Seleção dos Dados, Identificação de Problemas de Qualidade, Seleção das Dimensões de Qualidade e Avaliação da Qualidade dos Dados;

6. **Diagnóstico dos Dados:** Utilizando a metodologia proposta, foi realizado um diagnóstico dos dados do PBLE, com foco nos dados extraídos entre fevereiro de 2013 e junho de 2024. A análise revelou diversos campos com baixa qualidade. Os detalhes e as conclusões desse diagnóstico podem ser encontrados no Capítulo 6; e
7. **Solução Proposta:** Por fim, com base nos resultados da avaliação da qualidade dos dados, foram desenvolvidas soluções para aprimorar a gestão e o controle da qualidade no PBLE. Essas soluções incluem a definição de Regras de Qualidade de Dados, a implementação de Modelos de Auditoria de Dados e a criação de Relatórios de Auditoria de Dados.

Essa estrutura metodológica foi elaborada para garantir uma abordagem sistemática e prática, assegurando que todas as etapas da pesquisa estivessem alinhadas com os objetivos do estudo. O enfoque no uso de metodologias consolidadas e a análise aprofundada dos dados asseguram resultados com potencial de contribuição relevante para a área de Governança de Dados e Qualidade de Dados, especialmente no contexto do PBLE.

Capítulo 3

Mapeamento da Literatura

Este capítulo apresenta o mapeamento da literatura para identificar e avaliar estudos relevantes sobre governança de dados e qualidade de dados. O objetivo principal desta metodologia é selecionar e revisar publicações diretamente relacionadas aos temas centrais deste trabalho.

O processo de mapeamento foi estruturado em várias etapas, com buscas realizadas entre 8 de fevereiro e 24 de março de 2023, e atualizações posteriores em 20 de julho a 2 de agosto de 2023, 4 de março a 5 de abril de 2024, e 28 de fevereiro a 14 de março de 2025, garantindo a inclusão de fontes atualizadas e relevantes.

3.1 Protocolo de Pesquisa

O mapeamento da literatura nas áreas de Governança de Dados e Qualidade de Dados foi conduzido em três fases distintas: **Planejamento**, **Execução** e **Documentação**, conforme descrito em [28].

A fase inicial, denominada Planejamento, teve como objetivo a definição das questões de pesquisa e a identificação da necessidade do mapeamento. Nessa etapa, foram estabelecidas as diretrizes para a busca e seleção dos trabalhos relevantes, com a definição das questões de pesquisa e a elaboração dos critérios de inclusão e exclusão dos estudos.

Na fase de Execução, foi realizada a identificação dos artigos e estudos pertinentes, seguida pela seleção daqueles considerados mais relevantes. Esse processo incluiu a análise crítica da qualidade dos trabalhos e a extração dos dados necessários para consolidar as informações relevantes.

A última fase, Documentação, consistiu na organização e apresentação dos resultados da pesquisa, incluindo um resumo detalhado de todas as etapas e dos procedimentos seguidos.

De acordo com [29], o mapeamento da literatura deve ser realizado com base em uma estratégia de busca bem definida, incluindo um período determinado para a coleta dos trabalhos. Além disso, é crucial que os critérios de inclusão e exclusão sejam claros e alinhados aos objetivos da pesquisa, assegurando a consistência e a qualidade dos estudos selecionados.

3.1.1 Questões de Pesquisa

A definição das questões de pesquisa é uma etapa fundamental no planejamento de um mapeamento da literatura, pois orienta todas as fases subsequentes da investigação. Essas questões estabelecem o foco da pesquisa e direcionam os critérios de seleção dos estudos, garantindo que os objetivos do mapeamento sejam alcançados de forma eficaz. Assim, todas as atividades subsequentes decorrem diretamente dessas questões iniciais [29].

O objetivo principal deste mapeamento da literatura foi responder às seguintes questões de pesquisa (QP):

- **QP1:** Quais processos podem compor a Governança de Dados?
- **QP2:** Como promover a melhoria da Qualidade de Dados?
- **QP3:** Quais dimensões da Qualidade de Dados são mais frequentemente exploradas na literatura?

3.1.2 Identificação e Seleção dos Estudos

A fase de execução iniciou-se com a realização de uma pesquisa em três bases científicas amplamente reconhecidas e relevantes para os temas de Governança de Dados e Qualidade de Dados. As bases selecionadas foram:

1. Web of Science (Clarivate Analytics)¹;
2. Scopus (Elsevier)²;
3. IEEE Xplore³.

¹<https://clarivate.com>

²<https://www.scopus.com>

³<http://ieeexplore.ieee.org>

Para conduzir a pesquisa nas bases científicas, foram elaboradas duas *strings* de busca, estruturadas com o objetivo de atender a dois objetivos principais: identificar estudos relacionados à Governança e Qualidade de Dados em geral (para responder às questões QP1 e QP2) e mapear as dimensões da Qualidade de Dados mais exploradas na literatura acadêmica (para responder à questão QP3).

As *strings* de busca definidas foram as seguintes:

1. *(DATA AND GOVERNANCE) OR (DATA AND QUALITY)*
2. *(DATA AND QUALITY) AND (PARAMETERS OR ASSESSMENT OR INDICATORS OR METRICS OR STUDY OR ANALYSIS)*

Essas *strings* foram aplicadas nas bases bibliográficas selecionadas com o intuito de recuperar artigos que respondessem às questões de pesquisa formuladas neste estudo.

Durante a etapa de busca e seleção, foram aplicados os seguintes critérios de inclusão:

- Apenas artigos publicados a partir de 2019 foram considerados, embora trabalhos que apresentassem conceitos fundamentais ou referências clássicas sobre Governança de Dados e Qualidade de Dados também fossem incluídos;
- Trabalhos que propusessem modelos, técnicas ou abordagens para a implementação de Governança de Dados e Qualidade de Dados;
- Artigos que detalhassem explicitamente o uso de Qualidade de Dados, incluindo as áreas do conhecimento em que essa prática está sendo aplicada;
- Artigos com, no mínimo, três citações foram priorizados para garantir relevância na comunidade acadêmica; e
- O artigo deveria estar disponível para *download* na base de dados consultada.

Além disso, foram estabelecidos os seguintes critérios de exclusão:

- Artigos que se desviassem do escopo ou fossem considerados irrelevantes para os objetivos desta pesquisa foram descartados;
- Trabalhos publicados antes de 2019, exceto aqueles que apresentassem definições clássicas ou fossem amplamente reconhecidos como referências fundamentais no campo, foram excluídos;
- Artigos duplicados, ou que estivessem disponíveis em mais de uma das bases de dados consultadas, foram removidos da seleção; e
- Artigos com menos de três citações foram desconsiderados para garantir a qualidade e a relevância do material selecionado.

3.1.3 Análise de Qualidade da Pesquisa

Na segunda etapa da fase de Execução, foi realizada a análise da qualidade da busca e seleção dos artigos, com base nos critérios previamente definidos. A partir da aplicação das duas *strings* de pesquisa, foram extraídos 1.864 documentos das bases de dados consultadas. Após a aplicação rigorosa dos critérios de inclusão e exclusão, 1.809 documentos foram descartados, resultando em 55 trabalhos considerados relevantes para o desenvolvimento desta pesquisa.

Desses 55 artigos selecionados, 63% foram encontrados na base de dados Web of Science, 29% na Scopus e 4% na IEEE Xplore. Esta distribuição reflete a relevância e a especialização das fontes, garantindo a abrangência e a profundidade da revisão da literatura.

3.1.4 Extração de Dados da Pesquisa

Na última etapa da fase de Execução, denominada “extração de dados”, foram analisados os resumos e as introduções dos 55 artigos selecionados, com o objetivo de identificar as informações mais pertinentes para responder às questões de pesquisa. A prioridade foi assegurar que os artigos selecionados estivessem diretamente alinhados com os objetivos do estudo, conforme as três questões de pesquisa previamente formuladas.

Após a análise detalhada, 19 artigos foram considerados fora do escopo da pesquisa e, portanto, descartados. Ao final do mapeamento da literatura, obteve-se um total de 36 artigos relevantes, que forneceram as bases necessárias para o desenvolvimento e aprofundamento desta dissertação.

3.1.5 Síntese e Análise dos Dados da Pesquisa

A síntese e a análise dos dados extraídos dos artigos selecionados teve como objetivo identificar as soluções propostas e quantificar os estudos relacionados às áreas de Governança de Dados e Qualidade de Dados. Esta etapa foi essencial para fornecer as respostas às questões de pesquisa estabelecidas.

Em relação à questão de pesquisa QP1, foram identificados dezoito artigos relevantes, os quais contribuíram para a compreensão dos processos que compõem a Governança de Dados. Entre eles, seis estudos apresentaram soluções aplicadas em cenários reais, evidenciando a viabilidade prática dos conceitos abordados.

Além disso, dois artigos abordaram a implementação de Governança de Dados no setor governamental, destacando os desafios específicos enfrentados pelo setor público. Outras quatro pesquisas exploraram iniciativas em setores específicos, como saúde e finanças, evidenciando particularidades e obstáculos em contextos distintos.

Um resumo detalhado dos resultados obtidos pode ser visualizado na Tabela 3.1.

Tabela 3.1: Trabalhos Identificados sobre Governança de Dados - QP1

Contexto	Artigos	Total
Conceito de Governança de Dados	[30], [31], [32], [33], [34], [35]	6
Proposta de Implementação	[36], [37], [38], [39], [40], [41]	6
Implement. na Área Governamental	[42], [43]	2
Implement. na Área de Saúde ou Financeira	[44], [45], [46], [47]	4

Os resultados relacionados à questão de pesquisa QP1 demonstram que o tema “Governança de Dados” vem sendo amplamente investigado em diversos setores, evidenciando sua crescente relevância e o interesse contínuo da comunidade acadêmica. A variedade de abordagens identificadas reforça a percepção de que a Governança de Dados é um componente fundamental para a melhoria da gestão e da integridade dos dados em diversos contextos organizacionais.

Em relação à QP2, observou-se que o contexto de Metodologias para a Implementação da Qualidade de Dados foi abordado em treze artigos, representando 43% do total de trabalhos considerados relevantes. Os contextos de Conceito de Qualidade de Dados e Técnicas de Qualidade de Dados foram explorados em nove artigos cada (32%). Já os tópicos Dimensões de Qualidade de Dados apareceu em oito artigos (29%), enquanto Tomada de Decisão foi tratado em seis artigos (21%).

Cabe ressaltar que alguns artigos abordaram mais de um contexto, o que justifica a soma percentual superior a 100%.

A Tabela 3.2 apresenta os resultados detalhados relacionados à questão de pesquisa QP2, proporcionando uma visão mais aprofundada sobre a promoção da melhoria da Qualidade de Dados.

Tabela 3.2: Trabalhos sobre Melhoria da Qualidade de Dados - QP2

Contexto	Artigos	Total
Conceito de Qual. de Dados	[5], [48], [49], [50], [51], [52], [53], [54], [55]	9
Dimensões de Qual. de Dados	[5], [48], [56], [52], [53], [54], [57], [58]	8
Metodologias de Qual. de Dados	[5], [27], [48], [51], [59], [60], [61], [62], [63], [64], [65], [66], [67]	13
Técnicas de Qual. de Dados	[5], [48], [49], [68], [69], [70], [71], [72], [73]	9
Tomada de Decisão	[4], [5], [27], [48], [49], [59]	6

Adicionalmente, a Tabela 3.3 descreve as áreas mais abordadas nos artigos selecionados, oferecendo uma visão clara das tendências nas pesquisas sobre a melhoria da Qualidade de Dados.

Tabela 3.3: Áreas Abordadas sobre Melhoria da Qualidade de Dados - QP2

Área	Artigos	Total
Financeira	[68], [52], [61], [62], [65], [46]	6
Governamental	[53], [60], [73], [42], [45], [43]	6
Indústria	[27], [53], [69], [70], [66]	5
Saúde	[72], [47]	2
Tecnologia	[49], [51], [59], [54], [63], [71], [64], [57], [58]	9

Por fim, respondendo à QP3, os resultados dos artigos analisados revelaram que a dimensão Completude foi abordada em 9 estudos (45%), seguida por Consistência, presente em 8 artigos (40%), e Acurácia, tratada em 3 trabalhos (15%).

A Tabela 3.4 apresenta esses dados em detalhes e evidencia que alguns artigos exploraram mais de uma dimensão, o que reforça a complexidade e a inter-relação entre as diferentes dimensões da Qualidade de Dados.

Tabela 3.4: Dimensões da Qualidade de Dados mais Exploradas na Literatura - QP3

Dimensão	Artigos	Total
Acurácia	[5], [48], [57]	3
Completude	[5], [48], [56], [52], [53], [54], [57], [58], [67]	9
Consistência	[5], [48], [56], [52], [54], [57], [58], [67]	8

As informações obtidas na síntese e análise dos dados são essenciais para identificar lacunas na pesquisa, especialmente em áreas que ainda demandam maior investigação no campo da Governança de Dados e Qualidade de Dados, com ênfase no setor de Serviços de Telecomunicações.

Esses resultados podem, assim, servir como base consistente para o desenvolvimento de pesquisas futuras, fornecendo um direcionamento estratégico para novos estudos e promovendo novas abordagens que complementem os aspectos ainda pouco explorados sobre o tema.

Capítulo 4

Trabalhos Relacionados

Este capítulo apresenta a análise dos trabalhos relacionados, com base nos dados obtidos no capítulo anterior. Além disso, resume os estudos mais relevantes da literatura sobre o PBLE e a implementação de soluções voltadas à governança e à qualidade de dados.

4.1 Análise dos Trabalhos Relacionados

O processo de mapeamento da literatura resultou na identificação de 55 artigos relevantes, selecionados a partir de um universo de 1.864 documentos coletados em diferentes bases de pesquisa reconhecidas, como Web of Science, IEEE e Scopus.

Os resultados revelam uma crescente expansão nas áreas de Governança de Dados e Qualidade de Dados, indicando não apenas a existência de campos ainda pouco explorados, mas também desafios significativos a serem enfrentados. O tema Qualidade de Dados se destaca como uma questão transversal em diversas áreas do conhecimento, com um número considerável de estudos voltados à implementação de soluções específicas para problemas relacionados à qualidade dos dados em diferentes contextos.

Além disso, observou-se a publicação de diversos estudos que aplicam técnicas ou metodologias já consolidadas para a avaliação da qualidade de dados. Esse fenômeno ressalta a importância da temática, especialmente em áreas estratégicas como o setor de telecomunicações, que desempenha um papel fundamental no desenvolvimento nacional.

Outro ponto relevante do mapeamento foi a constatação da heterogeneidade das abordagens adotadas nas pesquisas sobre Qualidade de Dados, bem como a recorrente utilização das dimensões de completude, consistência e acurácia, amplamente discutidas nos artigos analisados.

No contexto específico do PBLE, observou-se uma lacuna expressiva na literatura científica. A revisão dos trabalhos publicados revelou a ausência de estudos sobre a governança e sobre a qualidade dos dados associados ao programa. Essa lacuna reforça a originalidade

e a relevância da presente dissertação, ao propor uma abordagem estruturada e inédita para a avaliação da qualidade dos dados do PBLE, contribuindo para o aprimoramento da gestão de políticas públicas voltadas à conectividade escolar.

4.2 Resumo dos Trabalhos Relacionados

Entre os estudos revisados, quatro se destacaram por sua contribuição relevante à compreensão dos processos de Governança de Dados e à avaliação da Qualidade de Dados. Dentre eles, dois trabalhos propuseram metodologias específicas para a avaliação da qualidade dos dados; um estudo apresentou um *framework* voltado à governança de dados no setor público; e outro propôs uma técnica voltada à avaliação da qualidade de dados.

No estudo [61], os autores propõem uma metodologia para a avaliação da Qualidade de Dados por meio de uma verificação experimental, utilizando diversas dimensões, como consistência, confiabilidade, rastreabilidade, compreensão e recuperação. A pesquisa começa com a avaliação do estado atual dos dados, com base nas características estruturais relevantes para os usuários. O trabalho destaca diversos problemas de Qualidade de Dados, como a ausência de gerenciamento das regras de geração de dados, o que pode resultar em inconsistências, dados incompletos e um alto número de valores nulos. Esses problemas comprometem a capacidade de análise da qualidade dos dados. Os autores enfatizam que a avaliação da qualidade dos dados deve ser baseada em evidências objetivas e científicas, obtidas diretamente do conjunto de dados a serem avaliados. Além disso, o estudo se distingue de outras pesquisas da área ao incluir abordagens como estimativa dedutiva, verificação interna, comparação com dados originais, inspeção por amostragem independente, verificação da sobreposição de polígonos e inspeção do valor efetivo.

O estudo [67] propôs uma metodologia para a avaliação da Qualidade de Dados em três cidades chinesas, nas quais o compartilhamento de dados governamentais abertos é uma parte central da estratégia nacional de *Big Data*. A metodologia apresentada envolve a análise de um conjunto de dados utilizando dimensões de Qualidade de Dados e métricas específicas para avaliar sua qualidade. O processo é estruturado em três etapas principais: identificação dos problemas de qualidade, classificação das dimensões e métricas de qualidade, e análise dos dados. Durante a pesquisa, foram identificados dezesseis tipos de problemas relacionados à qualidade dos dados, além de sete dimensões e métricas de qualidade em diferentes níveis. Os resultados da avaliação indicaram pontuações baixas nas dimensões de integridade, precisão e consistência. Segundo os autores, essas deficiências podem comprometer a disponibilidade dos dados e induzir os usuários a tomar decisões equivocadas. Com base nos resultados da avaliação, os autores sugerem que o governo

chinês adote medidas corretivas para superar as falhas observadas na qualidade dos dados abertos nas três cidades analisadas.

O artigo [42] propôs um novo *framework* para a Governança de Dados no setor público, fundamentado em uma plataforma intermediária de dados, com o objetivo de especificar os requisitos e funcionalidades de uma Governança de Dados adaptada ao contexto governamental. Utilizando o paradigma *Design Science Research* (DSR) e um ciclo de atividades estruturado em três etapas, o estudo formulou proposições com base em uma revisão sistemática da literatura sobre Governança de Dados no setor público e *frameworks* correlatos. Além disso, a eficácia da implementação foi avaliada por meio de estudos de caso práticos e entrevistas com especialistas da área. Os resultados indicaram que o *framework* desenvolvido foi eficaz em atender às necessidades dos serviços digitais governamentais, proporcionando melhorias na adaptação das operações organizacionais aos dados, valorizando os ativos de dados, aprimorando os processos de auditoria e supervisão, e facilitando a comunicação entre as partes envolvidas. O modelo demonstrou flexibilidade para se ajustar às novas características do governo digital, abrangendo desde a coleta de dados até a entrega de serviços governamentais. Este estudo contribuiu de maneira significativa para o campo da Governança de Dados, apresentando novas perspectivas sobre as “plataformas intermediárias de dados” e oferecendo diretrizes valiosas para o desenvolvimento de *frameworks* específicos para a Governança de Dados no setor público.

A pesquisa de [73] propõe uma técnica para a avaliação da qualidade de dados em cenários de *Big Data*, onde os algoritmos precisam ser adaptados para atender aos novos requisitos impostos pela variedade, volume e velocidade de processamento dos dados. O objetivo dessa abordagem é possibilitar que os gestores tomem decisões e realizem ações baseadas em dados corretos, confiáveis e completos. Realizado com dados reais coletados a partir de um estudo de caso de uma cidade inteligente, o trabalho enfatiza a importância de lidar com fontes de dados heterogêneas e a necessidade de uma abordagem adaptativa. Essa abordagem visa selecionar os métodos de avaliação de qualidade mais adequados, com base no tipo de dado e no contexto em que estão inseridos. O estudo sugere que a qualidade seja avaliada a partir de uma amostra do conjunto de dados, considerando a possível perda de precisão e introduzindo um fator de confiança para medir a confiabilidade do processo de avaliação. Adicionalmente, é proposta uma técnica para a construção de um Módulo Adaptador de Qualidade de Dados, que escolhe a melhor configuração de avaliação com base nos requisitos do usuário, como redução de tempo, redução de custos e maximização da confiança. Os experimentos realizados demonstraram que a avaliação da qualidade dos dados depende tanto dos recursos da fonte de dados quanto da dimensão escolhida para a avaliação. Isso implica que diferentes dimensões podem gerar avaliações distintas, mesmo dentro de um mesmo cenário de *Big Data*.

Adicionalmente, foram identificados quatro estudos relevantes diretamente relacionados ao PBLE. Essas pesquisas abordam a importância da continuidade e do aprimoramento das políticas públicas voltadas à conectividade nas instituições de ensino da rede pública, evidenciando o papel estratégico do PBLE na promoção da inclusão digital. Além disso, os trabalhos traçam um panorama abrangente sobre os impactos do programa nas atividades educacionais das escolas atendidas, destacando avanços, limitações e desafios persistentes.

O estudo [74] examina as principais iniciativas voltadas à expansão da internet banda larga no Brasil e em Portugal ao longo de um período de quase duas décadas, evidenciando os avanços obtidos e os desafios persistentes enfrentados por ambos os países. No contexto brasileiro, merecem especial atenção o Programa Banda Larga nas Escolas (PBLE) e o Programa Nacional de Banda Larga (PNBL), concebidos com o objetivo de ampliar o acesso à internet, com ênfase em áreas rurais, regiões urbanas periféricas e instituições de ensino público. A análise comparativa revela que, embora ambos os países tenham alcançado avanços expressivos na massificação do acesso à internet, o Brasil ainda enfrenta desigualdades estruturais – tanto regionais quanto socioeconômicas – que limitam a universalização do serviço. O estudo conclui ressaltando a importância da continuidade e do aprimoramento das políticas públicas de conectividade, especialmente diante dos desafios impostos pelas novas tecnologias, como o 5G, e pela incorporação de soluções baseadas em inteligência artificial e outras tecnologias emergentes.

Na pesquisa apresentada por [75], os autores analisaram a infraestrutura de redes nas escolas municipais de Santa Maria, no estado do Rio Grande do Sul, e evidenciaram diversos entraves técnicos que comprometem a efetividade do uso pedagógico das tecnologias digitais. Entre os principais desafios identificados, destaca-se a baixa velocidade de conexão. Ressalta-se que, embora o PBLE tenha desempenhado um papel relevante ao viabilizar o acesso inicial à internet em escolas públicas – incluindo aquelas situadas em Santa Maria –, a velocidade atualmente ofertada por escola permanece insuficiente diante das crescentes demandas do ensino mediado por tecnologias digitais. Essa limitação compromete a integração de recursos educacionais online e dificulta a adoção de práticas pedagógicas inovadoras. O documento aponta a limitação na expansão da conectividade nas escolas brasileiras, atribuindo esse cenário às políticas públicas de investimento restrito e à capacidade técnica reduzida das conexões fornecidas. Apesar do mérito do PBLE como política de inclusão digital, observa-se que o programa não tem acompanhado o ritmo dos avanços tecnológicos necessários para atender às exigências contemporâneas da educação digital. Como alternativa, propõe-se a criação de uma rede metropolitana privada como solução viável para garantir maior qualidade, velocidade e autonomia no uso de recursos digitais pelas instituições de ensino.

Já no trabalho [76], os autores avaliam o impacto do PBLE na qualidade educacional no Brasil, por meio da aplicação de métodos econométricos robustos. A análise foi segmentada por regiões do país e contempla diferentes áreas do conhecimento avaliadas no Exame Nacional do Ensino Médio (Enem), incluindo Matemática, Ciências Naturais, Ciências Humanas e Ciências da Vida. Os resultados obtidos indicam que o acesso à internet de banda larga nas escolas públicas exerce efeitos positivos ou neutros sobre o desempenho estudantil. Em especial, observam-se melhorias estatisticamente significativas em regiões com menores índices de desenvolvimento, evidenciando o potencial do programa para mitigar desigualdades educacionais. O trabalho também discute as metodologias utilizadas, destacando a relevância de se avaliar, de forma rigorosa, os efeitos de políticas públicas voltadas à ampliação do acesso à tecnologia no ambiente escolar. A conclusão reforça que o PBLE tem desempenhado um papel relevante na promoção do acesso equitativo à tecnologia educacional, sobretudo em regiões mais vulneráveis. Os resultados obtidos evidenciam o potencial transformador da conectividade escolar, não apenas como infraestrutura básica, mas como um vetor de melhoria nos indicadores de desempenho acadêmico e de fortalecimento da educação pública brasileira.

A pesquisa [77] realizou uma análise abrangente sobre o PBLE, explorando as etapas de formulação, implementação, monitoramento e avaliação da política pública. O estudo reconhece avanços significativos no provimento de acesso à internet nas escolas públicas, ao mesmo tempo em que evidencia limitações estruturais e operacionais. Entre os aspectos positivos apontados estão: o alto percentual de escolas conectadas, a importância do programa para a inclusão digital, a eficácia na implementação de seus objetivos gerais, a contribuição para a modernização do ensino público e o fortalecimento da educação conectada. Tais elementos reforçam o papel estratégico do PBLE como instrumento de expansão do acesso às TICs nas escolas públicas, promovendo a inclusão digital de uma parcela expressiva da população estudantil brasileira. Contudo, a pesquisa também identifica desafios relevantes, como a insuficiência da velocidade de conexão ofertada, a ausência de dados confiáveis acerca da execução do programa e a fragilidade dos mecanismos de governança e fiscalização. Além disso, são destacadas dificuldades na realização de ações de fiscalização presencial, na gestão das informações e no monitoramento da qualidade dos serviços prestados. O estudo conclui que, embora o PBLE represente uma iniciativa valiosa no avanço da conectividade escolar, persistem desafios para sua consolidação como política pública universal e de qualidade. Nesse sentido, os autores defendem o fortalecimento da governança do programa, a adoção de estratégias eficazes de fiscalização e a implementação de mecanismos robustos de monitoramento.

Capítulo 5

Fundamentação Teórica

Este capítulo apresenta os principais conceitos teóricos utilizados no desenvolvimento deste trabalho. A primeira seção aborda a definição de dados, enquanto as duas seções seguintes tratam, respectivamente, da governança e da qualidade de dados. A quarta seção aborda os dados governamentais. Por fim, a quinta seção discute as dimensões da qualidade e a última seção trata das metodologias, técnicas e modelos.

5.1 Dados

Os dados desempenham um papel crucial no nosso entendimento do mundo, pois nos permitem analisar e interpretar padrões, tendências e relações complexas em uma escala antes inimaginável. Além disso, os dados promovem a transparência, a inovação, a eficiência e o progresso em diversas áreas da sociedade.

Indubitavelmente, os dados são um ativo extremamente valioso para as organizações. Gestores dos mais diversos tipos de organizações no Brasil e no exterior usam-na para difundir a ideia de que os dados são tão valiosos quanto o petróleo, ou seja, as pessoas gestoras que souberem obter valor (*insights*) a partir dos dados para aprimorar suas decisões serão detentoras de um grande diferencial competitivo [78].

Com os dados adequados, cientistas, pesquisadores e tomadores de decisão são capazes de responder a questões essenciais, gerar *insights* valiosos e descobrir soluções para inúmeros desafios que ainda desafiam nossa compreensão.

Ainda, os dados desempenham um papel essencial na Era da Tecnologia da Informação e Comunicações, pois são fundamentais em todas as relações entre empresas, governos e cidadãos. Como os dados são amplamente utilizados em atividades da sociedade, a qualidade e os efeitos relacionados ao seu uso são cada vez mais importantes e críticos para as organizações [48].

Ao contrário dos tempos em que os dados eram simplesmente combustível para programas, hoje eles se transformaram em elementos essenciais do capital intelectual das empresas, atuando como geradores de informação e conhecimento [1].

Nesse contexto, diversas empresas passaram a tratar seus dados como ativos organizacionais, reconhecendo o valor intrínseco que eles possuem e o potencial que podem oferecer. Os dados são, sem dúvida, considerados fatores críticos em uma era em que os conceitos de *Big Data*, Internet das Coisas (IoT) e aprendizado de máquinas se tornam cada vez mais comuns em uma sociedade altamente digital e inovadora [1].

Além disso, ao transitarem do universo transacional para o informacional, os dados exigem uma avaliação cuidadosa de sua estrutura e conteúdo. Estatísticas indicam que dados de baixa qualidade podem resultar em perdas que variam de 15% e 25% da receita de uma empresa. Isso significa que, cada vez mais, os dados precisarão ser geridos de forma eficiente, com uma governança sólida e eficaz [1].

Dessa forma, os conceitos de Governança de Dados e Qualidade de Dados assumem novas dimensões, não apenas visando à organização desses acervos, mas também atendendo às exigências e desafios típicos desses novos tempos [1].

5.1.1 Conceito de Dados

Para fundamentar teoricamente este estudo, é imprescindível iniciar pela definição de “dados”. Em ciência de dados, dados são representações estruturadas ou semiestruturadas de observações, medições ou fatos, expressas por meio de valores numéricos, textos ou categorias.

Dados podem ser definidos como a “representação reinterpretável de informações de maneira formalizada adequada para comunicação, interpretação ou processamento” [48]. Essa definição destaca a flexibilidade dos dados, que podem ser analisados e manipulados conforme as necessidades do contexto.

Adicionalmente, os dados podem ser considerados a estrutura elementar da cadeia informacional, servindo como base para a transformação em informações nos estágios subsequentes [1].

Uma outra definição adicional destaca que os dados consistem em um conjunto de caracteres que, isoladamente, não possuem significado. Somente quando inseridos em um contexto específico e associados a um propósito, os dados adquirem significado, sendo então transformados em informações [35].

O termo “dados” é frequentemente empregado para se referir a dados estruturados em um Sistema Gerenciador de Banco de Dados (SGBD) e também aparece em expressões amplamente reconhecidas na literatura, como: fonte de dados, modelo de dados, integração de dados, dados abertos, dados da *Web* e *Big Data*. Por sua vez, o termo “Qualidade

de Dados” é utilizado para tratar de dimensões, modelos e técnicas diretamente relacionadas a dados estruturados [48].

Os dados representam objetos do mundo real em formatos que podem ser construídos, armazenados e recuperados por meio de softwares, além de serem transmitidos por redes de computadores. Esse processo de representação do mundo real por meio de dados é amplamente aplicável a diversos fenômenos, como medições, eventos, características de pessoas, ambiente, sons e até cheiros. Essa versatilidade torna os dados ferramentas indispensáveis para diversas finalidades [5].

Adicionalmente, outros tipos de informações também desempenham papéis importantes nos processos da vida real e nas operações empresariais, como informações baseadas em documentos físicos (papel) e informações transmitidas verbalmente (voz) [5].

5.1.2 Classificação dos Dados

Os dados, conforme representado na Figura 5.1, podem ser classificados em categorias específicas, o que facilita sua compreensão e contribui diretamente para o desenvolvimento de estratégias mais eficazes de gerenciamento, controle e aprimoramento contínuo das informações [1].

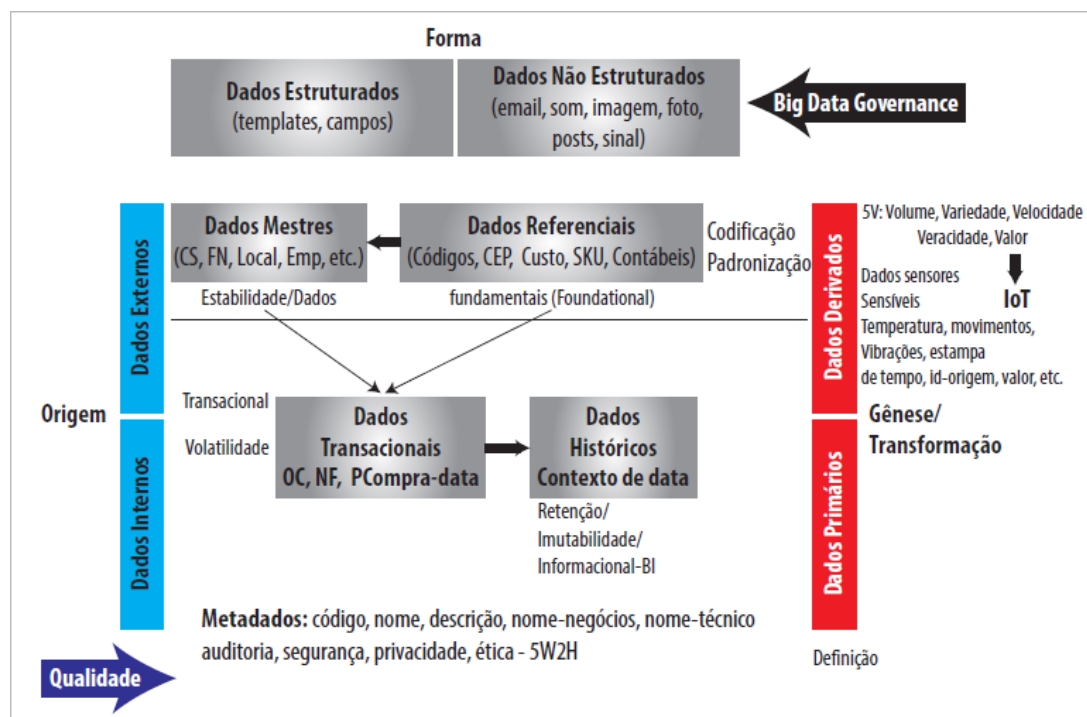


Figura 5.1: Classificação dos Dados

[1]

A partir dos conceitos representados na Figura 5.1, os dados podem ser organizados da seguinte maneira:

- **Mestres:** Representam os dados fundamentais ou pilares de qualquer organização. Caracterizam-se por sua estabilidade e por não serem altamente dependentes do tempo, funcionando como a base para as grandes transações corporativas;
- **Referenciais:** Estão associados aos Dados Mestres, mas possuem maior volatilidade, sendo suscetíveis a variações. Esses dados são frequentemente obtidos de fontes externas;
- **Transacionais:** São dinâmicos e resultam das atividades de negócio da empresa. Fortemente conectados à dimensão temporal, esses dados estabelecem vínculos diretos com os Dados Mestres;
- **Históricos:** Originam-se dos Dados Mestres, Referenciais e Transacionais, sendo armazenados em uma linha do tempo para referência e análise futuras.

Além disso, os dados podem ser classificados com base em sua estrutura, conforme descrito abaixo [5]:

- **Estruturado:** Cada elemento possui uma estrutura fixa definida, o que facilita sua organização e processamento. Tabelas relacionais em SGBD são o exemplo mais comum desse tipo;
- **Semiestruturado:** Possuem alguma estrutura, mas com flexibilidade. São frequentemente chamados de “dados sem esquema” ou “autodescritivos”. A linguagem XML (*eXtensible Markup Language*) é um exemplo amplamente utilizado para representar esse tipo de dado.

Os dados semiestruturados apresentam características específicas, como a possibilidade de conter campos desconhecidos, como ocorre em arquivos XML sem um esquema associado; a representação do mesmo tipo de dado de maneiras variadas, como datas armazenadas em formatos diferentes; e a ausência de valores explícitos em muitos campos conhecidos;

- **Não estruturado:** São expressos em linguagem natural e não apresentam uma estrutura ou domínio claramente definidos. Exemplos incluem textos livres, imagens, vídeos e outros formatos que não seguem padrões rígidos.

Essa classificação fornece um panorama abrangente das diferentes naturezas e usos dos dados, permitindo uma gestão mais eficiente e alinhada às necessidades organizacionais.

5.2 Governança de Dados

A Governança de Dados tem emergido como um componente central nas estratégias organizacionais, especialmente diante do avanço da transformação digital. Antes tratada como uma função operacional de importância secundária, passou a ocupar uma posição estratégica nas agendas de organizações públicas, privadas e instituições governamentais [35].

Inicialmente, a governança de dados era frequentemente confundida com conceitos correlatos, como qualidade, segurança e privacidade de dados. Entretanto, esforços recentes na literatura e na prática organizacional vêm buscando estabelecer distinções conceituais mais precisas. Ainda assim, observa-se uma lacuna semântica persistente entre governança de dados e conceitos adjacentes, como governança da informação, governança analítica e governança de tecnologia da informação [79].

De forma geral, a Governança de Dados envolve padrões, políticas, responsabilidades e relações que orientam a gestão e o uso adequado dos dados ao longo de seu ciclo de vida. É fundamental garantir sua conformidade com as legislações vigentes – como o GDPR e a Lei de Inteligência Artificial da União Europeia. Ao estabelecer papéis e responsabilidades claras, a governança contribui para o alinhamento entre os dados e os objetivos estratégicos da organização, além de viabilizar a conformidade regulatória [79].

Esse alinhamento é particularmente importante em ambientes corporativos e institucionais onde a tomada de decisão baseada em dados é valorizada. Nesse contexto, a governança de dados atua como fundamento organizacional que garante não apenas a confiabilidade da informação, mas também sua disponibilidade e relevância. Diversos *frameworks* de Governança de TI vêm sendo adaptados e aplicados à realidade do gerenciamento de dados, incorporando práticas voltadas à qualidade [80].

A crescente relevância da governança de dados nas organizações decorre, em grande parte, do crescimento exponencial na geração de dados. Estima-se que o volume global de dados, que era de aproximadamente 4,4 zettabytes, deverá atingir 44 zettabytes nos próximos anos. Diante desse cenário, organizações enfrentam o desafio de assegurar a confiabilidade dos dados utilizados, evitando decisões estratégicas baseadas em informações imprecisas, incompletas ou desatualizadas [35].

Nesse sentido, a Governança de Dados vem se consolidando como o eixo central da gestão de dados em grandes organizações. Sua implementação reflete a urgência de coordenar aspectos operacionais, atender às exigências legais e cumprir os requisitos de conformidade específicos de cada país. A Governança de Dados, portanto, não apenas assegura a qualidade e precisão das informações, mas também viabiliza a aderência a normas regulatórias, fortalecendo a tomada de decisões baseada em dados confiáveis [81].

Particularmente no setor público, há uma necessidade crescente de ampliar a compreensão e a adoção da Governança de Dados. A implementação de práticas de governança em diferentes níveis da administração pública pode contribuir significativamente para a modernização do Estado, a melhoria na entrega de serviços e o fortalecimento da transparência e da *accountability* [80]

Apesar do reconhecimento crescente de sua importância, a literatura acadêmica sobre Governança de Dados ainda apresenta limitações. Estudos recentes apontam que há pouca investigação sistemática sobre as atividades práticas associadas à governança de dados, especialmente no setor público [82]. Embora o tema venha amadurecendo como disciplina, persistem ambiguidades em sua definição, escopo e metodologias de implementação, o que dificulta sua consolidação teórica e prática [79].

Dessa forma, torna-se fundamental pesquisar o estado atual da Governança de Dados nas organizações, especialmente no contexto da gestão de dados públicos. Essa iniciativa não apenas contribuirá para o avanço científico da área, mas também fornecerá subsídios para o desenvolvimento de políticas públicas e modelos institucionais mais eficazes e sustentáveis [80].

5.2.1 Definição de Governança de Dados

A definição de Governança de Dados é ampla e plural porque é um conceito em evolução e que envolve a interseção de diversas disciplinas com foco central em qualidade de dados no sentido mais amplo deste conceito. Ainda, passa pela maturidade da empresa na gerência desses recursos, melhoria na valoração e produção dos dados e monitoramento de seu uso, além de aspectos de segurança, privacidade, ética e aderência às regras de *compliance* [1].

De forma prática, a Governança de Dados pode ser definida como um conjunto de práticas, dispostas em um *framework*, com o objetivo de organizar o uso e o controle adequado dos dados como um ativo organizacional, ou seja, é uma forma de organização em relação aos aspectos de dados, visando disponibilidade, integridade, consistência, usabilidade, segurança e controle [1].

O DAMA *Data Management Body of Knowledge* (DMBoK) define a Governança de Dados como o exercício de autoridade e controle – envolvendo planejamento, monitoramento e engajamento – sobre o gerenciamento de ativos de dados. Nesse contexto, ela orienta todas as demais funções relacionadas à gestão de dados e opera em um nível estratégico, situando-se na esfera executiva [83].

Como processo organizacional, a Governança de Dados define e estabelece políticas, padrões, processos, procedimentos e diretrizes corporativas, legislando sobre os dados e atribuindo papéis específicos para tratar esses elementos com responsabilidade e *accountability* [1].

A Governança de Dados também pode ser compreendida como o exercício de autoridade e controle sobre a gestão de dados, com o propósito de implementar uma agenda estratégica voltada à maximização do valor dos ativos informacionais dentro das organizações [35].

Mais do que estabelecer uma estrutura organizacional, a Governança de Dados garante que os dados e as informações sejam gerenciados de forma adequada, descrevendo processos, definindo responsabilidades e assegurando a consistência e confiabilidade das informações [35].

Além disso, a Governança de Dados desempenha um papel central na integração com diversas áreas e funções organizacionais, sendo responsável por estabelecer princípios de organização e controle de dados e informações [2]. Seu escopo abrange aspectos físicos, operacionais e estruturais, incluindo políticas de metadados, linhagem, qualidade e limpeza de dados, níveis de serviço, segurança e gestão de impactos de mudanças [34].

Assim, a Governança de Dados não apenas organiza os dados, mas também os prepara para atender às necessidades estratégicas e operacionais das empresas, promovendo maior eficiência e conformidade.

Atualmente, a Governança de Dados é vista como um subdomínio da Governança de Tecnologia da Informação (GTI), que, por sua vez, é um subdomínio da Governança Corporativa. Isso se deve à independência com que a Governança de Dados é vista em relação à GTI, na medida em que os dados caminham para ter o status de um ativo da organização, e não uma propriedade da tecnologia [1].

Nesse novo cenário, a área de Tecnologia da Informação ganha a importante missão de ser a “custodiadora” dos dados [1]. A Figura 5.2 ilustra as relações entre Governança Corporativa, Governança de Dados (GD) e Governança de TI (GTI).

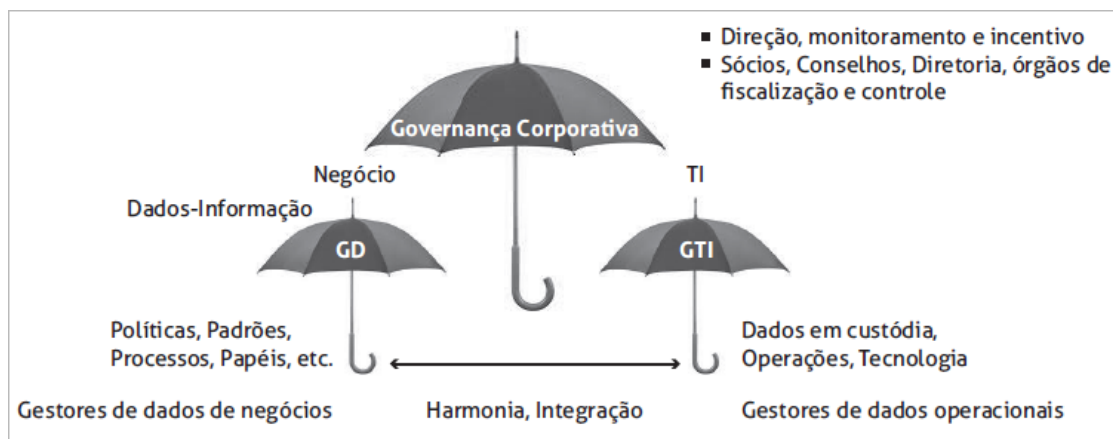


Figura 5.2: Relações Entre Tipos de Governança

[1]

5.2.2 Princípios da Governança de Dados

Os princípios de governança de dados são elementos mais filosóficos que envolvem regras definidas por consenso e aprovadas, atualizadas e respeitadas pela organização, e que legislam sobre os dados [1].

Estes princípios regem as práticas relacionadas aos dados e orientam a definição de estratégias e políticas que visam maximizar o valor dos dados. Os documentos que fundamentam a Governança de Dados são conhecidos como os “Ps” (Princípios, Padrões, Processos/Procedimentos e Papéis), conforme [34].

Nos últimos anos, diversos princípios de Governança de Dados têm sido discutidos e apresentados em estudos acadêmicos. No contexto das organizações públicas, alguns dos princípios mais relevantes incluem [30]:

- **Organização:** A Governança de Dados deve ser bem estruturada e organizada em dimensões específicas.

A primeira dimensão refere-se às metas, voltadas para o desempenho organizacional por meio da valorização dos dados. Envolve a definição de métricas de qualidade, especificação de metadados e o planejamento da arquitetura e do ciclo de vida dos dados.

A segunda dimensão diz respeito à estrutura organizacional, englobando a definição clara de responsabilidades, papéis e processos.

Por fim, a terceira dimensão trata da comunicação e dos padrões, essenciais para a formulação de políticas consistentes, agilidade na resolução de problemas e suporte à tomada de decisão.

- **Alinhamento:** A Governança de Dados deve assegurar que os dados estejam alinhados às necessidades do negócio, gerando valor para a organização. Quando bem utilizados, os dados se tornam um ativo estratégico que impulsiona os resultados da empresa.

Portanto, a Governança de Dados precisa garantir que os dados sejam “úteis”, o que implica em melhorar sua qualidade, atendendo assim aos objetivos estratégicos da organização. A qualidade dos dados é amplamente reconhecida como uma métrica essencial para avaliar o sucesso da Governança de Dados.

- **Conformidade:** Este princípio envolve a definição de uma autoridade responsável por criar procedimentos e políticas de conformidade para garantir que a organização atenda às exigências legais e regulatórias. A definição dessas políticas é geralmente feita de forma colaborativa entre as equipes de TI e de negócios, visando a construção de uma estrutura aplicável a toda a organização. Além disso, é crucial

estabelecer mecanismos que assegurem o cumprimento das responsabilidades legais e regulamentares.

- **Entendimento:** Para uma Governança de Dados eficaz, é essencial que os dados sejam plenamente compreendidos, incluindo seu valor e importância para a organização. A compreensão adequada dos dados é fundamental para qualquer desenvolvimento de aplicativos, armazenamento de dados ou arquitetura de serviços em uma empresa.

A melhor forma de evitar problemas causados por interpretações incorretas dos dados é a criação de um Modelo de Dados Organizacionais, que deve ser uma atividade central na Governança de Dados.

5.2.3 Cenário Atual da Governança de Dados

Atualmente, as organizações enfrentam um crescente volume de dados provenientes de múltiplas fontes, o que frequentemente resulta em inconsistências nas informações. Para evitar que decisões estratégicas sejam baseadas em dados imprecisos, é imprescindível identificar e corrigir essas discrepâncias [35].

Esse desafio também se aplica aos órgãos públicos, que lidam diariamente com grandes volumes de dados. Como esses dados são utilizados para decisões que impactam diretamente a sociedade, a atenção ao seu armazenamento e análise torna-se essencial [35].

Os objetivos formais da Governança de Dados em organizações públicas incluem aprimorar a tomada de decisões, garantir conformidade regulatória, aumentar a eficiência e eficácia organizacional e apoiar a integração de negócios [35].

Além disso, regulamentações como o *General Data Protection Regulation* (GDPR) intensificam a pressão para que as empresas adotem práticas rigorosas no recebimento, uso, tratamento e armazenamento de dados. Superar desafios relacionados à imprecisão e incompletude dos dados tornou-se crucial nesse contexto [35].

A implementação da Governança de Dados é indispensável, conforme evidenciado pelas seguintes justificativas [34]:

- Melhorar a qualidade das decisões e fomentar a inovação, garantindo acesso rápido a informações corretas e confiáveis;
- Propagar o conhecimento sobre os dados do negócio, permitindo que toda a organização esteja alinhada às políticas internas;
- Mitigar prejuízos causados pela baixa qualidade dos dados;

- Reduzir custos operacionais por meio de processos claros para criação, uso e disponibilização de dados governados;
- Atender a regulamentações específicas, como a lei Sarbanes-Oxley (SOx) e o *International Convergence of Capital Measurement and Capital Standards* (Acordo de Basileia), assegurando governança não apenas dos dados, mas de toda a empresa;
- Minimizar custos no desenvolvimento de aplicações, diminuindo retrabalho, redundâncias e disseminando boas práticas de gestão de dados;
- Prevenir fraudes, fortalecendo processos formais de controle e eliminando lacunas em práticas de Gestão de Dados.

5.2.4 Implementação da Governança de Dados

A implementação de uma estrutura abrangente de Governança de Dados, que seja capaz de abranger todos os níveis e áreas da organização, continua representando um desafio significativo, mesmo para instituições tecnologicamente avançadas e com maturidade digital elevada [81].

Isso se deve, em parte, à escassez de definições detalhadas e diretrizes práticas sobre como projetar e aplicar a Governança de Dados. Aspectos como funções, práticas, processos, métodos e ferramentas necessários para estabelecer um modelo de governança maduro ainda carecem de maior clareza e padronização [81].

Apesar desses desafios, diversos *frameworks* foram desenvolvidos por organizações e pela indústria para apresentar conceitos e formas de implementação da Governança de Dados. Esses *frameworks* têm desempenhado um papel essencial na formação inicial dos modelos de governança [1].

Entre eles, destaca-se o DAMA DMBOK V2, que servirá como referência para este estudo uma vez que é amplamente reconhecido como uma das melhores práticas globais em governança de dados.

Essa versão atualizada do *framework* apresenta novos conceitos e mantém as 11 disciplinas tradicionais, além de preencher lacunas de versões anteriores. Elementos como Governança de Dados, Classificação de Dados, Avaliação de Maturidade e Ética foram incorporados, trazendo maior modernidade ao seu Corpo de Conhecimento [1].

A Figura 5.3 ilustra o diagrama da versão evoluída do DAMA DMBOK V2, que introduz uma abordagem mais abrangente e atualizada, reforçando sua relevância para a prática contemporânea de Governança de Dados [1].

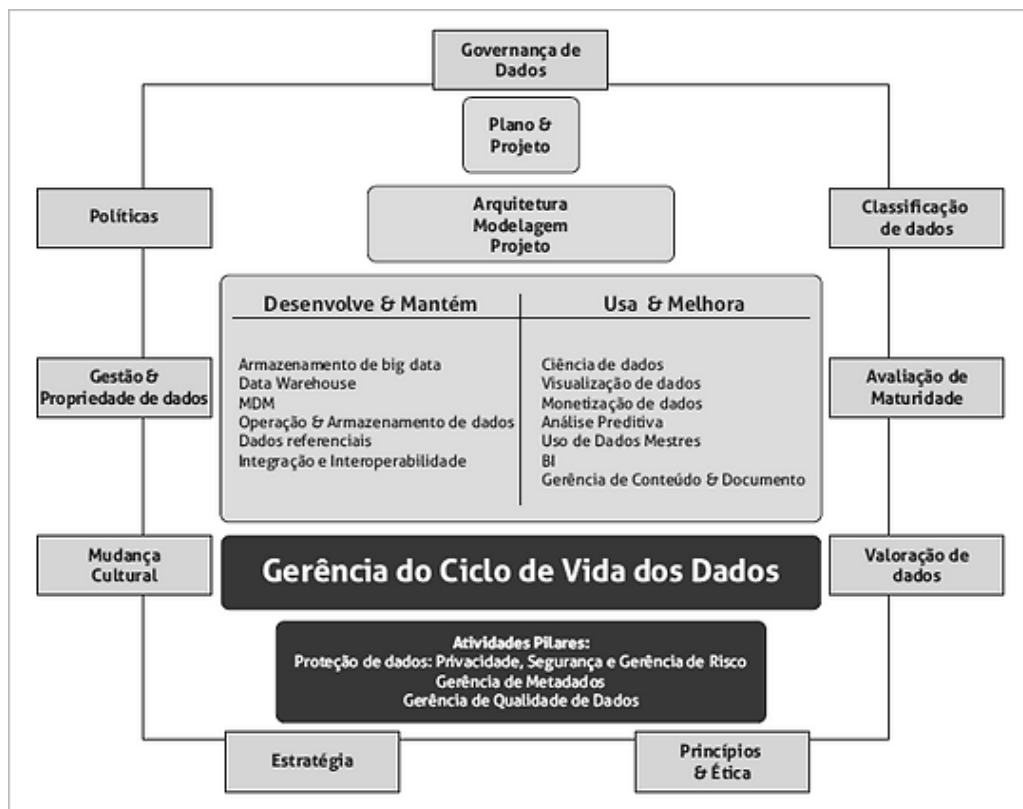


Figura 5.3: Diagrama do DAMA DMBOK V2 - Evoluído
[1]

Para implementar um programa de Governança de Dados em uma organização, é fundamental conhecer o seu estado atual em relação aos dados, sua gestão, aspectos de governança e gerência eventualmente já existentes e uma visão de propensão às mudanças culturais relacionadas a esses recursos [1].

É importante ressaltar que a Governança de Dados é um programa que se desenvolve ao longo do tempo com certa perenidade, podendo envolver várias etapas, e não um projeto que tem início, meio e fim [1].

O objetivo principal é assegurar que os dados estejam plenamente alinhados às necessidades da organização, garantindo que sua qualidade atenda aos padrões exigidos pelo negócio [30].

Para alcançar esse alinhamento, a implementação eficaz da Governança de Dados requer diretrizes e regras claras, rigorosas e vinculativas, orientadas para a gestão da qualidade, garantindo a adequação dos dados às demandas organizacionais [30].

5.3 Qualidade de Dados

As transformações trazidas pela era digital, como o advento da Indústria 4.0, colocaram os dados no centro das operações estratégicas e dos processos de negócios. Nesse contexto, é possível comparar o papel dos dados hoje ao que os motores a vapor representavam na Indústria 1.0: a força motriz fundamental para o sucesso das operações [59].

Assim como na Revolução Industrial, o êxito da era digital depende substancialmente da qualidade da matéria-prima utilizada – no caso, os dados [59]. A qualidade dos dados, portanto, tornou-se um fator determinante para o funcionamento eficaz e a competitividade das organizações no cenário atual.

Nas grandes corporações, a busca pela Qualidade de Dados tem sido impulsionada pela necessidade de atender a requisitos e padrões de conformidade cada vez mais rigorosos [81]. Como resultado, a gestão da Qualidade de Dados emergiu como um tema central e imprescindível para o sucesso organizacional na era digital [59].

5.3.1 Conceito da Qualidade de Dados

O conceito de qualidade, de forma geral, é frequentemente definido como a “totalidade das características de um produto relacionadas à sua capacidade de atender às necessidades declaradas ou implícitas”, ou seja, como a “conformidade com os requisitos” [48].

No contexto dos dados, a Qualidade de Dados é abordada pelo DAMA DMBOK como o planejamento, implementação e controle de atividades que aplicam técnicas de gestão de qualidade para garantir que os dados estejam disponíveis e atendam às necessidades de uso [83].

Para o DAMA DMBOK, o conceito de Qualidade de Dados pode ser considerado um dos processos mais importante em sua estrutura, pois é essencial para alcançar os objetivos de negócios almejados por meio do uso de dados [1].

Outra definição complementa esse entendimento ao conceituar a Qualidade de Dados como a medição da adequação dos dados para o uso pretendido [84], destacando que sua eficácia depende do nível de adequação necessário para que o usuário atinja os objetivos estabelecidos [85].

A baixa qualidade dos dados pode gerar decisões inadequadas, inconsistentes ou errôneas, resultando em custos computacionais elevados, redução de lucros e até mesmo a perda de clientes. Por isso, a Qualidade de Dados é um tema de grande relevância para pesquisadores e profissionais de diversas áreas [84].

Outro aspecto relevante é a relação da Qualidade de Dados com propriedades chamadas de dimensões, como precisão, integridade e consistência, que servem como critérios para avaliar a qualidade dos dados em uma organização [86].

É importante ressaltar que o conceito de Qualidade de Dados varia entre especialistas, com cada abordagem apresentando diferentes dimensões – algumas compartilhadas e outras específicas – para guiar a avaliação da qualidade nas empresas [1]. Nesse contexto, a análise da Qualidade de Dados deve considerar não apenas a estrutura física dos dados, mas também os fatores contextuais e os impactos negativos que a baixa qualidade pode acarretar para as organizações [1].

5.3.2 Cenário da Qualidade de Dados

A evolução dos Sistemas de Informação tornou o tema Qualidade de Dados ainda mais desafiador, devido à crescente complexidade dos processos que envolvem dados e informações. Com o tempo, a qualidade dos dados compartilhados entre os sistemas pode se degradar rapidamente caso as entradas de informação e a própria qualidade dos dados não sejam devidamente monitoradas e controladas [51].

Por outro lado, a ampliação das capacidades tecnológicas oferece novas oportunidades para a gestão da Qualidade de Dados. A disponibilidade de uma gama mais ampla de fontes e a possibilidade de comparar dados provenientes de diferentes origens permitem detectar e corrigir erros, contribuindo para a melhoria da qualidade geral [51].

Além disso, os Sistemas de Informação vêm migrando de estruturas hierárquicas e monolíticas para modelos baseados em redes, impulsionados por bancos de dados. Essa transformação, observada nos últimos 40 anos, trouxe um aumento exponencial no volume e na diversidade de fontes de dados utilizadas pelas organizações em seus modelos de negócio [48].

Como consequência das transformações na era digital, os problemas relacionados à Qualidade de Dados em bancos de dados tornaram-se mais complexos. Nos sistemas atuais, os dados são frequentemente obtidos de fontes externas, muitas vezes desconhecidas, o que aumenta a complexidade e o risco de inconsistências[48].

A falta de Qualidade de Dados gera custos elevados, afeta negativamente a tomada de decisões e pode prejudicar as relações com clientes e parceiros, além de resultar em perdas financeiras significativas [68].

Um fator crucial para a degradação da Qualidade de Dados é a ausência de regras de negócio claramente definidas durante o processo de coleta de dados [87]. A definição e a aplicação dessas regras são essenciais para garantir a integridade e a confiabilidade das informações, prevenindo falhas nos sistemas e nos processos organizacionais.

Exemplos históricos ilustram os graves impactos da baixa Qualidade de Dados. O “Bug do Milênio”, na transição de 1999 para 2000, revelou a inadequação de sistemas que usavam campos de dois dígitos para representar anos, resultando em custos estimados em 1,5 trilhão de dólares globalmente [48].

Da mesma forma, a explosão do ônibus espacial Challenger, em 1986, foi atribuída ao uso de dados imprecisos, incompletos e desatualizados, demonstrando como a má qualidade pode levar a consequências trágicas [48].

A importância da Qualidade de Dados é amplamente reconhecida por instituições internacionais, como o Sistema Estatístico Europeu, que promove iniciativas voltadas à sua melhoria. Um exemplo é *Data Quality Act*, promulgada pelo governo dos Estados Unidos em 2002, que destaca o papel central da Qualidade de Dados nos processos decisórios e operacionais [48].

Garantir a Qualidade de Dados é fundamental para evitar que elementos incompletos, campos sem significados, dados com erros ou sem precisões contaminem tomadas de decisões corretas, fundamentais no mundo informacional [1].

Além disso, as soluções voltadas à Qualidade de Dados devem garantir que as propriedades dos dados estejam em conformidade com as políticas organizacionais estabelecidas, bem como com os recursos disponíveis [84]. Isso implica na implementação de processos e tecnologias que assegurem a qualidade dos dados, alinhando-os às necessidades estratégicas da organização.

Por fim, a avaliação contínua da Qualidade de Dados nos processos de negócios permite que as organizações maximizem os benefícios no uso de dados, promovendo decisões gerenciais mais assertivas [49]. Dados de baixa qualidade impactam diretamente a eficácia dos processos de negócios, gerando redundâncias e anomalias ao longo do ciclo de vida dos dados [48].

5.3.3 Áreas de Pesquisa da Qualidade de Dados

No contexto corporativo e em diversas nações ao redor do mundo, a Qualidade de Dados tem adquirido uma relevância crescente, além de despertar um interesse significativo no meio acadêmico [68].

Embora seja uma área de pesquisa relativamente recente, a Qualidade de Dados tem sido explorada por diversas disciplinas da Ciência da Computação e outras áreas do conhecimento, conforme ilustrado na Figura 5.4 [48]. Isso reflete a importância multidisciplinar do tema, que abrange desde aspectos técnicos até questões organizacionais e estratégicas.

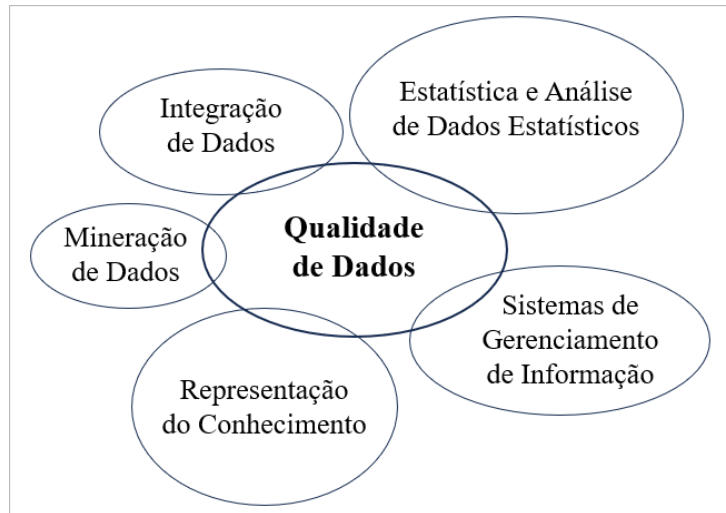


Figura 5.4: Áreas de Pesquisa em Qualidade de Dados - Adaptada [48]

Nas últimas décadas, essas áreas, utilizando conceitos da Estatística, desenvolveram paradigmas, modelos e metodologias que se mostraram de grande importância para o avanço da pesquisa em Qualidade de Dados [48].

A Estatística inclui um conjunto de métodos que são usados para coletar, analisar, apresentar e interpretar dados e informações. Essa ciência desenvolveu, nos últimos dois séculos, um amplo espectro de métodos e modelos que permitem expressar previsões e formular decisões em todos os contextos em que informações incertas e imprecisas estão disponíveis para o domínio de interesse [48].

A Estatística e a Metodologia Estatística, como base da análise de dados, estão focadas em dois tipos básicos de problemas: (a) resumir, descrever e explorar dados; e (b) usar dados amostrais para inferir a natureza do processo que produziu a informação. Com isso, vários métodos estatísticos foram desenvolvidos para avaliar e melhorar a Qualidade de Dados [48].

Ainda, é importante citar a Mineração de Dados, que é um processo analítico projetado para explorar, geralmente, grandes conjuntos de dados em busca de padrões consistentes e/ou relações sistemáticas entre atributos/variáveis [48].

Nessa área, temos a Mineração de Dados Exploratória como o processo preliminar de descoberta de estrutura em um conjunto de dados usando conceitos estatísticos e processo de descoberta de padrões para identificar dados de boa qualidade [48].

Nos últimos anos, diversos estudos sobre Qualidade de Dados têm sido conduzidos em diferentes campos, desde áreas voltadas para dados, como *Big Data* e Estatística, até áreas especializadas, como saúde, finanças e Sistemas de Informação [85].

A Figura 5.5 ilustra as principais questões de pesquisa e os domínios de aplicações relacionados à Qualidade de Dados [48].

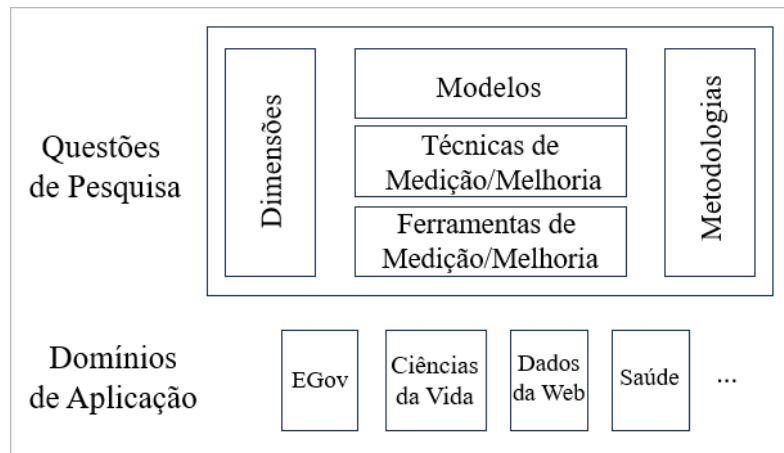


Figura 5.5: Pesquisas em Qualidade de Dados - Adaptada [48]

Apesar dos grandes esforços na área de pesquisa sobre Qualidade de Dados, as organizações continuam aquém no que se refere à qualidade dos seus dados. Observa-se que os processos de gestão da qualidade dos dados permanecem com um baixo nível de maturidade em muitas organizações [81].

Isso significa que a avaliação e a melhoria da qualidade dos dados são realizadas de forma *ad hoc* ou descentralizada dentro dos departamentos das organizações e, muitas vezes, existe uma falta de gestão centralizada da qualidade dos dados e de um alinhamento das necessidades estratégicas e operacionais da empresa [81].

5.3.4 Riscos da Baixa Qualidade de Dados

A baixa qualidade dos dados pode gerar impactos no desempenho e na aplicabilidade dos modelos de dados, comprometendo a confiabilidade e a tomada de decisões. Dados inconsistentes, incompletos ou imprecisos podem levar a análises equivocadas, reduzindo a eficácia dos modelos e aumentando o risco de interpretações errôneas [88].

Para mitigar esses riscos e garantir a confiabilidade dos modelos, é fundamental adotar práticas de análise da qualidade de dados. Isso inclui a realização de procedimentos para assegurar que os dados estejam completos, corretos e devidamente estruturados antes de serem utilizados [88].

5.3.5 Avaliação da Qualidade de Dados

Atualmente, a Qualidade de Dados é um tema de grande relevância, recebendo crescente atenção devido à sua influência direta sobre o valor dos dados. A qualidade está intrinsecamente relacionada à sua utilidade, refletindo na facilidade de acesso, interpretação e aplicação eficaz nas operações e estratégias organizacionais [85].

O uso de dados corretos e de alta qualidade pode auxiliar no planejamento, análise de informações e tomada de decisão nas organizações. Por outro lado, dados de baixa qualidade podem ser inadequados para a finalidade pretendida e trazer consequências graves às atividades das empresas [85].

As descobertas obtidas em recentes estudos sobre a Qualidade de Dados revelam a natureza crítica do tema e destacam a importância de garantir dados de alta qualidade e de avaliar a qualidade dos dados de forma eficaz. Ainda, essas pesquisas enfatizam a necessidade de estudos contínuos nessa área para que se possa avançar no conhecimento do campo de gerenciamento da qualidade dos dados e orientar pesquisadores e profissionais interessados em aprimorar seu entendimento [85].

O procedimento de avaliação da Qualidade de Dados é um componente essencial no procedimento de análise de dados, pois é um processo que determina se os dados atendem aos requisitos de informação dos usuários. Além disso, a avaliação envolve a medição das dimensões ou critérios de qualidade que são relevantes para as necessidades dos usuários [85].

No entanto – devido à sua natureza, variedade de dados e tipos Sistemas de Informação existentes nas organizações –, avaliar a Qualidade de Dados tem sido bastante desafiador para os pesquisadores, uma vez que se trata de uma área de investigação complexa e multidisciplinar [48].

Compreende-se que a avaliação da Qualidade de Dados pode variar de acordo com a finalidade específica a que os dados se destinam. Para avaliar efetivamente a qualidade dos dados, é essencial levar em conta o uso pretendido e os requisitos específicos dos usuários dos dados [85].

Ademais, ressalta-se a importância de se entender que uma abordagem padronizada para a avaliação da Qualidade de Dados pode não ser adequada em todos os casos, pois existem usos e critérios de avaliação distintos para diferentes tipos de dados [85]. Além disso, o tema envolve várias questões de pesquisa, como: dimensões, modelos e metodologias [48].

Nesse cenário, as dimensões da Qualidade de Dados – coleção de atributos dos dados que refletem uma característica ou um aspecto específico da sua natureza – vêm sendo frequentemente citadas na literatura como uma forma eficiente de avaliação da Qualidade de Dados [85].

Atualmente, observa-se que pesquisadores vêm propondo várias categorias e definições para as dimensões da Qualidade de Dados, o que tem resultado na definição de múltiplas categorias encontradas na literatura atualmente. Um exemplo disso é a classificação de dimensão por meio de critérios baseados em sistema, processos e dados [85], conforme ilustrado na Tabela 5.1.

Tabela 5.1: Classificação de Dimensão Por Critérios

Sistema	Processos	Dados
Confiabilidade, disponibilidade, integridade, segurança, desempenho e verificabilidade	Propriedades de segurança e propriedades de veracidade	Precisão, abrangência, oportunidade, credibilidade, custo-benefício e latência

Ainda, a utilização de modelos de qualidade tem se mostrado bastante apropriada, pois fornecem especificações detalhadas e orientação sobre a seleção e a aplicação de medidas apropriadas para realizar a avaliação da qualidade dos dados. Ademais, esses modelos buscam aumentar a eficácia e a precisão dos procedimentos de avaliação da Qualidade de Dados [85].

5.4 Dados Governamentais Abertos

Diversas pesquisas têm apontado que, em diversos países, muitos órgãos de governo têm integrado informações internas, além de procurar compartilhar informações com outros órgãos governamentais. O objetivo dessas ações é oferecer serviços públicos mais eficientes e inovadores aos cidadãos, além de buscar soluções para problemas sociais cada vez mais complexos [89].

De acordo com o *Open Government Working Group* – fórum criado em 2010 pelo governo dos Estados Unidos para profissionais de governo aberto com o objetivo de compartilhar as melhores práticas para promover transparência, participação e colaboração – [90], todos os dados governamentais devem ser abertos ao público, exceto se estiverem relacionados à segurança nacional, segredos comerciais, privacidade pessoal ou outras restrições legais [89].

Os Dados Governamentais Abertos (DGA) possuem importantes valores políticos, econômicos e sociais, o que os torna de grande relevância para muitos países em todo o mundo. Na prática, dados governamentais abertos representam o compartilhamento de informações entre os governos e a sociedade [89].

Além disso, esses dados podem incentivar o público a participar das atividades governamentais e, assim, melhorar a transparência de suas ações e mobilizar a sociedade para a criação de valores econômicos e sociais [89].

Atualmente, é consenso que fatores técnicos têm um impacto direto e significativo sobre as iniciativas de implementação de Dados Governamentais Abertos, mas alguns fatores dificultam a implementação, tais como: dados pouco estruturados, bancos de dados dispersos e dados sem padrão definido [89].

Com o desenvolvimento de pesquisas acerca do tema Dados Governamentais Abertos, a questão da Qualidade de Dados se torna mais relevante, dada a importância de se avaliar com precisão a qualidade dos dados antes de sua publicação e utilização [67].

5.4.1 Qualidade de Dados Governamentais Abertos

O “Princípio da Qualidade” é um dos princípios básicos dos Dados Governamentais Abertos no mundo. Em junho de 2013, os líderes do G8 assinaram a Carta de Dados Abertos na Cúpula da Irlanda do Norte e um dos cinco princípios dessa carta é “Garantir melhor qualidade e maior quantidade de dados abertos”. E, para cumprir esse compromisso, os países do G8 implementaram um Plano de Ação Nacional [67].

A Qualidade de Dados é um dos principais fatores que determina o valor de Dados Governamentais Abertos. As disposições de qualidade dos dados dos documentos da política de dados abertos incluem a verificação dos dados, a avaliação da qualidade dos dados, o gerenciamento da qualidade do ciclo de vida dos dados e o tratamento de fraudes nos dados em todos os níveis de um governo [67].

Adicionalmente, a avaliação da qualidade de Dados Governamentais Abertos deve se concentrar nas características dos dados abertos, objetivos, aplicações e reutilização. Além disso, a disponibilização dos metadados de forma detalhada é fundamental para melhorar a adequação dos dados abertos e para auxiliar os usuários que irão utilizá-los [89].

Desse modo, a Qualidade de Dados é a chave para agregar valor aos dados abertos – inclusive os Dados Governamentais Abertos –, que possuem diversas dimensões, incluindo: originalidade, completude, precisão, pontualidade, disponibilidade, usabilidade, conformidade e compatibilidade [89].

5.5 Dimensões da Qualidade de Dados

O conceito de dimensões de dados é fundamental quando abordamos o tema Qualidade de Dados, uma vez que cada dimensão de um dado reflete um aspecto específico e suas características, especialmente em dados estruturados, típicos de bancos de dados relacionais, que apresentam múltiplas dimensões associadas [5].

Especificamente, as dimensões da Qualidade de Dados referem-se tanto à extensão dos dados (ou seja, aos valores dos dados), quanto à sua finalidade (isto é, ao esquema

de dados). Além disso, as dimensões da Qualidade de Dados geralmente são definidas de forma qualitativa, referindo-se às propriedades dos dados [5].

Apesar da relevância reconhecida dos esquemas de dados, a maior parte da pesquisa sobre as Dimensões da Qualidade de Dados tem se concentrado no estudo dos valores dos dados, uma vez que estes são mais utilizados em comparação ao uso de esquemas em processos de negócios[5].

As dimensões podem ser caracterizadas por uma estrutura de classificação comum que nos permite comparar dimensões entre diferentes tipos de informações [5].

Ainda, as dimensões Qualidade de Dados oferecem uma estrutura para entender os atributos-chaves que definem e medem a qualidade dos dados. Embora o número de dimensões possa variar dependendo da fonte, uma estrutura comumente adotada inclui seis dimensões principais [91]:

1. **Completeness:** Avalia se todos os dados necessários estão disponíveis. Se houver campos obrigatórios ausentes em um conjunto de dados, ele não será considerado completo;
2. **Consistência:** Os dados precisam ser consistentes, o que significa que não devem existir discrepâncias quando comparados entre diferentes conjuntos de dados ou diferentes partes do mesmo conjunto de dados;
3. **Conformidade:** Refere-se ao grau em que os dados aderem a padrões especificados, convenções e regras de negócios. Por exemplo, um número de telefone deve estar em um formato válido;
4. **Integridade:** Diz respeito à validade e consistência de relacionamentos entre entidades e registros de dados, como referências cruzadas entre tabelas em um banco de dados relacional;
5. **Precisão:** Relaciona-se ao grau em que os dados representam a realidade ou a verdade. A precisão pode ser difícil de medir, pois requer um ponto de referência verdadeiro;
6. **Atualidade:** Refere-se à relevância dos dados no tempo. Isso pode variar dependendo do contexto. Por exemplo, dados de vendas de um dia atrás podem ser considerados atuais para algumas empresas, enquanto que, para outras empresas, esses dados podem ser considerados desatualizados.

5.6 Metodologias, Técnicas e Modelos

Para a abordagem do tema Qualidade de Dados, é fundamental compreender alguns conceitos e definições-chave. Esta pesquisa discutirá, entre esses conceitos, as Metodologias,

Técnicas e Modelos, que desempenham papel crucial na avaliação e aprimoramento da qualidade dos dados.

As Metodologias podem ser definidas como um conjunto de práticas que fornecem diretrizes para que seja possível identificar e selecionar, a partir das técnicas disponíveis, o procedimento mais adequado e eficiente a ser utilizado em um determinado projeto, como, por exemplo, realizar a avaliação e a melhoria da Qualidade de Dados em um Sistema de Informação específico [48].

As Técnicas correspondem ao conjunto de algoritmos, heurísticas, procedimentos baseados em conhecimento e processos de aprendizagem que fornecem uma solução para um problema específico ou para uma determinada atividade como a avaliação da Qualidade de Dados [48].

Um exemplo de técnica para realizar a avaliação da Qualidade de Dados é o procedimento para identificar a fonte de dados mais confiável de alguma informação específica. Outro exemplo é o mecanismo de identificação de registros de bancos de dados diferentes que representam o mesmo objeto do mundo real [48].

Os Modelos são usados em bancos de dados para representar dados e esquemas de dados. Eles também são usados em Sistemas de Informação para representar os processos de negócios de uma organização. No contexto da avaliação da Qualidade de Dados, os modelos de dados precisam ser adequados para representar as dimensões de qualidade [48].

5.6.1 Modelos para Avaliação da Qualidade de Dados

No contexto da avaliação da Qualidade de Dados, os modelos de qualidade desempenham um papel crucial ao fornecer procedimentos detalhados que orientam a medição da qualidade dos dados [85].

Esses modelos não apenas orientam a seleção e aplicação de métricas apropriadas para avaliar a qualidade dos dados, mas também permitem a criação de procedimentos padronizados, o que contribui para a eficácia e precisão dos processos de avaliação da Qualidade de Dados [85].

Além disso, os modelos de avaliação de qualidade incluem elementos de modelagem que representam requisitos específicos de qualidade dos dados. Esses elementos estão intimamente relacionados às dimensões fundamentais necessárias para garantir a qualidade dos dados [5].

Portanto, os modelos de avaliação de dados especificam quais dados são críticos, pois sua qualidade é essencial para o sucesso dos fluxos de informações e processos dentro de uma organização [5].

5.6.2 Metodologias para Avaliação da Qualidade de Dados

Uma metodologia para avaliação da Qualidade de Dados pode ser definida como um conjunto estruturado de diretrizes e técnicas que, a partir de informações relacionadas a uma realidade específica, estabelece um processo para medir e aprimorar a qualidade dos dados em uma organização. Este processo é desenvolvido por meio de fases, etapas e pontos de decisão [5].

No livro “Data Quality: Concepts, Methodologies and Techniques” [5], é apresentada uma metodologia que adota como premissa o uso de estruturas de entrada e saída, fundamentadas no conhecimento prévio dos dados a serem avaliados, com o objetivo de mensurar e melhorar o processo de Qualidade de Dados.

Essa abordagem metodológica baseia-se na comparação entre a avaliação objetiva e a avaliação subjetiva dos dados, por meio da classificação das dimensões da qualidade. Isso permite verificar a conformidade e a qualidade dos dados, de acordo com as especificações e expectativas dos usuários.

A estrutura e as etapas dessa metodologia são apresentadas na Figura 5.6.

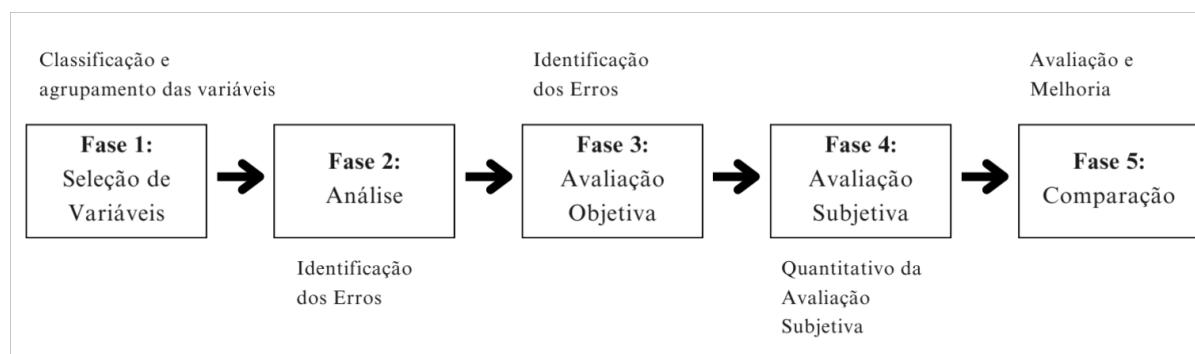


Figura 5.6: Fases da Metodologia de Avaliação da Qualidade de Dados - Adaptada [5]

As fases dessa metodologia [5] são descritas a seguir:

- **Fase 1: Seleção das Variáveis:** Esta fase envolve a identificação, descrição, classificação e agrupamento das variáveis, que representam os principais atributos dos dados a serem avaliados. Ela é crucial para garantir que os dados selecionados sejam relevantes para as etapas subsequentes de análise;
- **Fase 2: Análise:** O objetivo desta fase é identificar as principais dimensões dos dados e as restrições de integridade a serem avaliadas. O foco principal é detectar as causas subjacentes dos erros nos dados, proporcionando uma base sólida para as fases de avaliação subsequentes;

- **Fase 3: Avaliação Objetiva:** Esta fase se concentra na definição dos índices para quantificar a qualidade dos dados. Para tanto, são aplicados métodos estatísticos que determinam o número de observações erradas em relação às diferentes dimensões da qualidade dos dados, permitindo uma análise quantitativa da situação;
- **Fase 4: Avaliação Subjetiva:** A avaliação subjetiva envolve três análises: a análise dos dados sob a ótica do processo de negócios, a análise do usuário que interage com os dados no dia a dia e a análise de um especialista em qualidade de dados, que examina os aspectos técnicos e operacionais da qualidade dos dados; e
- **Fase 5: Comparação:** Por fim, realiza-se a comparação entre as avaliações objetiva e subjetiva para cada variável e dimensão da qualidade. As discrepâncias identificadas pelo especialista são essenciais para a detecção das causas dos erros, permitindo a proposição de soluções para a correção desses problemas.

Ademais, um outro estudo intitulado “Quality Assessment for Open Government Data in China” [67] apresentou uma metodologia para a avaliação da Qualidade de Dados em três cidades chinesas: Pequim, Guangzhou e Harbin. Nessas cidades, o compartilhamento de dados governamentais fazem parte na estratégia nacional de *Big Data*.

A metodologia proposta por esse estudo baseia-se na análise de um conjunto de dados das três cidades, utilizando dimensões e métricas específicas para avaliar a qualidade dos dados. Além disso, inclui a identificação dos principais problemas relacionados à qualidade dos dados, uma etapa crucial que impacta diretamente a disponibilidade e a confiabilidade das informações para os usuários.

Em síntese, o estudo chinês delinea os seguintes passos para a avaliação da Qualidade de Dados, conforme descrito abaixo [67]:

1. **Identificação dos Problemas de Qualidade:** A partir da perspectiva dos usuários dos dados, foram identificados os tipos de problemas de qualidade presentes em cada conjunto de dados nos sites de dados abertos das cidades de Pequim, Guangzhou e Harbin;
2. **Classificação de Dimensões e Métricas de Qualidade:** Após mapear os problemas de qualidade, foi realizada a classificação de cada um dos problemas identificados, levando em consideração as dimensões e métricas da Qualidade de Dados;
3. **Análise dos Dados:** Com base nos cálculos das métricas aplicadas aos dados das três cidades, foram calculadas a média estatística e o desvio padrão das dimensões de qualidade, possibilitando uma análise quantitativa do estado da qualidade dos dados.

Após a análise e pesquisa das metodologias apresentadas, constatou-se que a abordagem descrita no estudo “Quality Assessment for Open Government Data in China” [67] é a mais adequada para os objetivos deste trabalho, especialmente no que diz respeito à Avaliação da Qualidade dos Dados do PBLE.

Portanto, a metodologia proposta no estudo chinês será adotada como base para a presente pesquisa, pois se mostrou apropriada e alinhada aos objetivos do nosso estudo. Contudo, é importante destacar que, embora a metodologia tenha sido adotada como estrutura principal, adaptações foram realizadas para ajustá-la às características e especificidades dos dados do PBLE.

Dessa forma, ajustamos a metodologia à realidade deste estudo, e as etapas do desenvolvimento desta pesquisa serão detalhadas no próximo capítulo.

Capítulo 6

Desenvolvimento do Trabalho

Este capítulo apresenta o desenvolvimento da pesquisa, organizado em nove seções. As duas primeiras descrevem a arquitetura da informação. A terceira aborda os dados do PBLE, seguida da quarta, que detalha a metodologia adotada. A quinta seção expõe o desenvolvimento da metodologia. A sexta apresenta o diagnóstico da Qualidade dos Dados. A sétima trata dos resultados finais, enquanto a oitava propõe uma solução. Por fim, a nona seção discute a análise de riscos.

6.1 Arquitetura da Informação

O conceito Arquitetura da Informação surgiu na década de 1990 como uma metáfora que estabelece uma analogia entre o processo de coleta, organização, estruturação e representação de informações em ambientes digitais e os desafios enfrentados por um arquiteto ao projetar edificações [92].

De maneira geral, a Arquitetura da Informação é definida como o processo de criação de estruturas que facilitam a compreensão e a utilização das informações de maneira mais eficaz e eficiente. Seu papel é essencial para a organização e acessibilidade dos dados, contribuindo diretamente para a otimização da gestão do conhecimento e a recuperação de informações, especialmente em ambientes corporativos e digitais [92].

Além disso, a Arquitetura da Informação está diretamente ligada ao fluxo da informação. Enquanto a primeira estabelece a estrutura necessária para organizar os dados, o segundo define como essa estrutura é utilizada e gerenciada no cotidiano. Uma Arquitetura da Informação bem planejada é fundamental para um fluxo informacional eficiente, e a interação entre esses dois elementos influencia diretamente a qualidade da organização, do acesso e do uso das informações [92].

6.2 Arquitetura da Informação do PBLE

A Arquitetura da Informação do PBLE é composta por um conjunto de estruturas interconectadas que possibilitam a geração, atualização, coleta, armazenamento e análise das informações do programa. Essas estruturas desempenham um papel essencial na organização e no fluxo das informações, garantindo que os dados sejam acessíveis e confiáveis para realizar as atividades de acompanhamento e controle e de fiscalização do PBLE.

As principais estruturas e funções da Arquitetura da Informação do PBLE estão organizadas em três etapas fundamentais:

- **Geração de Informações das Escolas:** O processo tem início com o Censo Escolar Anual realizado pelo INEP, que coleta e consolida os dados das instituições de ensino. Posteriormente, o FNDE processa essas informações e disponibiliza às prestadoras de serviços de telecomunicações a lista de escolas elegíveis ao PBLE, conforme os critérios estabelecidos pelo programa;
- **Geração de Dados do PBLE:** Com a lista de escolas elegíveis em mãos, as prestadoras elaboram um cronograma anual para a instalação da conexão nas unidades de ensino. Após a efetivação das conexões, as prestadoras geram relatórios detalhados sobre a execução do serviço e enviam esses dados à Anatel, garantindo o acompanhamento da implementação do programa; e
- **Coleta e Armazenamento de Dados do PBLE:** A Anatel utiliza o Sistema de Coleta de Informações (SICI) para receber, armazenar e gerenciar os dados enviados pelas operadoras. Esse sistema desempenha um papel central nas atividades de fiscalização e no acompanhamento e controle do cumprimento das obrigações do PBLE, assegurando a transparência e a eficiência das ações da Agência.

A Figura 6.1 apresenta a estrutura da Arquitetura da Informação do PBLE, evidenciando a organização lógica dos seus componentes, fluxos de dados e inter-relações entre os elementos que compõem o ecossistema informacional do programa.

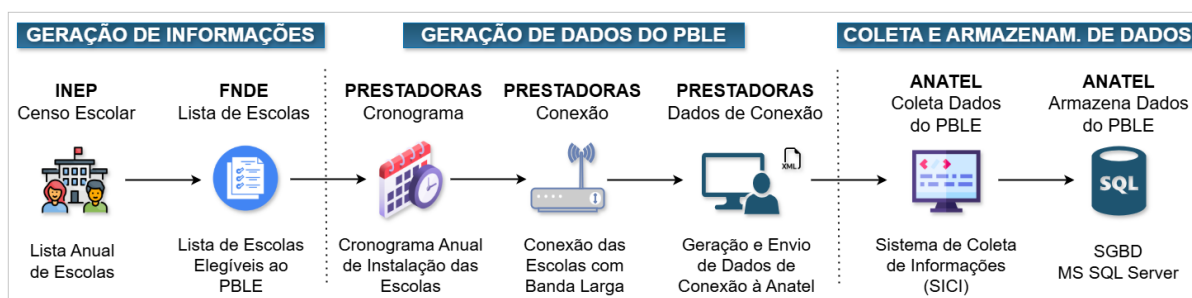


Figura 6.1: Arquitetura da Informação do PBLE

A Arquitetura da Informação do PBLE é um elemento essencial para garantir a eficiência e a transparência na gestão do programa pela Anatel. No entanto, aprimorar a governança e a qualidade dos dados é fundamental para aumentar a confiabilidade das informações e garantir a efetividade das ações do programa.

6.3 Dados do Estudo

Os dados utilizados neste estudo têm como base o fluxo completo de informações do PBLE, abrangendo desde sua coleta até sua utilização para a gestão do programa. O processo tem início com a coleta dos dados do programa, realizada por meio do SICI, que recebe periodicamente as informações enviadas pelas prestadoras.

Uma vez coletados, os dados são armazenados em um SGBD mantido pela Anatel. Esse armazenamento estruturado assegura a integridade, a disponibilidade e a confiabilidade dos dados, permitindo que sejam utilizados com eficiência em atividades de fiscalização, acompanhamento e controle e tomada de decisão no âmbito da Agência.

Com base nesse fluxo, o presente estudo propõe a realização de uma avaliação criteriosa da qualidade dos dados do PBLE. Esse procedimento é essencial para identificar e corrigir possíveis inconsistências, promovendo maior precisão e fidedignidade das informações e, conseqüentemente, assegurando que os dados reflitam com rigor a realidade da conectividade das escolas públicas atendidas.

A estrutura geral desse fluxo de informações é ilustrada na Figura 6.2.

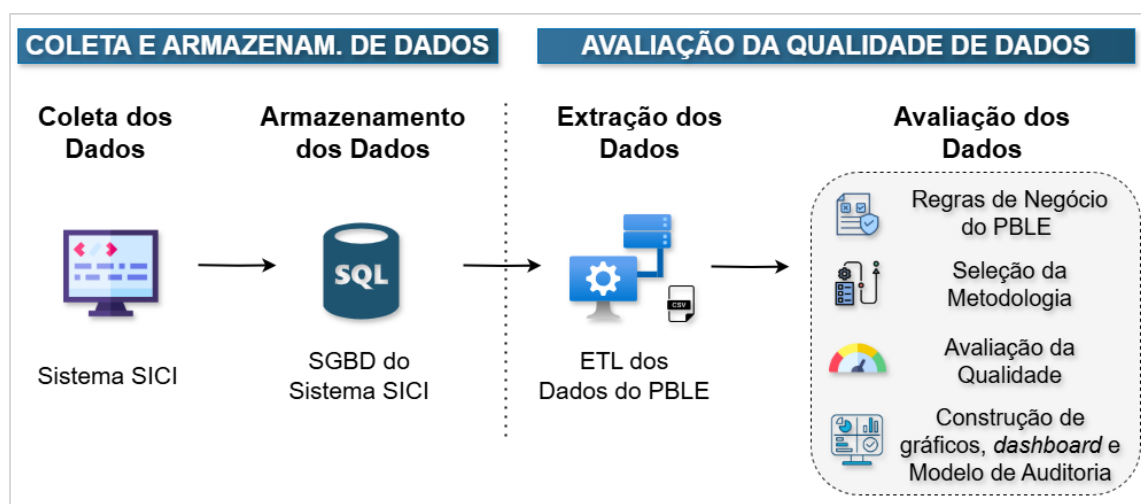


Figura 6.2: Fluxo dos Dados do PBLE

6.3.1 Coleta dos Dados

A coleta dos dados do PBLE é realizada pela Anatel por meio do sistema SICI, uma plataforma desenvolvida em linguagem de programação ASP e mantida pela área de Tecnologia da Informação da Agência. Esse sistema desempenha um papel essencial na centralização e processamento das informações relacionadas à conectividade nas escolas públicas urbanas.

O SICI é um sistema de acesso exclusivo das prestadoras de serviços de telecomunicações, que, por meio de seus representantes legais, encaminham periodicamente os dados do PBLE à Anatel. As informações enviadas incluem detalhes sobre a instalação e manutenção das conexões de internet nas escolas atendidas pelo programa.

Entre os dados encaminhados pelas prestadoras, estão informações técnicas da conexão, velocidade instalada, tecnologia utilizada, endereços das instituições atendidas, além de previsões para novas instalações. Esses registros permitem à Anatel acompanhar o cumprimento das metas estabelecidas pelo programa e identificar possíveis falhas na execução dos serviços.

Com base nessas informações, a Anatel realiza atividades de acompanhamento e controle, verificando o cumprimento das obrigações de instalação da conexão das escolas públicas, conforme os regramentos do PBLE. Além disso, os dados armazenados no sistema subsidiam as ações fiscalizatórias da Agência, garantindo que as prestadoras cumpram os requisitos técnicos e operacionais estabelecidos.

A avaliação da qualidade dos dados coletados é fundamental para assegurar que o PBLE atinja seu objetivo principal: garantir o acesso à internet de qualidade para os estudantes das escolas públicas urbanas, promovendo a inclusão digital e impulsionando a transformação tecnológica no ambiente escolar.

6.3.2 Armazenamento dos Dados

Atualmente, os dados do PBLE estão armazenados em um servidor de banco de dados Microsoft SQL Server 2014 (versão 12.0.6449.1 – SGBD), licenciado pela Anatel. Esse servidor está hospedado em um data center localizado na infraestrutura do ambiente computacional interno da autarquia, garantindo maior controle e segurança sobre as informações.

A base de dados do PBLE possui um volume significativo, pois armazena os dados enviados mensalmente pelas prestadoras de serviços de telecomunicações desde fevereiro de 2013. Ao longo dos anos, esse repositório cresceu substancialmente, refletindo a evolução do programa e a necessidade contínua de monitoramento da conectividade nas escolas públicas urbanas.

Em uma consulta realizada em junho de 2024, verificou-se que a base de dados do PBLE já contabilizava aproximadamente doze milhões e setecentos mil registros, evidenciando tanto a complexidade da gestão dessas informações e a importância de sua manutenção eficiente.

Para garantir a segurança, disponibilidade e acessibilidade dos dados do programa, a Anatel adota uma estratégia de redundância dos servidores e do sistema de armazenamento. Esse mecanismo assegura que, em caso de falha, os dados permaneçam acessíveis, minimizando riscos operacionais.

Além disso, a Agência mantém uma rotina diária de *backup* de suas bases de dados, incluindo os registros do PBLE. Esse procedimento é realizado tanto localmente, utilizando dispositivos próprios, quanto por meio da contratação de serviços em nuvem, garantindo uma camada extra de proteção contra eventuais incidentes.

Essas ações são fundamentais para proteger os dados do programa contra possíveis perdas ou danos decorrentes de falhas de hardware ou software, desastres naturais ou ataques cibernéticos.

Dessa forma, busca-se assegurar alta disponibilidade dos serviços, mitigação dos riscos de perda de dados e continuidade do acesso às informações do PBLE, elementos essenciais para a efetiva fiscalização e aprimoramento do programa.

6.3.3 Fiscalização e Qualidade dos Dados

A qualidade dos dados exerce um papel imprescindível para as atividades de fiscalização do PBLE. Os dados provenientes do sistema SICI integram a etapa de Planejamento do Procedimento de Fiscalização do programa, servindo como base para as diretrizes operacionais e para o embasamento técnico das ações fiscalizatórias da Anatel. Dessa forma, a confiabilidade, a consistência e a integridade das informações extraídas do SICI são determinantes para o sucesso do processo fiscalizatório conduzido pela Agência.

Dados imprecisos, incompletos ou desatualizados comprometem a precisão das análises realizadas pela área de fiscalização da Agência, dificultando a identificação de possíveis irregularidades, o apontamento do descumprimento das obrigações das prestadoras e a aferição da efetividade da conectividade nas escolas públicas, além de resultar na alocação inadequada de recursos de fiscalização, reduzindo a eficiência das ações regulatórias.

A influência da qualidade dos dados se estende a todas as fases do ciclo fiscalizatório, desde o planejamento e a priorização das fiscalizações de campo até a elaboração dos Relatórios de Fiscalização. Esses relatórios, por sua vez, constituem instrumentos técnicos fundamentais para subsidiar decisões institucionais no âmbito da Anatel, impactando diretamente a gestão e o acompanhamento do PBLE.

Diante desse cenário, torna-se evidente que a avaliação da qualidade dos dados do PBLE constitui condição fundamental para assegurar a efetividade do procedimento de fiscalização, especialmente na etapa de planejamento, na qual esse tipo de análise possibilita à Agência identificar fragilidades nos fluxos informacionais, corrigir inconsistências e aprimorar os processos internos de fiscalização e acompanhamento do programa.

6.3.4 Avaliação da Qualidade dos Dados

Esta pesquisa concentra-se na análise dos dados das escolas públicas atendidas pelo PBLE, com o objetivo de avaliar a qualidade dessas informações e identificar possíveis inconsistências. Essa análise é fundamental para assegurar a precisão dos registros e a eficácia das ações vinculadas ao programa. O processo envolve diversas etapas, incluindo a análise das regras de negócio do PBLE e a seleção da metodologia de avaliação, conforme ilustrado na Figura 6.2.

O estudo será conduzido em fases sequenciais, iniciando-se pela análise minuciosa das regras de negócio do PBLE, a fim de compreender seus critérios operacionais e normativos. Em seguida, será realizada a seleção da metodologia mais adequada para avaliar a qualidade dos dados, com base em práticas consolidadas na literatura científica, garantindo rigor técnico e aplicabilidade prática.

A avaliação da qualidade dos dados do PBLE assume papel estratégico para a Anatel, configurando-se como um instrumento fundamental para a gestão e monitoramento do programa. Esse processo permite não apenas o acompanhamento e controle do cumprimento das obrigações assumidas pelas prestadoras de serviços de telecomunicações, mas também subsidia o aperfeiçoamento das ações fiscalizatórias conduzidas pela Agência.

Além disso, a avaliação da qualidade dos dados viabiliza a implementação de melhorias nas atividades operacionais da Anatel, contribuindo para a adequada execução do PBLE e para a efetividade das conexões de internet nas escolas públicas. Esse processo também promove a revisão e o aprimoramento contínuo dos procedimentos de coleta de dados, tornando-os mais confiáveis, eficientes e alinhados às melhores práticas de governança de dados.

Adicionalmente, os resultados obtidos poderão ser utilizados como subsídio para a atualização das metodologias de fiscalização, otimizando a atuação da Anatel junto às prestadoras responsáveis pela instalação e manutenção da conectividade nas escolas. A melhoria da qualidade dos dados também impacta diretamente a publicação das informações do PBLE, tanto no Portal da Anatel quanto no Portal de Dados Abertos do Governo Federal, assegurando maior transparência, acessibilidade e confiabilidade dos dados disponibilizados à sociedade.

Por fim, o aprimoramento da qualidade das informações fortalecerá a capacidade analítica da Agência, oferecendo suporte qualificado à formulação de políticas públicas e à definição de diretrizes estratégicas voltadas à promoção da conectividade nas escolas públicas e à ampliação da inclusão digital dos estudantes brasileiros.

6.4 Metodologia da Avaliação da Qualidade de Dados

Conforme exposto no capítulo anterior, este estudo adota como referência a metodologia de Avaliação da Qualidade de Dados apresentada no artigo “Quality Assessment for Open Government Data in China” [67], a qual demonstrou ser a mais adequada para atingir os objetivos propostos nesta pesquisa, especialmente no contexto do PBLE.

A metodologia proposta organiza o processo de avaliação da qualidade dos dados em três etapas fundamentais, que se complementam de forma lógica e estruturada: i) Identificação dos Problemas de Qualidade, ii) Classificação de Dimensões e Métricas de Qualidade, e iii) Análise dos Dados.

No entanto, diante das particularidades dos dados provenientes do PBLE, aliadas às exigências específicas do contexto regulatório e operacional da Anatel, foi necessário adaptar a metodologia original para melhor refletir as especificidades do conjunto de dados analisado.

A Figura 6.3 sistematiza os cinco passos que compõem a metodologia adaptada e proposta neste trabalho, organizados sequencialmente de modo a assegurar consistência, rigor analítico e alinhamento com os objetivos da pesquisa.



Figura 6.3: Metodologia de Avaliação da Qualidade de Dados

A seguir, são descritos, de forma detalhada, os cinco passos que compõem a metodologia proposta neste estudo para a avaliação da qualidade dos dados do PBLE:

1. **Identificação dos Dados do PBLE:** Esta etapa consiste na análise dos esquemas, metadados e tabelas disponíveis no SGBD que armazenam os dados do PBLE. A partir dessa análise, são identificados e documentados os nomes e descrições de todos os campos que integram a estrutura de dados do programa;
2. **Seleção dos Dados do PBLE:** Com base nos objetivos definidos para a pesquisa, procede-se à seleção dos campos mais relevantes para a análise. Essa escolha considera critérios técnicos e estratégicos relacionados ao contexto do acompanhamento, controle e fiscalização do PBLE realizadas pela Anatel. Nessa etapa, são também definidos a natureza (tipo de dado) e o tamanho de cada campo selecionado;
3. **Identificação dos Problemas de Qualidade dos Dados do PBLE:** A partir dos campos selecionados, realiza-se a identificação dos principais tipos de problemas de qualidade que afetam as informações do PBLE. Essa identificação é orientada pelos objetivos específicos da pesquisa e busca reconhecer os fatores que mais comprometem os dados;
4. **Seleção das Dimensões de Qualidade de Dados:** Com base nos problemas de qualidade previamente mapeados, esta etapa compreende a seleção das dimensões de qualidade mais adequadas à realidade dos dados do PBLE. A análise é conduzida à luz das boas práticas da literatura especializada, permitindo a construção de métricas que possibilitem a mensuração objetiva da qualidade dos dados;
5. **Avaliação da Qualidade dos Dados do PBLE:** Por fim, realiza-se a avaliação propriamente dita da qualidade dos dados, por meio do cálculo das métricas definidas na etapa anterior. A análise quantitativa é conduzida com base nos campos selecionados, e os resultados obtidos são sintetizados em médias estatísticas das dimensões de qualidade avaliadas.

6.5 Desenvolvimento da Pesquisa

6.5.1 Identificação dos Dados do PBLE

A partir da análise da estrutura das tabelas relacionais que compõem a base de dados do PBLE, identificou-se um total de trinta e um campos que integram a arquitetura de dados empregada no acompanhamento e controle das ações vinculadas ao programa.

Esses dados são fornecidos pelas prestadoras de serviços de telecomunicações, configurando-se como a principal fonte de dados utilizada para subsidiar o monitoramento da execução do PBLE, bem como para a análise da efetividade da conectividade ofertada às escolas públicas urbanas.

A Tabela 6.1 apresenta de forma sistematizada os campos identificados na base de dados, contendo o identificador sequencial (ID), o nome técnico do campo (Campo) e sua respectiva definição (Descrição). Essa estruturação visa proporcionar uma visão clara e objetiva da composição informacional que sustenta o PBLE, servindo de base para análises consistentes no contexto da avaliação da qualidade dos dados.

Tabela 6.1: Campos do PBLE

ID	Campo	Descrição
1	NumFistel	Código identificador da autorização do serviço de telecomunicação
2	DataMovimentoAno	Ano de referência da carga de dados
3	DataMovimentoMes	Mês de referência da carga de dados
4	CodINEP	Código do INEP que identifica uma escola
5	NomeEscola	Nome da escola
6	CodMunicipio	Código do IBGE que identifica um município
7	EnderecoEscola	Endereço da localização da escola
8	BairroEscola	Bairro da localização da escola
9	ComplementoEscola	Complemento da localização da escola
10	NumeroEscola	Número da localização da escola
11	CEPEscola	Código dos Correios que identifica a localização da escola
12	IdtTipoEscola	Código que identifica o tipo de escola
13	codTipoObrigacao	Código que identifica o PBLE
14	IdtSituacao	Código que identifica a situação da conexão da escola
15	DataAtivacao	Data de ativação da conexão da escola
16	VelAcessoObrigacao	Valor da velocidade de obrigação da conexão da escola
17	VelAcessoInstalada	Valor da velocidade de conexão da escola
18	IP	Endereço IP que identifica a conexão da escola
19	IdtTecnologia	Código que identifica a tecnologia de conexão da escola
20	Circuito	Código que identifica o circuito da prestadora
21	IdentificacaoCentral	Código que identifica a central da prestadora
22	IdentificacaoDSLAM	Código que identifica a DSLAN da prestadora
23	IdtTipoPendencia	Código que identifica o tipo de pendência de conexão da escola
24	DataPrevisao	Data de previsão de conexão da escola
25	DataUltimaVistoria	Data da última vistoria realizada na escola
26	DataDesativacao	Data de desativação da escola
27	DataHoraSolicitacao	Data e Hora da solicitação de reparo da escola
28	DataHoraReparo	Data e Hora da execução do reparo da escola
29	PrzAtendimento	Prazo de atendimento de reparo da escola
30	MotivoReclamado	Código que identifica o motivo de reclamação da escola
31	DefeitoConstatado	Código que identifica o defeito constatado

6.5.2 Seleção dos Dados do PBLE

Após a identificação dos campos que compõem a estrutura das tabelas do SGBD do PBLE, responsáveis por armazenar os dados e informações do programa, iniciou-se o processo de seleção dos campos mais relevantes a serem considerados nesta pesquisa. Essa etapa

foi fundamental para garantir que a análise fosse conduzida com base em informações alinhadas aos objetivos do estudo e às necessidades institucionais da Anatel.

A seleção dos campos foi guiada por critérios técnicos e estratégicos, alinhados aos objetivos centrais do estudo e ao papel institucional da Anatel na fiscalização, no acompanhamento e no controle do PBLE. Essa escolha criteriosa é essencial para assegurar análises consistentes, baseadas em informações relevantes e representativas da realidade operacional do programa.

Foram priorizados os campos diretamente relacionados às atividades de fiscalização do cumprimento das obrigações assumidas pelas prestadoras de serviços de telecomunicações no âmbito do PBLE. Essa decisão está em conformidade com as competências regimentais da Anatel e evidencia a importância de apoiar as ações fiscalizatórias da Agência, que visam garantir a efetiva implementação da conectividade nas instituições de ensino da rede pública urbana.

Com esse direcionamento, foram descartados campos relacionados à infraestrutura interna das redes de telecomunicações das prestadoras, como *IdentificacaoDSLAM*, *IdentificacaoCentral* e *Circuito*, por não estarem diretamente vinculados às atividades de fiscalização da Anatel no âmbito do PBLE. Da mesma maneira, também foram excluídos campos que não são objeto de acompanhamento pela Agência, como *DataUltimaVistoria*, *DataHoraSolicitacao* e *DataHoraReparo*.

Assim, a partir do total de trinta e um campos inicialmente identificados na base de dados do PBLE – conforme descrito na etapa anterior desta metodologia – foi conduzido um processo criterioso de análise, que resultou na seleção de vinte e dois campos considerados mais relevantes para os objetivos deste estudo.

Adicionalmente, essa seleção representa um passo fundamental para viabilizar a aplicação da metodologia de avaliação da qualidade de dados sob uma ótica prática e aplicada, contribuindo para o fortalecimento das ações de governança e para a melhoria contínua dos processos de fiscalização institucional.

A Tabela 6.2 apresenta os campos selecionados, contendo o identificador sequencial (ID), nome (Campo), contendo o identificador sequencial (ID), o nome do campo (Campo), a respectiva definição (Descrição), o tipo de dado (Tipo) e o tamanho máximo permitido (Tamanho).

Tabela 6.2: Dados do PBLE Selecionados

ID	Campo	Descrição	Tipo	Tamanho
1	NumFistel	Código que identifica a autorização de um serviço de telecomunicação	Numérico	11
2	DataMovimentoAno	Ano de referência da carga de dados	Numérico	4
3	DataMovimentoMes	Mês de referência da carga de dados	Numérico	2
4	CodINEP	Código do INEP que identifica uma escola	Numérico	8
5	NomeEscola	Nome da escola	Texto	255
6	CodMunicipio	Código do IBGE que identifica um município	Numérico	7
7	EnderecoEscola	Endereço da localização da escola	Texto	255
8	BairroEscola	Bairro da localização da escola	Texto	255
9	ComplementoEscola	Complemento da localização da escola	Texto	255
10	NumeroEscola	Número da localização da escola	Numérico	6
11	CEPEscola	Código dos Correios que identifica a localização da escola	Texto	8
12	IdtTipoEscola	Código que identifica o tipo de escola	Numérico	2
13	codTipoObrigacao	Código que identifica o PBLE	Numérico	2
14	IdtSituacao	Código que identifica a situação da conexão da escola	Numérico	2
15	DataAtivacao	Data de ativação da conexão da escola	Data	10
16	VelAcessoObrigacao	Valor da velocidade de obrigação da conexão da escola	Numérico	4
17	VelAcessoInstalada	Valor da velocidade de conexão da escola	Numérico	4
18	IP	Endereço IP que identifica a conexão da escola	Texto	9 a 15
19	IdtTecnologia	Código que identifica a tecnologia de conexão da escola	Numérico	2
20	IdtTipoPendencia	Código que identifica o tipo de pendência de conexão da escola	Numérico	2
21	DataPrevisao	Data de previsão de conexão da escola	Data	7
22	DataDesativacao	Data de desativação da escola	Data	10

6.5.3 Identificação dos Problemas de Qualidade dos Dados

A partir da seleção dos campos considerados mais relevantes da base de dados do PBLE, foi conduzida uma análise sistemática com o objetivo de identificar os principais tipos de problemas de qualidade que podem comprometer negativamente a eficácia das atividades desenvolvidas pela Anatel na fiscalização e no acompanhamento e controle do atendimento às escolas públicas pelo programa.

Esse processo analítico resultou na identificação de oito categorias de problemas de qualidade de dados recorrentes no PBLE, descritas a seguir:

- **Problema 1 (P1):** Ausência de dados obrigatórios;
- **Problema 2 (P2):** Divergência entre os dados do PBLE e a base histórica da Anatel;
- **Problema 3 (P3):** Divergência entre os dados do PBLE e os registros do INEP;
- **Problema 4 (P4):** Divergência entre os dados do PBLE e as informações da base dos Correios;
- **Problema 5 (P5):** Não conformidade com as regras de negócio estabelecidas pela Anatel;

- **Problema 6 (P6):** Alterações indevidas nos dados ao longo de seu ciclo de vida;
- **Problema 7 (P7):** Presença de valores inadequados ou em formato inconsistente com o padrão definido;
- **Problema 8 (P8):** Desatualização dos dados fornecidos.

Observou-se que esses problemas se manifestam em diferentes níveis de granularidade dentro da estrutura dos dados: no nível de Campo (coluna), Registro (linha) e Conjunto de Dados (arquivo), o que reforça a necessidade de abordagens específicas para tratamento em cada um desses níveis.

A Tabela 6.3 apresenta um resumo dos oito tipos de problemas identificados, classificando-os de acordo com seu nível de granularidade, com o intuito de apoiar uma análise mais precisa e orientar o desenvolvimento das métricas de qualidade de dados adotadas nas etapas subsequentes desta pesquisa.

Tabela 6.3: Tipos de Problemas de Dados do PBLE

ID	Problema de Qualidade	Granularidade
P1	Os dados obrigatórios estão ausentes	Registro
P2	Os dados históricos estão divergentes da base da Anatel	Conjunto de Dados
P3	Os dados estão divergentes da base da INEP	Conjunto de Dados
P4	Os dados estão divergentes da base dos Correios	Conjunto de Dados
P5	Os dados em inconformidade com regras de negócio da Anatel	Campo
P6	Os dados são alterados durante o seu ciclo de vida	Registro
P7	Os dados possuem valores inadequados ou formato inconsistente	Registro
P8	Os dados estão desatualizados	Registro

6.5.4 Seleção das Dimensões de Qualidade de Dados

Neste estudo, diversas dimensões de qualidade de dados, amplamente abordadas na literatura especializada, foram investigadas e analisadas com o objetivo de avaliar a qualidade dos dados do PBLE.

As dimensões que se mostraram mais adequadas ao contexto do estudo foram: **Completude**, **Consistência**, **Conformidade**, **Integridade**, **Precisão** e **Atualidade**. A seleção dessas dimensões baseou-se tanto na recorrência de problemas identificados nos dados do PBLE quanto na relevância dessas dimensões em relação aos campos selecionados para análise.

A partir dessa definição, iniciou-se a identificação dos critérios específicos – denominados Regras – para cada um dos vinte e dois campos selecionados. Essas regras foram formuladas com base nas diretrizes do PBLE, que estabelecem os requisitos e condições para cada uma das seis dimensões de qualidade adotadas neste estudo.

A Tabela 6.4 apresenta os campos selecionados e os critérios definidos para a dimensão **Compleitude**.

Tabela 6.4: Critérios da Dimensão Compleitude

Campo	Compleitude - Regras
NumFistel	O dado é obrigatório e não poderá haver valor ausente
DataMovimentoAno	O dado é obrigatório e não poderá haver valor ausente
DataMovimentoMes	O dado é obrigatório e não poderá haver valor ausente
CodINEP	O dado é obrigatório e não poderá haver valor ausente
NomeEscola	O dado é obrigatório e não poderá haver valor ausente
CodMunicipio	O dado é obrigatório e não poderá haver valor ausente
EnderecoEscola	O dado é obrigatório e não poderá haver valor ausente
BairroEscola	O dado é obrigatório e não poderá haver valor ausente
ComplementoEscola	<i>A dimensão não se aplica ao campo</i>
NumeroEscola	O dado é obrigatório e não poderá haver valor ausente
CEPEscola	O dado é obrigatório e não poderá haver valor ausente
IdtTipoEscola	O dado é obrigatório e não poderá haver valor ausente
codTipoObrigacao	O dado é obrigatório e não poderá haver valor ausente
IdtSituacao	O dado é obrigatório e não poderá haver valor ausente
DataAtivacao	O dado é obrigatório e não poderá haver valor ausente
VelAcessoObrigacao	O dado é obrigatório e não poderá haver valor ausente
VelAcessoInstalada	O dado é obrigatório e não poderá haver valor ausente
IP	O dado é obrigatório e não poderá haver valor ausente
IdtTecnologia	O dado é obrigatório e não poderá haver valor ausente
IdtTipoPendencia	O dado é obrigatório e não poderá haver valor ausente
DataPrevisao	O dado é obrigatório e não poderá haver valor ausente
DataDesativacao	<i>A dimensão não se aplica ao campo</i>

Para atender a essa dimensão, os vinte e dois campos selecionados devem conter valores preenchidos, ou seja, não podem apresentar ausência de dados. A presença de valores vazios ou nulos compromete diretamente a completude dos dados.

Exceções se aplicam apenas aos campos *ComplementoEscola* e *DataDesativacao*, cujo preenchimento não é obrigatório pelas prestadoras de serviços de telecomunicações, conforme estabelecido pelas regras do PBLE. Nesses casos, a ausência de dados é considerada justificável.

Na sequência, a Tabela 6.5 apresenta os critérios definidos para a dimensão de **Consistência**, que exige que os dados estejam livres de contradições ou discrepâncias em relação a informações provenientes de outras fontes confiáveis.

Tabela 6.5: Critérios da Dimensão Consistência

Campo	Consistência - Regras
NumFistel	O dado deverá ser o mesmo quando comparado com a base de dados da Anatel
DataMovimentoAno	O dado deverá ser o mesmo quando comparado com a base de dados da Anatel
DataMovimentoMes	O dado deverá ser o mesmo quando comparado com a base de dados da Anatel
CodINEP	O dado deverá ser o mesmo quando comparado com a base de dados do INEP
NomeEscola	O dado deverá ser o mesmo quando comparado com a base de dados do INEP
CodMunicipio	O dado deverá ser o mesmo quando comparado com a base de dados do IBGE
EnderecoEscola	O dado deverá ser o mesmo quando comparado com a base de dados do INEP
BairroEscola	O dado deverá ser o mesmo quando comparado com a base de dados do INEP
ComplementoEscola	O dado deverá ser o mesmo quando comparado com a base de dados do INEP
NumeroEscola	O dado deverá ser o mesmo quando comparado com a base de dados do INEP
CEPEscola	O dado deverá ser o mesmo quando comparado com a base de dados dos Correios
IdtTipoEscola	O dado deverá ser o mesmo quando comparado com a base de dados do INEP
codTipoObrigacao	O dado deverá ser o mesmo quando comparado com a base de dados do INEP
IdtSituacao	O dado deverá ser o mesmo quando comparado com a base de dados da Anatel
DataAtivacao	O dado deverá ser o mesmo quando comparado com a base de dados da Anatel
VelAcessoObrigacao	O dado deverá ser o mesmo quando comparado com a base de dados da Anatel
VelAcessoInstalada	O dado deverá ser o mesmo quando comparado com a base de dados da Anatel
IP	O dado deverá ser o mesmo quando comparado com a base de dados da Anatel
IdtTecnologia	O dado deverá ser o mesmo quando comparado com a base de dados da Anatel
IdtTipoPendencia	O dado deverá ser o mesmo quando comparado com a base de dados da Anatel
DataPrevisao	O dado deverá ser o mesmo quando comparado com a base de dados da Anatel
DataDesativacao	O dado deverá ser o mesmo quando comparado com a base de dados da Anatel

Para garantir a Consistência dos dados do PBLE, é fundamental que todos os dados estejam coerentes com outras fontes de dados externas utilizadas neste estudo.

A verificação da consistência considerou, portanto, as seguintes fontes externas para os campos selecionados:

- Base de Dados dos Correios: *CEPEscola*;
- Base de Dados do IBGE: *CodMunicipio*;
- Base de Dados do Censo Escolar do INEP: *CodINEP*, *NomeEscola*, *EnderecoEscola*, *BairroEscola*, *ComplementoEscola*, *NumeroEscola*, *IdtTipoEscola* e *codTipoObrigacao*.

Na sequência, a Tabela 6.6 apresenta os critérios estabelecidos para a dimensão **Conformidade**, a qual exige que os dados estejam em conformidade com padrões definidos, convenções estabelecidas e regras de negócio aplicáveis.

Tabela 6.6: Critérios da Dimensão Conformidade

Campo	Conformidade - Regras
NumFistel	Número Fistel válido, Tipo numérico e Tamanho 11
DataMovimentoAno	Ano válido, Tipo numérico e Tamanho 4
DataMovimentoMes	Mês válido, Tipo numérico e Tamanho 2
CodINEP	Código do INEP válido, Tipo numérico e Tamanho 8
NomeEscola	Tipo texto e Tamanho 255
CodMunicipio	Código do IBGE válido, Tipo numérico e Tamanho 7
EnderecoEscola	Tipo texto e Tamanho 255
BairroEscola	Tipo texto e Tamanho 255
ComplementoEscola	Tipo texto e Tamanho 255
NumeroEscola	Tipo numérico e Tamanho 6
CEPEscola	Código dos Correios válido, Tipo numérico e Tamanho 8
IdtTipoEscola	Tipo numérico e Tamanho 2
codTipoObrigacao	Tipo numérico e Tamanho 2
IdtSituacao	Tipo numérico e Tamanho 2
DataAtivacao	Tipo data e Tamanho 10
VelAcessoObrigacao	Tipo numérico e Tamanho 4
VelAcessoInstalada	Tipo numérico e Tamanho 4
IP	Tipo texto e Tamanho 9 a 15
IdtTecnologia	Tipo numérico e Tamanho 2
IdtTipoPendencia	Tipo numérico e Tamanho 2
DataPrevisao	Tipo data e Tamanho 7
DataDesativacao	Tipo data e Tamanho 10

Os critérios da dimensão Conformidade avaliam se os dados dos campos selecionados seguem padrões previamente definidos, como tipo, formato e tamanho.

Além disso, é essencial que os dados estejam em conformidade com as regras de negócio estabelecidas pelo Sistema de Coleta de Dados da Anatel para o envio das informações do PBLE.

Também é necessário que os dados estejam alinhados às normas aplicadas em bases externas utilizadas pela Anatel, como as do IBGE, Correios e INEP, garantindo uniformidade e compatibilidade entre as informações.

A Tabela 6.7 apresenta os critérios adotados neste estudo para a dimensão **Integridade**, que se refere à validade e à consistência dos relacionamentos entre entidades e registros de dados.

Tabela 6.7: Critérios da Dimensão Integridade

Campo	Integridade - Regras
NumFistel	O dado deverá ser completo e inalterado ao longo do tempo
DataMovimentoAno	<i>A dimensão não se aplica ao campo</i>
DataMovimentoMes	<i>A dimensão não se aplica ao campo</i>
CodINEP	O dado deverá ser completo e inalterado ao longo do tempo
NomeEscola	<i>A dimensão não se aplica ao campo</i>
CodMunicipio	O dado deverá ser completo e inalterado ao longo do tempo
EnderecoEscola	<i>A dimensão não se aplica ao campo</i>
BairroEscola	<i>A dimensão não se aplica ao campo</i>
ComplementoEscola	<i>A dimensão não se aplica ao campo</i>
NumeroEscola	<i>A dimensão não se aplica ao campo</i>
CEPEscola	<i>A dimensão não se aplica ao campo</i>
IdtTipoEscola	O dado deverá ser completo e inalterado ao longo do tempo
codTipoObrigacao	O dado deverá ser completo e inalterado ao longo do tempo
IdtSituacao	<i>A dimensão não se aplica ao campo</i>
DataAtivacao	O dado deverá ser completo e inalterado ao longo do tempo
VelAcessoObrigacao	<i>A dimensão não se aplica ao campo</i>
VelAcessoInstalada	<i>A dimensão não se aplica ao campo</i>
IP	<i>A dimensão não se aplica ao campo</i>
IdtTecnologia	<i>A dimensão não se aplica ao campo</i>
IdtTipoPendencia	<i>A dimensão não se aplica ao campo</i>
DataPrevisao	<i>A dimensão não se aplica ao campo</i>
DataDesativacao	O dado deverá ser completo e inalterado ao longo do tempo

Assegurar a integridade implica manter os dados completos, inalterados e coerentes, de modo a garantir sua precisão e confiabilidade, além de evitar conflitos com outros sistemas internos da Anatel.

Como destacado na Tabela 6.7, a dimensão Integridade não se aplica a alguns campos selecionados, devido à natureza específica desses dados, conforme detalhado a seguir:

- **Dados Temporais:** *DataMovimentoAno*, *DataMovimentoMes* e *DataPrevisao* podem ser alterados ao longo do tempo com os envios periódicos das prestadoras;
- **Dados da Escola:** *NomeEscola*, *EnderecoEscola*, *BairroEscola*, *NumeroEscola*, *ComplementoEscola*, e *CEPEscola* são dados passíveis de alteração, caso ocorra mudança na localização da escola;
- **Dados da Conexão:** *IdtSituacao*, *VelAcessoObrigacao*, *VelAcessoInstalada*, *IP*, *IdtTecnologia* e *IdtTipoPendencia* são dados que podem ser atualizados pelas prestadoras, devido à instalação da conexão ou à evolução tecnológica do processo de instalação ao longo do tempo.

A Tabela 6.8 apresenta os critérios definidos para a dimensão **Precisão**, que se refere ao grau em que os dados representam com exatidão a realidade.

Tabela 6.8: Critérios da Dimensão Precisão

Campo	Precisão - Regras
NumFistel	O dado deverá ter, obrigatoriamente, 11 dígitos
DataMovimentoAno	O dado deverá ter, obrigatoriamente, 4 dígitos
DataMovimentoMes	O dado deverá ter, obrigatoriamente, 2 dígitos
CodINEP	O dado deverá ter, obrigatoriamente, 8 dígitos
NomeEscola	<i>A dimensão não se aplica ao campo</i>
CodMunicipio	O dado deverá ter, obrigatoriamente, 7 dígitos
EnderecoEscola	<i>A dimensão não se aplica ao campo</i>
BairroEscola	<i>A dimensão não se aplica ao campo</i>
ComplementoEscola	<i>A dimensão não se aplica ao campo</i>
NumeroEscola	<i>A dimensão não se aplica ao campo</i>
CEPEscola	O dado deverá ter, obrigatoriamente, 7 dígitos
IdtTipoEscola	O dado deverá ter, obrigatoriamente, 2 dígitos
codTipoObrigacao	O dado deverá ter, obrigatoriamente, 2 dígitos
IdtSituacao	O dado deverá ter, obrigatoriamente, 2 dígitos
DataAtivacao	O dado deverá ter, obrigatoriamente, o formato de data com 10 dígitos
VelAcessoObrigacao	O dado deverá ter, obrigatoriamente, 4 dígitos
VelAcessoInstalada	O dado deverá ter, obrigatoriamente, 4 dígitos
IP	O dado deverá ter, obrigatoriamente, entre 9 e 15 dígitos
IdtTecnologia	O dado deverá ter, obrigatoriamente, 2 dígitos
IdtTipoPendencia	O dado deverá ter, obrigatoriamente, 2 dígitos
DataPrevisao	O dado deverá ter, obrigatoriamente, o formato de data com 7 dígitos
DataDesativacao	O dado deverá ter, obrigatoriamente, o formato de data com 10 dígitos

Para atender a essa dimensão, os dados devem estar corretos e refletir fielmente as informações do campo correspondente, além de apresentar o nível de detalhamento adequado conforme as exigências do PBLE.

Alguns campos, no entanto, não se enquadram nessa dimensão, conforme indicado na Tabela 6.8 e descrito a seguir:

- **Dados Textuais:** *NomeEscola*, *EnderecoEscola*, *ComplementoEscola*, *CEPEscola* e *NumeroEscola* são dados que podem ser modificados ao longo do tempo, caso haja alterações no nome ou na localização geográfica da escola atendida pelo PBLE.

Isso se deve ao fato de conterem informações que não requerem precisão ou cuja natureza impede a avaliação precisa desse aspecto.

Por fim, a Tabela 6.9 apresenta os critérios definidos para a dimensão **Atualidade**, que se refere à relevância e validade dos dados ao longo do tempo.

Tabela 6.9: Critérios da Dimensão Atualidade

Campo	Atualidade - Regras
NumFistel	O dado deverá ser o mais atual, conforme base de dados da Anatel
DataMovimentoAno	O dado deverá ser do ano mais atual
DataMovimentoMes	O dado deverá ser do mês mais atual
CodINEP	O dado deverá ser o mais atual, conforme base de dados do INEP
NomeEscola	O dado deverá ser o mais atual, conforme base de dados do INEP
CodMunicipio	O dado deverá ser o mais atual, conforme base de dados do IBGE
EnderecoEscola	O dado deverá ser o mais atual, conforme base de dados do INEP
BairroEscola	O dado deverá ser o mais atual, conforme base de dados do INEP
ComplementoEscola	O dado deverá ser o mais atual, conforme base de dados do INEP
NumeroEscola	O dado deverá ser o mais atual, conforme base de dados do INEP
CEPEscola	O dado deverá ser o mais atual, conforme base de dados do Correios
IdtTipoEscola	O dado deverá ser o mais atual, conforme base de dados do INEP
codTipoObrigacao	O dado deverá ser o mais atual, conforme base de dados do INEP
IdtSituacao	O dado deverá ser o mais atual, conforme base de dados da Anatel
DataAtivacao	O dado deverá ser o mais atual, conforme base de dados da Anatel
VelAcessoObrigacao	O dado deverá ser o mais atual, conforme base de dados da Anatel
VelAcessoInstalada	O dado deverá ser o mais atual, conforme base de dados da Anatel
IP	O dado deverá ser o mais atual, conforme base de dados da Anatel
IdtTecnologia	O dado deverá ser o mais atual, conforme base de dados da Anatel
IdtTipoPendencia	O dado deverá ser o mais atual, conforme base de dados da Anatel
DataPrevisao	O dado deverá ser o mais atual, conforme base de dados da Anatel
DataDesativacao	O dado deverá ser o mais atual, conforme base de dados da Anatel

Para atender a essa dimensão, os campos selecionados do PBLE devem conter informações atualizadas, em conformidade com a temporalidade e o contexto estabelecidos pelo programa.

Além disso, os dados devem refletir com precisão a situação vigente das informações fornecidas pelas prestadoras responsáveis pelo PBLE, assegurando sua publicação atualizada nos canais oficiais da Anatel, no Portal de Dados Abertos e no Portal do Governo Federal representem fielmente a realidade vigente.

6.5.5 Avaliação da Qualidade dos Dados do PBLE

Com base na identificação dos principais problemas de qualidade que afetam os campos selecionados do PBLE, bem como na definição das dimensões de qualidade mais adequadas aos objetivos deste estudo, torna-se possível avançar para a etapa de construção das métricas que irão subsidiar a avaliação da Qualidade dos Dados do programa.

Antes da aplicação dessas métricas, considera-se relevante a realização de uma análise preliminar exploratória em um subconjunto dos dados do PBLE. Essa análise inicial tem como objetivo verificar a aplicabilidade das regras estabelecidas, avaliar a consistência das métricas propostas e identificar possíveis ajustes antes da execução da avaliação em larga escala.

6.5.5.1 Análise Preliminar da Qualidade dos Dados

Inicialmente, foi estabelecida uma relação entre as dimensões de qualidade selecionadas e os problemas de qualidade identificados na base de dados do PBLE, considerando-se a granularidade e a descrição conceitual de cada dimensão, conforme sistematizado na Tabela 6.10.

Tabela 6.10: Relação entre Dimensões e Problemas de Qualidade do PBLE

Dimensão	Problema de Qualidade	Granularidade	Descrição
COMPLETUDE	P1	Registro	Os registros obrigatórios estão presentes
CONSISTÊNCIA	P2, P3 e P4	Conjunto de Dados	Os campos são idênticos a outras fontes de dados
CONFORMIDADE	P5	Campo	Os campos estão em conformidade com as regras de negócios
INTEGRIDADE	P6	Registro	Os registros estão completos e inalterados ao longo do tempo
PRECISÃO	P7	Registro	Os registros não possuem erros e inconsistências
ATUALIDADE	P8	Registro	Os registros refletem o estado mais atual dos dados

Na sequência, foi extraído um subconjunto de dados históricos do PBLE, correspondente aos últimos doze meses, com o objetivo de realizar uma análise exploratória inicial. Para essa etapa, foram selecionados quatro campos estratégicos: *CodINEP*, *IP*, *VelAcessoInstalada* e *IdtTecnologia*.

Esses campos permitem identificar as unidades escolares atendidas pelo PBLE, bem como as características da conexão instalada – especificamente a velocidade e a tecnologia utilizada pelas prestadoras de serviços de telecomunicações. Com base nesses campos, foram definidos três problemas de qualidade de dados a serem investigados na análise preliminar, com base na tipologia apresentada na Seção 6.5.3.

- **Problema 1:** Ausência de dados obrigatórios;
- **Problema 5:** Não conformidade com as regras de negócio estabelecidas pela Anatel;
e
- **Problema 7:** Presença de valores inadequados ou em formato inconsistente com o padrão definido.

Os resultados da análise preliminar da qualidade dos dados do PBLE, com base em um total de 882.938 registros, foram organizados e apresentados na Tabela 6.11.

Tabela 6.11: Conjunto de Dados Históricos do PBLE

Campo	Conj. de Dados	Dados Não Qualificados	Proporção
CodINEP	882.938	29.604	3,4%
IP	882.938	36.386	4,1%
VelAcessoInstalada	882.938	25.259	2,9%
IdtTecnologia	882.938	74.724	8,5%

A Tabela 6.11 mostra que, entre os quatro campos analisados, *IdtTecnologia* apresentou o maior percentual de dados não qualificados (8,5%), seguido por *IP* (4,1%) e *CodINEP* (3,4%). O campo *VelAcessoInstalada* teve o menor índice de problemas (2,9%). Mesmo com igual volume de dados, observam-se variações na qualidade entre os campos.

As Figuras 6.4, e 6.5 detalham a distribuição dos problemas identificados por campo.

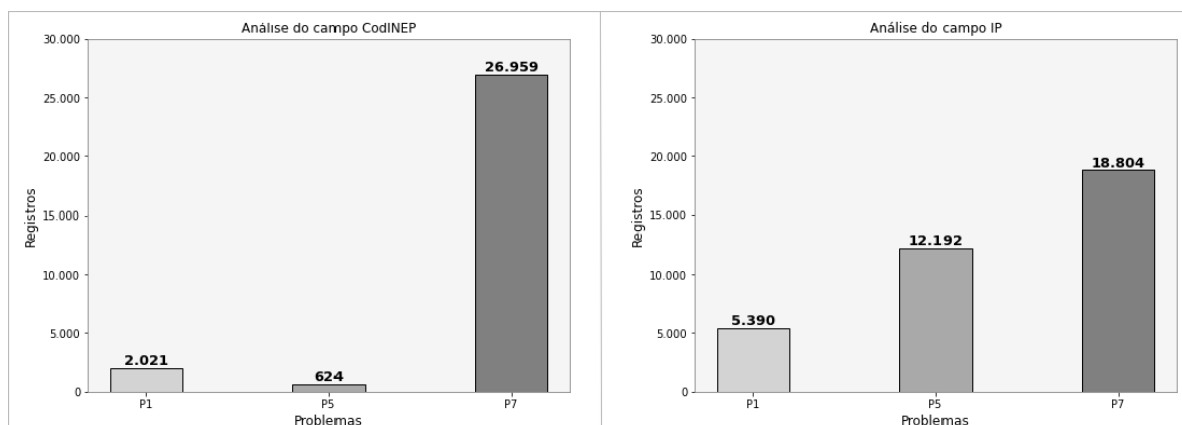


Figura 6.4: Resultado dos campos *CodINEP* e *IP*

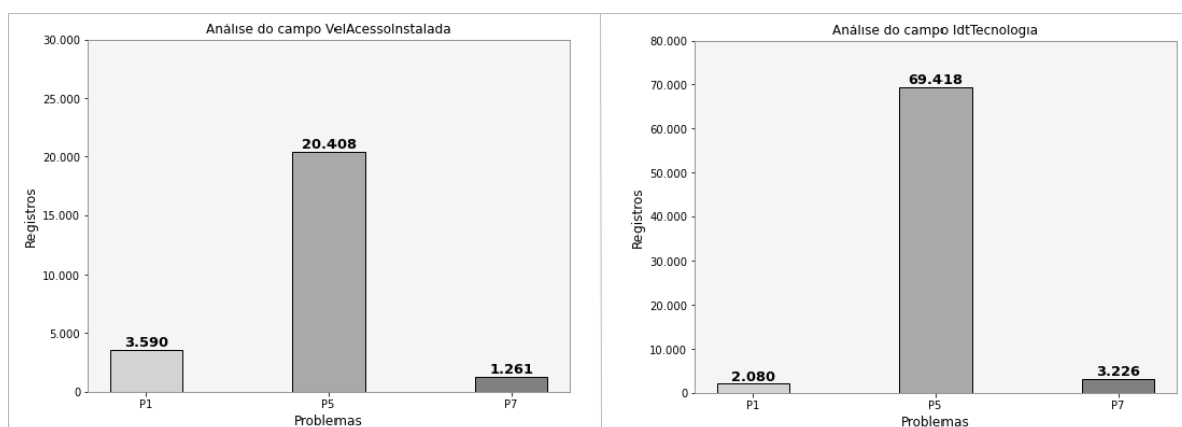


Figura 6.5: Resultado dos campos *VelAcessoInstalada* e *IdtTecnologia*

A Figura 6.4 exibe os resultados para os campos *CodINEP* e *IP*. Observa-se que o maior volume de registros não qualificados para *CodINEP* está associado ao Problema 7 (valores inadequados ou em formato inconsistente), com 26.959 ocorrências. O mesmo padrão se repete no campo *IP*, onde temos 18.804 registros com o Problema 7.

Já a Figura 6.5 apresenta os resultados referentes aos campos *VelAcessoInstalada* e *IdtTecnologia*. Nesse caso, destaca-se a predominância do Problema 5 (não conformidade com as regras de negócio), com 20.408 registros em *VelAcessoInstalada* e 69.418 registros em *IdtTecnologia*. Esses números indicam que as principais falhas de qualidade desses dois campos estão relacionadas à violação das regras de negócio estabelecidas.

A análise desse subconjunto de dados revelou a existência de registros com problemas de qualidade, evidenciando a ocorrência de dados incompletos, inconsistentes ou fora do padrão especificado. Essa constatação reforça a importância da adoção de mecanismos sistemáticos para monitoramento e tratamento da qualidade dos dados no contexto do PBLE.

Conclui-se, portanto, que mesmo uma análise restrita a um número limitado de campos e registros já é suficiente para evidenciar a presença de problemas de qualidade nos dados do PBLE. Esse diagnóstico inicial demonstra a necessidade de aprofundamento da análise por meio de métricas específicas e bem definidas, que possibilitem uma avaliação mais robusta e precisa da qualidade dos dados em todo o conjunto informacional do programa.

6.5.5.2 Métricas de Avaliação da Qualidade de Dados

Após a análise preliminar dos problemas de qualidade identificados nos dados do PBLE, iniciou-se uma etapa fundamental desta pesquisa: a construção das métricas para a Avaliação da Qualidade dos Dados do programa.

A definição dessas métricas baseou-se nos indicadores associados a cada uma das dimensões de qualidade de dados previamente selecionadas, buscando garantir uma avaliação sistemática, objetiva e alinhada aos objetivos da pesquisa.

Além disso, para garantir sua efetividade, as métricas foram desenvolvidas com base em três premissas essenciais [67]:

1. **Mensuração:** As métricas devem ser expressas em valores normalizados, preferencialmente escalonados em intervalos contínuos ou discretos;
2. **Interpretação:** As métricas devem apresentar clareza conceitual e ser de fácil compreensão, permitindo análises objetivas; e
3. **Agregação:** As métricas devem permitir a avaliação da qualidade dos dados em diferentes níveis de granularidade, tais como campo (coluna), registro (linha), conjunto de dados (arquivo) e base de dados como um todo.

Com base nesses princípios, foram elaborados seis indicadores que compõem o modelo de avaliação proposto. A Tabela 6.12 apresenta os seis indicadores, associadas às respectivas dimensões de qualidade de dados, juntamente com suas métricas, fórmulas de cálculo e escalas de interpretação.

Tabela 6.12: Métricas dos Indicadores de Avaliação de Qualidade

Indicador	Métrica	Fórmula	Escala
COMPLETUDE DOS REGISTROS	Média da relação entre a qtade. de registros obrigatórios disponíveis e a qtade. total de registros	$\text{Comp} = \frac{1}{n} \sum_{i=1}^n \left(\frac{R_i^{\text{ObrigDisp}}}{R_i^{\text{Total}}} \right)$, onde: $R^{\text{ObrigDisp}}$ é nº de regs. obrig. disp. e R^{Total} é nº total de registros	$[0, 1]$
CONSISTÊNCIA DOS REGISTROS	Média da relação entre a qtade. de registros idênticos a outras bases de dados e a qtade. total de registros	$\text{Cons} = \frac{1}{n} \sum_{i=1}^n \left(\frac{R_i^{\text{IgualOtBases}}}{R_i^{\text{Total}}} \right)$, onde: $R^{\text{IgualOtBases}}$ é nº de regs. iguais e R^{Total} é nº total de registros	$[0, 1]$
CONFORMIDADE DOS REGISTROS	Média da relação entre a qtade. de registros em conformidade com as regras de negócio e a qtade. total de registros	$\text{Conf} = \frac{1}{n} \sum_{i=1}^n \left(\frac{R_i^{\text{ConfRegNeg}}}{R_i^{\text{Total}}} \right)$, onde: $R^{\text{ConfRegNeg}}$ é nº de regs. em conform. e R^{Total} é nº total de registros	$[0, 1]$
INTEGRIDADE DOS REGISTROS	Média da relação entre a qtade. de registros inalterados ao longo do tempo e a qtade. total de registros	$\text{Intg} = \frac{1}{n} \sum_{i=1}^n \left(\frac{R_i^{\text{Inalterados}}}{R_i^{\text{Total}}} \right)$, onde: $R^{\text{Inalterados}}$ é nº de regs. inalterados e R^{Total} é nº total de registros	$[0, 1]$
PRECISÃO DOS REGISTROS	Média da relação entre a qtad. de registros sem erros ou inconsistências e a qtade. total de registros	$\text{Prec} = \frac{1}{n} \sum_{i=1}^n \left(\frac{R_i^{\text{SemErro}}}{R_i^{\text{Total}}} \right)$, onde: R^{SemErro} é nº de regs. sem erros/incons. e R^{Total} é nº total de registros	$[0, 1]$
ATUALIDADE DOS REGISTROS	Média da relação entre a qtad. de registros no estado mais atual e a qtade. total de registros	$\text{Atlz} = \frac{1}{n} \sum_{i=1}^n \left(\frac{R_i^{\text{Atuais}}}{R_i^{\text{Total}}} \right)$, onde: R^{Atuais} é nº de regs. mais atuais e R^{Total} é nº total de registros	$[0, 1]$

O primeiro indicador de qualidade de dados do PBLE, denominado **Compleitude**, é calculado com base na proporção média de registros obrigatórios disponíveis em cada campo selecionado, em relação ao total de registros do respectivo campo. A métrica

consiste em apurar, mensalmente, a razão entre a quantidade de registros válidos e o número total de registros para cada campo, no período de fevereiro de 2013 a junho de 2024. Em seguida, calcula-se a média dessas razões ao longo de todo o intervalo analisado.

Essa abordagem permite uma visão consolidada do grau de completude dos dados, viabilizando a identificação de padrões e possibilitando uma avaliação objetiva da qualidade da informação disponibilizada pelas prestadoras.

Os demais cinco indicadores – **Consistência, Conformidade, Integridade, Precisão e Atualidade** – seguem uma estrutura metodológica semelhante, com métricas específicas para cada dimensão da qualidade, conforme definido na Tabela 6.12. Cada métrica é aplicada mensalmente, e os resultados são agregados ao longo do tempo, permitindo comparações e análises evolutivas.

A aplicação sistemática e padronizada dessas métricas tem como objetivo proporcionar uma avaliação prática, mensurável e comparável da qualidade dos dados do PBLE. Essa avaliação contribui diretamente para o fortalecimento das práticas de fiscalização e da governança de dados no âmbito da Anatel, além de subsidiar a tomada de decisões baseada em evidências.

6.5.5.3 Avaliação dos Dados do PBLE

Concluída a etapa de definição das métricas associadas aos indicadores de qualidade, deu-se início ao processo de aplicação sistemática dessas métricas, com o objetivo de avaliar a qualidade dos vinte e dois campos de dados do PBLE selecionados neste estudo. Essa fase foi conduzida de forma criteriosa, com base nas seis dimensões de qualidade previamente estabelecidas: Completude, Consistência, Conformidade, Integridade, Precisão e Atualidade.

A aplicação das métricas permitiu mensurar, de forma quantitativa, o desempenho de cada campo em relação a essas dimensões, possibilitando a identificação de padrões, tendências e inconsistências ao longo do tempo. Esse procedimento resultou em uma análise estruturada, objetiva e comparativa, fundamental para compreender o nível de qualidade dos dados gerados pelas prestadoras de serviços de telecomunicações e coletados pela Anatel.

Os resultados dessa avaliação estão apresentados de forma consolidada na Tabela 6.13, que reúne os valores apurados para cada um dos indicadores de qualidade, individualmente por campo de dado analisado. Essa sistematização dos resultados oferece uma base sólida para diagnósticos técnicos, subsidiando tanto a identificação de fragilidades quanto a proposição de ações corretivas e preventivas voltadas ao aprimoramento da qualidade dos dados do PBLE.

Adicionalmente, essa avaliação contribui para o fortalecimento da governança de dados no âmbito da Anatel, ao fornecer evidências empíricas que podem orientar decisões estratégicas e operacionais relacionadas às atividades de fiscalização, acompanhamento e controle, transparência e publicação dos dados.

Tabela 6.13: Avaliação dos Campos do PBLE

Campo	Compleitude	Consistência	Conformidade	Integridade	Precisão	Atualidade
NumFistel	1	1	1	1	1	1
DataMovimentoAno	1	1	1	N/A	1	1
DataMovimentoMes	1	1	1	N/A	1	1
CodINEP	0,99	0,99	0,96	0,98	0,56	0,97
NomeEscola	1	0,99	0,99	N/A	N/A	0,75
CodMunicipio	1	1	1	0,99	0,99	0,97
EnderecoEscola	1	0,98	0,99	N/A	N/A	0,80
BairroEscola	0,99	0,94	0,91	N/A	N/A	0,83
ComplementoEscola	N/A	0,47	0,90	N/A	N/A	0,82
NumeroEscola	0,99	0,40	0,41	N/A	N/A	0,81
CEPEscola	0,99	0,99	0,99	N/A	0,91	0,89
IdtTipoEscola	1	0,99	0,99	0,98	1	1
codTipoObrigacao	1	0,72	0,76	1	0,75	1
IdtSituacao	1	1	1	N/A	1	1
DataAtivacao	0,99	0,99	0,99	0,62	0,73	0,88
VelAcessoObrigacao	0,98	0,98	0,96	N/A	0,69	0,82
VelAcessoInstalada	0,99	0,99	0,99	N/A	0,51	0,79
IP	0,99	0,99	0,99	N/A	0,63	0,99
IdtTecnologia	1	1	1	N/A	0,69	1
IdtTipoPendencia	0,99	0,99	0,99	N/A	0,99	0,99
DataPrevisao	0,86	0,86	0,86	N/A	0,66	0,99
DataDesativacao	N/A	0,08	0,08	0,09	0,21	0,85

N/A: Não Aplicável.

6.6 Diagnóstico dos Dados do PBLE

Com a conclusão das etapas iniciais da metodologia adotada nesta pesquisa – que incluiu a identificação e seleção dos dados do PBLE, a identificação dos principais problemas de qualidade, a escolha criteriosa das dimensões de qualidade de dados mais adequadas ao escopo do estudo e, por fim, a aplicação das métricas definidas para avaliação – tornou-se possível elaborar um diagnóstico detalhado sobre a qualidade dos dados do PBLE.

Esse diagnóstico tem como objetivo central fornecer uma visão abrangente e fundamentada sobre o estado atual da qualidade dos dados mantidos pela Anatel, no contexto específico do PBLE. A análise baseou-se em um volume expressivo de informações: ao todo, foram examinados 12.731.899 registros enviados pelas prestadoras de serviços de

telecomunicações, no período de fevereiro de 2013 a junho de 2024, os quais estão armazenados em um SGBD interno da Agência, utilizado especificamente para esse fim.

A elaboração do diagnóstico fundamentou-se nos resultados obtidos a partir das métricas de avaliação desenvolvidas neste estudo, cujos valores estão consolidados na Tabela 6.13. Essa tabela apresenta, de forma clara e estruturada, os níveis de qualidade dos dados por campo, considerando seis indicadores: Completude, Consistência, Conformidade, Integridade, Precisão e Atualidade. Cada indicador foi aplicado a 22 campos selecionados com base em sua relevância operacional e aderência às regras de negócio do programa.

A análise dos indicadores possibilitou a identificação de campos que apresentaram alto desempenho em termos de qualidade de dados, evidenciando completude elevada, aderência às regras de negócio, consistência entre fontes e formatos válidos e atualizados.

Por outro lado, também foi possível identificar campos com desempenho inferior, que concentram a maior parte dos problemas detectados – tais como ausência de dados obrigatórios, inconsistências com bases externas, formatos incorretos ou desatualização das informações.

Com base nesse conjunto de informações, o diagnóstico elaborado oferece subsídios para o aprimoramento dos processos de coleta, validação e integração dos dados no âmbito do PBLE. Além disso, destaca os pontos críticos que demandam atenção prioritária por parte dos gestores de dados e das áreas técnicas da Anatel, com o objetivo de promover a melhoria contínua da qualidade dos dados utilizados nas ações fiscalizatórias e nas atividades de acompanhamento e controle do programa.

A seção subsequente apresenta os principais achados do diagnóstico da qualidade dos dados do PBLE. A análise evidencia os campos que obtiveram os melhores e os piores desempenhos em cada uma das dimensões de qualidade avaliadas, conforme resultados gerados a partir da aplicação das métricas definidas neste estudo. Adicionalmente, discutem-se os impactos da baixa qualidade dos dados sobre a efetividade das ações do programa.

6.6.1 Campos com Alta Qualidade de Dados

A análise dos dados da Tabela 6.13 permitiu identificar quatro campos do PBLE – *DataMovimentoAno*, *DataMovimentoMes*, *NumFistel* e *IdtSituacao* – que se destacam por apresentar excelente desempenho em todos os seis indicadores de qualidade avaliados: Completude, Consistência, Conformidade, Integridade, Precisão e Atualidade.

A performance consistente desses campos demonstra a robustez e a confiabilidade das informações, tornando-os adequados para os objetivos de monitoramento e análise do PBLE. Essa alta qualidade é essencial para a eficácia das ações de fiscalização e para a tomada de decisões da Anatel no acompanhamento das prestadoras que atendem às escolas por meio do programa.

As Figuras 6.6, 6.7, 6.8 e 6.9 ilustram graficamente os resultados individuais desses campos, apresentando a distribuição dos valores obtidos em cada dimensão de qualidade. Esses gráficos facilitam a visualização dos dados e reforçam a compreensão dos bons resultados observados.

A Figura 6.6, em particular, demonstra o desempenho do campo *DataMovimentoAno*, que atingiu pontuação máxima nas dimensões aplicáveis. O valor zero na dimensão Integridade indica que essa métrica não se aplica ao campo.

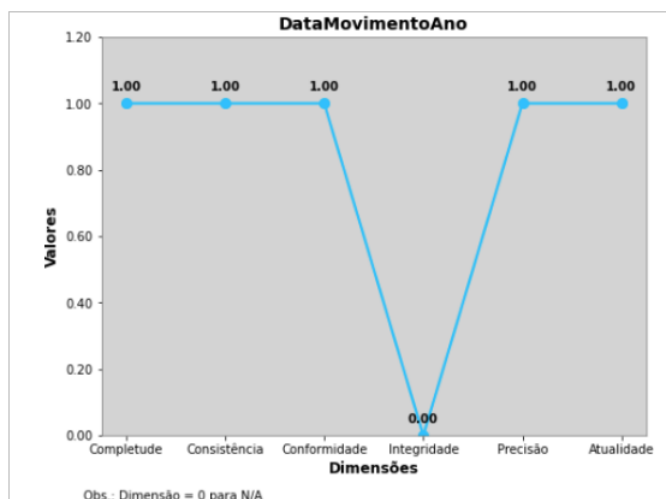


Figura 6.6: *DataMovimentoAno* - Resultado da Avaliação da Qualidade

A Figura 6.7 mostra que o campo *DataMovimentoMes* também alcançou desempenho máximo nas dimensões pertinentes, sendo a pontuação zero em Integridade justificada por sua não aplicabilidade.

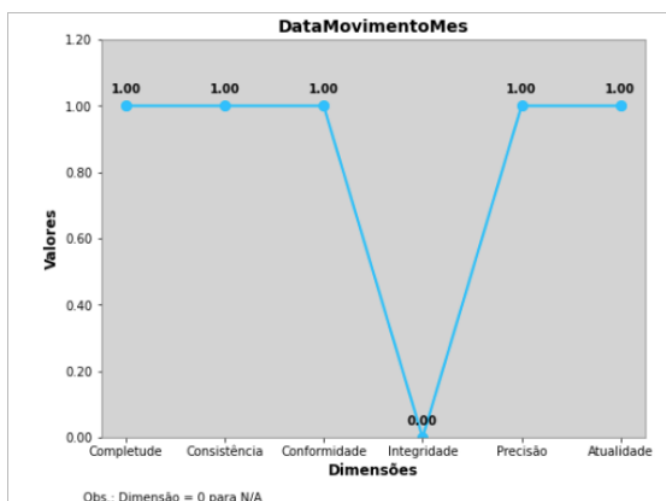


Figura 6.7: *DataMovimentoMes* - Resultado da Avaliação da Qualidade

Na Figura 6.8, o campo *NumFistel* apresenta nota máxima (1.00) em todas as dimensões de qualidade consideradas, evidenciando dados completos, consistentes, conformes, íntegros, precisos e atualizados.

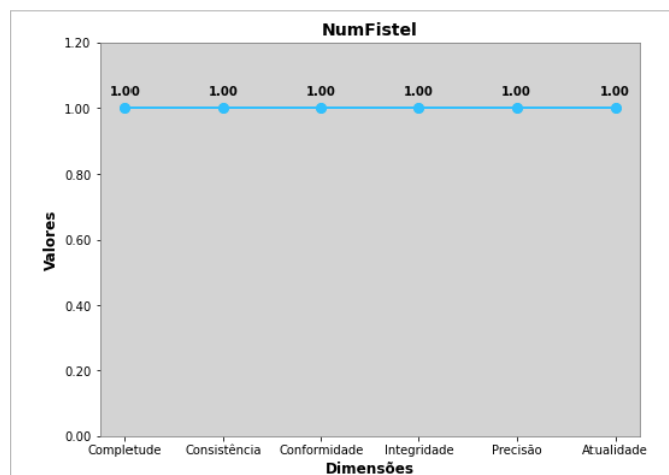


Figura 6.8: *NumFistel* - Resultado da Avaliação da Qualidade

Por fim, a Figura 6.9 revela que o campo *IdtSituacao* obteve desempenho máximo nas dimensões aplicáveis. Assim como nos casos anteriores, o valor zero em Integridade reflete apenas a não aplicabilidade dessa dimensão ao campo avaliado.

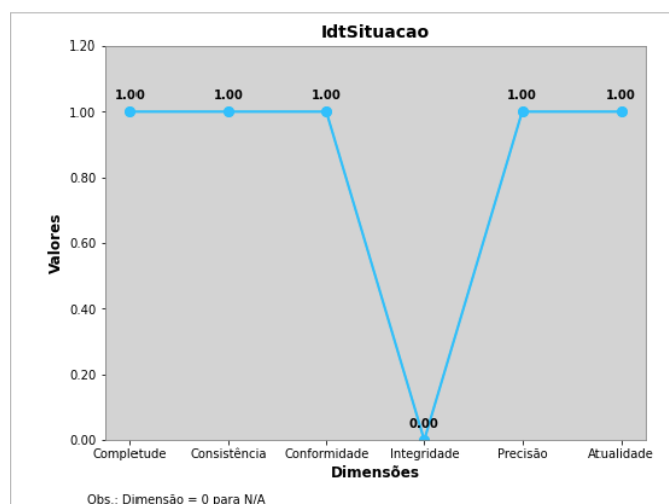


Figura 6.9: *IdtSituacao* - Resultado da Avaliação da Qualidade

Adicionalmente, destaca-se que o indicador de Conformidade apresentou desempenho bastante satisfatório, com valores superiores a 0,86 em todos os campos analisados. Esse resultado demonstra que os dados estão alinhados às regras de negócio e às exigências regulatórias do PBLE, aspecto fundamental para a integridade do processo de fiscalização

e para o monitoramento eficaz das obrigações das prestadoras de serviços de telecomunicações no programa.

6.6.2 Campos com Baixa Qualidade de Dados

Um dos objetivos desta pesquisa foi identificar os campos do PBLE que apresentam baixa qualidade de dados, de modo a embasar o desenvolvimento de ações corretivas. A correta classificação destes campos é fundamental para aprimorar a confiabilidade, a consistência e a utilidade dos dados gerenciados pela Anatel, garantindo suporte adequado às atividades de monitoramento e tomada de decisão.

Ao examinar os resultados apresentados na Tabela 6.13, verificou-se que os campos *NumeroEscola*, *DataPrevisao* e *codTipoObrigacao* registraram desempenhos insatisfatórios em diversas dimensões de qualidade – em particular nas métricas de Consistência, Conformidade, Integridade e Precisão. Tais deficiências refletem a ocorrência de disparidades nos valores esperados, o descumprimento de regras de negócio, a presença de entradas inválidas e eventuais falhas de preenchimento.

As Figuras 6.10, 6.11 e 6.12 apresentam o desempenho individual de cada campo nas diversas dimensões de qualidade, evidenciando de forma clara os pontos críticos que demandam ação corretiva.

Na Figura 6.10, o campo *NumeroEscola* registrou valores muito baixos em Consistência (0,40) e Conformidade (0,41), refletindo deficiências na padronização e no atendimento às regras de negócio. Os valores nulos em Integridade e Precisão indicam a não aplicabilidade dessas métricas a esse campo.

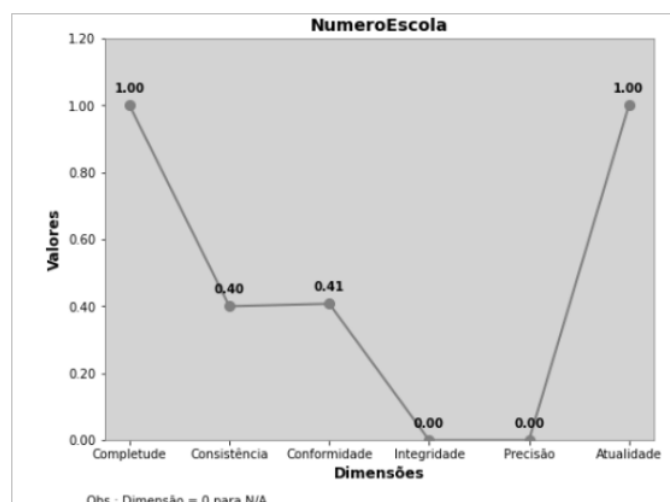


Figura 6.10: *NumeroEscola* - Resultado da Avaliação da Qualidade

Conforme ilustrado na Figura 6.11, o campo *DataPrevisao* apresentou baixo valor em Precisão (0,66), o que indica incerteza na exatidão dos dados. A dimensão Integridade, com valor zero, não se aplica à natureza do campo.

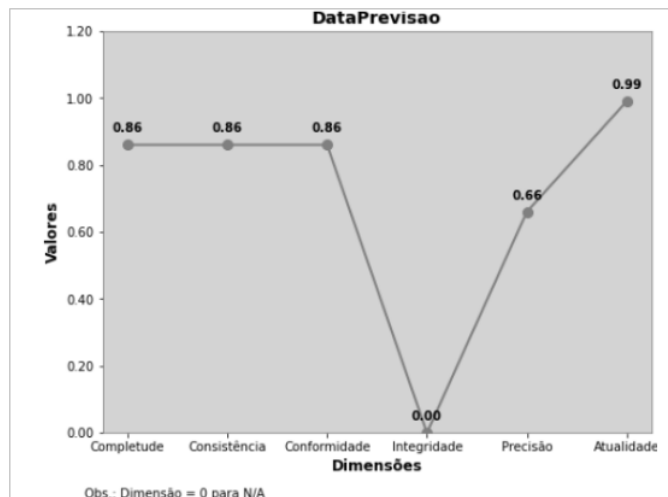


Figura 6.11: *DataPrevisao* - Resultado da Avaliação da Qualidade

A Figura 6.12 revela que o campo *codTipoObrigacao* apresenta fragilidades nas dimensões de Consistência (0,72), Conformidade (0,76) e Precisão (0,75), indicando necessidade de revisão nos controles de regras de negócios e nas verificações de exatidão dos valores.

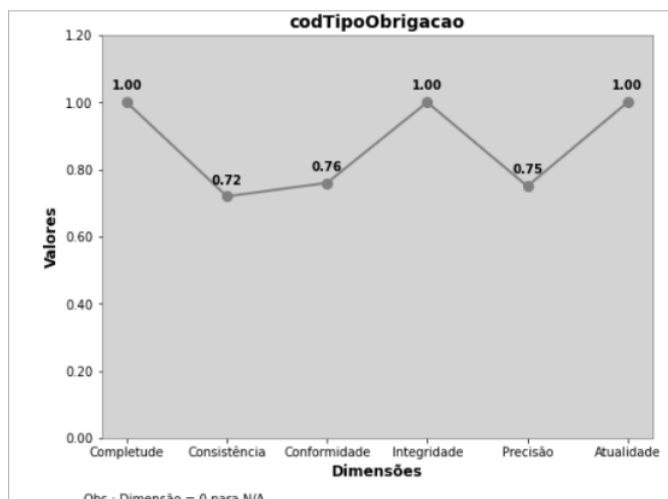


Figura 6.12: *codTipoObrigacao* - Resultado da Avaliação da Qualidade

Adicionalmente, o campo *DataDesativacao* destacou-se negativamente, conforme ilustra a Figura 6.13. As dimensões Consistência (0,08), Conformidade (0,08), Integridade (0,09) e Precisão (0,21) atingiram níveis críticos, refletindo baixa confiabilidade dos dados. O valor zero indica que a dimensão Completude não é aplicável ao campo.

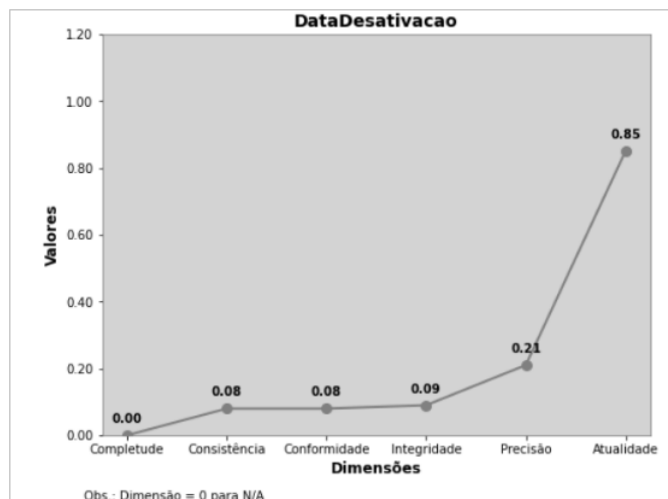


Figura 6.13: *DataDesativacao* - Resultado da Avaliação da Qualidade

Observou-se também uma tendência preocupante nos indicadores de Precisão e Atualidade, em que nove e dez campos, respectivamente, apresentaram resultados aquém do esperado. Esse padrão evidencia a presença de informações imprecisas ou desatualizadas, o que compromete a tomada de decisão e o monitoramento eficaz do programa.

A baixa qualidade dos dados verificada em determinados campos impacta diretamente a confiabilidade das informações do PBLE divulgadas pela Anatel e no Portal de Dados Abertos do Governo Federal. Esse cenário evidencia a necessidade de adoção de mecanismos contínuos de monitoramento, controle e aprimoramento da qualidade dos dados coletados. Tais ações são fundamentais para assegurar a integridade, a utilidade e a fidedignidade das informações disponibilizadas à sociedade e aos órgãos de controle.

6.7 Resultado da Avaliação da Qualidade dos Dados

A qualidade dos dados constitui um fator determinante para sua utilidade e eficácia em processos analíticos gerenciais e de suporte à tomada de decisão. Assim, a avaliação sistemática dessa qualidade assume papel fundamental no gerenciamento e na disponibilização de informações, assegurando o atendimento a requisitos e padrões preestabelecidos para finalidades que englobam monitoramento, fiscalização, controle e planejamento estratégico.

A partir dos resultados da avaliação e do diagnóstico realizados para os dados do PBLE, observou-se que uma parte significativa dos dados apresenta baixos níveis de qualidade. Esse cenário compromete os processos de acompanhamento e controle, além de prejudicar o planejamento e a execução das ações fiscalizatórias essenciais para o cumprimento das normas do programa. Ressalta-se que os dados do PBLE constituem a principal base de informações para essas atividades regulatórias conduzidas pela Anatel.

Além disso, a baixa qualidade dos dados impacta diretamente a confiabilidade das informações publicadas no Portal da Anatel, amplamente acessadas e utilizadas por diversos órgãos da Administração Pública, empresas privadas e organizações não governamentais. Essas informações são de grande relevância, especialmente para as iniciativas voltadas à promoção da conectividade e inclusão digital em escolas públicas urbanas em todo o país, uma área de interesse estratégico tanto para o governo quanto para a sociedade.

Outro ponto crítico identificado foi o sistema de coleta de dados do PBLE, que serve como o principal canal de entrada das informações na Anatel. Foi constatado que esse sistema apresenta limitações em relação à qualidade dos dados coletados. Não há mecanismos que permitam a verificação prévia da qualidade dos dados no momento da coleta, antes de seu armazenamento nas bases de dados da Agência. Isso contribui para a inclusão de dados imprecisos ou incompletos, comprometendo a qualidade das informações.

Diante desse cenário, o estudo propõe uma solução para aprimorar o processo de coleta de dados do PBLE. A implementação de melhorias nesse processo visa não apenas elevar a qualidade dos dados coletados, mas também garantir maior confiabilidade e consistência nas informações gerenciadas pela Anatel. Essa abordagem é fundamental para que o programa desempenhe seu papel de forma mais eficiente e para que os dados publicados pela Agência possam ser utilizados com maior precisão e transparência.

6.8 Solução Proposta

A partir dos resultados sobre a Qualidade de Dados do PBLE, foram elaboradas três propostas de solução com potencial para aprimorar a gestão e o controle da qualidade dos dados no âmbito do programa: **Regras de Qualidade de Dados, Modelos de Auditoria de Dados e Relatórios de Auditoria de Dados.**

As Regras de Qualidade de Dados consistem em critérios técnicos e lógicos aplicáveis diretamente nos sistemas de coleta, com o intuito de evitar a inserção de registros incorretos, incompletos ou fora dos padrões esperados. A implementação dessas regras no momento da entrada dos dados nos sistemas da Anatel permitirá garantir maior conformidade com os requisitos de qualidade definidos, atuando preventivamente na redução de erros e aumentando a confiabilidade da base de dados do PBLE.

Os Modelos de Auditoria de Dados, por sua vez, consistem na representação das relações entre as tabelas que compõem o modelo de dados. Essa abordagem permite a auditoria do modelo atual, com a identificação de fragilidades estruturais, incoerências nos relacionamentos e lacunas informacionais, além de viabilizar a proposição de melhorias que contribuam para a integridade, consistência e eficiência no uso dos dados.

Por fim, os Relatórios de Auditoria de Dados são produtos resultantes do processo de verificação da qualidade das bases de dados, reunindo análises quantitativas e qualitativas, além da descrição das principais inconsistências identificadas nos campos mais críticos. Esses documentos subsidiam a tomada de decisão pelos gestores de dados, orientando ações de correção e aprimoramento do modelo de dados, além de contribuírem para o fortalecimento da governança de dados.

6.8.1 Regras de Qualidade de Dados

O modelo de Regras de Qualidade proposto para o PBLE foi desenvolvido em três partes, levando em consideração as especificidades de cada campo do programa selecionado nesta pesquisa, assim como sua aderência ao modelo de negócios e às diretrizes necessárias para atender aos objetivos do programa.

As partes do modelo proposto para as Regras de Qualidade de Dados do PBLE são descritas a seguir:

1. **Regras de Qualidade - Dados de Identificação de Escolas:** Esta categoria abrange os campos destinados a caracterizar as escolas atendidas pelo PBLE. Seu objetivo é garantir a qualidade das informações relacionadas a cada unidade escolar, assegurando que os dados reflitam corretamente as informações das escolas beneficiadas;
2. **Regras de Qualidade - Dados de Controle de Instalação:** Esta seção inclui os campos relacionados ao acompanhamento e controle das obrigações das prestadoras. Por meio desses dados, é possível verificar a execução e o cumprimento dos compromissos relativos à instalação da conexão, conforme os parâmetros definidos pelas regras do PBLE; e
3. **Regras de Qualidade - Dados Técnicos de Telecomunicações:** Refere-se aos campos que contêm informações técnicas sobre a infraestrutura de telecomunicações envolvida no PBLE. Esses dados são essenciais para a correta análise, interpretação e avaliação do programa, pois fornecem informações fundamentais para entender o contexto técnico das instalações no âmbito do programa.

As Tabelas 6.14, 6.15 e 6.16 apresentam de forma detalhada cada uma das partes da proposta das Regras de Qualidade dos Dados do PBLE, organizando os critérios e diretrizes específicos para cada categoria de dados. Essas tabelas permitem uma compreensão clara das regras necessárias para assegurar a qualidade e integridade das informações do programa.

Tabela 6.14: Regras de Qualidade - Dados de Identificação de Escolas

Campo	Descrição	Tipo (Tamanho)	Obrigatório	Regra de Qualidade
CodINEP	Código INEP de até 10 dígitos referentes ao código único da Escola	Texto com limitação de caracteres (10)	SIM	Verificar se CodINEP corresponde a um valor que está na base de dados do INEP
NomeEscola	Nome da Escola	Texto com limitação de caracteres (255)	SIM	-
EnderecoEscola	Endereço de localização da Escola	Texto com limitação de caracteres (500)	SIM	-
NumeroEscola	Número do endereço da Escola	Número inteiro (8)	Não	-
BairroEscola	Bairro do endereço da Escola	Texto com limitação de caracteres (255)	SIM	-
ComplementoEscola	Complemento do endereço da Escola	Texto com limitação de caracteres (500)	Não	-
CEPEscola	Código de Endereçamento Postal da Escola com exatamente 8 dígitos, incluindo zeros à esquerda e sem traço	Texto com limitação de caracteres (8)	SIM	Verificar se CEPEscola corresponde a um valor que está na base de dados dos Correios
CodMunicipio	Código IBGE de identificação do Município, com 7 dígitos, onde a Escola está localizada	Número inteiro (7)	SIM	Verificar se CodMunicipio corresponde a um valor que está na base de dados do IBGE
IdtTipoEscola	Código do Tipo de Escola	Número inteiro (2)	SIM	Verificar se IdtTipoEscola corresponde a um dos valores válidos: 01 - Federal 02 - Estadual 03 - Municipal 11 - UCA (Um Computador por Aluno) 12 - UAB (Universidade Aberta do Brasil) 13 - Sistema S 91 - INTRAGOV 92 - PRODAM

Tabela 6.15: Regras de Qualidade - Dados de Controle de Instalação

Campo	Descrição	Tipo (Tamanho)	Obrigatório	Regra de Qualidade
DataMovimentoAno	Ano referente aos dados com 4 dígitos numéricos	Número inteiro (4)	SIM	Verificar se DataMovimentoAno está no formato: AAAA
DataMovimentoMes	Mês referente aos dados com 2 dígitos numéricos	Número inteiro (2)	SIM	Verificar se DataMovimentoMes está no formato: MM
codTipoObrigacao	Código do Tipo de de Obrigação para a conexão da Escola	Número inteiro (1)	SIM	Verificar se codTipoObrigacao corresponde a um dos valores válidos: 0 - PBLE; 1 - Rural - Edital 4G
IdtSituacao	Tipo de Situação da conexão da Escola com 2 dígitos	Texto com limitação de caracteres (2)	SIM	Verificar se IdtSituacao corresponde a um dos valores válidos: 01 - Instalada 02 - Pendência da Escola 03 - Pendência da prestadora 05 - Previsão 09 - Desinstalada
DataAtivacao	Data de ativação da conexão da Escola	Data (10)	Não	Verificar se DataAtivacao está no formato DD/MM/AAAA; Verificar se DataAtivacao é superior a abril/2008 e igual ou inferior aos campos DataMovimentoAno e DataMovimentoAno
VelAcessoInstalada	Velocidade (em MB) instalada na conexão da Escola com uma casa decimal	Número decimal (5,1)	Não	Verificar se o campo IdtSituacao está preenchido com os valores 1 ou 9
VelAcessoObrigacao	Velocidade (em MB) obrigatória na conexão da Escola com uma casa decimal	Número decimal (5,1)	SIM	Verificar se VelAcessoObrigacao está preenchido com valor maior ou igual a 2,0
IdtTipoPendencia	Código do Tipo de Pendência de instalação da Escola	Número inteiro (2)	Não	Verificar se IdtTipoPendencia corresponde a um dos valores válidos: 00 - Sem Impedimentos; 10 - Em planejamento; 12 - Disponibilidade técnica; 13 - Escola sem energia; 50 - Divergência de informações da Escola; 51 - Escola desativada; 52 - Escola não localizada; 53 - Escola não elegível ao PBLE; 56 - Escola retirada do PBLE; 57 - Escola em reforma; 66 - Escola inacessível
DataPrevisao	Data de previsão de conexão da Escola	Data (7)	Não	Verificar se o campo IdtSituacao está preenchido com valor 5; Verificar se DataPrevisao está no formato MM/AAAA; Verificar se DataPrevisao é superior aos campos DataMovimentoAno e DataMovimentoAno
DataDesativacao	Data de desativação da conexão da Escola	Data (10)	Não	Verificar se DataDesativacao está no formato DD/MM/AAAA; Verificar se DataDesativacao é superior a abril/2008 e igual ou inferior aos campos DataMovimentoAno e DataMovimentoAno

Tabela 6.16: Regras de Qualidade - Dados Técnicos de Telecomunicações

Campo	Descrição	Tipo (Tamanho)	Obrigatório	Regra de Qualidade
NumFistel	Número do Fistel ativo na Anatel e correspondente à identificação da prestadora	Número inteiro (11)	SIM	Verificar se NumFistel corresponde a um valor que está na base de dados da Anatel
IP	Endereço de IP referente à conexão da Escola	Texto com limitação de caracteres (15)	Não	Verificar se IP está no formato IPv4: XXX.XXX.XXX.XXX
IdtTecnologia	Código do tipo de tecnologia de conexão da Escola	Número inteiro (2)	Não	Verificar se IdtTecnologia corresponde a um dos valores válidos: 01 - Cabo Coaxial; 02 - Fibra; 03 - Cabo Metálico; 04 - Rádio; 05 - Satélite

6.8.2 Modelos de Auditoria de Dados

6.8.2.1 Modelo de Auditoria do SICI

O primeiro modelo desenvolvido nesta pesquisa é o **Modelo de Auditoria do SICI**, estruturado com base no modelo relacional da base de dados do sistema SICI, conforme ilustrado na Figura 6.14.

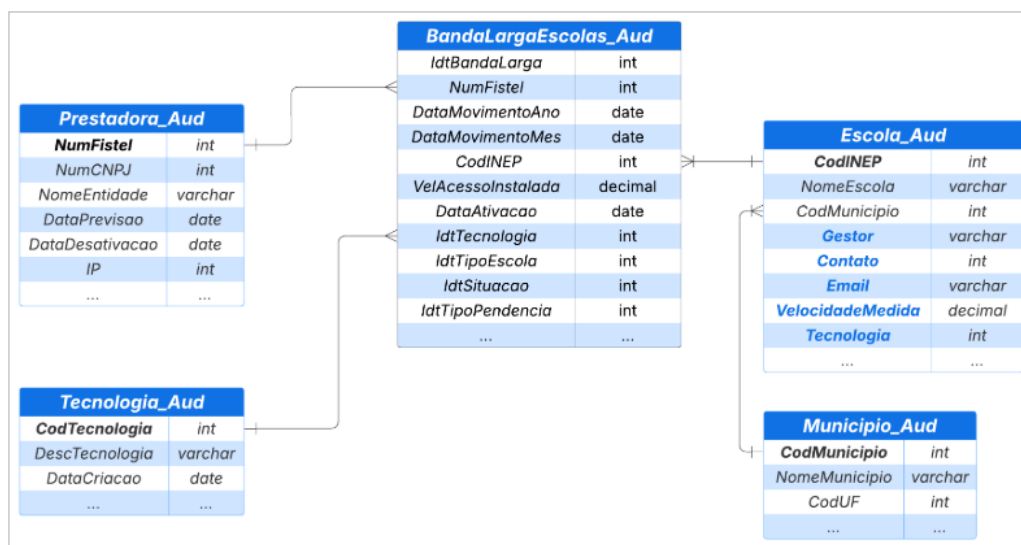


Figura 6.14: Modelo de Auditoria do Sistema SICI

Esse modelo é composto por cinco tabelas, cujas relações estabelecem uma estrutura integrada e coesa. A tabela central, **BandaLargaEscolas__Aud**, é responsável por consolidar os registros das escolas e desempenha um papel central na organização dos dados do programa. Essa configuração relacional possibilita a rastreabilidade dos dados informados pelas prestadoras. Dentre os principais relacionamentos, destacam-se:

- A relação com a tabela **Prestadora__Aud**, por meio do campo *NumFistel*, que identifica a prestadora responsável pela conexão;
- O relacionamento com a tabela **Escola__Aud**, via o campo *CodINEP*, que associa os dados técnicos às informações cadastrais das escolas;
- A ligação com a tabela **Tecnologia__Aud**, por meio do campo *IdtTecnologia*, que descreve a tecnologia utilizada no acesso; e
- A conexão com a tabela **Municipio__Aud**, a partir da **Escola__Aud**, permitindo identificar a localização da unidade escolar.

Além disso, a **BandaLargaEscolas__Aud** reúne campos críticos para a avaliação da qualidade dos dados, como a velocidade e a situação da conexão, que viabilizam o monitoramento do programa. O *script* de criação da tabela encontra-se no Anexo I.

6.8.2.2 Modelo de Auditoria dos Sistemas Internos

O segundo modelo de auditoria é apresentado na Figura 6.15.

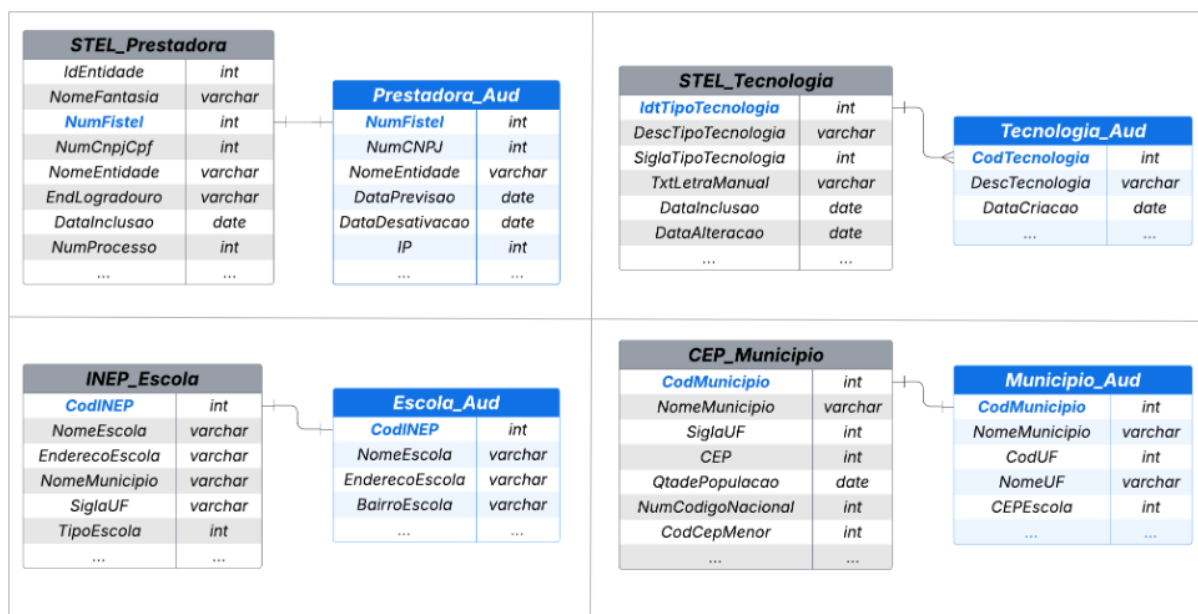


Figura 6.15: Modelo de Auditoria do Sistemas Internos

Esse modelo elaborado, denominado **Modelo de Auditoria dos Sistemas Internos**, corresponde a um modelo relacional que integra as tabelas das bases de dados dos sistemas internos da Anatel – STEL, INEP e CEP (em cinza) – que recebem, das prestadoras, os mesmos dados enviados ao sistema SICI (em azul).

A tabela **STEL_Prestadora** reúne informações provenientes do sistema STEL, responsável por armazenar dados cadastrais e regulatórios das prestadoras de serviços de telecomunicações. Essa tabela contém diversos campos que também são registrados no sistema SICI, como o *NumFistel*. Esse campo identifica a prestadora responsável pela conexão da escola atendida pelo PBLE e permite validar a consistência entre os dados declarados nos diferentes sistemas da Anatel.

De forma análoga, a tabela **INEP_Escola** consolida dados do sistema INEP. Utilizando o campo *CodINEP* como identificador único, é possível correlacionar os registros do SICI com os dados oficiais do INEP. A tabela inclui ainda atributos como nome da escola, endereço e município, permitindo avaliar a exatidão das informações fornecidas pelas prestadoras em comparação aos registros do sistema educacional.

Com o objetivo de viabilizar a auditoria entre sistemas, foram criadas tabelas com o sufixo “_Aud”, permitindo a comparação entre os dados declarados no SICI e aqueles recebidos por sistemas internos distintos da Anatel. Essa abordagem possibilita avaliar a exatidão e a coerência das informações entre múltiplas fontes.

Os *scripts* de criação das quatro tabelas que compõem o Modelo de Auditoria dos Sistemas Internos estão disponibilizados no Anexo I desta dissertação.

6.8.2.3 Modelo de Auditoria de Fiscalização e das Escolas

E, por fim, o terceiro é o **Modelo de Auditoria de Fiscalização e das Escolas** é apresentado pela Figura 6.16.

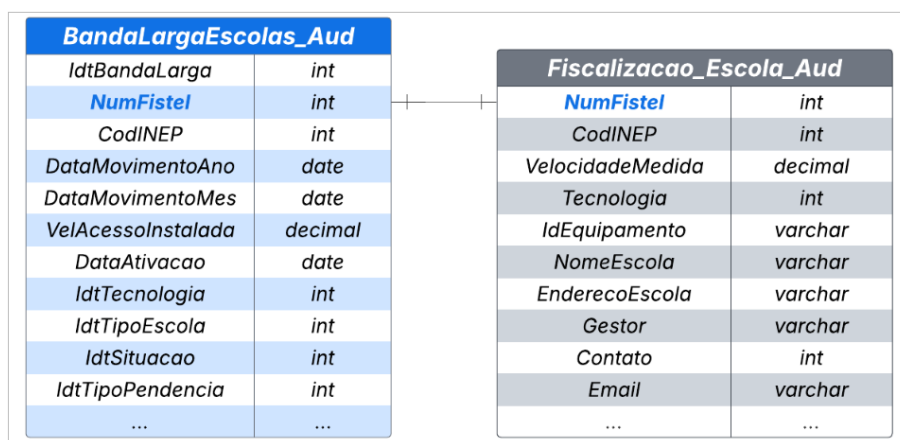


Figura 6.16: Modelo de Auditoria de Fiscalização e das Escolas

O modelo proposto foi concebido com o objetivo de viabilizar a coleta estruturada de dados durante o processo de inspeção *in loco*, conduzido pela equipe técnica da Anatel durante a fiscalização do PBLE. Nessa etapa, em que o agente da Agência se encontra fisicamente na escola, abre-se uma oportunidade para a obtenção de informações técnicas diretamente no local, assegurando maior precisão e confiabilidade aos dados coletados.

Adicionalmente, nesse mesmo contexto, o diretor ou responsável pela unidade escolar visitada poderá, de forma voluntária, fornecer informações institucionais complementares. Essa participação colaborativa contribui para o enriquecimento do processo de coleta, ampliando a base de dados do PBLE com informações relevantes e atualizadas, provenientes diretamente das instituições beneficiárias do programa.

Dessa forma, o **Modelo de Auditoria de Fiscalização e das Escolas**, ilustrado na Figura 6.16, foi desenvolvido com o propósito de consolidar informações técnicas e descritivas coletadas durante as atividades de fiscalização presencial nas unidades escolares. A estrutura relacional é representada pela tabela **Fiscalizacao_Escola_Aud** (destacada em azul).

A estrutura dessa tabela contempla dois grupos distintos de atributos. O primeiro refere-se às informações de natureza técnica, coletadas diretamente pelos fiscais durante as visitas às escolas. Entre esses campos, destacam-se *NumFistel*, *CodINEP*, *VelocidadeMedida* e *Tecnologia*, que são essenciais para validar a infraestrutura de conectividade instalada, conforme informações coletadas pelo sistema SICI.

O segundo grupo de atributos refere-se a informações descritivas da unidade escolar, as quais são geralmente fornecidas pelo responsável local. Incluem-se nesse conjunto os campos *NomeEscola*, *EnderecoEscola*, *Gestor*, *Contato* e *Email*, que contribuem para contextualizar o ambiente escolar em que a conectividade está sendo ofertada.

A incorporação desses dois grupos de dados – técnicos e descritivos – em um mesmo modelo permite uma avaliação mais ampla e qualificada da situação de conectividade nas escolas, contribuindo para a verificação da integridade, precisão e consistência dos dados declarados pelas prestadoras no sistema SICI.

O *script* de criação da tabela destinada ao armazenamento dessas informações encontra-se disponível no Anexo I desta dissertação, possibilitando a reprodução e aplicação prática do modelo proposto nos processos de auditoria e fiscalização da Anatel.

6.8.3 Relatório de Auditoria de Dados

Com a elaboração dos Modelos de Auditoria de Dados propostos nesta pesquisa, viabiliza-se a construção de três **Relatórios de Auditoria de Dados** distintos e complementares, cada um voltado para a análise de uma fonte específica de informação dos dados do PBLE. Esses relatórios têm o objetivo de consolidar os resultados das auditorias a partir

dos modelos desenvolvidos, permitindo uma avaliação mais abrangente e detalhada da qualidade dos dados do programa.

6.8.3.1 Relatório de Auditoria de Dados do SICI

O Relatório de Auditoria de Dados do SICI tem como objetivo principal avaliar a qualidade dos dados provenientes diretamente do sistema SICI, responsável por coletar as informações declaradas pelas prestadoras encarregadas de prover conexão de banda larga às escolas públicas no âmbito do PBLE.

A elaboração do relatório foi baseada nos resultados obtidos durante o processo de avaliação da qualidade dos dados do PBLE, conforme detalhado na Seção 6.5.5.3 deste trabalho. Essa avaliação forneceu insumos essenciais para identificar os campos mais críticos e orientar a estruturação do relatório, destacando os principais pontos de atenção.

Para aprimorar a apresentação dos resultados e facilitar a análise, utilizou-se a ferramenta Qlik Sense – plataforma de visualização de dados licenciada pela Anatel. A solução viabilizou a criação de *dashboards* interativos, que permitem a exploração das informações e a geração de *insights* para melhorar a qualidade dos dados coletados no SICI.

A Figura 6.17 apresenta a tela do *dashboard* desenvolvido, com visão consolidada do Relatório de Auditoria de Dados do SICI. A interface destaca principais indicadores de qualidade e campos mais críticos, facilitando a identificação de inconsistências nos dados analisados.

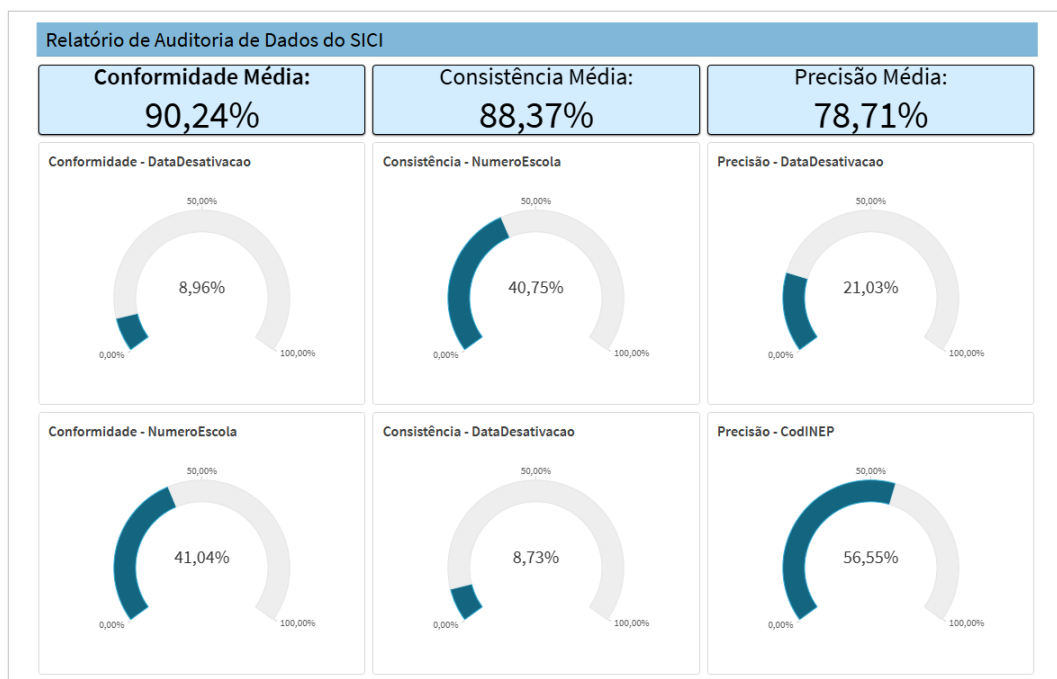


Figura 6.17: Relatório de Auditoria de Dados do SICI

6.8.3.2 Relatório de Auditoria de Dados dos Sistemas Internos

O segundo relatório, intitulado Relatório de Auditoria de Dados dos Sistemas Internos, fundamenta-se nos dados registrados em sistemas institucionais da Anatel, como o STEL, o INEP e o CEP. Esses sistemas recebem dados das mesmas prestadoras que alimentam o SICI, o que permite realizar comparações diretas entre as bases. Essa abordagem viabiliza uma análise do grau de exatidão entre diferentes fontes institucionais, contribuindo para a identificação de inconsistências decorrentes de eventuais divergências nos registros.

Com o objetivo de facilitar a compreensão do processo de elaboração de relatório de auditoria, foram gerados relatórios individualizados para cada sistema analisado. Essa separação possibilita uma visualização mais clara das particularidades de cada base de dados, bem como das discrepâncias observadas. Os *scripts* completos utilizados na geração dos três relatórios encontram-se disponíveis no Anexo III desta dissertação.

Tabela 6.17: Relatório de Auditoria de Dados do Sistema STEL

Campo	Regs. Base SICI	Diverg. Base SICI x Base STEL	Perc. Divergência
NumFistel	12.731.899	0	0%
CodTecnologia	12.731.899	228.606	1,80%

A Tabela 6.17 apresenta o relatório de auditoria de dados, comparando as bases dos sistemas SICI e STEL, especificamente para os campos *NumFistel* e *CodTecnologia*. O objetivo desta análise foi identificar eventuais divergências entre as duas bases, considerando o volume de registros auditados.

Para o campo *NumFistel*, observa-se que, dos 12.731.899 registros auditados na base do SICI, não foram encontradas divergências em comparação à base do sistema STEL. Assim, o percentual de divergência para este campo foi de 0%, indicando plena consistência entre os sistemas no que se refere à identificação das prestadoras.

Por outro lado, no campo *CodTecnologia*, foram identificadas 228.606 divergências entre as bases, em um universo igualmente composto por 12.731.899 registros. Esse número corresponde a um percentual de divergência de aproximadamente 1,80%. Esse resultado evidencia a existência de inconsistências específicas no atributo relacionado à tecnologia de acesso informada pelas prestadoras.

Tabela 6.18: Relatório de Auditoria de Dados do Sistema INEP

Campo	Regs. Base SICI	Diverg. Base SICI x Base INEP	Perc. Divergência
CodINEP	12.731.899	6.460.844	50,75%
NomeEscola	12.731.899	7.476.192	58,72%

Com base na análise dos dados apresentados na Tabela 6.18, foram comparadas as bases do sistema SICI e do sistema INEP, com foco nos campos *CodINEP* e *NomeEscola*. A auditoria foi realizada sobre o total de registros provenientes de ambas as fontes, com o objetivo de avaliar a consistência das informações registradas nos dois sistemas.

Em relação ao campo *CodINEP*, foram identificados 6.460.844 registros divergentes entre as duas bases, em um universo de 12.731.899 registros auditados. O percentual de divergência para este campo é de aproximadamente 50,75%, evidenciando uma significativa discrepância na correspondência do código de identificação das escolas entre os sistemas SICI e INEP.

No campo *NomeEscola*, a divergência foi ainda mais expressiva, com 7.476.192 registros inconsistentes, o que corresponde a 58,72% do total de registros auditados. Essa inconsistência reforça a necessidade de revisão e padronização das informações referentes às escolas, a fim de assegurar a precisão dos dados entre os dois sistemas.

Esses resultados indicam a existência de áreas críticas que requerem atenção especial, particularmente no que diz respeito à identificação das escolas e seus respectivos códigos. O elevado percentual de divergência encontrado sugere a necessidade de uma investigação mais aprofundada para identificar as possíveis causas dessas discrepâncias.

Tabela 6.19: Relatório de Auditoria de Dados do Sistema CEP

Campo	Regs. Base SICI	Diverg. Base SICI x Base CEP	Perc. Divergência
CodMunicipio	12.731.899	0	0%
CEPEscola	12.731.899	2.752.560	21,62%

Com base nos dados apresentados na Tabela 6.19, foi realizada uma auditoria comparativa entre os registros das bases dos sistemas SICI e CEP, com o objetivo de avaliar a consistência das informações relacionadas aos campos *CodMunicipio* e *CEPEscola*.

No campo *CodMunicipio*, foram analisados 12.731.899 registros, não sendo identificada nenhuma divergência entre as bases. Esse resultado indica total conformidade e integridade desse campo específico entre os dois sistemas, resultando em um percentual de divergência de 0%.

Por outro lado, o campo *CEPEscola* apresentou 2.752.560 registros divergentes em relação ao total de 12.731.899 registros avaliados, o que representa um percentual de divergência de aproximadamente 21,62%. Esse índice evidencia uma inconsistência significativa nos dados de CEP das escolas, sugerindo a necessidade de intervenções para a correção e padronização desses valores.

A elevada taxa de divergência identificada no campo *CEPEscola* reforça a importância da implementação de mecanismos sistemáticos de verificação e validação dos dados no sis-

tema SICI, especialmente em campos críticos para a localização geográfica e identificação das unidades escolares vinculadas ao PBLE.

6.8.3.3 Relatório de Auditoria de Fiscalização e das Escolas

O Relatório de Auditoria de Fiscalização e das Escolas consolida os dados obtidos durante as inspeções presenciais conduzidas por fiscais da Anatel, bem como as informações fornecidas diretamente pelos gestores das escolas visitadas de forma voluntária. Esse relatório permitirá uma verificação empírica da realidade observada nas unidades escolares, fortalecendo o processo de validação dos dados e contribuindo significativamente para a credibilidade e a confiabilidade das informações utilizadas no PBLE.

A elaboração desse relatório configura-se como uma proposta técnica a ser encaminhada à área de Tecnologia da Informação da Anatel, com o objetivo de viabilizar a implementação de uma nova forma de coleta de dados. Essa coleta seria realizada durante as inspeções presenciais conduzidas por fiscais da Agência, complementada pelas informações fornecidas, de forma voluntária, pelos gestores das escolas visitadas.

A proposta visa integrar os dados técnicos coletados *in loco* com as informações administrativas fornecidas pelas unidades escolares, ampliando a abrangência e a precisão das informações reunidas. A adoção desse relatório permitirá uma verificação empírica da realidade observada nas escolas, possibilitando a comparação entre os dados declarados e aqueles efetivamente verificados no local.

Essa abordagem fortalece o processo de validação dos dados utilizados pelo PBLE e contribui diretamente para a melhoria da qualidade das informações, promovendo maior confiabilidade e transparência na gestão do programa. Além disso, os dados coletados poderão subsidiar ações corretivas e aprimoramentos no processo de fiscalização e monitoramento da conectividade nas escolas públicas.

Como não há dados disponíveis oriundos das inspeções presenciais, apresenta-se, na Tabela 6.20, a estrutura sugerida para o Relatório de Auditoria de Fiscalização e das Escolas. Essa estrutura contempla os principais campos a serem coletados e serve como base para o desenvolvimento futuro da solução, possibilitando a padronização dos dados e a posterior análise de sua qualidade no contexto do PBLE.

Tabela 6.20: Relatório de Auditoria de Dados de Fiscalização e Escolas

Campo	Regs. Base SICI	Diverg. Base SICI x Coleta Fisc./Escolas	Perc. Divergência
NumFistel	12.731.899	-	-
CodINEP	12.731.899	-	-
Tecnologia	12.731.899	-	-
Velocidade	12.731.899	-	-

6.9 Análise de Riscos

Embora os resultados desta pesquisa indiquem avanços significativos no diagnóstico e na melhoria da Qualidade de Dados no âmbito do PBLE, sua efetiva aplicação na prática institucional da Anatel dependerá da aceitação e posterior implementação pela área de Tecnologia da Informação TI. Essa etapa representa um ponto crítico no ciclo de vida da proposta desenvolvida, e, por isso, exige atenção especial quanto aos riscos envolvidos.

O principal risco identificado é a resistência institucional à mudança, especialmente quando se trata da adoção de novas metodologias que alteram processos estabelecidos. Embora esta dissertação utilize metodologias e ferramentas amplamente consolidadas na literatura científica, seu uso direcionado à avaliação da qualidade dos dados ainda é inédito na Anatel. Assim, há o risco de que a área de TI priorize outras demandas estratégicas ou questione a viabilidade técnica sem uma análise dos benefícios demonstrados.

Outro risco relevante é subestimar a Qualidade de Dados como vetor de governança e eficiência regulatória. Caso os resultados desta pesquisa sejam vistos apenas como exercício acadêmico, e não como uma ferramenta estratégica para aprimorar a fiscalização e as políticas públicas, perde-se a oportunidade de institucionalizar práticas que garantam a melhoria dos dados da Agência, mantendo fragilidades informacionais já diagnosticadas.

Há, ainda, o risco técnico relacionado à necessidade de integração da metodologia proposta aos sistemas legados da Anatel, o que poderá demandar adaptações tecnológicas, investimentos em capacitação técnica ou ajustes em fluxos de trabalho. Esses fatores podem ser interpretados como entraves operacionais, caso não haja envolvimento colaborativo entre as áreas técnicas e as lideranças institucionais.

Assim, mitigar esses riscos exigirá uma abordagem propositiva e articulada com os objetivos estratégicos da Anatel. Para tanto, os resultados desta pesquisa serão apresentados às áreas envolvidas, destacando os ganhos práticos em termos de eficiência fiscalizatória e alocação racional de recursos. A incorporação da proposta contribuirá não apenas para a melhoria da qualidade dos dados do PBLE, mas também para a maturidade e uso de boas práticas de governança de dados na Agência.

A Tabela 6.21 apresenta a Matriz de Impactos de Riscos construída a partir da análise realizada, sintetizando os riscos identificados, suas respectivas probabilidades e impactos, bem como as estratégias propostas para sua mitigação de forma clara e sistematizada.

Tabela 6.21: Matriz de Impactos de Riscos

Risco	Impacto	Probabilidade	Severidade	Plano de Mitigação
Resistência institucional à mudança	Rejeição da adoção da metodologia proposta ou atrasos na implementação	ALTA	ALTA	Apresentar as evidências científicas e práticas que demonstrem os benefícios da proposta; e Sensibilizar e capacitar as equipes envolvidas.
Subvalorização da Qualidade de Dados	Rejeição ou adoção incompleta da metodologia proposta	ALTA	ALTA	Reforçar a importância estratégica da qualidade de dados; e Alinhar a proposta aos objetivos estratégicos da Anatel.
Integração com sistemas legados da Anatel	Custos adicionais, retrabalho ou falhas técnicas, dificultando a implementação da metodologia	Média	ALTA	Mapear os sistemas legados; Definir um plano de integração adequado; e Garantir a participação da área de TI na fase de planejamento.
Falta de recursos para implementação	Comprometimento da execução da proposta devido à alocação inadequada de orçamento e pessoal.	Média	Média	Elaborar um plano detalhado de alocação de recursos, priorizando a execução da metodologia como uma melhoria estratégica de governança de dados.
Falta de envolvimento das áreas-chave (TI, fiscalização, Gestão)	Dificuldade na integração das áreas e adoção de um processo fragmentado, prejudicando a eficácia da metodologia proposta.	ALTA	ALTA	Realizar reuniões de alinhamento e de integração com todas as áreas-chave para garantir a adesão à proposta.

Capítulo 7

Conclusão e Trabalhos Futuros

Este capítulo apresenta as conclusões obtidas a partir da avaliação da qualidade dos dados do PBLE, destacando as principais contribuições e os impactos desta pesquisa para a Anatel. Além disso, são indicados os próximos passos para o aprimoramento e a evolução da metodologia proposta, com vistas à sua aplicação em estudos futuros e ao aperfeiçoamento contínuo da gestão da qualidade e da governança de dados.

7.1 Conclusão

O número de publicações relacionadas ao tema da Qualidade de Dados tem apresentado crescimento significativo nos últimos anos, refletindo a crescente relevância dessa área em diversos domínios do conhecimento, como educação, telecomunicações, letramento digital, *Big Data* e Inteligência Artificial.

Nesse contexto, a Qualidade de Dados consolidou-se como um elemento essencial para organizações públicas e privadas, especialmente diante do expressivo volume de dados que é gerado, processado e consumido diariamente por cidadãos e usuários de redes digitais. Garantir a qualidade dessas informações tornou-se indispensável para a tomada de decisões baseadas em evidências e para a formulação de políticas públicas mais eficazes.

Assim, o estudo sobre esse tema assume caráter estratégico, sobretudo para órgãos governamentais como a Anatel, diante da crescente demanda da sociedade por informações confiáveis e da pressão exercida sobre os órgãos de controle quanto à disponibilização e publicação de dados governamentais de qualidade, em conformidade com as diretrizes de dados abertos definidas por organismos internacionais.

Esta pesquisa concentrou-se na análise da Governança de Dados e da Qualidade de Dados. Para isso, foram examinados mais de 12 milhões de registros do Programa Banda Larga nas Escolas (PBLE), compostos por dados enviados pelas prestadoras de serviços de telecomunicações à Anatel, no período de fevereiro de 2013 a junho de 2024.

A partir da criação e aplicação de uma metodologia estruturada em cinco etapas – Identificação dos Dados, Seleção dos Dados, Identificação dos Problemas de Qualidade, Seleção das Dimensões de Qualidade e Avaliação da Qualidade dos Dados – foi possível alcançar o objetivo geral da pesquisa, além de atender aos cinco objetivos específicos estabelecidos.

Como contribuição institucional, este estudo representa um avanço relevante para a Anatel na área de Qualidade de Dados, uma vez que, até então, a Agência não dispunha de estudos sistemáticos voltados à avaliação da qualidade dos dados recebidos das prestadoras de serviços de telecomunicações.

Adicionalmente, propôs-se um modelo descritivo para a implementação de regras de qualidade no sistema de coleta de dados da Anatel, com o objetivo de aprimorar as informações relacionadas ao PBLE, promovendo uma gestão mais eficiente e alinhada às boas práticas de qualidade de dados.

Consequentemente, a pesquisa também oferece subsídios para o aprimoramento das atividades de fiscalização, acompanhamento e controle do cumprimento das obrigações de instalação e universalização da banda larga nas escolas públicas, com vistas à conclusão dessas ações até o final do programa.

Por fim, o modelo desenvolvido neste estudo poderá servir como referência estratégica para a construção e implementação de políticas de qualidade de dados em outros sistemas corporativos da Anatel, proporcionando suporte eficaz às suas atividades regulatórias no setor de telecomunicações no Brasil.

7.2 Contribuições

Esta dissertação representa contribuições relevantes em três dimensões: tecnológica, inovadora e regulatória. No aspecto tecnológico, propõe o uso de metodologias consolidadas da Ciência de Dados para avaliar a qualidade dos dados do PBLE. No campo da inovação, promove o fortalecimento de uma cultura institucional orientada à Qualidade de Dados. Já na dimensão regulatória, reforça a capacidade da Agência em conduzir uma fiscalização mais eficaz, precisa e alinhada às diretrizes normativas.

7.2.1 Contribuição Tecnológica

Esta pesquisa representa uma contribuição tecnológica significativa para o avanço científico na área de Ciência de Dados aplicada à Anatel, ao propor a utilização de metodologias e ferramentas amplamente reconhecidas no meio acadêmico, mas ainda pouco exploradas no contexto institucional da Agência. A proposta visa à avaliação da qualidade dos dados do PBLE e ao fortalecimento da governança de dados em suas práticas regulatórias.

Embora existam diversas metodologias disponíveis para a avaliação da qualidade de dados, sua aplicação integrada e orientada à governança ainda não foi incorporada aos processos operacionais da Anatel. Essa lacuna evidenciou a necessidade de desenvolver uma metodologia adaptada à realidade do PBLE, utilizando tecnologias já existentes de forma inovadora e com foco na melhoria da efetividade das políticas públicas educacionais.

A inovação tecnológica desta dissertação reside na forma como as metodologias e técnicas conhecidas foram organizadas, sistematizadas e aplicadas de maneira inovadora em um processo institucional de avaliação da qualidade dos dados, proporcionando maior rigor, rastreabilidade e padronização às análises.

Mais do que atender a uma demanda específica do PBLE, a metodologia proposta neste estudo estabelece as bases para a adoção de uma abordagem replicável em outros sistemas da Anatel, representando um avanço expressivo na maturidade dos processos internos e contribuindo diretamente para a transformação digital da Agência, sob a perspectiva estratégica da governança de dados.

7.2.2 Contribuição de Inovação

Esta pesquisa se destaca por sua contribuição inovadora ao introduzir, no âmbito da Anatel, uma abordagem sistematizada e estratégica voltada à Qualidade de Dados – tema ainda incipiente na cultura organizacional da Agência. Ao investigar a qualidade dos dados do PBLE e propor uma metodologia de avaliação alinhada às melhores práticas da Ciência de Dados, o estudo transcende sua aplicação pontual e atua como catalisador de uma mudança cultural necessária.

A adoção da metodologia proposta tem o potencial de transformar o modo como a Anatel trata seus dados, ao promover maior controle de qualidade nos fluxos de informação. Esse movimento impulsiona a maturidade digital da Agência e prepara o terreno para a replicação dessa abordagem em outros programas regulatórios, consolidando um ecossistema de dados mais robusto e confiável. Ao integrar a Qualidade de Dados à rotina institucional, a pesquisa propõe não apenas um ganho operacional, mas um novo paradigma de gestão da informação.

Desse modo, a inovação desta pesquisa reside não apenas na aplicação inédita de metodologias consolidadas da Ciência de Dados ao contexto da Anatel, mas, sobretudo, no estímulo à construção de uma cultura institucional voltada à Qualidade de Dados. Trata-se de um passo fundamental para o fortalecimento da governança de dados, contribuindo para uma atuação mais eficiente, responsiva e alinhada às exigências atuais de gestão pública orientada por dados confiáveis e transparentes.

7.2.3 Contribuição para a Fiscalização

Esta pesquisa, ao aprofundar a análise da qualidade dos dados no contexto do PBLE, contribui para a melhoria das atividades de fiscalização conduzidas pela Anatel. A partir da proposição de uma metodologia de avaliação da qualidade dos dados oriundos do sistema SICI, o estudo permite à Agência identificar inconsistências, lacunas e padrões que possam comprometer a confiabilidade dos dados utilizados nas diversas etapas das ações fiscalizatórias do programa.

Os resultados obtidos fornecem subsídios técnicos robustos que ampliam a capacidade analítica da Agência, possibilitando a elaboração de diagnósticos mais precisos e fundamentados. Tais informações qualificam a tomada de decisões estratégicas nas atividades de acompanhamento e controle do PBLE, inclusive no que se refere à aplicação de medidas sancionatórias às prestadoras, conforme previsto na regulamentação vigente.

Ademais, a pesquisa promove o aprimoramento do acompanhamento do programa, possibilitando uma atuação fiscalizatória mais proativa, eficaz e alinhada às diretrizes regulatórias da Anatel. Ao integrar a avaliação da qualidade dos dados aos processos de fiscalização, o estudo fortalece a capacidade institucional da Agência de acompanhar com exatidão o cumprimento das obrigações assumidas pelas prestadoras, assegurando, assim, maior efetividade na implementação do programa.

Portanto, esta dissertação representa uma contribuição relevante para o avanço científico no campo da qualidade de dados da Anatel, ao mesmo tempo em que fortalece as práticas institucionais da Agência, propondo soluções operacionais que visam tornar as ações fiscalizatórias mais eficientes, responsivas e fundamentadas em evidências.

7.2.4 Geração de *Scripts*

Como parte prática da presente pesquisa, foram desenvolvidos *scripts* em linguagem SQL para a extração de informações da base de dados do PBLE e de outros sistemas internos da Anatel. A criação desses *scripts* visa não apenas à replicabilidade da metodologia proposta, mas também à incorporação de boas práticas de análise de dados no contexto regulatório da Anatel, fortalecendo a capacidade analítica da Agência.

Os *scripts* desenvolvidos, apresentados nos Anexos I, II e III desta dissertação, têm como objetivo facilitar a aplicação da metodologia proposta por outras áreas da Anatel. Além disso, contribuem para sua reutilização, adaptação e aperfeiçoamento em futuras iniciativas de avaliação da qualidade dos dados no âmbito da Agência.

7.3 Impactos para a Anatel

Os resultados desta pesquisa apresentam um grande potencial para gerar impactos significativos nas atividades da Anatel, especialmente nas áreas de fiscalização, no acompanhamento e controle eficaz das políticas públicas, além da otimização da tomada de decisões estratégicas. Esses impactos não apenas aprimoram a eficiência e a eficácia dos processos da Agência, como também fortalecem sua atuação regulatória orientada por dados.

Um dos principais impactos é a ampliação da capacidade analítica da Anatel ao lidar com os dados reportados pelas prestadoras. Com a adoção de critérios de avaliação baseados em dimensões da qualidade – como completude, consistência, conformidade, integridade, precisão e atualidade – é possível identificar, com maior acurácia, erros, inconsistências e omissões nos dados fornecidos pelas prestadoras.

Esse aprimoramento na qualidade dos dados do PBLE beneficia diretamente o desempenho da equipe de fiscalização, promovendo uma atuação mais proativa, eficiente e fundamentada em dados corretos e confiáveis. Com isso, torna-se possível otimizar a alocação de recursos, elevar o grau de cumprimento das obrigações pelas prestadoras e fortalecer a efetividade das ações regulatórias da Anatel, com base em evidências.

Além disso, ao propor mecanismos de controle e avaliação da Qualidade de Dados, este estudo aprimora a capacidade institucional da Anatel de promover uma regulação mais eficiente e ampliar a efetividade das políticas públicas voltadas à conectividade nas escolas públicas brasileiras. Essa contribuição é fundamental para reduzir desigualdades educacionais e fomentar a inclusão digital no país.

Portanto, esta pesquisa se alinha aos esforços de transformação digital da Anatel, ao incentivar a adoção de práticas inovadoras e sustentáveis de gestão da qualidade dos dados. Sua aplicabilidade transcende o escopo do PBLE, oferecendo uma abordagem replicável a outros sistemas e processos da Agência. Com isso, reforça-se a capacidade institucional da Anatel de cumprir sua missão pública com eficiência e transparência.

7.4 Trabalhos Futuros

Os resultados obtidos neste estudo de mestrado representam um avanço expressivo no fortalecimento da qualidade e da governança de dados no âmbito da Anatel, contribuindo para a construção de bases informacionais mais robustas e confiáveis no contexto da gestão pública de dados no setor de telecomunicações.

Como desdobramento futuro, propõe-se a ampliação da avaliação da qualidade dos dados coletados tanto pelos fiscais quanto pelos responsáveis pelas escolas, bem como sua comparação com as informações fornecidas pelas prestadoras, conforme delineado

nesta pesquisa. Essa abordagem permitirá uma análise mais abrangente e precisa da consistência dos dados provenientes de diferentes fontes.

Adicionalmente, sugere-se a expansão do escopo dos dados analisados no âmbito do PBLE, mediante a incorporação de novas dimensões de qualidade e a integração de dados oriundos de fontes externas, como IBGE, Correios e INEP. Essa ampliação tem o potencial de enriquecer a avaliação, proporcionando uma mensuração mais completa e contextualizada da qualidade das informações utilizadas pelo programa.

Outro direcionamento relevante consiste na aplicação de técnicas avançadas de preparação, limpeza e transformação de dados, com o objetivo de promover o aprimoramento contínuo da metodologia desenvolvida para a avaliação da qualidade das informações no PBLE.

Recomenda-se, ainda, a realização de uma nova rodada de avaliação após a implementação do modelo proposto neste trabalho, de modo a permitir a verificação empírica dos impactos gerados na qualidade dos dados do programa.

Por fim, destaca-se como linha promissora de investigação a incorporação de abordagens baseadas em Inteligência Artificial, com vistas à automação, refinamento e evolução do processo de avaliação da qualidade dos dados, ampliando sua escalabilidade e eficiência no contexto institucional da Anatel.

Referências

- [1] Barbieri, C.: *Governança de Dados - Práticas, Conceitos e Novos Caminhos*. 2019. 1, 2, 32, 33, 36, 37, 38, 40, 41, 42, 43, 44
- [2] Rêgo, B. L.: *Gestão e Governança de Dados - Promovendo dados como ativo de valores nas empresas*. 2013, ISBN 9788574525891. 1, 37
- [3] McGilvray, D.: *Executing Data Quality Projects: Ten Steps to Quality Data and Trusted Information*. 2021. 1
- [4] Côrte-Real, N., P. Ruivo e T. Oliveira: *Leveraging internet of things and big data analytics initiatives in european and american firms: Is data quality a way to extract business value?* janeiro 2019. 2, 24
- [5] Batini, C. e M. Scannapieco: *Data Quality: Concepts, Methodologies and Techniques*. 2012. 2, 24, 25, 33, 34, 49, 50, 51, 52
- [6] Júnior, I. F. M.: *O programa 'banda larga nas escolas' no contexto das políticas públicas de inclusão digital*, março 2018. "<https://repositorio.enap.gov.br/handle/1/3356>", [Online; acessado em 20-janeiro-2023]. 3
- [7] Grimes, A. e W. Townsend: *Effects of (ultra-fast) fibre broadband on student achievement*. Information Economics and Policy, 44:8–15, setembro 2018. 3
- [8] Nordrum, E. e P. Gracia: *Impacts of broadband internet on adolescents' academic outcomes: heterogeneous effects among lower secondary school students in norway*. Information Communication Society, dezembro 2023. 3
- [9] Kende, M., S. Minehane, M. Minges, S. Molloy e G. Sciadas: *Universal and Meaningful Connectivity: The New Imperative*. 2022. 3, 4
- [10] MegaEdu: *Estudo de conectividade das escolas públicas*, 2022. "https://www.megaedu.org.br/_files/ugd/1a73c8_6cc96307d4cf428bb2756b70104512a8.pdf", [Online; acessado em 12-abril-2023]. 4
- [11] Pesquisas Educacionais Anísio Teixeira, Instituto Nacional de Estudos e: *Sinopse estatística da educação básica 2021*, 2022. "<https://www.gov.br/inep/pt-br/areas-de-atuacao/pesquisas-estatisticas-e-indicadores/censo-escolar/resultados/2021>", [Online; acessado em 4-março-2023]. 4

- [12] República, Presidência da: *Decreto nº 6.424, de 4 de abril de 2008*, 2021. https://www.planalto.gov.br/ccivil_03/_ato2007-2010/2008/decreto/d6424.htm, [Online; acessado em 19-janeiro-2023]. 4
- [13] Anatel: *Temo aditivo nº 01/2008/spv-anatel*, 2021. <https://www.gov.br/anatel/pt-br/regulado/universalizacao/plano-banda-larga-nas-escolas>, [Online; acessado em 19-janeiro-2023]. 5, 9
- [14] Cerqueira, L. A. M.: *Panorama de política pública federal de integração entre escolas públicas e tic: estudos sobre proinfo, prouca, pble*. Universidade do Estado do Rio de Janeiro, 2016. 5
- [15] Anatel: *Programa banda larga nas escolas*, 2024. <https://www.gov.br/anatel/pt-br/regulado/universalizacao/plano-banda-larga-nas-escolas>, [Online; acessado em 5-abril-2024]. 6, 7
- [16] Anatel: *Procedimento de fiscalização do pble*, 2023. <https://informacoes.anatel.gov.br/legislacao/procedimentos-de-fiscalizacao/1906-portaria-2740>, [Online; acessado em 5-fevereiro-2025]. 9
- [17] Rizun, N., C. Alexopoulos, S. Saxena, F. Kleiman e R. Matheus: *Do personality traits influence the user's behavioral intention to adopt and use open government data (ogd)? an empirical investigation*. Telematics and Informatics 2024, novembro 2023. 12
- [18] Brys, C., I. Navas-Delgado e J. Aldana Montes M. Roldán-García: *Lego: Linked electronic government ontology*. Journal of Information Science, páginas 64–73, junho 2023. 12
- [19] OKF: *Open knowledge foundation*, 2023. <https://www.okfn.org/en>, [Online; acessado em 12-novembro-2023]. 12
- [20] Broomfield, H.: *Where is open data in the open data directive?* Information Polity, 12:175–188, julho 2023. 13
- [21] Maanen, G. van: *Studying open government data: Acknowledging practices and politics*. Data Policy 2023, 5:300–322, fevereiro 2023. 13
- [22] Lnenicka, M., M. Lnenicka, A. Nikiforova, M. Luterek, P. Milic, D. Rudmark, S. Neumaier, C. Santoro, C. Flores, M. Janssen e M. Bolívar: *Identifying patterns and recommendations of and for sustainable open data initiatives: A benchmarking-driven analysis of open government data initiatives among european countries*. Government Information Quarterly, 41:64–73, março 2014. 13
- [23] Fan, B. e X. Meng: *Moderating effects of governance on open government data quality and open government data utilization: Analysis based on the resource complementarity perspective*. Journal of Global Information Technology Management, 26:300–322, outubro 2023. 13

- [24] República, Presidência da: *Decreto nº 8.777, de 11 de maio de 2016*, 2016. https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2016/decreto/d8777.htm, [Online; acessado em 10-novembro-2023]. 13
- [25] Brasil: *Portal de dados abertos*, 2024. <https://www.dados.gov.br>, [Online; acessado em 3-abril-2024]. 13
- [26] Pesquisas Educacionais Anísio Teixeira, Instituto Nacional de Estudos e: *Sinopse estatística da educação básica 2024*, 2024. "<https://www.gov.br/inep/pt-br/areas-de-atuacao/pesquisas-estatisticas-e-indicadores/censo-escolar/resultados/2024>", [Online; acessado em 24-setembro-2024]. 14
- [27] Guenther, L. C., E. Colangelo, H. Wiendahl e C. Bauer: *Data quality assessment for improved decision-making: a methodology for small and medium-sized enterprises*. Volume 29, páginas 583–591, setembro 2019. 14, 15, 24, 25
- [28] Kitchenham, B. e S. Charters: *Guidelines for performing systematic literature reviews in software engineering*. Keele University and Durham University Joint Report, julho 2007. 20
- [29] Dermeval, D., J. A. P. Coelho e I. I. Bittencourt: *Mapeamento sistemático e revisão sistemática da literatura em informática na educação*. Sociedade Brasileira de Computação, novembro 2020. 21
- [30] Brous, P., M. Janssen e R. Vilminko-Heikkinen: *Coordinating decision-making in data management activities: A systematic review of data governance principles*. Electronic Governmet, 9820:115–125, agosto 2019. 24, 38, 41
- [31] Behringer, G. e M. Hizli: *Data governance: State-of-the-art*. Innovation Through Information Systems, 47:687–699, agosto 2021. 24
- [32] Bachtiar, A., S. Suhardi e W. Muhamad: *Literature review of open government data*. International Conference on Information Technology Systems and Innovation (IC-ITSI), 9820:329–334, outubro 2020. 24
- [33] Al-Ruithe, M., E. Benkhelifa e K. Hameed: *A systematic literature review of data governance and cloud data governance*. Personal and Ubiquitous Computing, 23:839–859, novembro 2019. 24
- [34] Administração Pública, Escola Nacional de: *Governança de Dados na Transformação Digital*. 2022. 24, 37, 38, 39
- [35] Abraham, R., J. Schneider e J. vom Brocke: *Data governance: A conceptual framework, structured review, and research agenda*, volume 49. outubro 2019. 24, 32, 35, 37, 39
- [36] Al-Badia A., A. Tarhinia e A. Khan: *Exploring big data governance frameworks*. 8th International Conference On Current And Future Trends Of Information And Communication Technologies In Healthcare, 141:271–277, julho 2019. 24

- [37] Setiawan, A., A. Syamsudin e Y. Rosmansyah: *Assessment of information technology security governance for supervisory control and data acquisition (scada) on the smart grid electricity*. International Conference On Information Technology Systems And Innovation (Icitsi), setembro 2016. 24
- [38] Lis, D e B. Otto: *Data governance in data ecosystems – insights from organizations*. Conference of the Association-for-Information-Systems (AMCIS), páginas 10–14, setembro 2020. 24
- [39] Bertin, P., J. Meireles, A. Silva, M. Okawachi e M. Cardoso: *A política de governança de dados, informação e conhecimento da embrapa como mecanismo para a gestão de dados de pesquisa agropecuários*. Liinc em Revista, 15:194–204, novembro 2019. 24
- [40] Buss, C., D. Rosa, F. Metzner e J. Júnior: *Gestão de dados acadêmicos: a aplicação dos conceitos de governança de dados para melhoria nos processos de utilização das informações do sistema de gestão de dados acadêmicos*. Revista Ibero-Americana de Ciência da Informação, 14:401–419, janeiro 2021. 24
- [41] Santos, N. e E. Streit: *O processo decisório de governança de dados*. Brazilian Journal of Information Studies: Research Trends, 12:64–73, maio 2018. 24
- [42] Mao, Z., J. Wu, Y. Qiao e H. Yao: *Government data governance framework based on a data middle platform*. Aslib Journal of Information Management, 74:289–310, fevereiro 2022. 24, 25, 28
- [43] Benfeldt, O., J. Stouby e S. Madsen: *Data governance as a collective action problem*. Information Systems Frontiers, 22:299–313, abril 2020. 24, 25
- [44] Zhang, Q., X. Sun e M. Zhang: *Data matters: A strategic action framework for data governance*. Information Management, 59, junho 2022. 24
- [45] Jang, K. e W. Kim: *Development of data governance components using dematel and content analysis*. Journal of Supercomputing, 77:3695–3709, abril 2021. 24, 25
- [46] Karkoskova, S.: *Data governance model to enhance data quality in financial institutions*. Information Systems Management, 40:90–110, janeiro 2023. 24, 25
- [47] Nadal, S., P. Jovanovic, B. Bilalli e O. Romero: *Operationalizing and automating data governance*. Journal of Big Data, 9, dezembro 2022. 24, 25
- [48] Batini, C. e M. Scannapieco: *Data and Information Quality - Dimensions, Principles and Techniques*. 2016, ISBN 978-3-319-24104-3. 24, 25, 31, 32, 33, 42, 43, 44, 45, 46, 47, 51
- [49] Valencia-Parra, A., L. Parody, A. J. Varela-Vaca, I. Caballero e M. T. Gómez-López: *Dmn for data quality measurement and assessment*. Volume 1, páginas 362–374, setembro 2019. 24, 25, 44
- [50] Richard, Y. W., M. Ziad e Y. W. Lee: *Data Quality*. 2002, ISBN 0-792-37215-8. 24

- [51] Batini, C., C. Cappiello, C. Francalanci e A. Maurino: *Methodologies for data quality assessment and improvement*. ACM Computing Surveys, 41, julho 2009. 24, 25, 43
- [52] Pipino, L. L., Y. W. Lee e R. Y. Wang: *Data quality assessment*. Communications of the ACM, 45:211–218, abril 2002. 24, 25
- [53] Wang, R. Y. e D. M. Strong: *Beyond accuracy: What data quality means to data consumers*. Journal of Management Information Systems, 12:5–33, dezembro 2015. 24, 25
- [54] Ballou, D. P. e H. L. Pazer: *Modeling completeness versus consistency tradeoffs in information decision contexts*. IEEE Transactions on Knowledge and Data Engineering, 15, janeiro 2003. 24, 25
- [55] Sadiq, S., T. Dasu, X. L. Dong, J. Freire, I. F. Ilyas, S. Link, M. J. Miller, F. Naumann, X. Zhou e D. Srivastava: *Data quality: The role of empiricism*. ACM SIGMOD Record, 46:35–43, dezembro 2017. 24
- [56] Nasr, M., E. Shaaban e M. I. Gabr: *Data quality dimensions*. Internet of Things — Applications and Future, 114, outubro 2020. 24, 25
- [57] Šlibar, B., D. Oreški e N. B. Redep: *Importance of the open data assessment: An insight into the (meta) data quality dimensions*. SAGE Open, 1:1–18, junho 2021. 24, 25
- [58] Amrapalia, Z., A. Rula, A. Maurino, R. Pietrobon, J. Lehmann e S. Auer: *Quality assessment for linked data: A survey*. Semantic Web Journal, 7:63–93, março 2016. 24, 25
- [59] Valencia, A., L. Parody, A. J. Varela-Vaca, I. Caballero e M. T. Gomez-Lopez: *Dmn4dq: When data quality meets dmn*. 141, fevereiro 2021. 24, 25, 42
- [60] Yusheng, Z., Q. Rufu, H. Xu, S. Sadiq e Y. Yu: *A data quality control method for seafloor observatories: The application of observed time series data in the east china sea*. National Center of Biotechnology Information, 18, agosto 2018. 24, 25
- [61] Liu, Z., Q. Chen e L. Cai: *Application of requirement-oriented data quality evaluation method*. 19th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD), 18:407–412, julho 2018. 24, 25, 27
- [62] Liu, Q., G. Feng, X. Zhao e W. Wang: *Minimizing the data quality problem of information systems: A process-based method*. Decision Support Systems, 137, agosto 2020. 24, 25
- [63] Timmerman, Y. e A. Bronselaer: *Measuring data quality in information systems research*. Decision Support Systems, 126:113–138, maio 2019. 24, 25
- [64] Mouhamed D., M. S. Camara, A. Bah e I. Fall: *Prior management of temporal data quality in a data mining process : an implementation architecture*. Procedia Computer Science, 148:273–282, junho 2019. 24, 25

- [65] Ngueilbaye, A., H. Wang, D. A. Mahamat, I. A. Elgendy e S. B. Junaidu: *Methods for detecting and correcting contextual data quality problems*. Intelligent Data Analysis, 25:763–787, novembro 2021. 24, 25
- [66] Gürdür, D., J. El khoury e M. Nyberg: *Methodology for linked enterprise data quality assessment through information visualizations*. Journal of Industrial Information Integration, 15:191–200, setembro 2019. 24, 25
- [67] Xiao-Tong, L., J. Zhai, G. Zheng e Y. Chang-Feng: *Quality Assessment for Open Government Data in China*, volume 1. setembro 2018. 24, 25, 27, 49, 53, 54, 61, 75
- [68] Ngueilbaye, A., H. Wang, D. A. Mahamat, I. A. Elgendy e S. B. Junaidu: *Methods for detecting and correcting contextual data quality problems*. Intelligent Data Analysis, 25:763–787, junho 2021. 24, 25, 43, 44
- [69] Yang, S., W. Fan, S. Tang, X. Gao, B. Yang e G. Chen: *On designing data quality-aware truth estimation and surplus sharing method for mobile crowdsensing*. IEEE Journal on Selected Areas in Communications, 35, março 2017. 24, 25
- [70] Byabazaire, J., G. O’Hare e D. Delaney: *Data quality and trust: Review of challenges and opportunities for data sharing in iot*. MPDI Journal, dezembro 2020. 24, 25
- [71] Hao, Y., M. Wang, J. H. Chow, E. Farantatos e M. Patel: *Modelless data quality improvement of streaming synchrophasor measurements by exploiting the low-rank hankel structure*. IEEE Transactions on Power Systems, 33:6966–6977, novembro 2018. 24, 25
- [72] Kurniati, A. P., E. Rojas, D. Hogg, G. Hall e J. Owen: *The assessment of data quality issues for process mining in healthcare using medical information mart for intensive care iii, a freely available e-health record database*. Health Informatics Journal, página 1878–1893, novembro 2018. 24, 25
- [73] Ardagna, D., C. Cappiello, W. Samá e M. Vitali: *Context-aware data quality assessment for big data*. Future Generation Computer Systems, 89:548–562, dezembro 2018. 24, 25, 28
- [74] Leal, A. P.: *Políticas de expansão da internet banda larga no brasil e em portugal: Análise e perspectivas futuras*. Centro de Investigação de Direito Privado da Faculdade de Direito da Universidade de Lisboa, 6:167–213, janeiro 2024. 29
- [75] Giuliani, G. S., C. S. Santos, F. M. Schaf e K. M. Rocha: *Escolas conectadas? a possibilidade de novas redes computacionais*. Revista Iberoamericana de Tecnología en Educación y Educación en Tecnología, 30:59–65, junho 2021. 29
- [76] Lima, A. F. R., H. K. Barbosa; e A. Sachsida: *Avaliando o impacto do programa banda larga nas escolas sobre a qualidade educacional*. Instituto de Pesquisa Econômica Aplicada, dezembro 2018. 30
- [77] Júnior, I. F. M.: *O programa ‘banda larga nas escolas’ no contexto das políticas públicas de inclusão digital*. Escola Nacional de Administração Pública, março 2018. 30

- [78] Victorino, M.: *Alfabetização de Dados*. Caso do Código, 1ª edição, 2024, ISBN 978-85-5519-387-3. 31
- [79] Bližnák, K., M. Munk e A. Pilková: *A systematic review of recent literature on data governance (2017–2023)*. IEEE Access, 12:149875–149888, outubro 2024. 35, 36
- [80] Nielsen, O. B.: *A comprehensive review of data governance literature*. Association for Information Systems, 8, dezembro 2017. 35, 36
- [81] Karkoskova, S.: *Data governance model to enhance data quality in financial institutions*. Information Systems Management, 40:90–110, março 2023. 35, 40, 42, 46
- [82] Alhassan, I., D. Sammon e M. Daly: *Data governance activities: an analysis of the literature*. Journal of Decision Systems, 71:64–75, junho 2016. 36
- [83] INTERNATIONAL, DAMA: *DAMA DMBOK: Data Management Body of Knowledge*. 2017. 36, 42
- [84] Fadlallah, H., R. Kilany, H. Dhayne, R. El Haddad, R. Haque, Y. Taher e A. Jaber: *Context-aware big data quality assessment: A scoping review*. ACM Journal of Data Information Quality, 15, setembro 2023. 42, 44
- [85] Rueda, O., N. Benabdellah e A. Zellou: *A systematic literature review on data quality assessment*. Bulletin of Electrical Engineering and Informatics, 12:3736–3757, dezembro 2023. 42, 45, 47, 48, 51
- [86] Oliveira, G., B. Mendes, C. Bacha, L. Costa, L. Gomide, M. Silva, M. Brandão, A. Lacerda e G. Pappa: *Assessing data quality inconsistencies in brazilian governmental data*. Journal of Information and Data Management, 14, outubro 2023. 42
- [87] Gupta, N.: *Optimising data quality of a data warehouse using data purgation process*. International Journal of Data Mining, Modelling and Management, 15:102–131, abril 2023. 43
- [88] Youdi, G., L. Guangzhen, X. Yunzhi, L. Rui e M. Lingzhong: *A survey on dataset quality in machine learning*. Information and Software Technology, 162, outubro 2023. 46
- [89] Fan, B. e Y. Zhao: *The moderating effect of external pressure on the relationship between internal organizational factors and the quality of open government data*, volume 34. setembro 2017. 48, 49
- [90] OGWG: *Open government working group*, 2023. <https://opengovdata.org>, [Online; acessado em 12-novembro-2023]. 48
- [91] Academy, Equipe Data Science: *As 6 dimensões da qualidade de dados (data quality)*, 2023. <https://blog.dsacademy.com.br/as-6-dimensoes-da-qualidade-de-dados-data-quality/>, [Online; acessado em 22-julho-2023]. 50

- [92] Almeida, M., E. Felipe e R. Barcelos: *Toward a document-centered ontological theory for information architecture in corporations*. Journal of the Association for Information Science and Technology, 71, dezembro 2020. 55

Anexo I

Scripts de Tabelas de Auditoria

I.1 Criação da Tabela BandaLargaEscolas__Aud

```
1 CREATE TABLE BandaLargaEscolas__Aud (
2     IdtBandaLarga int IDENTITY(1,1) NOT NULL,
3     NumFistel int NOT NULL,
4     DataMovimentoAno date NOT NULL,
5     DataMovimentoMes date NOT NULL,
6     CodINEP int NOT NULL,
7     NomeEscola varchar(255) NOT NULL,
8     CodMunicipio int NOT NULL,
9     EnderecoEscola varchar(255) NOT NULL,
10    BairroEscola varchar(255) NOT NULL,
11    ComplementoEscola varchar(255) NOT NULL,
12    NumeroEscola int NOT NULL,
13    CEPescola varchar(8) NOT NULL,
14    IdtTipoEscola int NOT NULL,
15    codTipoObrigacao int NOT NULL,
16    IdtSituacao int NOT NULL,
17    DataAtivacao date NOT NULL,
18    VelAcessoInstalada decimal(5,2) NOT NULL,
19    VelAcessoObrigacao decimal(5,2) NOT NULL,
20    IP varchar(15) NOT NULL,
21    IdtTecnologia int NOT NULL,
22    IdtTipoPendencia int NOT NULL,
23    DataPrevisao date NOT NULL,
24    DataDesativacao date NOT NULL
25
26    CONSTRAINT PK_BandaLargaEscolas__Aud PRIMARY KEY CLUSTERED
27    (
28        IdtBandaLarganasEscolas ASC
29    ) WITH (PAD_INDEX = OFF, STATISTICS_NORECOMPUTE = OFF,
```

```

30     IGNORE_DUP_KEY = OFF, ALLOW_ROW_LOCKS = ON,
31     ALLOW_PAGE_LOCKS = ON) ON [PRIMARY]
32 ) ON [PRIMARY]

```

I.2 Criação da Tabela Prestadora__Aud

```

1  CREATE TABLE Prestadora__Aud (
2      IdtPrestadora int IDENTITY(1,1) NOT NULL,
3      NumFistel int NOT NULL,
4      NumCNPJ int NOT NULL,
5      NomeEntidade varchar(255) NOT NULL,
6      EnderecoEntidade varchar(255) NOT NULL,
7      ResponsavelEntidade varchar(100) NOT NULL,
8      IP varchar(15) NOT NULL,
9      DataPrevisao date NOT NULL,
10     DataDesativacao date NOT NULL
11
12     CONSTRAINT PK_Prestadora__Aud PRIMARY KEY CLUSTERED
13     CONSTRAINT FK_Prest_BandaLarga FOREIGN KEY (NumFistel)
14         REFERENCES BandaLargaEscolas__Aud (NumFistel)
15     (
16         IdtPrestadora ASC
17     ) WITH (PAD_INDEX = OFF, STATISTICS_NORECOMPUTE = OFF,
18         IGNORE_DUP_KEY = OFF, ALLOW_ROW_LOCKS = ON,
19         ALLOW_PAGE_LOCKS = ON) ON [PRIMARY]
20 ) ON [PRIMARY]

```

I.3 Criação da Tabela Escola__Aud

```

1  CREATE TABLE Escola__Aud (
2      IdtEscola int IDENTITY(1,1) NOT NULL,
3      CodINEP int NOT NULL,
4      NomeEscola varchar(255) NOT NULL,
5      CodMunicipio int NOT NULL,
6      EnderecoEscola varchar(255) NOT NULL,
7      BairroEscola varchar(255) NOT NULL,
8      ComplementoEscola varchar(255) NOT NULL,
9      NumeroEscola int NOT NULL,
10     CEPEscola varchar(8) NOT NULL,
11     NomeGestor varchar(100) NOT NULL,
12     Contato varchar(50) NOT NULL,
13     Email varchar(50) NOT NULL,

```

```

14     VelocidadeMedida decimal (5,2) NOT NULL,
15     Tecnologia int NOT NULL,
16
17     CONSTRAINT PK_Escola_Aud PRIMARY KEY CLUSTERED
18     CONSTRAINT FK_Escola_BandaLarga FOREIGN KEY (CodINEP)
19         REFERENCES BandaLargaEscolas_Aud (CodINEP)
20     (
21         IdtEscola ASC
22     ) WITH (PAD_INDEX = OFF, STATISTICS_NORECOMPUTE = OFF,
23         IGNORE_DUP_KEY = OFF, ALLOW_ROW_LOCKS = ON,
24         ALLOW_PAGE_LOCKS = ON) ON [PRIMARY]
25 ) ON [PRIMARY]

```

I.4 Criação da Tabela Tecnologia__Aud

```

1  CREATE TABLE Tecnologia_Aud (
2      IdtTecnologia int IDENTITY(1,1) NOT NULL,
3      CodTecnologia int NOT NULL,
4      DescTecnologia varchar(255) NOT NULL,
5      DataCriacao date NOT NULL,
6
7      CONSTRAINT PK_Tecnologia_Aud PRIMARY KEY CLUSTERED
8      CONSTRAINT FK_Tecnologia_BandaLarga FOREIGN KEY (CodTecnologia)
9          REFERENCES BandaLargaEscolas_Aud (CodTecnologia)
10     (
11         IdtTecnologia ASC
12     ) WITH (PAD_INDEX = OFF, STATISTICS_NORECOMPUTE = OFF,
13         IGNORE_DUP_KEY = OFF, ALLOW_ROW_LOCKS = ON,
14         ALLOW_PAGE_LOCKS = ON) ON [PRIMARY]
15 ) ON [PRIMARY]

```

I.5 Criação da Tabela Municipio__Aud

```

1  CREATE TABLE Municipio_Aud (
2      IdtMunicipio int IDENTITY(1,1) NOT NULL,
3      CodMunicipio int NOT NULL,
4      NomeMunicipio varchar(255) NOT NULL,
5      CodUF int NOT NULL,
6      NomeUF varchar(255) NOT NULL,
7      CEPescola int NOT NULL,
8      CodRegiao int NOT NULL,
9      NomeRegiao varchar(255) NOT NULL,

```

```

10
11     CONSTRAINT PK_Municipio_Aud PRIMARY KEY CLUSTERED
12     CONSTRAINT FK_Municipio_Escola FOREIGN KEY (CodMunicipio)
13         REFERENCES Escola_Aud(CodMunicipio)
14     (
15         IdtMunicipio ASC
16     ) WITH (PAD_INDEX = OFF, STATISTICS_NORECOMPUTE = OFF,
17         IGNORE_DUP_KEY = OFF, ALLOW_ROW_LOCKS = ON,
18         ALLOW_PAGE_LOCKS = ON) ON [PRIMARY]
19 ) ON [PRIMARY]

```

I.6 Criação da Tabela Fiscalizacao__Escola__Aud

```

1  CREATE TABLE Tabela_Fiscalizacao_Escola_Aud (
2      IdtFiscalizacaoEscola int IDENTITY(1,1) NOT NULL,
3      NumFistel int NOT NULL,
4      CodINEP int NOT NULL,
5      VelocidadeMedida decimal (5,2) NOT NULL,
6      Tecnologia int NOT NULL,
7      IdEquipamento int NOT NULL,
8      NomeEscola varchar(255) NOT NULL,
9      EnderecoEscola varchar(255) NOT NULL,
10     NumeroEscola int NOT NULL,
11     ComplementoEscola varchar(255) NOT NULL,
12     BairroEscola varchar(255) NOT NULL,
13     MunicipioEscola varchar(255) NOT NULL,
14     CEPescola varchar (8) NOT NULL,
15     Gestor varchar(100) NOT NULL,
16     Contato varchar(100) NOT NULL,
17     Email varchar(100) NOT NULL
18
19     CONSTRAINT PK_Fiscalizacao_AudEscola_Aud PRIMARY KEY CLUSTERED
20     (
21         IdtFiscalizacaoEscola ASC
22     ) WITH (PAD_INDEX = OFF, STATISTICS_NORECOMPUTE = OFF,
23         IGNORE_DUP_KEY = OFF, ALLOW_ROW_LOCKS = ON,
24         ALLOW_PAGE_LOCKS = ON) ON [PRIMARY]
25 ) ON [PRIMARY]

```

Anexo II

Scripts de Extração de Dados

II.1 Base de Dados: PBLE Sistema Interno: SICI

```
1 USE SITARWEB
2
3 ;WITH censo AS (
4     SELECT ble.NumFistel, ble.DataMovimentoAno, ble.DataMovimentoMes,
5         ISNULL(ble.CodINEP, 'N/I') AS CodINEP, ble.NomeEscola,
6         ble.CodMunicipio, ble.EnderecoEscola, ble.BairroEscola,
7         ble.ComplementoEscola, ble.NumeroEscola, ble.CEPEscola,
8         ble.IdtTipoEscola, ble.codTipoObrigacao, ble.IdtSituacao,
9         ble.DataAtivacao, ble.VelAcessoInstalada, ble.VelAcessoObrigacao,
10        ble.IP, ble.IdtTecnologia, ble.IdtTipoPendencia, ble.DataPrevisao,
11        ble.DataDesativacao
12 FROM SICI_INEP inep RIGHT JOIN SICI_BandaLarganasEscolas ble
13     ON ble.CodINEP = inep.codINEP
14 ), escolas AS (
15     SELECT inep.CodINEP, inep.Operadora AS NumFistel, inep.CodMunicipio,
16         inep.IdtTipoPendencia, IdtTecnologia, IdtTipoEscola, IdtSituacao,
17         inep.NomeEscola, inep.EnderecoEscola, inep.BairroEscola,
18         inep.ComplementoEscola, inep.NumeroEscola, inep.CEPEscola,
19         VelAcessoInstalada, VelAcessoObrigacao, IP, DataAtivacao,
20         DataPrevisao, DataDesativacao, DataMovimentoAno, DataMovimentoMes,
21         codTipoObrigacao
22 FROM SICI_INEP inep
23     WHERE inep.CodINEP NOT IN (SELECT c.CodINEP FROM censo c)
24 UNION ALL SELECT ble.CodINEP, ble.NumFistel, ble.CodMunicipio,
25     ble.IdtTipoPendencia, ble.IdtTecnologia, ble.IdtTipoEscola,
26     ble.IdtSituacao, ble.NomeEscola, ble.EnderecoEscola,
27     ble.BairroEscola, ble.NumeroEscola, ble.ComplementoEscola,
```

```

28     ble.CEPEscola, ble.VelAcessoInstalada, ble.VelAcessoObrigacao,
29     ble.IP, ble.DataAtivacao, ble.DataPrevisao, ble.DataDesativacao,
30     ble.DataMovimentoAno, ble.DataMovimentoMes, ble.codTipoObrigacao
31 FROM censo ble
32 )
33 SELECT DISTINCT es.NumFistel, es.DataMovimentoAno,
34     es.DataMovimentoMes, es.CodINEP, es.NomeEscola, es.CodMunicipio,
35     es.EnderecoEscola, es.BairroEscola, es.ComplementoEscola,
36     es.NumeroEscola, es.CEPEscola, es.IdtTipoEscola,
37     es.codTipoObrigacao, es.IdtSituacao,
38     CONVERT (VARCHAR, es.DataAtivacao, 103) as DataAtivacao,
39     es.VelAcessoInstalada, es.VelAcessoObrigacao, es.IP,
40     es.IdtTecnologia, es.IdtTipoPendencia,
41     CONVERT (VARCHAR, es.DataPrevisao, 103) as DataPrevisao,
42     CONVERT (VARCHAR, es.DataDesativacao, 103) as DataDesativacao
43 FROM escolas es
44     LEFT JOIN SICI_NomeFantasia nf ON nf.NumFistel = es.NumFistel
45     LEFT JOIN HABILITACAO h ON h.NumFistel = es.NumFistel
46     LEFT JOIN ENTIDADE e ON e.IdtEntidade = h.IdtEntidade
47     LEFT JOIN SICI_TipoPendencia tp
48         ON tp.CodTipoPendencia = es.IdtTipoPendencia
49     LEFT JOIN SICI_Tecnologia t ON t.CodTecnologia = es.IdtTecnologia
50     LEFT JOIN SICI_TipoEscola te ON te.CodTipoEscola = es.IdtTipoEscola
51     LEFT JOIN SICI_Situacao s ON s.CodSituacao = es.IdtSituacao
52     LEFT JOIN MUNICIPIO m ON m.CodMunicipio = es.CodMunicipio
53 WHERE 1 = 1

```

II.2 Base de Dados: Prestadoras

Sistema Interno: STEL

```

1 USE SITARWEB
2
3 SELECT
4     E.IdtEntidade as IdtPrestadora,
5     E.NomeFantasia,
6     H.NumFistel as Fistel,
7     E.NumCnpjCpf as CNPJ,
8     E.NomeEntidade as NomePrestadora,
9     En.EndLogradouro as Endereco,
10    H.DataInclusao as DtOutorga,
11    H.NumProcesso
12 FROM ENTIDADE E
13     inner join HABILITACAO H on H.IdtEntidade = E.IdtEntidade

```

```
14      inner join ENDERECO En on En.IdtEntidade = E.IdtEntidade
```

II.3 Base de Dados: Escolas

Sistema Interno: INEP

```
1 USE SITARWEB
2
3 SELECT
4     CodINEP ,
5     NomeEscola ,
6     EnderecoEscola ,
7     NomeMunicipio ,
8     SiglaUF ,
9     TipoEscola
10  FROM SICI_INEP
```

II.4 Base de Dados: Tecnologia

Sistema Interno: STEL

```
1 USE SITARWEB
2
3 SELECT
4     IdtTipoTecnologia as IdTecnologia ,
5     DescTipoTecnologia as NomeTecnologia ,
6     SiglaTipoTecnologia as SiglaTecnologia ,
7     TxtLetraManual as Manual ,
8     DataInclusao ,
9     DataAlteracao
10
11     CONVERT (VARCHAR, DataInclusao , 103) as DtRegulamento
12  FROM TECNOLOGIA
```

II.5 Base de Dados: Municípios

Sistema Interno: CEP

```
1 USE SITARWEB
2
3 SELECT
4     CodMunicipio ,
5     NomeMunicipio ,
6     SiglaUF ,
7     QtdePopulacao ,
8     NumCodigoNacional ,
9     CodCepMenor ,
10    CodCepMaior
11 FROM MUNICIPIO
```


Anexo III

Scripts de Geração de Relatórios

III.1 Relatório SICI x Prestadora__Aud

```
1 USE SITARWEB
2
3  /* Analise do campo 'NumFistel' */
4
5  SELECT count(A.NumFistel) as Regs_Divergentes
6  FROM BandaLarganasEscolas__Aud A
7  WHERE A.NumFistel IN (
8      SELECT A.NumFistel FROM BandaLarganasEscolas__Aud A
9      EXCEPT
10     SELECT B.NumFistel from Prestadora__Aud B
11 )
12
13
14  /* Analise do campo 'CodTecnologia' */
15
16  SELECT count(A.CodTecnologia) as Regs_Divergentes
17  FROM BandaLarganasEscolas__Aud A
18  WHERE A.CodTecnologia IN (
19      SELECT A.CodTecnologia FROM BandaLarganasEscolas__Aud A
20      EXCEPT
21     SELECT B.CodTecnologia from Prestadora__Aud B
22 )
```

III.2 Relatório SICI x Escola_Aud

```
1 USE SITARWEB
2
3  /* Analise do campo 'CodINEP' */
4
5  SELECT count(A.CodINEP) as Regs_Divergentes
6  FROM BandaLarganasEscolas_Aud A
7  WHERE A.CodINEP IN (
8      SELECT A.CodINEP FROM BandaLarganasEscolas_Aud A
9      EXCEPT
10     SELECT B.CodINEP from Escola_Aud B
11 )
12
13
14  /* Analise do campo 'NomeEscola' */
15
16  SELECT count(A.NomeEscola) as Regs_Divergentes
17  FROM BandaLarganasEscolas_Aud A
18  WHERE A.NomeEscola IN (
19      SELECT A.NomeEscola FROM BandaLarganasEscolas_Aud A
20      EXCEPT
21      SELECT B.NomeEscola from Escola_Aud B
22  )
```

III.3 Avaliação SICI x Municipio_Aud

```
1 USE SITARWEB
2
3  /* Analise do campo 'CodMunicipio' */
4
5  SELECT count(A.CodMunicipio) as Regs_Divergentes
6  FROM BandaLarganasEscolas_Aud A
7  WHERE A.CodMunicipio IN (
8      SELECT A.CodMunicipio FROM BandaLarganasEscolas_Aud A
9      EXCEPT
10     SELECT B.CodMunicipio from Municipio_Aud B
11 )
12
13
14  /* Analise do campo 'CEPEscola' */
15
16  SELECT count(A.CEPEscola) as Regs_Divergentes
```

```

17 FROM BandaLarganasEscolas_Aud A
18 WHERE A.CEPEscola IN (
19     SELECT A.CEPEscola FROM BandaLarganasEscolas_Aud A
20     EXCEPT
21     SELECT B.CEPEscola from Municipio_Aud B
22 )

```

III.4 Relatório SICI x Tecnologia_Aud

```

1 USE SITARWEB
2
3 /* Analise do campo 'IdtTecnologia' */
4
5 SELECT count(A.IdtTecnologia) as Regs_Divergentes
6 FROM BandaLarganasEscolas_Aud A
7 WHERE A.IdtTecnologia IN (
8     SELECT A.IdtTecnologia FROM BandaLarganasEscolas_Aud A
9     EXCEPT
10    SELECT B.IdtTecnologia from Tecnologia_Aud B
11 )

```

III.5 Relatório SICI x Fiscalizacao_Escola_Aud

```

1 USE SITARWEB
2
3 /* Analise do campo 'NumFistel' */
4
5 SELECT count(A.NumFistel) as Regs_Divergentes
6 FROM BandaLarganasEscolas_Aud A
7 WHERE A.NumFistel IN (
8     SELECT A.NumFistel FROM BandaLarganasEscolas_Aud A
9     EXCEPT
10    SELECT B.NumFistel from Fiscalizacao_Escola_Aud B
11 )
12
13
14 /* Analise do campo 'CodINEP' */
15
16 SELECT count(A.CodINEP) as Regs_Divergentes
17 FROM BandaLarganasEscolas_Aud A
18 WHERE A.CodINEP IN (
19     SELECT A.CodINEP FROM BandaLarganasEscolas_Aud A

```

```

20             EXCEPT
21         SELECT B.CodINEP from Fiscalizacao_Escola_Aud B
22     )
23
24
25     /* Analise do campo 'Tecnologia' */
26
27     SELECT count(A.IdtTecnologia) as Regs_Divergentes
28     FROM BandaLarganasEscolas_Aud A
29     WHERE A.IdtTecnologia IN (
30         SELECT A.IdtTecnologia FROM BandaLarganasEscolas_Aud A
31         EXCEPT
32         SELECT B.Tecnologia from Fiscalizacao_Escola_Aud B
33     )
34
35
36     /* Analise do campo 'Velocidade' */
37
38     SELECT count(A.VelAcessoInstalada) as Regs_Divergentes
39     FROM BandaLarganasEscolas_Aud A
40     WHERE A.VelAcessoInstalada IN (
41         SELECT A.VelAcessoInstalada FROM BandaLarganasEscolas_Aud A
42         EXCEPT
43         SELECT B.VelocidadeMedida from Fiscalizacao_Escola_Aud B
44     )

```