

SIMONE CRISTINA CERON RIPOLI

**ESTUDO E ANÁLISE DA QUALIDADE DOS DADOS GOVERNAMENTAIS
ABERTOS (DGA) NO CONTEXTO DO ACESSO, USO E REÚSO DOS DADOS**

Marília, SP
2026

SIMONE CRISTINA CERON RIPOLI

**ESTUDO E ANÁLISE DA QUALIDADE DOS DADOS GOVERNAMENTAIS
ABERTOS (DGA) NO CONTEXTO DO ACESSO, USO E REÚSO DOS DADOS**

Tese apresentada ao Programa de Pós-Graduação em Ciência da Informação (PPGCI) como parte das exigências para a obtenção do título de Doutora em Ciência da Informação pela Universidade Estadual Paulista “Júlio de Mesquita Filho” (Unesp), Faculdade de Filosofia e Ciências (FFC), Campus de Marília, SP.

Área de Concentração: Informação, Tecnologia e Conhecimento.

Orientador (a): Prof. Dr. José Eduardo Santarém Segundo.

Marília, SP
2026

R592e Ripoli, Simone Cristina Ceron
Estudo e análise da qualidade dos Dados Governamentais
abertos (DGA) no contexto do acesso, uso e reúso dos dados /
Simone Cristina Ceron Ripoli. -- Marília, 2026
332 p.

Tese (doutorado) - Universidade Estadual Paulista (UNESP),
Faculdade de Filosofia e Ciências, Marília
Orientadora: José Eduardo Santarém Segundo

1. dados governamentais abertos. 2. qualidade dos dados. 3.
dimensões de Qualidade. 4. informação Pública. 5. métricas de
qualidade de dados. I. Título.

Impacto potencial desta pesquisa

O impacto potencial desta pesquisa é multidimensional, abrangendo desde a eficiência da gestão pública até o fortalecimento da democracia e a inovação tecnológica. A disponibilização de dados de qualidade impacta diretamente a governança, a transparência, a eficiência na tomada de decisões baseadas em evidências, o fomento à inovação, o desenvolvimento econômico e o avanço científico e metodológico. Os Dados Governamentais Abertos (DGA) de qualidade são aceleradores fundamentais para a Agenda 2030 da ONU. A relação entre a qualidade desses dados e os Objetivos de Desenvolvimento Sustentável (ODS) é direta: dados imprecisos tornam-se obstáculos para mensurar progressos, identificar populações vulneráveis ou aplicar recursos com eficiência. Ao contribuir diretamente para a Ciência da Informação e Administração Pública a pesquisa alinha-se aos objetivos de ODS. Destaca-se a contribuição no âmbito do ODS 9 (Indústria, inovação e Infraestrutura) dados constituem a infraestrutura digital do século XXI. A gestão urbana moderna (*Smart Cities*) depende inteiramente da integridade dos dados abertos contribuindo para o alcance do objetivo 11 (Cidades e comunidades Sustentáveis). A qualidade dos dados é o que transforma a “transparência de fachada” em transparência real, instituições eficazes, tomada de decisão inclusiva e acesso à informação, são metas do objetivo 16 (Paz, Justiça e Instituições Eficazes). O objetivo 17 (Parcerias e Meios de Implementação) apresenta como meta específica (17.18) aumentar significativamente a disponibilidade de dados de alta qualidade, atuais e confiáveis, reforçando a necessidade de pesquisas que contribuam para o aumento da qualidade dos dados públicos.

Potential impact of this research

The potential impact of this research is multidimensional, ranging from the efficiency of public management to the strengthening of democracy and technological innovation. The availability of quality data directly impacts governance, transparency, efficiency in evidence-based decision-making, the promotion of innovation, economic development, and scientific and methodological advancement. High-quality Open Government Data (OGD) are fundamental accelerators for the UN's 2030 Agenda. The relationship between the quality of this data and the Sustainable Development Goals (SDGs) is direct: inaccurate data become obstacles to measuring progress, identifying vulnerable populations, or applying resources efficiently. By directly contributing to Information Science and Public Administration, the research aligns with the SDG objectives. Its contribution to SDG 9 (Industry, Innovation and Infrastructure) is particularly noteworthy, as data constitute the digital infrastructure of the 21st century. Modern urban management (*Smart Cities*) depends entirely on the integrity of open data, contributing to the achievement of goal 11 (Sustainable Cities and Communities). Data quality is what transforms "facade transparency" into real transparency; effective institutions, inclusive decision-making, and access to information are goals of objective 16 (Peace, Justice and Strong Institutions). Objective 17 (Partnerships for the Goals) presents as a specific target (17.18) significantly increasing the availability of high-quality, current and reliable data, reinforcing the need for research that contributes to increasing the quality of public data.

SIMONE CRISTINA CERON RIPOLI

**ESTUDO E ANÁLISE DA QUALIDADE DOS DADOS GOVERNAMENTAIS
ABERTOS (DGA) NO CONTEXTO DO ACESSO, USO E REÚSO DOS DADOS**

Tese apresentada ao Programa de Pós-graduação em Ciência da Informação (PPGCI) da Universidade Estadual Paulista “Júlio de Mesquita Filho” (Unesp), como requisito para a obtenção do título de Doutora em Ciência da Informação.

Área de concentração: Informação, Tecnologia e Conhecimento.

Linha de pesquisa: Informação e Tecnologia.

Banca Examinadora

Prof. Dr. José Eduardo Santarém Segundo

Universidade Estadual Paulista Júlio de Mesquita Filho (Unesp), Faculdade de Filosofia e Ciências (FFC), Campus de Marília, SP

Profa. Dra. Natália Marinho do Nascimento

Universidade Estadual Paulista Júlio de Mesquita Filho (Unesp), Faculdade de Filosofia e Ciências (FFC), Campus de Marília, SP

Prof. Dr. Caio Saraiva Coneglian

Universidade Estadual Paulista Júlio de Mesquita Filho (Unesp), Faculdade de Filosofia e Ciências (FFC), Campus de Marília, SP

Profa. Dra. Ananda Fernanda de Jesus

Universidade Federal de São Carlos, Centro de Educação e Ciências Humanas, Departamento de Ciência da Informação, São Carlos, SP

Profa. Dra. Priscila Machado Borges Sena

Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT), Rio de Janeiro, RJ

Marília, 17 de março de 2026.

Aos meus pais, que sempre me inspiraram com seu amor, dedicação e exemplo de vida, mostrando-me a importância do esforço e da perseverança.

AGRADECIMENTOS

Agradeço, primeiramente, ao meu marido, Danilo, pelo amor, paciência e incentivo incondicional em todos os momentos. Sua presença constante, mesmo nos dias mais difíceis, foi essencial para que eu seguisse firme e confiante. Obrigada por compreender as ausências, celebrar as pequenas conquistas e acreditar neste sonho tanto quanto eu. Obrigada por estar ao meu lado ao longo desta jornada.

Aos meus pais, Zelindo e Nadir, minha eterna gratidão por tudo o que sou. Pela base sólida, pelos valores transmitidos, pela educação e pelo apoio inabalável em cada etapa de minha trajetória, mesmo quando a distância se fez presente.

Agradeço a Deus pelas oportunidades e pela resiliência encontrada nos dias mais difíceis.

À Universidade Estadual Paulista “Júlio de Mesquita Filho” e ao Programa de Pós-Graduação em Ciência da Informação, pela oportunidade da realização deste doutorado e pela infraestrutura concedida.

Ao meu orientador, Prof. Dr. José Eduardo Santarém Segundo, agradeço pelas orientações, pelo rigor acadêmico, pela escuta atenta e por acreditar neste projeto desde o início. Agradeço pelo incentivo e dedicação do seu tempo à minha pesquisa.

Agradeço aos membros da banca de defesa: Prof. Dr. Caio Saraiva Coneglian, Profa. Dra. Priscila Machado Borges Sena, Profa. Dra. Natália Marinho do Nascimento, Profa. Dra. Ananda Fernanda de Jesus, por aceitarem fazer parte deste momento tão importante na minha vida.

A jornada foi longa, mas o aprendizado foi maior que o cansaço.

[...] os dados abertos não existem com um fim em si mesmo. As promessas de entrega de dados abertos estão principalmente do lado da legitimação dos órgãos de governo na adesão a uma boa prática de transparência, controle social, participação social e economia digital (Barbalho, 2018).

RESUMO

Em um contexto no qual a informação e o conhecimento assumem papel central no desenvolvimento socioeconômico, fortemente influenciado pelas tecnologias digitais e pela internet, os dados tornam-se instrumentos fundamentais para sustentar a estrutura organizacional. Nesse sentido, as organizações apoiam-se em dados para subsidiar a tomada de decisões, impulsionar a inovação e gerar valor. O avanço das Tecnologias da Informação e Comunicação tem possibilitado novas formas de interação, superando barreiras de tempo e espaço. No setor público, iniciativas como a Lei de Acesso à Informação e o Movimento de Dados Governamentais Abertos ampliaram a transparência, embora a qualidade desses dados ainda apresente desafios de representação e de uso. Esse cenário motivou o estudo e a análise da qualidade dos Dados Governamentais Abertos no contexto do acesso, uso e reúso dos dados. O objetivo geral desta pesquisa é desenvolver um instrumento de avaliação da qualidade dos dados governamentais abertos, fundamentado em normas, legislações e boas práticas nacionais e internacionais, com o propósito de aprimorar o acesso, o uso e o reúso. Trata-se de um estudo de natureza aplicada, com caráter exploratório e abordagem qualitativa. Utiliza como procedimentos metodológicos a Revisão de Literatura, a Revisão Sistemática da Literatura e pesquisa documental que permitem a construção de uma base sólida e a identificação de práticas e padrões relacionados à qualidade dos Dados Governamentais Abertos. A análise dos dados foi realizada por meio da análise de conteúdo. O instrumento de avaliação da qualidade dos dados proposto define dimensões mínimas a serem consideradas na construção e disponibilização de conjuntos de dados, de modo a elevar seu nível de qualidade e, conseqüentemente, ampliar seu aproveitamento social. Complementarmente, conduziu-se uma pesquisa documental com análise de conjuntos de dados governamentais abertos. Com base nesses levantamentos, o estudo definiu um conjunto de dimensões e métricas essenciais para a avaliação de dados. Como resultados, a pesquisa apresenta um instrumento validado que contempla as dimensões de acessibilidade, completude, atualidade, formato do arquivo, disponibilidade, precisão, recuperabilidade, inconsistência, compreensibilidade, conformidade, licença aberta, usabilidade, não discricionário, primário e interoperabilidade. A aplicação do instrumento por meio da Prova de Conceito demonstrou que, das cinquenta e sete métricas avaliadas, trinta (52,63%) foram classificadas como excelentes, dezessete (29,82%) como aceitáveis e dez (17,54%) como críticas, evidenciando um desempenho predominantemente positivo do conjunto de dados analisado, indicando aspectos que demandam atenção e aprimoramento. Conclui-se que o instrumento proposto é tecnicamente viável e efetivo para mensurar a qualidade dos Dados Governamentais Abertos, oferecendo ao gestor público um meio para elevar a qualidade dos dados disponibilizados e ampliar seu aproveitamento social.

Palavras-chave: Dados Governamentais Abertos (DGA); qualidade dos dados; dimensões de qualidade; informação pública; métricas da qualidade de dados.

ABSTRACT

In a context where information and knowledge play a central role in socioeconomic development, heavily influenced by digital technologies and the internet, data become fundamental instruments for sustaining organizational structure. In this sense, organizations rely on data to support decision-making, drive innovation, and generate value. Advances in Information and Communication Technologies have enabled new forms of interaction, overcoming barriers of time and space. In the public sector, initiatives such as the Access to Information Law and the Open Government Data Movement have increased transparency, although the quality of this data still presents challenges in terms of representation and use. This scenario motivated the study and analysis of the quality of Open Government Data in the context of access, use, and reuse of data. The overall objective of this research is to develop an instrument for evaluating the quality of open government data, based on national and international standards, legislation, and best practices, with the purpose of improving access, use, and reuse. This is an applied study, with an exploratory character and a qualitative approach. This study utilizes methodological procedures such as Literature Review, Systematic Literature Review, and documentary research, allowing for the construction of a solid foundation and the identification of practices and standards related to the quality of Open Government Data. Data analysis was performed using content analysis. The proposed data quality assessment instrument defines minimum dimensions to be considered in the construction and dissemination of datasets, in order to raise their quality level and, consequently, expand their social use. Complementarily, a documentary research was conducted with the analysis of open government datasets. Based on these surveys, the study defined a set of essential dimensions and metrics for data evaluation. As a result, the research presents a validated instrument that encompasses the dimensions of accessibility, completeness, timeliness, file format, availability, accuracy, recoverability, inconsistency, comprehensibility, compliance, open license, usability, non-discretionary nature, primary nature, and interoperability. The application of the instrument through Proof of Concept demonstrated that, of the fifty-seven metrics evaluated, thirty (52.63%) were classified as excellent, seventeen (29.82%) as acceptable, and ten (17.54%) as critical, evidencing a predominantly positive performance of the analyzed dataset, indicating aspects that demand attention and improvement. It is concluded that the proposed instrument is technically feasible and effective for measuring the quality of Open Government Data, offering public managers a means to improve the quality of the data made available and expand its social use.

Keywords: Open Government Data (OGD); data quality; quality dimensions; public information; data quality metrics.

LISTA DE FIGURAS

Figura 1 -	Síntese dos procedimentos metodológicos.....	39
Figura 2 -	Esquema das Cinco Estrelas em grau crescente de abertura do formato dos dados	61
Figura 3 -	Comparação entre as diferenças dos dados estruturados, semiestruturados e não estruturados	68
Figura 4 -	Representação gráfica dos tipos de dados	69
Figura 5 -	Evolução das tecnologias digitais no governo	71
Figura 6 -	Facetas de um setor público fundamentado em dados	72
Figura 7 -	Trajetória histórica do governo digital no Brasil	74
Figura 8 -	Contextualização histórica dos marcos relevantes dos DGA.....	83
Figura 9 -	Contextualização histórica da legislação acerca da Política de Dados Abertos no Brasil.....	92
Figura 10 -	O CVD.....	99
Figura 11 -	Valores ausentes	136
Figura 12 -	Células corrompidas	156

LISTA DE QUADROS

Quadro 1 –	Apresentação esquemática do processo de RSL	41
Quadro 2 –	Protocolo da RSL.....	42
Quadro 3 –	Combinações das palavras-chave	43
Quadro 4 –	Apresentação dos artigos científicos selecionados na RSL.....	47
Quadro 5 –	Dimensões para avaliação de qualidade dos DGA.....	50
Quadro 6 –	Métricas aplicadas às dimensões de qualidade.....	53
Quadro 7 –	Representação de dados estruturados	65
Quadro 8 –	Representação de dados não estruturados	66
Quadro 9 –	Representação de dados semiestruturados.....	67
Quadro 10 –	Seis dimensões para um governo digital	75
Quadro 11 –	Conceito de Governo Aberto por organismos internacionais	79
Quadro 12 –	Conceitos de DGA	85
Quadro 13 –	Oito Princípios de <i>Sebastopol</i>	87
Quadro 14 –	Princípios de <i>Sunlight Foundation</i>	88
Quadro 15 –	Os Quatros Ecossistemas dos DGA	88
Quadro 16 –	Principais legislações da Política de dados abertos no Brasil.....	90
Quadro 17 –	Legislação brasileira sobre DGA.....	93
Quadro 18 –	Consolidação das 15 dimensões de QD selecionadas	107
Quadro 19 –	Recursos de acessibilidade	133
Quadro 20 –	Critérios de referência para avaliação da qualidade acessibilidade	135
Quadro 21 –	Critérios de referência para avaliação dos valores ausentes.....	136
Quadro 22 –	Critérios de referência da cobertura dos dados coletados	138
Quadro 23 –	Critérios de referência dos campos obrigatórios.....	139
Quadro 24 –	Critérios de avaliação da comparabilidade	140
Quadro 25 –	Critérios de avaliação de metadados	141
Quadro 26 –	Critérios de avaliação da distribuição temporal.....	142
Quadro 27 –	Frequência da atualização dos dados abertos.....	144
Quadro 28 –	Critérios de avaliação da frequência de atualização dos dados abertos	145
Quadro 29 –	Cálculo da qualidade do tempo de registro.....	146
Quadro 30 –	Tempo entre a ocorrência do evento e do registro no sistema	146

Quadro 31 –	Critérios para avaliação da atualidade dos dados abertos.....	146
Quadro 32 –	Critérios para avaliação da atualização de metadados.....	147
Quadro 33 –	Critérios de avaliação para formato dos dados acessíveis	148
Quadro 34 –	Critérios de avaliação da multiplicidade dos formatos	149
Quadro 35 –	Critérios de avaliação de dados abertos em formato compactado	150
Quadro 36 –	Critérios de avaliação da documentação do formato utilizado	151
Quadro 37 –	Frequência dos períodos disponíveis	152
Quadro 38 –	Avaliação dos <i>hiperlinks</i> de <i>downloads</i> funcionais	152
Quadro 39 –	Critérios de avaliação da disponibilidade de canais de comunicação	153
Quadro 40 –	Critérios de avaliação de múltiplos idiomas	154
Quadro 41 –	Exemplos de taxas de dados corrompidos	155
Quadro 42 –	Taxa de dados corrompidos.....	156
Quadro 43 –	Avaliação de dados duplicados.....	157
Quadro 44 –	Avaliação do grau de correspondência dos dados abertos com a realidade	159
Quadro 45 –	Critérios de avaliação da ausência de ambiguidades	160
Quadro 46 –	Critérios de avaliação da função de recuperação de dados	161
Quadro 47 –	Grau de eficiência dos processos e sistemas de recuperação	162
Quadro 48 –	Avaliação da métrica navegação	163
Quadro 49 –	Taxa de falhas na recuperação de dados abertos	164
Quadro 50 –	Metadados recuperáveis.....	165
Quadro 51 –	Critérios de avaliação dos metadados recuperáveis.....	166
Quadro 52 –	Critérios para avaliação dos vocabulários controlados	167
Quadro 53 –	Critérios de avaliação da coerência lógica.....	168
Quadro 54 –	Critérios de avaliação da consistência entre conjuntos de dados..	169
Quadro 55 –	Critério de avaliação da coerência em dados agregados	170
Quadro 56 –	Critérios de avaliação do uso de classificações padronizadas	171
Quadro 57 –	Critérios de avaliação da presença de dicionários, notas explicativas ou legendas	172
Quadro 58 –	Critérios de avaliação da consistência na linguagem	173
Quadro 59 –	Critérios de avaliação das descrições claras e detalhadas dos dados abertos	175

Quadro 60 –	Recursos de linguagem adequada a diferentes públicos.....	176
Quadro 61 –	Critérios de referência para avaliação da linguagem adequada	177
Quadro 62 –	Recursos de avaliação de valores padronizados.....	178
Quadro 63 –	Critérios de referência para avaliação dos valores padronizados (datas, moedas, medidas e outros).....	180
Quadro 64 –	Proteção de dados pessoais e sensíveis	181
Quadro 65 –	Conformidade com a LGPD	181
Quadro 66 –	Avaliação da métrica proteção de dados pessoais e sensíveis	182
Quadro 67 –	Origem dos dados abertos.....	182
Quadro 68 –	Grau de identificação de origem dos dados abertos.....	183
Quadro 69 –	Responsáveis legais por dados pessoais ou sensíveis disponibilizados.....	184
Quadro 70 –	Grau de clareza institucional.....	184
Quadro 71 –	Existência de licenças.....	186
Quadro 72 –	Critérios de avaliação da existência de licenças.....	186
Quadro 73 –	Critérios de avaliação dos tipos de licenças	187
Quadro 74 –	Critérios para avaliação de permissões para reutilização de licenças	188
Quadro 75 –	Critérios de avaliação do período de validade para acesso e uso dos dados abertos	189
Quadro 76 –	Critérios de avaliação da ausência de restrições (Domínio Público)	190
Quadro 77 –	Critérios de avaliação da declaração de Domínio Público	191
Quadro 78 –	Requisitos para acesso aos dados disponíveis para <i>download</i> direto	192
Quadro 79 –	Critérios de avaliação para dados abertos disponíveis por <i>download</i>	193
Quadro 80 –	Grau de interface amigável	194
Quadro 81 –	Critérios de avaliação de dados abertos disponíveis diretamente aos usuários.	196
Quadro 82 –	Critérios de avaliação dos custos de acesso aos dados abertos...	196
Quadro 83 –	Fonte de dados abertos	198
Quadro 84 –	Processo dos dados documentais	198

Quadro 85 –	Rastreabilidade dos dados abertos.....	199
Quadro 86 –	Critérios de avaliação da capacidade de reutilização dos dados abertos	200
Quadro 87 –	Critérios de avaliação da integração sem sistemas externos.....	201
Quadro 88 –	Critérios de avaliação de cada dimensão.....	203
Quadro 89 –	Critérios de avaliação geral do conjunto de dados	204
Quadro 90 –	Síntese das métricas com qualificação excelente	209
Quadro 91 –	Síntese das métricas com qualificação aceitável	211
Quadro 92 –	Síntese das métricas com qualificação crítico	243
Quadro 93 –	Sugestões para melhoria da avaliação das métricas	216

LISTA DE TABELAS

Tabela 1 -	Resultado da RSL nas bases de dados selecionadas	44
Tabela 2 -	Diferença entre completude e precisão.....	117
Tabela 3 -	Cálculo e fórmula da qualidade acessibilidade	134
Tabela 4 -	Cálculo e fórmula da qualidade valores ausentes.....	136
Tabela 5 -	Cálculo e fórmula da qualidade cobertura de dados coletados.....	137
Tabela 6 -	Cálculo e fórmula da qualidade campos obrigatórios	139
Tabela 7 -	Fórmula da qualidade idade dos dados	143
Tabela 8 -	Cálculo e fórmula da qualidade taxa de dados corrompidos.....	156
Tabela 9 -	Cálculo e fórmula da qualidade de dados duplicados	157
Tabela 10 -	Cálculo e fórmula da qualidade grau de correspondência	159
Tabela 11 -	Cálculo e fórmula da qualidade taxa de falhas na recuperação de dados	163
Tabela 12 -	Fórmula da qualidade coerência lógica.....	168
Tabela 13 -	Fórmula da qualidade coerência entre conjunto de dados.....	169
Tabela 14 -	Fórmula da qualidade coerência em dados agregados	170
Tabela 15 -	Fórmula da qualidade uso de classificações.....	171
Tabela 16 -	Fórmula da qualidade consistência na linguagem	173
Tabela 17 -	Fórmula da qualidade descrições claras e detalhadas dos dados.	174
Tabela 18 -	Fórmula da qualidade linguagem adequada a diferentes públicos	177
Tabela 19 -	Cálculo e fórmula da qualidade valores padronizados (datas, moedas, medidas e outros).....	179
Tabela 20 -	Fórmula da qualidade conjunto de dados disponíveis para <i>download</i> sem necessidade de autenticação.....	193
Tabela 21 -	Cálculo e fórmula da qualidade das dimensões do conjunto de dados.....	202
Tabela 22 -	Cálculo e fórmula para avaliação geral do conjunto de dados.....	203
Tabela 23 -	Cálculo e fórmula para avaliação do peso efetivo de cada dimensão	204
Tabela 24 -	Cálculo e fórmula para avaliação da contribuição da dimensão do índice geral.....	205

LISTA DE ABREVIATURAS E SIGLAS

.gz	Acrônimo de Gzip
.tar	Tape ARchiver
.tar.gz	<i>Tape Archive</i>
.zip	<i>Zip File Format</i>
2D	Duas dimensões
3D	Três dimensões
a.C.	Antes de Cristo
AA/AAA	<i>Authentication, Authorization, and Accounting</i>
ABNT	Associação Brasileira de Normas Técnicas
Acess	Acessibilidade
ANPD	Autoridade Nacional de Proteção de Dados
API RESTful	API REST
API	<i>Application Programming Interface</i>
art.	Artigo
Atual	Atualidade
BI	<i>Business Intelligence</i>
BID	Banco Interamericano de Desenvolvimento
Cadúnico	Cadastro Único para Programas Sociais
CC BY – ND	<i>Creative Commons Attribution NoDerivatives</i>
CC BY– NC	<i>Creative Commons Attribution Não Comercial</i>
CC BY	<i>Creative Commons Attribution</i>
CC	<i>Creative Commons</i>
CC0 1.0	<i>Creative Commons Zero</i>
CC0	<i>Creative Commons Zero</i>
CE	Ceará
CEPAL	Comissão Econômica para a América Latina e o Caribe
CERN	Centro Europeu de Pesquisas Nucleares
CF	Constituição Federal
CGU	Controladoria Geral da União
CI	Ciência da Informação
CNAE	Classificação Nacional de Atividades Econômicas

Comp	Completude
Compr	Comprensabilidade
Cons	Consistência
CPF	Cadastro de Pessoa Física
CRM	<i>Customer Relationship Management</i>
CSV	<i>Comma-Separated Values</i>
CVD	Ciclo de Vida dos Dados
DCAT	<i>Data Catalog Vocabulary</i>
DCAT-AP	<i>Application Profile for Data Portals in Europe</i>
DF	Distrito Federal
DGA	Dados Governamentais Abertos
DOC	<i>Document</i>
DPO	<i>Data Protection Officer</i>
Dra.	Doutora
Dr.	Doutor
DWG	<i>Drawing</i>
ENAP	Escola Nacional de Administração Pública
ERP	<i>Enterprise Resource Planning</i>
ES	Espírito Santo
EUA	Estados Unidos da América
Ex.	Exemplo
FFC	Faculdade de Filosofia e Ciências
G8	Grupo dos oito
GI	Gestão da Informação
Gzip	GNU zip.
h	Horas
HTML	<i>Hypertext Markup Language</i>
HTTP	<i>Hypertext Transfer Protocol</i>
IBGE	Instituto Brasileiro de Geografia e Estatística
IBRAOP	Instituto Brasileiro de Auditoria de Obras Públicas
ID	Identidade
INDA	Infraestrutura Nacional de Dados Abertos
INDE	Infraestrutura Nacional de Dados Espaciais

INEP	Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira
IPCA	Índice Nacional de Preços ao Consumidor Amplo
ISO	<i>International Organization for Standardization</i>
JSON	<i>JavaScript Object Notation</i>
LAI	Lei de Acesso à Informação
LGPD	Lei Geral de Proteção de Dados Pessoais
Libras	Língua Brasileira de Sinais
LISTA	<i>Library, Information Sciênciã & Technology Abstracts with Full Text</i>
MB	<i>Megabyte</i>
MG	Minas Gerais
MIT	<i>Massachusetts Institute of Technology</i>
n.º	Número
NA	Não Aplicável
NBR	Norma Brasileira
NoSQL	<i>Not Only Structured Query Language</i>
OCD-BY	<i>Open Data Commons Attribution License</i>
OCDE	Organização para Cooperação e o Desenvolvimento Econômico
OCI	<i>Oracle Cloud Infrastructure</i>
ODS	<i>Planilha OpenDocument</i>
OEA	Organização dos Estados Americanos
OGL	<i>Open Government Licence</i>
OGP	<i>Open Government Directive</i>
OKBR	<i>Open Knowledge Brasil</i>
OKF	<i>Open Knowledge Foundation</i>
PB	Paraíba
PDA	Plano de Dados Abertos
PDDL	<i>Open Data Commons Public Domain Dedication and Licence</i>
PDF	<i>Portable Document Format</i>
PI	Profissional da Informação
PNDA	Política Nacional de Dados Abertos
PoC	Prova de Conceito
PPGCI	Programa de Pós-Graduação em Ciência da Informação
PR	Paraná

Prec	Precisão
Prof.	Professor
Profa.	Professora
PSD	Acrônimo de <i>Photoshop Document</i>
QD	Qualidade dos Dados
RJ	Rio de Janeiro
RDF	<i>Resource Description Framework</i>
REST	<i>Representational State Transfer</i>
RG	Registro Geral
RL	Revisão de Literatura
RS	Rio Grande do Sul
RSL	Revisão Sistemática da Literatura
RSS	<i>Really Simple Syndication</i>
SC	Santa Catarina
SC	Sem classificação
SciELO	<i>Scientific Electronic Library Online</i>
SIC	Sistema de Informação ao Cidadão
SP	São Paulo
SPARQL	<i>Protocol and RDF Query Language</i>
SQL	<i>Strutured Query Language</i>
T atual	Tupla atualidade
Temp	Tempo
TI	Tecnologia da Informação
TIC	Tecnologia da Informação e Comunicação
Tmax	Tupla máxima
UF	Unidade Federativa
UFG	Universidade Federal de Goiás
UFRGS	Universidade Federal do Rio Grande do Sul
Unesp	Universidade Estadual Paulista “Júlio de Mesquita Filho”
URI	<i>Unified Resource Identifier</i>
URL	<i>Uniform Resource Locator</i>
Usab	Usabilidade
W3C	<i>World Wide Web Consortium</i>
WCAG	<i>Web Content Accessibility Guidelines</i>

WoS	<i>Web of Science</i>
WWW	<i>Word Wide Web</i>
XLS	<i>Excel Spreadsheet</i>
XML	<i>Extensible Markup Language</i>

LISTA DE SÍMBOLOS

%	Porcentagem ou Porcentagem
*	Asterisco
§	Parágrafo
KG	Quilograma
Km	Quilômetro
R\$	Símbolo do Real

SUMÁRIO

1	INTRODUÇÃO	24
1.1	Problema da pesquisa.....	27
1.2	Tese e hipótese	29
1.3	Objetivos	30
1.3.1	Objetivo Geral.....	30
1.3.2	Objetivos Específicos.....	30
1.4	Justificativa	31
1.5	Estrutura da tese	35
2	PROCEDIMENTOS METODOLÓGICOS	37
2.1	Revisão Sistemática da Literatura.....	39
2.2	Pesquisa Documental.....	45
3	REVISÃO SISTEMÁTICA DA LITERATURA (RSL)	47
4	DADOS ABERTOS	54
4.1	Origem dos Dados Abertos	54
4.2	Conceito e classificação de Dados Abertos.....	58
4.3	Tipos de dados	63
4.4	Evolução histórica da política de governo digital no Brasil	68
4.5	Evolução dos DGA	76
4.6	Definição de DGA	83
4.7	Legislações que contribuíram para a abertura dos Dados Governamentais	88
5	A QUALIDADE DOS DADOS GOVERNAMENTAIS ABERTOS (DGA)	95
5.1	A qualidade dos Dados Abertos.....	96
5.2	Dimensões para avaliar a qualidade dos dados abertos	102
5.2.1	Acessibilidade.....	108
5.2.2	Compleitude	110
5.2.3	Atualidade.....	112

5.2.4	Formato de Arquivo	113
5.2.5	Disponibilidade	114
5.2.6	Precisão.....	116
5.2.7	Recuperabilidade	118
5.2.8	Inconsistência	120
5.2.9	Compreensibilidade	121
5.2.10	Conformidade	122
5.2.11	Licença Aberta/Não Proprietária	124
5.2.12	Usabilidade.....	125
5.2.13	Não Discricionário	126
5.2.14	Primário	127
5.2.15	Interoperabilidade.....	128
6	DEFINIÇÃO DO INSTRUMENTO PARA MENSURAÇÃO DA QUALIDADE DOS DADOS GOVERNAMENTAIS ABERTOS (DGA)	130
6.1	Dimensão: Acessibilidade	131
6.2	Dimensão: Completude.....	134
6.3	Dimensão: Atualidade	142
6.4	Dimensão: Formato De Arquivo	146
6.5	Dimensão: Disponibilidade.....	150
6.6	Dimensão: Precisão.....	154
6.7	Dimensão: Recuperabilidade.....	159
6.8	Dimensão: Inconsistência.....	166
6.9	Dimensão: Compreensibilidade	170
6.10	Dimensão: Conformidade	179
6.11	Dimensão: Licença Aberta/Não Proprietária	183
6.12	Dimensão: Usabilidade	190
6.13	Dimensão: Não Discricionário.....	194
6.14	Dimensão: Primário	195
6.15	Dimensão: Interoperabilidade	198
7	PROVA DE CONCEITO (PoC).....	205
7.1	Análise dos resultados da avaliação	207

8	CONSIDERAÇÕES FINAIS	216
	REFERÊNCIAS	220
	APÊNDICES	235
	Apêndice A – Instrumento para avaliação da qualidade dos dados abertos.	235
	Apêndice B – Prova DE Conceito	284

1 INTRODUÇÃO

Todas as organizações, sejam elas públicas ou privadas, precisam de dados. Alguns setores, em uma frequência e quantidade mais intensas que outros. Em um contexto geral, dados são um “[...] conjunto de fatos distintos e objetivos, relativos a eventos [...]” (Davenport; Prusak, 2018, p. 4). Em um contexto organizacional, “[...] dados são descritos como registros estruturados de transações [...]” (Davenport; Prusak, 2018, p. 4).

Na Ciência da Informação (CI), dados são definidos como as representações convencionais, codificadas, de uma informação de forma a permitir seu armazenamento, transmissão e processamento eletrônico (Le Coadic, 1996). No mesmo sentido, Kitchin (2014, p. 1) define dados “como a matéria-prima produzida pela abstração do mundo em categorias, medidas e outras formas representacionais - números, caracteres, imagens, sons, bits que constituem os blocos de construção a partir dos quais informação e conhecimento são criados”. Assim, a informação é um conjunto de dados que permite extrair algum significado, podendo favorecer a obtenção de conhecimento (Lima; Alvares, 2012, p. 24).

No contexto dos dados governamentais, dados são definidos como sequência de valores ou símbolos, apresentados por quaisquer meio, produzidos ou acumulados pelos órgãos públicos como resultados de um processo natural ou artificial, que não estejam sob sigilo ou restrição de acesso (Brasil, 2016b).

Os dados estão transformando a forma como as organizações públicas e privadas estruturam e conduzem seus negócios. A crescente demanda por dados públicos por diversos setores da sociedade, em geral, na condução de suas atividades intensifica a necessidade de sistemas capazes de conectar, de maneira eficiente, quem disponibiliza os dados e quem necessita deles.

Assim como um edifício depende de uma fundação sólida para sustentar sua estrutura, as organizações na era digital apoiam-se em dados para orientar a tomada de decisões, promover inovação e gerar valor econômico. Nesse contexto, Mahanti (2019) destaca que os dados deixaram de ser meros subprodutos dos sistemas de Tecnologia da Informação (TI), passando a construir ativos estratégicos dotados de valor real e mensurável para as organizações. Empresas dos setores público e privado utilizam dados para analisar comportamentos, prever demandas e otimizar recursos.

O avanço contínuo da Tecnologia da Informação e Comunicação (TIC) tem possibilitado novas formas de interação e de mediação da informação, modificando a maneira como os dados são produzidos, compartilhados e consumidos. Conforme destaca Sant’Ana (2009, p. 12), as TICs vêm “[...] quebrando as barreiras de espaço e tempo, propiciando novas formas de interação [...]”. Mahanti (2019) observa que a expansão na oferta de dados transformou não apenas os negócios, mas também a forma como realizamos transações, pesquisas e comunicações.

No setor público, a relevância dos dados intensificou-se por meio de iniciativas como a Lei n.º 12.527/2011 - Lei de Acesso à Informação (LAI) e através do fortalecimento de políticas de governo digital¹, que impulsionam a ampliação da disponibilização de dados sob sua guarda. Contudo, até o reconhecimento do acesso à informação como direito fundamental, ocorrido após o ano de 2000, muitos desses dados eram acessíveis aos cidadãos apenas por meio de relatórios estatísticos ou mediante transações formais (Nikiforova; McBride, 2021).

Nesse cenário, a combinação entre avanços tecnológicos, aumento da demanda social e iniciativas governamentais deu origem ao Movimento de DGA. O propósito do DGA é fortalecer a relação entre Estado e sociedade no compartilhamento de dados governamentais e públicos. Segundo Silva, Monteiro e Reis (2020, p. 32), a abertura de dados públicos busca “[...] disponibilizar informações referentes à gestão governamental em formato aberto e acessível, fornecendo acesso e instrumentos para que os cidadãos passem de meros espectadores a sujeitos ativos no processo da ação pública”.

Segundo o *Open Knowledge Foundation* (OKF, 2010), os dados são abertos quando “[...] podem ser acessados, utilizados, modificados e compartilhados livremente para qualquer finalidade por qualquer pessoa (sujeito, no máximo, a requisitos que preservem a proveniência e a abertura)”. A definição proposta pelo OKF (2015) expressa que os dados são considerados abertos quando podem ser

1

1988 – Constituição Federal (CF) e o direito de acesso à informação pública;
 2000 – Lei de Responsabilidade Fiscal;
 2009 – Lei da Transparência;
 2011 – LAI;
 2016 – Política de Dados Abertos do Poder Executivo Federal, também conhecida por Política Nacional de Dados Abertos (PNDA);
 2016 – Política de Governança Digital;
 2016 – Portal de Dados Abertos;
 2019 – Política Nacional de Governo Aberto.

acessados, utilizados, modificados e compartilhados livremente por qualquer pessoa, para qualquer finalidade, sem restrições indevidas.

A abertura de dados públicos pode gerar benefícios significativos, respondendo a demandas de múltiplos atores sociais (Vetrò *et al.*, 2016, p. 3). Contudo, a simples disponibilização dos dados não é suficiente para garantir seu efetivo uso e reúso. Para que os dados abertos possam de fato cumprir sua função pública, é necessário que apresentem níveis adequados de qualidade, possibilitando que sejam “[...] utilizados de maneira e conveniência do interessado de tal forma que eles possam ser misturados e combinados para agregar mais valor aos dados” (Diniz, 2010, p. 1). Os dados devem apresentar elevada qualidade, uma vez que, “[...] enquanto bons dados são uma fonte de inúmeras oportunidades, dados ruins são um fardo tremendo e só apresentam problemas [...]” para a organização (Mahanti, 2019, p. 29).

A Qualidade dos Dados (QD) é concebida como um construto multidimensional, com uma definição popular de “adequação para uso” que pode depender de inúmeros fatores (Zaveri *et al.*, 2012, p. 6). Sebastian-Coleman (2013, p. 123), complementa ao afirmar que “[...] a qualidade dos dados é definida por dois fatores relacionados: quão bem eles atendem às expectativas dos consumidores de dados e quão bem eles representam os objetos, eventos, e conceitos que são criados para representar”.

A QD representa a busca por informações confiáveis e precisas, capazes de servir de apoio em decisões, análises científicas e orientação em políticas públicas. A QD é um compromisso fundamental, que deve ser firmado por instituições produtoras e gestoras de dados, e assegurar que suas informações sejam completas, consistentes, acessíveis e atualizadas permitindo o uso e reúso. Diante da importância atribuída aos dados públicos, a qualidade destes dados pode ser considerada um valor institucional.

No contexto dos DGA e demais realidades de dados abertos, a qualidade não se resume à disponibilidade. É necessário que os dados estejam acessíveis, compreensíveis e úteis, com valor público agregado (Silva *et al.*, 2021). Falhas no fornecimento de dados de boa qualidade podem comprometer não apenas a reutilização dos dados, mas também a eficácia e a usabilidade dos portais institucionais. Vetrò *et al.* (2016, p. 3) ressaltam que os portais de DGA só cumprem sua função, quando os dados disponibilizados atendem aos requisitos que permitem

acesso fácil, consulta, processamento e integração com outros dados, sem restrições.

A avaliação da qualidade dos conjuntos de dados é realizada a partir de dimensões de qualidade. Wang e Strong (1996, p. 6) definem dimensão como “[...] um conjunto de atributos de qualidade de dados que representam um único aspecto ou construção da qualidade de dados”.

O termo dimensão foi escolhido por estudiosos da DQ e informação para “[...] identificar os aspectos dos dados que podem ser medidos e por meio dos quais sua qualidade pode ser quantificada” (Sebastian-Coleman, 2013, p. 125). As dimensões fornecem um meio de avaliar e gerenciar a QD. Ao definir as dimensões de qualidade é necessário que o objetivo seja o acesso, uso e reúso dos dados pelo usuário, e sem complicações para isso. Batini e Scannapieco (2006, p. 19) afirmam que “[...] cada dimensão captura um aspecto específico incluído sob o guarda-chuva geral da qualidade dos dados”.

Por exemplo, a dimensão “completude” assegura que os dados estejam completos; a dimensão “atualidade” garante a capacidade dos dados refletirem de forma fidedigna os eventos ou fenômenos mais recentes, a acessibilidade garante à inclusão de pessoas portadoras de deficiência. Desta forma, cada dimensão atua de forma específica para que os DGA sejam confiáveis para uso e reúso.

Após expor brevemente o contexto da pesquisa, a seguir apresenta-se a delimitação do problema que a motivou no âmbito da qualidade dos DGA.

1.1 Problema da pesquisa

Apesar dos avanços significativos nas políticas de abertura de dados e na implementação e fortalecimento das plataformas de dados governamentais, o ato de disponibilizá-los não assegura seu uso efetivo pela sociedade.

Além do aumento da transparência pública, a disponibilização dos dados do setor público proporciona uma oferta que pode responder a todos os tipos de usuários (Vetrò *et al.*, 2016). Os “[...] dados governamentais são particularmente importantes devido à sua escala, amplitude e status como fonte confiável de informação sobre uma ampla gama de assuntos” (Máchová; Lnenicka, 2017, p. 22).

No entanto, disponibilizar dados por si só não é suficiente para torná-los valiosos e úteis; é a qualidade com que esses dados são apresentados que fazem a

diferença no seu uso e reúso. Conforme apontado por Bovee, Srivastava e Mak (2003, p. 312) “[...] dependemos implicitamente da qualidade da informação que usamos nas decisões, mas informações de baixa qualidade são uma fonte de perda de produtividade ou de fracasso empresarial”.

Os DGA frequentemente apresentam problemas de qualidade que comprometem seu potencial de gerar valor público e impactam diretamente na capacidade da sociedade acessar, compreender, integrar e reutilizar os dados de forma eficaz. Entre alguns problemas na publicação de dados de qualidade estão a falta de padronização, profissionais habilitados para elaboração e análise, ausência de protocolos claros, falta de governança dos dados, dados incompletos, desatualizados, inconsistentes, entre outros.

Nesse contexto, o profissional da informação tem papel fundamental. Segundo Valentim (2004, p. 140), o profissional da informação é o “[...] indivíduo que recebeu formação específica para trabalhar com dados, informação e conhecimento, bem como sua mediação”. Os profissionais da informação utilizam-na de “maneira estratégica em seu ambiente de trabalho, desenvolvendo recursos informacionais e estabelecendo serviços”; a sua atuação vai além da gestão documental ou tecnológica, trata-se de um agente mediador entre a informação e o usuário (Abels et al., 2003).

Esses fatores levam os usuários de dados a enfrentarem problemas como a dificuldade de acesso, problemas de compreensibilidade limitada, de localização, de barreiras técnicas e de integração entre sistemas.

Embora existam iniciativas e recomendações internacionais sobre práticas para publicação de dados, ainda há carência de instrumentos metodológicos específicos que permitam avaliar, de forma sistemática e mensurável, a adequação dos dados governamentais para acesso, uso e reúso (Bovee; Srivastava; Mak, 2003, p. 311). A falta de instrumentos de avaliação da QD dificulta não apenas o diagnóstico de suas deficiências, mas também a implementação de ações corretivas e o monitoramento da evolução da qualidade ao longo do tempo.

Para que os portais de DGA sejam mais do que apenas eficientes, é necessário que alcancem níveis de efetividade e relevância social, o que exige uma avaliação sistemática da qualidade desses dados. Esse processo possibilita não apenas compreendê-los de forma mais aprofundada, mas também analisar os diferentes tipos de valor que podem gerar (Máchová; Lnenicka, 2017).

Ao oferecer critérios claros e padronizados na avaliação da QD busca-se apoiar a gestão pública na melhoria dos seus conjuntos de dados, fortalecer a confiabilidade, disponibilizar e ampliar a capacidade de geração de valor público por meio do acesso, uso e reutilização desses objetos. Para alcançar esse objetivo, este estudo busca responder às seguintes questões de pesquisa:

Como estabelecer parâmetros de avaliação da qualidade de DGA, com base em normas e legislações para o acesso, uso e reúso pela sociedade?

1.2 Tese e hipóteses

Esta **tese** defende que a qualidade dos DGA é fator determinante que pode transformar a publicação de dados oficiais e pode ser avaliada de forma mais objetiva e consistente por meio da proposição de um instrumento fundamentado em normas, legislações e boas práticas nacionais e internacionais, estruturado a partir de quinze dimensões e um conjunto de métricas operacionais pode transformar a gestão pública e fortalecer a interação entre governo e a sociedade. A abertura dos dados governamentais busca ampliar a transparência pública e promover o controle social, especialmente por meio do acesso amplo, facilitado e confiável às informações públicas.

Entretanto, no Brasil, o uso efetivo dos DGA enfrenta barreiras significativas por seus usuários, uma vez que a qualidade dos dados públicos disponibilizados é um dos principais desafios. Dados com lacunas, inconsistências, ausência de metadados ou formatos inadequados comprometem não apenas o acesso, mas também a interpretação e análise, o que reduz substancialmente o potencial de reúso e os benefícios esperados do “Governo Aberto”, como a transparência e o controle social, *accountability* e integridade, participação social, inovação e desenvolvimento econômico, melhorias na gestão pública, inclusão e equidades de direitos e fortalecimento democrático.

Nesse sentido, compreender e assegurar a QD é condição indispensável para que esses conjuntos de dados públicos atendam de forma adequada às demandas de diferentes perfis de usuários. Somente com níveis consistentes e verificáveis de qualidade será possível transformar dados brutos em insumo para pesquisas, formulação de políticas públicas, inovação e participação social.

A pesquisa fundamenta a **hipótese** de que, ao estabelecer critérios para análise da qualidade dos DGA, é possível avaliar de forma objetiva tanto a adequação da publicação dos dados quanto o potencial de uso e reúso. As falhas na qualidade dos conjuntos de dados reduzem significativamente sua confiabilidade e comprometem o aproveitamento efetivo dos DGA. A ausência de padrões consolidados de qualidade compromete a capacidade de utilização dos dados, dificultando sua integração, interpretação e aplicação prática. A adoção de parâmetros de qualidade bem definidos pode permitir que os dados atendam às expectativas e necessidades informacionais de diferentes perfis de usuários, ampliando o impacto, utilidade e valor público.

1.3 Objetivos

Considerando a tese e a hipótese apresentada, o objetivo geral e os específicos da pesquisa são apresentados na próxima seção.

1.3.1 Objetivo Geral

Desenvolver um instrumento de avaliação da qualidade dos dados governamentais abertos, fundamentado em normas, legislações e boas práticas nacionais e internacionais, com o propósito de aprimorar o acesso, o uso e o reúso.

1.3.2 Objetivos Específicos

Para alcançar o objetivo geral, os seguintes objetivos específicos são propostos:

- a) identificar os principais critérios utilizados na literatura científica e em documentos normativos que possibilitem a avaliação da qualidade de DGA;
- b) selecionar os critérios mais relevantes para a análise da QD, considerando todas as etapas desde sua produção, publicação, distribuição, disseminação e reúso;
- c) construir um instrumento para avaliar a qualidade dos DGA;
- d) testar sua aplicabilidade e consistência;

e) aplicar como prova de conceito o instrumento desenvolvido, em busca de sua validação, adequação metodológica e efetividade na mensuração da qualidade diante dos DGA.

1.4 Justificativa

O interesse em desenvolver um instrumento de avaliação da qualidade dos DGA está diretamente relacionado à trajetória profissional e acadêmica da pesquisadora. Como funcionária pública, o convívio cotidiano com questões que envolvem a produção, a gestão e o acesso à informação pública reforça a compreensão sobre a relevância social sobre o tema.

Além disso, a experiência adquirida durante o mestrado, cuja dissertação teve como objeto a organização das informações públicas para o amplo acesso à informação, consolidou o envolvimento da pesquisadora com estudos voltados à informação pública. Nesse sentido, o estudo sobre a qualidade dos DGA representa a continuidade e aprofundamento do percurso acadêmico e profissional da pesquisadora, alinhando-se ao compromisso de contribuir para o acesso à informação e para a melhoria dos serviços públicos.

A necessidade das pesquisas voltadas à qualidade dos DGA acentua-se diante da lacuna crítica na literatura, que carece de estudos aprofundados referenciais teóricos capazes de orientar o usuário quanto à qualidade dos dados. Essa escassez de conhecimento prejudica principalmente a construção de um conjunto de dimensões e métricas, fundamentais para diagnosticar e assegurar um padrão mínimo de qualidade que sustente a utilidade pública desses dados. Sem critérios de avaliação bem definidos, a abertura dos dados governamentais corre o risco de tornar-se meramente um instrumento de conformidade legal ao direito de acesso à informação, sendo incapaz de promover a transparência, a inovação social ou o embasamento de políticas públicas. Assim, o estudo da qualidade dos DGA torna-se essencial para transformar os dados públicos em ativos digitais de valor.

A escolha pelo PPGCI da Universidade Estadual Paulista “Júlio de Mesquita Filho” se justifica pela excelência acadêmica e científica consolidada da instituição, reconhecida nacional e internacionalmente pela produção de conhecimento inovador na área.

O Programa oferece uma sólida base teórico-metodológica e epistemológica, alinhada às demandas da sociedade da informação estruturando-se em três linhas de pesquisa, entre elas, a de “Informação e Tecnologia”. Essa linha se dedica aos estudos teóricos, epistemológicos e práticos relacionados à produção, ao processamento, à representação, ao acesso, à recuperação, à transferência, à visualização, ao design, à arquitetura, à utilização, à gestão e a preservação de dados, informação e de documentos em ambientes digitais (Unesp, 2025).

Essas pesquisas consideram os sistemas informacionais em contextos tecnológicos, organizacionais e sociais, articulando metodologias, instrumentos e estratégias de uso das TICs (Unesp, 2025). Nesse cenário, o objeto de estudo encontra consonância com a proposta e com linhas de pesquisa do PPGCI. Destaca-se, ainda, a relevância da atuação do Prof. Dr. José Eduardo Santarém Segundo, orientador desta pesquisa, cuja experiência acadêmica e produção científica nas linhas de pesquisa de Tecnologias e Web Semântica aplicadas na disponibilização de Dados Abertos e Ciência de Dados na CI conferem vasta experiência e conhecimento para o desenvolvimento da pesquisa.

Considerando as transformações vivenciadas pela sociedade contemporânea, impulsionadas pelo grande volume de informações que circulam em tempo real, a Administração Pública também precisou se adaptar a essas novas mudanças. Até a promulgação da CF de 1988, os dados públicos eram produzidos, coletados, processados, armazenados e descartados sem exercer impacto significativo sobre a sociedade.

Com o avanço das TICs, a Administração Pública passou a utilizar seus dados de forma estratégica, disponibilizando-os de maneira aberta e acessível. O *Open Data Movement* (OKFN) tem sido progressivamente adotado por governos em todo o mundo, estimulando o uso inteligente de dados públicos e a criação de iniciativas voltadas à melhoria da qualidade de vida dos cidadãos (Ruvalcaba-Gómez, 2019).

Além de favorecer a gestão pública, os dados abertos “[...] também são valiosos para todas as áreas da sociedade na tomada de decisões e criação de produtos e serviços inovadores” (Janssen, 2011, p. 446). Quando combinados entre si, ou entre conjunto de dados “[...] permitem a geração de novos conjuntos de dados, novas informações e conhecimentos” (Penteado, 2020, p. 27).

No entanto, grande parte destes dados não estão estruturados para facilitar a extração, compreensão e manipulação por parte dos usuários (Isotani; Bittencourt, 2015). Como resultado, “[...] a extração de informações e a produção de conhecimentos que poderiam ser úteis para a sociedade não acontecem com a agilidade e a eficácia necessária para lidar com as questões sociais e econômicas do século XXI” (Isotani; Bittencourt, 2015, p. 16).

Segundo Mahanti (2019), apenas 3% dos dados disponibilizados atendem aos padrões básicos de qualidade, e sua baixa qualidade traz consequências que afetam a eficiência, a credibilidade governamental, a economia e a sociedade.

Entre essas consequências, destaca-se a perda de confiança e credibilidade, pois nenhum usuário confiaria em um novo *insight* empolgante se for baseado em dados de baixa qualidade. No entanto, essa vulnerabilidade é mais evidente quando o usuário não possui competência informacional necessária, ou seja, quando não possuem habilidades para avaliar a origem, a confiabilidade e a consistência dos dados. Os usuários críticos em informação são capazes de identificar falhas, questionar resultados e validar evidências (Mahanti, 2019).

A baixa QD também acarreta impactos econômicos, como dados de má qualidade que possuem uma estimativa de custo para a organização de aproximadamente 20% da sua receita (Mahanti, 2019). Isso pode acarretar desperdício de recursos financeiros e de oportunidades de inovação e investimento, especialmente no poder público.

A eficiência operacional também é prejudicada, pois cerca de 40% do tempo das equipes responsáveis em tratar os dados é gasto na identificação e correção de problemas encontrados (Moses; Gavish; Vorwerck, 2024).

A falta de QD pode levar ainda a decisões equivocadas e políticas públicas ineficazes, comprometendo a efetividade da gestão pública. Essas falhas podem gerar alocação inadequada de recursos, falhas em ações emergenciais e riscos sociais e legais, como aumento da desinformação, perda de direitos ou violação de normas legais ou regulamentos (Bovee; Srivastava; Mak, 2003).

Dados de alta qualidade são aqueles que atendem efetivamente às expectativas e necessidades dos usuários, possuindo atributos como confiabilidade, completude e acurácia (Moses; Gavish; Vorwerck, 2024; Sebastian-Coleman, 2013). Como ressalta Mahanti (2019, p. 29), “[...] o valor dos dados não reside apenas nos dados em si, mas também nas ações que surgem a partir deles e em seu uso”.

Reconhecer que grande parte dos dados governamentais disponibilizados possui problemas de qualidade que dificultam seu acesso, uso, reúso e distribuição é fator essencial para uma gestão desses objetos (Tayi; Ballou, 1998). Para Kitchin (2014, p. 1), dados de boa qualidade devem ser inteligíveis, acompanhados de metadados, capazes de se integrar a outros conjuntos de dados e de gerar descobertas inéditas. Alcançar padrões como esses exige a definição de estratégias, instrumentos e ferramentas que apoiem a produção de DGA com atributos de qualidade.

O primeiro passo para garantir dados de boa qualidade é compreender que eles possuem múltiplos atributos e dimensões, e que atingir a qualidade ideal para todos os perfis de usuário é tarefa complexa. Por isso, é fundamental contar com metodologias e instrumentos capazes de avaliar e assegurar a qualidade, desde a publicação até o acesso, uso, reúso e distribuição.

O desenvolvimento de um instrumento de avaliação da qualidade dos DGA revela-se imprescindível para orientar e apoiar os profissionais da informação na execução de práticas mais precisas, consistentes e alinhadas aos princípios de transparência e governança informacional. A ausência de protocolos claros e padronizados na avaliação dos dados de pesquisa abertos tem conduzido a avaliações fragmentadas, baseadas em percepções empíricas e sem respaldo técnico uniforme.

Nesse sentido, a criação de um instrumento de avaliação estruturado contribui para operacionalizar o conceito de qualidade da informação, oferecendo critérios objetivos, mensuráveis e replicáveis que possibilitam o diagnóstico, o monitoramento e a melhoria contínua dos dados públicos. A proposta não apenas fortalece o papel estratégico dos profissionais da informação na gestão do conhecimento público, como também promove uma mudança paradigmática, ao transformar a avaliação da qualidade informacional em um processo técnico-científico e sistemático, capaz de sustentar decisões.

Considerando o cenário apresentado, esta pesquisa propõe o desenvolvimento de um instrumento de avaliação intuitivo, capaz de atender tanto a especialistas da informação quanto ao público geral.

Nesse cenário, a Ciência da Informação (CI) desempenha papel central, por ser o campo científico dedicado ao estudo da informação, preocupando-se com a análise, coleta, classificação, manipulação, armazenamento, recuperação e

disseminação da informação. No contexto digital, a CI concentra-se na organização e representação da informação e do conhecimento, atuando de forma interdisciplinar para aperfeiçoar os sistemas de recuperação da informação na *Web* (Macedo; Lemos, 2021). No caso dos dados e, especialmente, os dados governamentais, não é diferente, pois a CI reconhece que os dados são recursos estratégicos para a sociedade e para o Estado.

Nesse contexto, Macedo e Lemos (2021) destacam que a CI enfatiza princípios teóricos e metodológicos voltados à organização da informação, por meio de suas representações descritivas e temáticas. No cenário digital, em constante evolução, essa abordagem propicia organização, acesso e compartilhamento eficiente dos dados.

A aplicação dos conhecimentos da CI à gestão dos DGA contribui diretamente para promover a transparência, acessibilidade, compreensibilidade e reúso dos dados. Além disso, estudos sobre o tema na área da CI permitem o desenvolvimento de instrumentos, padrões e protocolos capazes de estruturar e representar informações de forma eficiente no ambiente digital. Tais contribuições efetivam-se na construção de mecanismos de padronização, na definição de requisitos de metadados, na promoção da interoperabilidade dos dados e na adoção de práticas de organização e descrição da informação orientada aos usuários.

1.5 Estrutura da tese

Para compreensão do desenvolvimento desta tese, será apresentada a estruturação do conteúdo abordado. Na primeira seção, intitulada “Introdução”, é exposto o tema, a problemática, os objetivos e a justificativa, que orientam a pesquisa. Nesta seção, também são apresentados os procedimentos metodológicos utilizados para realizar a pesquisa.

Na segunda seção, “Revisão Sistemática da Literatura (RSL)” são apresentados os resultados obtidos a partir da pesquisa realizada por meio da RSL, cujo propósito foi identificar quais as dimensões empregadas na avaliação da qualidade dos DGA. Nesta seção são descritos os estudos recuperados nas bases de dados de referência da CI, bem como as dimensões de qualidade apontadas por cada autor, permitindo uma análise comparativa e o reconhecimento das principais abordagens teóricas e metodológicas adotadas na literatura científica sobre o tema.

A terceira seção, intitulada “Dados Abertos”, teve como finalidade abordar o conceito de dados abertos, discorrendo sobre seu marco teórico e aspectos relevantes da temática. Aborda-se também como o movimento da transparência das informações públicas influenciou na adoção de portais de dados abertos na administração pública, promovendo acesso facilitado à informação pública.

A quarta seção, “A Qualidade dos dados governamentais abertos (DGA)”, apresenta o referencial teórico acerca da QD, sua definição e os aspectos que a envolvem. Apresentam-se quinze dimensões de QD obtidas na RSL.

A seção cinco, intitulada “Definição do instrumento para mensuração da qualidade dos Dados Governamentais Abertos (DGA)” é dedicada à construção do instrumento metodológico voltado à avaliação dos conjuntos de dados governamentais abertos, disponível no Portal de Dados Abertos do Governo Federal. Nessa etapa, são identificadas e descritas as métricas de qualidade aplicáveis na prática, considerando sua pertinência, operacionalidade e relevância para o contexto da gestão pública de dados. A análise foi conduzida a partir da aplicação das quinze dimensões de qualidade apresentadas na seção quatro, com o propósito de verificar a eficácia e a coerência dos critérios propostos na mensuração da QD.

Na seção seis será realizada a Prova de Conceito (PoC), etapa em que se procede a análise dos conjuntos de DGA com propósito de validação prática do instrumento desenvolvido para avaliar a qualidade. A partir dessa aplicação prática, será desenvolvido um guia/protocolo de referência, composto pelas dimensões e métricas de qualidade definidas na etapa anterior. O guia/protocolo visa demonstrar como a adoção das métricas propostas pode conferir mais qualidade aos conjuntos de dados.

Na seção sete, intitulada “Considerações finais”, são apresentadas as reflexões e conclusões decorrentes da pesquisa desenvolvida, reunindo os resultados decorrentes da qualidade dos DGA. A seção demonstra ainda, possíveis contribuições da pesquisa para o campo da CI e para o aperfeiçoamento das políticas públicas de gestão e abertura dos dados públicos.

2 PROCEDIMENTOS METODOLÓGICOS

Trata-se de um estudo de natureza aplicada, com caráter exploratório e abordagem qualitativa. Utiliza como procedimentos metodológicos a Revisão Sistemática da Literatura (RSL), Pesquisa Documental que permite a construção de uma base sólida e identificar práticas e padrões relacionados à qualidade dos DGA. A análise dos dados foi realizada por meio da análise de conteúdo.

Por ser de natureza aplicada, esta pesquisa busca gerar conhecimento científico com foco na aplicação prática, especialmente na resolução de problemas relacionados à qualidade dos DGA.

Do ponto de vista dos objetivos propostos, trata-se de uma pesquisa do tipo exploratória, desenvolvida “[...] com o objetivo de proporcionar visão geral, de tipo aproximativo, acerca de determinado fato” (Gil, 2024, p. 26). Em uma pesquisa exploratória, “[...] os pesquisadores primeiro coletam dados qualitativos, os analisam e depois usam as informações para desenvolver sua pesquisa [...]” (Creswell; Clark, 2013, p. 165).

Segundo Marcone (2025, p. 91), na pesquisa exploratória, uma variedade de procedimentos de coletas de dados pode ser empregada. Dessa forma, a pesquisa exploratória pode ser dividida em: estudos exploratórios descritivos combinados, com finalidade de descrever completamente um determinado fenômeno; estudos que usam procedimentos específicos de coleta de dados para o desenvolvimento de ideias, geralmente utilizam exclusivamente um único método, como a análise de conteúdo, para extrair generalizações com o objetivo de produzir categorias conceituais; e, estudos de manipulação experimental, que tem por finalidade manipular uma variável independente.

Assim, a escolha pela abordagem exploratória justifica-se por possibilitar uma análise profunda e interpretativa, adequada à natureza qualitativa e aplicada do estudo. A abordagem exploratória possibilita também, compreender o fenômeno e propor instrumentos e soluções práticas voltadas à melhoria da qualidade dos DGA.

Em relação à abordagem, optou-se por uma pesquisa qualitativa, por considerar que esse tipo de perspectiva é mais adequado à investigação de significados, interpretações e percepções. Segundo Ternoski (2022), a pesquisa qualitativa valoriza a subjetividade do objeto de estudo. Nesse sentido, Flick (2009, p. 20) aponta que a “[...] pesquisa qualitativa é de particular relevância ao estudo das

relações sociais devido à pluralização das esferas da vida [...] o que exige uma nova sensibilidade para a investigação empírica das questões”. Assim, a abordagem qualitativa permite explorar significados, percepções e interpretações. A escolha adequada de métodos e teorias, a consideração de diferentes perspectivas e a reflexão crítica do pesquisador são aspectos fundamentais dessa abordagem (Flick, 2009).

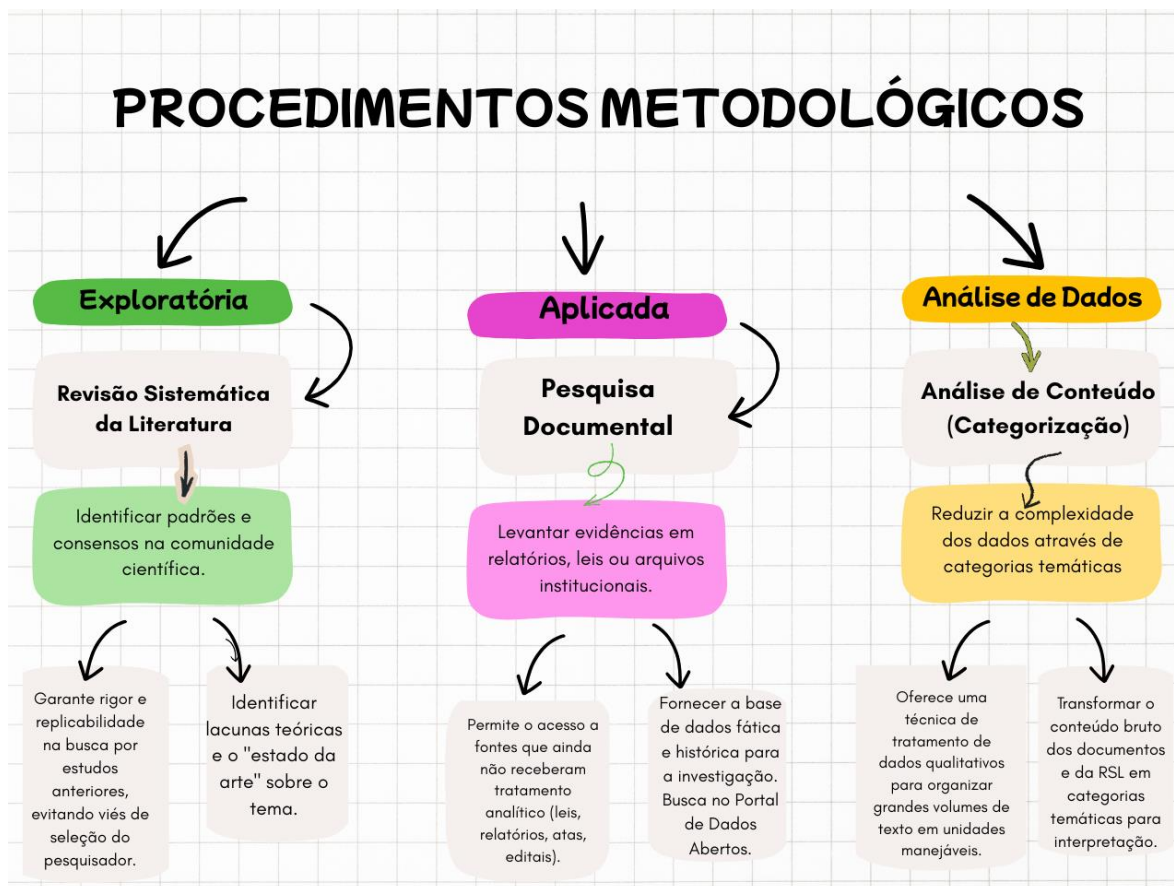
Para maior aprofundamento com o objeto de estudo e construção do referencial teórico, foi realizada uma RL mediante consulta a múltiplas fontes. Segundo Gil (2018, p. 43), a “[...] familiaridade com o tema proposto é essencial para que o problema seja formulado de maneira clara e precisa”. A RL é um elemento essencial da pesquisa acadêmica, pois, “[...] não apenas fornece uma compreensão abrangente de estudos anteriores sobre um tópico, mas também produz uma interpretação integrativa dos resultados” (Ermel *et al.*, 2022, p. 41).

Para a realização da RL foram utilizadas fontes disponíveis no Portal de Periódicos da Capes, que reúne artigos, livros, periódicos, instituições e pesquisadores de relevância. O Portal permite acesso a conteúdos indexados (mesmo que não assinados pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Capes), ampliando as possibilidades de consulta a materiais de diversas fontes (Capes, 2024) e em diversos idiomas.

Por se tratar de um tema pouco explorado na literatura, não foi definido um recorte temporal específico para o levantamento bibliográfico. Para a RL, foram utilizados os termos “*data quality*”, “qualidade dos dados”, “*data quality dimensions*” e “dimensões de qualidade dos dados”, o que permitiu a busca e análise das publicações.

Para uma maior compreensão, a figura 1 apresenta uma síntese dos procedimentos metodológicos utilizados nesta pesquisa.

Figura 1 – Síntese dos procedimentos metodológicos



Fonte: Elaborado pela autora (2026).

A partir da sistematização apresentada, as subseções seguintes descrevem detalhadamente cada uma das etapas, fundamentando as escolhas que permitiram o desenvolvimento da pesquisa.

2.1 Revisão Sistemática da Literatura

Com o objetivo de identificar, analisar e sintetizar estudos relevantes acerca da QD foi realizada uma RSL. De acordo com Ermel *et al.* (2022, p. 9), a RSL “[...] visa mapear, avaliar e agregar os resultados de estudos sobre um determinado tema ou área do conhecimento”. Por meio de seus procedimentos sistemáticos, a RSL permitiu identificar estudos relevantes, estabelecer critérios de inclusão e exclusão de textos, além de possibilitar a definição de métodos de síntese e análise, garantindo a consistência e a credibilidade do levantamento.

A escolha pelo método justifica-se por sua capacidade de contextualizar a pesquisa realizada e fornecer uma análise aprofundada do estado da arte a respeito do tema. A RSL possibilita contribuições originais e confiáveis para a comunidade científica e sociedade em geral (Faria, 2016), a partir de sua análise minuciosa das fontes.

Com base nos resultados obtidos por meio da RSL, foi possível formular o problema e os objetivos da pesquisa, bem como elaborar um planejamento metodológico estruturado. Como destaca Severino (2017, p. 1606) “[...] a ciência se faz quando o pesquisador aborda os fenômenos aplicando recursos técnicos, seguindo um método e apoiando-se em fundamentos epistemológicos”.

Nesse sentido, a escolha do método baseou-se na necessidade de responder aos objetivos propostos e garantir a cientificidade da investigação. Segundo Gil (2024, p. 9) para que um conhecimento seja considerado científico, “[...] torna-se necessário identificar as operações mentais e técnicas que possibilitam a sua verificação, ou seja, determinar o método que possibilitou chegar a esse conhecimento”.

Segundo Ermel *et al.* (2022, p. 09), a realização de revisões sistemáticas para investigar o conhecimento científico e tecnológico é uma condição necessária para a produção de pesquisas originais. A RSL é definida por Lopes e Fracolli (2008, p. 772), como uma “[...] síntese de estudos primários que contém objetivos, materiais e métodos claramente explicitados e que foi conduzida de acordo com uma metodologia clara e reproduzível”.

A RSL é um meio de identificar, selecionar, avaliar, interpretar e sumarizar todas as pesquisas disponíveis para uma determinada questão de pesquisa, área ou fenômeno de interesse (Kitchenham, 2004, p. 1). Neste contexto, com a finalidade de “[...] construir um caminho conceitual e metodológico com critérios definidos, cujo intuito será o de atribuir validade científica às obras selecionadas, a fim de que não se construa um corpus de referencial bibliográfico incompleto e ineficiente [...]”, adotou-se a RSL (Faria, 2016, p. 17).

Kitchenham (2004c) indica três razões para a realização de uma RSL: I – resumir as evidências existentes relativas a um tratamento ou tecnologia; II – identificar qualquer lacuna na investigação atual; III – fornecer uma estrutura para posicionar adequadamente novas atividades de pesquisa.

Nesse sentido, a revisão da literatura é sistemática, “[...] porque segue um conjunto determinado de passos dentro das bases de dados científicas, mas não ignora todo o acervo de publicações em que se verifique um reconhecimento por parte da comunidade acadêmica” (Faria, 2016, p. 20). Ao seguir passos determinados e processos explícitos e sistemáticos para revisar a literatura existente, os dados fornecidos tornam-se confiáveis.

O resultado de uma RSL realizada com base em todas as etapas é a construção de um “[...] conhecimento novo e integrativo que pode servir de base para a construção de novos modelos ou teorias, ao invés de uma mera descrição narrativa das pesquisas atuais” (Ermel *et al.*, 2022, p. 43).

No Quadro 1 é apresentado o **desenho** metodológico com a descrição das fases a serem seguidas na RSL.

Quadro 1 - Apresentação esquemática do processo de RSL

Objetivos
Definir a problemática a ser estudada, sintetizada em uma questão ou problema
Equações de Pesquisa
Encontrar expressões/palavras-chave que traduzem o assunto da pesquisa (por meio de um tesauro) utilizando AND, OR, NOT (*, ?)
Âmbito da Pesquisa
Bases selecionadas e variantes intrínsecas.
Critérios de Inclusão
Definem que o estudo é aceitável em determinado contexto.
Critérios de Exclusão
Excluem os estudos que não obedecem ao âmbito definido.
Critérios de Validade Metodológica
Verificação da objetividade da pesquisa pela replicação do processo de dois investigadores.
Resultados
Devem ser registrados todas as etapas da pesquisa.
Tratamento dos Dados e Resultados
Filtrar, analisar e descrever criticamente os resultados com apoio de <i>software</i> de gestão

bibliográfica.

Fonte: Faria (2016, p. 22).

O processo de RSL deve ser alinhado à problemática e aos seus objetivos da pesquisa. Segundo Faria (2016), cabe ao pesquisador a construção de um modelo de etapas, que permita selecionar, avaliar e sintetizar a literatura mais relevante para o tema. Sob essa perspectiva metodológica, foi elaborado o protocolo de pesquisa detalhado no Quadro 2, que serviu como guia para a condução da presente RSL.

Quadro 2 – Protocolo da Revisão Sistemática da Literatura

Protocolo de pesquisa	
Pergunta da Pesquisa	Quais as dimensões utilizadas para a avaliação da qualidade dos DGA?
Objetivos	Verificar estudos já realizados acerca da qualidade dos DGA.
Equações de pesquisa	“qualidade da informação” AND “dados governamentais abertos”; “qualidade” AND “dados governamentais abertos”; “qualidade dos dados” AND “dados governamentais abertos”.
Âmbitos da pesquisa	Portal de Periódicos da Capes; Library; LISTA; WoS; SciELO; Scopus e ScienceDirect.
Período abrangido	Sem restrição temporal
Idiomas	Português, inglês e espanhol
Tipos de documentos	Artigos publicados em periódicos científicos revisados por pares.
Critérios de inclusão	Foco na temática da qualidade dos dados; textos completos; textos disponíveis para download de forma gratuita.
Critérios de exclusão	Estar em idioma diverso do estabelecido na pesquisa; não tratar especificamente da temática; o documento não estar disponível gratuitamente na web.
Critérios de validade metodológica	Na primeira rodada foram eliminados os artigos duplicados. Na segunda rodada foi realizada leitura técnica dos resumos e palavras-chave. Na terceira rodada foi realizada uma análise crítica dos artigos resultantes.
Formulário de extração	Identificação e contexto do estudo; aspectos metodológicos; núcleo da pesquisa: qualidade de DGA.
Critérios para avaliação de qualidade dos trabalhos extraídos	Foram considerados apenas artigos científicos publicados em periódicos com revisão por pares, visando garantir a qualidade metodológica e a confiabilidade das informações extraídas para a análise da qualidade dos DGA.
Tratamento dos dados extraídos	a) Categorização por dimensão de qualidade, estabelecidas <i>a posteriori</i> ; b) Análise comparativa de dimensões; c) Agrupamento das dimensões de acordo com o objetivo de cada uma;

Fonte: Elaborado pela Autora (2025).

A aplicação do protocolo de pesquisa visou responder à seguinte questão norteadora: quais as dimensões utilizadas para a avaliação da qualidade dos dados governamentais abertos?

Após a definição do objetivo, foi realizado um teste de aderência, em que foram selecionadas as palavras-chave: “qualidade da informação”, “dados governamentais abertos” e “qualidade dos dados”. A utilização das palavras-chave na construção da string de busca contribuiu para que a pesquisa alcançasse trabalhos compatíveis ao objetivo da pesquisa e, que descartasse trabalhos que não estavam de acordo com a temática do estudo.

A *string* da pesquisa foi definida com uso do operador booleano AND, conforme demonstra o Quadro 2:

Quadro 3 - Combinações das palavras-chave

Combinações	
1	“qualidade da informação” AND “dados governamentais abertos”
2	“qualidade” AND “dados governamentais abertos”
3	“qualidade dos dados” AND “dados governamentais abertos”
4	“ <i>information quality</i> ” AND “ <i>open government data</i> ”
5	“ <i>quality</i> ” AND “ <i>open government data</i> ”
6	“ <i>data quality</i> ” AND “ <i>open government</i> ”

Fonte: Ripoli e Santarém Segundo (2025).

Das bases de dados referenciais da CI e campos correlatos, optou-se em realizar as pesquisas nas bases eletrônicas: Portal de Periódicos da Capes, *Library*, *Information Sciência & Technology Abstracts with Full Text* (LISTA), *Web of Science* (WoS), *Scientific Electronic Library Online* (SciELO), *Scopus* e *ScienceDirect*. A RSL foi realizada no mês de fevereiro de 2024.

Considerando que cada base de dados utilizada na pesquisa possui características próprias, a estratégia de busca foi ajustada de acordo com a aplicabilidade de cada uma, sem alterar o sentido e o âmbito das palavras-chave definidas, sempre utilizando o processo de busca avançada para determinar melhor o objeto de estudo.

Para a etapa de triagem e análise dos artigos científicos recuperados, foram definidos e aplicados critérios de elegibilidade, visando assegurar rigor e alinhamento ao escopo da pesquisa. Dessa maneira, os critérios de inclusão foram:

- foco na temática da qualidade dos dados;
- textos completos (*Full text*);
- textos disponíveis para *download* de forma gratuita;

Para a exclusão dos artigos científicos recuperados, foram utilizados os seguintes critérios:

- estar em idioma diverso do estabelecido na pesquisa;
- não tratar especificamente da temática;
- o documento não estar disponível integralmente na *web*;

A varredura nas bases de dados recuperou 1671 artigos científicos. Após a eliminação de 81 artigos científicos duplicados, restaram 1.563. Na etapa seguinte foi realizada uma leitura técnica dos resumos e palavras-chave, aplicando-se os critérios de inclusão e exclusão estabelecidos. Dessa forma, 1507 artigos que não estavam alinhados com a temática da pesquisa, foram descartados. Por fim, foi realizada a análise crítica de 56 artigos científicos resultantes. Apenas 23 artigos científicos, que possuíam alinhamento com o objetivo do estudo, foram selecionados.

A Tabela 1 apresenta a quantidade de artigos recuperados em cada base de dados utilizada.

Tabela 1 – Resultado da RSL nas bases de dados selecionadas

Fonte	Artigos recuperados
Portal de Periódicos Capes	88
LISTA	57
WoS	94
SciELO	51
Scopus	1002
ScienceDirect	379
Total	1671

Fonte: Ripoli e Santarém Segundo (2025).

Para a análise dos dados coletados foram utilizadas premissas da Análise de Conteúdo por meio de uma categorização realizada a *posteriori* em razão das características comuns entre os estudos. Por meio de análise dos estudos

selecionados, a categorização foi composta pelas dimensões de qualidade na recuperação de DGA.

2.2 Pesquisa documental

A pesquisa documental foi necessária para a coleta e análise das informações do conjunto de dados “Taxa de distorção Idade-série na Educação Básica” do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP), disponível no Portal de Dados Abertos do governo federal. A Pesquisa Documental serviu para base no aprofundamento dos conjuntos de dados disponibilizados pelo Portal de Dados Abertos.

2.3 Análise dos Dados

A análise dos dados foi conduzida por meio do método de Análise de Conteúdo proposto por Bardin (2016). A Análise de Conteúdo consiste em um conjunto de técnicas destinadas à análise das comunicações. De acordo com Bardin (2016, p. 37), a análise do conteúdo, “[...] não se trata de um instrumento, mas de um leque de apetrechos; ou, com maior rigor, será um único instrumento, mas marcado por uma grande disparidade de formas e adaptável a um campo de aplicação muito vasto: as comunicações”.

Segundo Bardin (2016, p. 44) “[...] a intenção da análise de conteúdo é a inferência de conhecimentos relativos às condições de produção, inferência esta que recorre a indicadores”. Dessa forma, foi aplicada a etapa de categorização, “[...] operação de classificação de elementos constitutivos de um conjunto por diferenciação e, em seguida, por reagrupamento segundo o gênero (analogia), com critérios previamente definidos” (Bardin, 2016, p. 147).

A Análise de Conteúdo foi adotada por sua adequação à natureza qualitativa da pesquisa e à necessidade de interpretação sistemática das informações coletadas na etapa da RSL. Conforme Bardin (2016), a Análise de Conteúdo permite que o pesquisador examine e interprete comunicações de modo a revelar significados implícitos. A Análise de Conteúdo possibilitou identificar dimensões e indicadores a partir de critérios previamente definidos.

Para a pesquisa, a Análise de Conteúdo mostrou-se o método mais adequado para identificar, sistematizar e classificar as dimensões e métricas de qualidade dos DGA. A aplicação da Análise de Conteúdo contribuiu para transformar dados descritivos em informações analíticas, estabelecendo conexão entre a teoria e a prática.

Na etapa de categorização, as unidades de análise foram identificadas e organizadas, permitindo ao pesquisador expressar suas compreensões sobre o fenômeno investigado. Para tanto, adotou-se como unidades centrais a QD qualidade dos dados e suas respectivas dimensões de qualidade. No contexto das pesquisas qualitativas, a análise categorial é uma das técnicas mais utilizadas por possibilitar a organização e interpretação dos dados em unidades de significados (Bardin, 2016). Dentre as formas de análise categorial, optou-se pela análise temática, por se tratar de um método que possibilita o desmembramento do texto em unidades.

3 REVISÃO SISTEMÁTICA DA LITERATURA (RSL)

Nesta seção serão apresentados os resultados obtidos por meio da RSL método empregado para identificar e analisar, de forma simultânea, o estado da arte referente às dimensões de qualidade dos DGA.

Após a conclusão das etapas de seleção dos artigos científicos, foi iniciada análise dos dados coletados, com o objetivo de identificar, organizar e interpretar as informações relevantes para a pesquisa.

O Quadro 4 apresenta os artigos científicos selecionados na RSL com o nome dos autores, ano de publicação e título.

Quadro 4 – Apresentação dos artigos científicos selecionados na RSL

AUTOR	ANO	TÍTULO
Moreira et al.	2020	A qualidade na recuperação de dados governamentais: um estudo sobre dados de políticas públicas na internet.
Máchová e Lnenicka	2017	Evaluating the quality of open data portals on the National Level.
Jetzek	2016	Managing complexity across multiple dimensions of liquid open data: the case of the Danish basic data program.
Vetrò et al.	2016	Open data quality measurement framework: definition and application to open government data.
Quarati	2017	Open government data: usage trends and metadata quality.
Fan e Zhao	2017	The moderating effect of external pressure on the relationship between internal organizational factors and the quality of open government data.
Kubler et al.	2018	Comparison of metadata quality in open data portals using the analytic hierarchy process.
Royo-Montañés e Benítez-Gómez	2019	Portales de datos abiertos. Metodología de análisis y aplicación a municipios españoles.
Herrera-Melo et al.	2019	Proposal for the evaluation of open data portals.
Dorobat; Posea	2021	Open data indicator: an Accumulative methodology for measuring the quality of open government data.
Silva; Monteiro; Reis	2020	Qualidade da Informação dos dados governamentais abertos: análise do portal de dados abertos brasileiro.
Zhang e Xiao	2020	Quality assessment framework for open government data.
Marmier e Mettler	2020	Developing an index for measuring OGD publisher compliance to good practice standards: insights from opendata swiss.
Raca et al.	2021	Application-based framework for analysis, monitoring and evaluation of national open data portals.
Noguera-Issó et al.	2021	Quality of metadata in open data portals.

Penteado et al.	2021	Process model with quality control or the production of high quality linked open government data.
Nikiforova e McBride	2021	Open government data portal usability: a user-centred usability analysis of 41 government data portals.
Slibar, Oreski e Begicevic	2021	Importance of the open data assessment: an insight into the (meta) data quality dimensions.
Ureña-Cámara; Nogueras-Iso; Lacasta	2021	Revisión de la calidad de los conjuntos de datos abiertos sobre presupuestos.
Raca et al.	2022	A framework for evaluation and improvement of open government data quality: application to the Western Balkans National Open Data Portals.
Martinez et al.	2022	Metrics proposal to measure the quality of governmental datasets.
Moradi, Mazoochi e Ahmadi	2022	A comprehensive method for improving the quality of open government data and increasing citizens willingness to use data by analyzing the complex system of citizens and organizations.
Maestre-Góngora, Rangel-Carrillo e Osorio-Sanabria	2023	Exploring open government data and big data from a quality perspective.

Fonte: Ripoli e Santarém Segundo (2025).

Após análise dos 23 artigos científicos e aplicação das categorias do estudo de caso, constatou-se a ausência de um consenso entre os autores quanto às dimensões e métricas de qualidade a serem utilizadas na disponibilização dos DGA.

No que diz respeito à avaliação da qualidade dos dados, Zaveri *et al.* (2015, p. 6) afirmam que “[...] a avaliação da qualidade dos dados envolve a medição de dimensões ou critérios de qualidade que são relevantes para o usuário”. Uma métrica de “[...] avaliação da qualidade dos dados é um procedimento para medir a dimensão da qualidade da informação” (Zaveri *et al.*, 2015, p. 6).

Conforme Penteado (2020, p. 87), a dimensão é o “[...] aspecto principal pelo qual a qualidade pode ser observada, comportando diversos critérios de qualidade para caracterizá-lo”. De forma complementar, Zaveri *et al.* (2015, p. 6), caracterizam as dimensões como conceitos abstratos, ou seja, “[...] as métricas de avaliação dependem de indicadores de qualidade que permitem a avaliação da qualidade de uma fonte de dados com relação aos critérios”.

No mesmo sentido, Moreira *et al.* (2020, p. 109), explicam que “[...] o conceito de qualidade é variável e subjetivo, e considera as necessidades, ações e objetivos

de cada domínio. Assim, a qualidade funciona como um conjunto de parâmetros ou requisitos que ajudam a verificar se aquele conteúdo atende aquilo que é necessário para um determinado uso” para que o conteúdo acessado esteja de acordo com os requisitos”.

A análise dos artigos científicos selecionados identificou 41 dimensões utilizadas pelos autores para avaliar a qualidade dos conjuntos de dados. No Quadro 5, são apresentadas as dimensões aplicadas pelos autores para avaliação da qualidade dos conjuntos de dados disponibilizados pelo setor público:

Como é possível observar por meio do Quadro 5, as dimensões com maiores incidências entre os autores foram: acessibilidade, completude, atualidade, formato do arquivo, disponibilidade, precisão, recuperabilidade, compreensibilidade, consistência, conformidade, livre de licença, usabilidade, não proprietário, não discricionário, primário, visualização e interoperabilidade.

Autores como Moreira *et al.* (2016) consideram que, no processo de verificação da qualidade dos dados, deve ser considerada a estrutura do sistema composta por dimensões de qualidade. Cada uma dessas dimensões é, constituída por métricas específicas, responsáveis por mensurar aspectos particulares do dado e que exercem influência direta sobre o nível de qualidade avaliado. Essas métricas podem ser classificadas em diferentes categorias, conforme o foco da análise:

- métricas relacionadas ao acesso, nas quais se encontram as relações de acesso dos usuários aos portais governamentais;
- métricas relacionadas ao conteúdo, que abrangem os elementos que compõem a informação disponibilizada nos portais;
- métricas relacionadas à estrutura dos conjuntos de dados, voltadas especificamente à estrutura dos conjuntos de dados;
- métricas de qualidade, também são denominadas de atributos ou critérios. Penteado (2020, p. 87) as define como “[...] uma característica particular dos dados, tanto objetiva (Ex.: acurácia do dado), quanto subjetiva (Ex.: reputação da fonte de dados)”.

Para exemplificar a aplicação das métricas nas dimensões de qualidade, apresenta-se o Quadro 6:

Quadro 6 – Métricas aplicadas às dimensões de qualidade

Dimensões	Sigla	Definição de métricas
Precisão (Prec)	Prec01	Precisão sintática: é a medida da distância entre o valor armazenado no banco de dados correto. Precisão sintática = número de valores corretos/número de valores totais.
	Prec02	Número de tuplas precisas entregues.
	Prec03	Questionário de pesquisa do usuário.
Completude (Comp)	Comp01	Número de valores não nulos/número total de valores.
	Comp02	Número de tuplas entregues/número esperado.

	Comp03	Completeness dos dados da Web = $\text{Tupla máxima (Tmax)} - \text{Tupla atualidade (T atual)}^*$ (Completeness Máxima. Completeness Atual)/2
	Comp04	Questionário de pesquisa do usuário.
Consistência (Cons)	Cons01	Número de valores consistentes/número de valores totais.
	Cons02	Número de tuplas violando restrições, número de diferenças de codificações.
	Cons03	Número de páginas com desvio guiado de estilo.
	Cons04	Questionário de pesquisa do usuário.
Tempo (Temp)	Temp01	Pontualidade.
	Temp02	Porcentagem de execuções de processos capazes de serem realizadas no prazo necessário.
	Temp03	Questionário de pesquisa do usuário.
Atualidade (Atual)	Atual01	Tempo em que os dados são armazenados no sistema -tempo em que os dados são atualizados no mundo físico.
	Atual02	Horário da última atualização.
	Atual03	Horário da solicitação - última atualização.
	Atual04	Idade + (Horário da entrega - Horário da entrada)
Acessibilidade (Acess)	Acess01	$\text{Max}(0; 1 - (\text{Tempo de entrega} - \text{tempo de solicitação}) / (\text{Tempo de prazo} - \text{tempo de solicitação}))$.
	Acess02	Número de <i>hyperlinks</i> quebrados-número de âncoras quebradas.
Usabilidade (Usab)	Usab01	Pesquisa de usuário – Questionário.
Compreensibilidade (Compr)	Compr01	Compreensão do consumidor de dados quanto ao que é representado no conjunto de dados.

Fonte: Batini *et al.* (2009, p. 19).

Por meio do Quadro 6 é possível observar que uma dimensão pode conter uma ou várias métricas, a depender da complexidade de cada dimensão a ser avaliada. Segundo Vetrò *et al.* (2016), as métricas distribuídas por dimensões possibilitam uma avaliação mais completa de um conjunto de dados abrangendo seus aspectos gerais, isto é, pode avaliar desde a estrutura do portal até seus conjuntos de dados.

4 DADOS ABERTOS

Esta seção aborda o tema dados abertos. Inicialmente, para uma melhor compreensão mais clara a respeito da temática, será apresentada contextualização sobre sua origem. Em seguida, discutem-se aspectos relevantes dos dados abertos, desde sua concepção até a sua adoção por diversos governos mundiais, que passaram a disponibilizar DGA ampliando a transparência pública e o acesso à informação.

A diversidade de dados abertos disponibilizados possibilita o armazenamento, o acesso e o *download* de uma grande quantidade de dados e informações para diferentes finalidades de pesquisa, análise e tomada de decisões. Esses dados atendem ao conceito de Dados Abertos, uma vez que são disponibilizados em formato e tecnologias abertas, ou seja, que permitem uso, modificação e distribuição livre, sem imposição de restrições proprietárias, os quais compõem uma plataforma eletrônica que viabiliza o acesso, o uso e reúso dos mesmos (Schulte; Chunpir; Vob, 2016, p. 197).

Com o avanço tecnológico, o compartilhamento de dados abertos tornou-se essencial para o cotidiano de indivíduos e organizações. Desta forma, o crescimento de dados disponível em formato aberto traz a necessidade de estudos que proporcionem o melhoramento da qualidade dos dados disponibilizados.

4.1 Origem dos dados abertos

A coleta, a organização, o processamento, o armazenamento e a manutenção de dados diversos tiveram início anterior a invenção dos computadores. O uso de dados na forma de registros é conhecido desde 2800 antes de Cristo (a.C.), no antigo Egito. Organizações religiosas e governos foram as primeiras organizações a coletar e explorar ativamente os dados (Lake; Crowther, 2013). Nesse sentido, a gestão de dados enquanto prática social é milenar.

Todo esse processo era realizado manualmente, os dados eram registrados em argila, pedras, paredes, papiro, pergaminho somente na Idade Média e Moderna passaram a ser registrados no papel e armazenados em arquivos físicos, o que tornava a recuperação um processo difícil e moroso.

Com os avanços proporcionados pela Revolução Industrial, entre o final do século XVII e o início do século XIX, o aumento da escala produtiva das fábricas passou a exigir registros detalhados para o gerenciamento das operações (Lake; Crowther, 2013). Nesse período, a tecnologia de cartão perfurado, inventada em 1890 por Herman Hollerith, foi amplamente utilizada por ser uma forma rápida de registrar e recuperar os dados (Mahanti, 2019).

Posteriormente, os cartões perfurados foram substituídos por fitas de papel. No entanto, por ainda apresentarem um custo elevado e inacessível para muitas organizações, essas fitas de papel foram rapidamente substituídas pelas fitas magnéticas baseada em discos, que permitiam armazenar um volume maior de dados e de forma mais acessível (Mahanti, 2019).

Para o armazenamento dos dados em discos, foi necessário organizá-los. As práticas de organização possibilitaram o desenvolvimento de estruturas teóricas e práticas que se tornaram banco de dados, permitindo o acesso a grandes volumes de dados de forma aleatória e rápida (Lake; Crowther, 2013).

Segundo Mahanti (2019, p. 97), um banco de dados pode ser definido como uma coleção estruturada de dados inter-relacionados, organizados e armazenados em sistemas computacionais de forma a facilitar a recuperação de maneira eficiente. Com os avanços tecnológicos, as dimensões, capacidades e desempenhos dos bancos de dados e seus respectivos sistemas de gerenciamento, cresceram em ordens de magnitude. A chamada “revolução dos bancos de dados”, ocorrida na década de 1980, tornou o acesso às informações contidas nos dados significativamente mais fácil e eficiente (Mahanti, 2019).

Desde 1960, os bancos de dados estão evoluindo significativamente. No início surgiram os bancos de dados hierárquicos e de rede, que organizavam as informações em estruturas rígidas e interligadas, adequadas às necessidades das primeiras aplicações empresariais e governamentais. Na década de 1970, foram introduzidos os bancos de dados orientados a objetos, que buscavam integrar os princípios da programação orientada a objetos ao armazenamento de dados. Em 1980, os bancos de dados relacionais ganham destaque por substituir as estruturas rígidas dos sistemas hierárquicos e de rede por uma abordagem baseada em tabelas inter-relacionadas, seguidos, na década de 1990, por versões relacionais mais sofisticadas. Já nos anos 2000, surgiram os bancos de dados *Not Only*

Structured Query Language (NoSQL)², e o banco de dados em nuvem. Atualmente, esses bancos de dados coexistem e são amplamente utilizados os modelos *Strutured Query Language* (SQL)³, NoSQL e banco de dados em nuvem (Mahanti, 2019).

Já o primeiro registro conhecido para o termo dados abertos remonta a 1995, em um documento emitido pela agência científica norte-americana *National Science Foundation* (NSF), no contexto da divulgação de dados geofísicos e ambientais (Carossi; Teixeira Filho, 2020). Segundo Isotani e Bittencourt (2015, p. 21), o documento expedido pela agência científica norte-americana, permitiu “[...] um completo e aberto intercâmbio de informações científicas entre diferentes países como um pré-requisito para a análise e a compreensão dos fenômenos naturais globais”.

Contudo, em 1942, Robert King Merton, já destacava a importância dos resultados de pesquisas serem abertos e acessíveis a todos. Essa perspectiva deu origem à chamada Teoria de Merton, que ressalta os benefícios dos dados científicos abertos (Chignard, 2013).

O século XX foi marcado também, por diversas revoluções, entre elas a revolução da informação, da comunicação e do conhecimento. Esse período foi responsável por grandes transformações tecnológicas, que fortaleceram a criação e a disseminação de diversos meios de comunicação técnica-científica, ampliando os recursos eletrônicos para a divulgação de informações (Araújo; Levino, 2012). Nesse momento, as TICs deram ênfase à filosofia dos bens comuns, com destaque a *World Wide Web* (WWW), que se tornou protagonista na rotina entre diversos ramos da sociedade e seus indivíduos (Bandeira *et al.*, 2015, p. 7; Chignard, 2013).

A *World Wide Web* ou *web* foi inventada em 1989, pelo físico inglês Sir Timothy John Berners-Lee, a partir da integração de três tecnologias: o *Hypertext Markup Language* (HTML), o *Hypertext transfer Protocol* (HTTP) e o (*Unified Resource Identifie* (URI). O objetivo inicial da criação da tecnologia era facilitar o compartilhamento de documentos entre os pesquisadores que participavam da conferência no *Conseil Européen pour la Recherche Nucléaire* (CERN). A realização

² Termo genérico que representa os bancos de dados não relacionais. Esses bancos de dados armazenam dados em um formato diferente das tabelas relacionais (Oracle Cloud Infrastructure - OCI, 2025).

³ São bancos de dados relacionais, é a base para consulta estruturada, que permite aos usuários acessar e manipular dados em tabelas altamente estruturadas (OCI, 2025).

de um congresso, juntamente com a primeira conferência internacional acerca da WWW realizada em Genebra, contribuiu significativamente para a popularização da *web* (Isotani; Bittencourt, 2015), especialmente em como é conhecida atualmente.

Dessa forma, os primeiros a reconhecerem os benefícios da abertura e do compartilhamento dos dados foram os membros da comunidade científica; muito antes dos dados abertos se tornarem um objeto técnico ou movimento científico (Chignard, 2013). Assim, acredita-se que a ciência ocupa uma posição de destaque em uma sociedade inovadora. Segundo Sena (2023, p. 7) “[...] transformações na ciência têm impacto no mercado e o mercado influencia na ciência”. Foi o encontro entre a ideia dos bens comuns com os ideais do *software* livre e de código aberto que moldou os dados abertos como é conhecido atualmente (Chignard, 2013).

Um importante acontecimento para a abertura dos dados ocorreu em 2009, quando Elinor Ostrom, por meio de sua pesquisa, demonstrou a especificidade dos bens comuns de informação, o que lhe rendeu o Prêmio Nobel de Economia daquele ano. Segundo Chignard (2013), os bens comuns de informação “[...] são muito semelhantes aos bens públicos, porque seu uso por uma pessoa não impede seu uso por outras. No entanto, esses são bens públicos de um novo tipo: não apenas seu uso não esgota o estoque comum, mas o enriquece”.

Outro marco da revolução digital que impactou na abertura dos dados, foi a digitalização. A digitalização de documentos afetou diretamente em duas características importantes dos dados abertos: I - quando os dados são acessíveis a mais de uma pessoa ao mesmo tempo, eles adquirem a característica de não rivalidade; e II - quando os custos na produção e distribuição dos dados são radicalmente reduzidos, a reutilização de dados se torna economicamente viável (Jetzek; Avital; Bjorn-Andersen, 2014).

Após a concepção do termo aberto, dados que anteriormente eram de domínio de um número limitado de pessoas, passaram a se tornar cada vez mais acessíveis. Com a ampla disponibilização de dados, as possibilidades de uso pela sociedade crescem continuamente, podendo consistir em simples consultas até fundamento para tomada de decisões.

Johnson (2014) eleva o debate sobre o movimento de dados abertos para um nível crucial, pois argumenta que apenas a disponibilização de informações não garante equidade. Nesse sentido, Johnson (2014) propõe o conceito de justiça informacional. Essa teoria está alicerçada em três pilares:

[...] a incorporação de privilégios sociais em conjuntos de dados à medida que os dados são construídos; as capacidades diferenciais dos usuários de dados; e, as normas que os sistemas de dados impõem por meio de sua função como sistemas disciplinares (Johnson, 2014, p. 263).

A justiça informacional, segundo Mathiesen (2016) é um conceito multifacetado que analisa a relação entre indivíduos e a informação sob três diferentes perspectivas de justiça. A primeira delas é denominada de justiça distributiva, pois considera que as pessoas desempenham papéis duplos: são buscadoras de informação e, ao mesmo tempo, fontes de informações. Nesse contexto, para que recebam um tratamento justo, as pessoas necessitam acessar a informação de forma equitativa. A segunda, denominada justiça participatória, enfatiza que, enquanto fontes de informação, os indivíduos devem receber oportunidades iguais de participar da produção e do fortalecimento do conhecimento. Por fim, a terceira, descrita como a justiça do reconhecimento, estabelece que, para serem sujeitos informacionais, as pessoas necessitam ser justamente representadas no conjunto geral de informações (Silva *et al.*, 2021).

Para melhor compreensão do termo “dados abertos”, serão detalhadas a seguir sua definição e suas principais classificações.

4.2 Conceito e classificação de dados abertos

Para clarificar o conceito de dados abertos, a seguir será apresentada sua definição e suas principais classificações.

Uma das principais definições do termo dados abertos é apresentada pela OFK (2010), que é descrita como “[...] dados que podem ser livremente usados, reutilizados e redistribuídos por qualquer pessoa - sujeitos, no máximo, à exigência de atribuição da fonte e compartilhamento pelas mesmas regras”. Com base nessa definição, Carossi e Teixeira Filho (2020, p. 15), explicam que a única exigência quanto ao uso dos dados abertos “[...] é que sua fonte original seja referenciada e que as licenças sejam preservadas quando esses dados foram compartilhados”. Destaca-se também “[...] que os dados podem ser utilizados e redistribuídos tanto por pessoas físicas quanto jurídicas, na geração de novos conhecimentos ou de novos negócios” (Carossi; Teixeira Filho, 2020, p. 15).

Segundo Isotani e Bittencourt (2015), o objetivo dos dados abertos é evitar a imposição de restrições sobre os dados a serem publicados, permitindo que todos os usuários possam explorá-los livremente. No mesmo sentido, Gurin (2014, p. 9) afirma que os dados abertos são “[...] projetados para fornecer dados gratuitos, abertos e transparentes, que podem transformar a maneira como fazemos negócio, administramos o governo e gerenciamos todos os tipos de transações”.

A OFK (2010) apresenta três normas fundamentais para a definição de dados abertos:

- **disponibilidade de acesso:** os dados devem estar disponíveis na íntegra, preferencialmente para *download*, e sob custo não maior que um custo razoável de reprodução. Além disso, os dados devem estar disponíveis também de uma forma conveniente e modificável, permitindo seu uso prático e adaptável;
- **reúso e redistribuição:** os dados devem ser fornecidos sob termos que permitam a sua reutilização e redistribuição, inclusive na combinação com outros conjuntos de dados, promovendo a criação de novos produtos, serviços e conhecimentos;
- **participação universal:** todos os indivíduos devem ser capazes de usar, reutilizar e redistribuir os dados – não deve haver discriminação contra campos do conhecimento, pessoas ou grupos. Exemplos de restrições podem ser verificados no uso não comercial que impediriam o uso comercial, ou ainda para fins diversos, como para educação. Estas ações restritivas excluem dados do conceito de abertos.

Quando as principais diretrizes que compõem o conceito de dados abertos são cumpridas, torna-se possível que diferentes organizações e sistemas colaborem de maneira conjunta no compartilhamento de dados (Isotani; Bittencourt, 2015).

O ativista em dados abertos, David Eaves⁴ sintetiza que as características que conceituam os dados abertos podem ser descritas em três leis fundamentais a serem observadas (Pinto, c2023, p. 22):

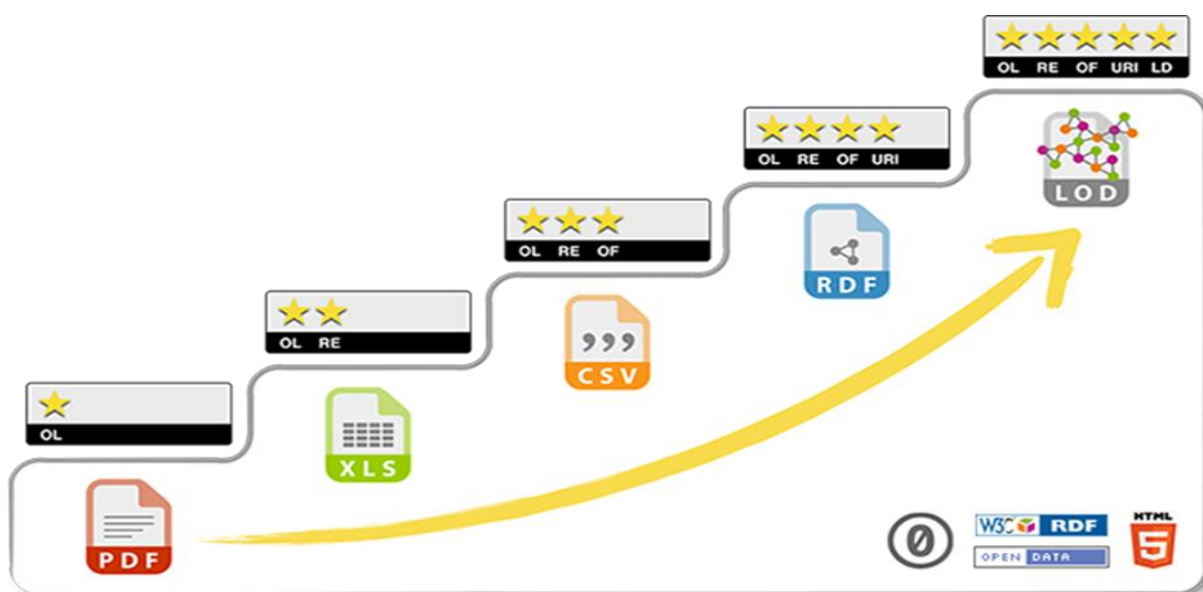
1. se o dado não pode ser encontrado ou indexado na *web*, não existe;
2. se o dado não está disponível em um formato aberto e legível por máquina, não pode ser reutilizado;
3. se dispositivos legais não permitem que ele seja compartilhado, não é útil.

⁴ David Eaves é considerado um dos principais ativistas e líderes globais do movimento dados abertos por sua atuação política, técnica, acadêmica e comunitária que auxiliou a transformar a agenda internacional de governo aberto no século XXI.

Desde o início da disponibilização dos dados abertos, diversas diretrizes foram propostas com o intuito de aprimorar o compartilhamento e a utilização desses objetos. Entre essas iniciativas, destaca-se o Esquema das Cinco Estrelas, proposto em 2006, por Tim Berners-Lee. Esse modelo estabelece uma classificação progressiva baseada em requisitos representados por estrelas. A cada nova estrela obtida, os dados se tornam progressivamente mais potentes e mais fáceis de serem utilizados (Open Knowledge Brasil - OKBR, 2013). À medida que ocorre a progressão da primeira a quinta estrela, os dados tornam-se progressivamente mais abertos quanto ao acesso, uso e reutilização, promovendo maior transparência e participação pública.

A Figura 1 ilustra o grau de abertura dos dados no Esquema das Cinco Estrelas, de acordo com seus cinco níveis. Nessa representação, observa-se que a cada estrela adquirida, um novo nível de abertura é alcançado, e as características do nível anterior se incorporam com a atual até completar o esquema de cinco níveis.

Figura 1 – Esquema das Cinco Estrelas em grau crescente de abertura do formato dos dados



Fonte: OKBR (2013).

No primeiro nível, correspondente a uma estrela, os dados estão disponíveis na *web* sob uma licença aberta. No entanto, estão inseridos em documentos em que não é possível a extração das informações de modo automático e facilitado, como

em *Portable Document Format* (PDF)⁵ ou *Hypertext Markup Language* (HTML)⁶ (Penteado, 2020).

O uso de formatos proprietários, como o PDF, pode dificultar a utilização dos dados em algumas ocasiões. Ainda assim, para que um conjunto de dados se enquadre no nível 1, “[...] basta que o dado esteja disponível e de forma livre na *web*” (Carossi; Teixeira Filho, 2020, p. 19). Assim, mesmo sendo apresentado em licença aberta, não pode ser manipulado e utilizado facilmente em alguns contextos ou situações, fazendo com que o uso e acesso aos dados por máquinas seja prejudicado (Isotani; Bittencourt, 2015). Essa abordagem do Esquema das Cinco Estrelas, com implementação simples e rápida, é a mais comum entre os governos, justamente para possibilitar um atendimento mais ágil às demandas de informação (Penteado, 2020).

No segundo nível (duas estrelas), as características do nível anterior permanecem, mas “[...] com uma licença de uso especificada e em forma estruturada (linhas e colunas)” (Penteado, 2020, p. 90). Contudo, os dados podem ser disponibilizados em formatos proprietários, como o *Excel Spreadsheet* (XLS)⁷, desde que se respeite a licença livre (Carossi; Teixeira Filho, 2020, p. 20).

Neste estágio, o acesso aos dados por máquinas é menos limitado, permitindo que sejam utilizados com facilidade tanto por pessoas, quanto por máquinas (Isotani; Bittencourt 2015). Entretanto, apesar de o formato XLS facilitar o consumo de dados, Isotami e Bittencourt (2015, p. 48) aponta dois problemas neste contexto: “[...] os dados continuam trancados em um documento, e; o acesso ao dado dentro do documento XLS necessita de software proprietário”.

No terceiro nível (três estrelas), “[...] os dados estão disponíveis em formatos estruturados e não proprietários, como os *Comma-separated values* (CSV)⁸, *Extensible Markup Language* (XML)⁹, *JavaScript Object Notation* (JSON)¹⁰, sem a

⁵ Significa formato de documento portátil. Formato de arquivo que permite compartilhar e visualizar documentos (Adobe Acrobat, 2025a).

⁶ É uma linguagem de marcação de hipertexto utilizada na criação de documentos e páginas da *web* (Melo, 2021).

⁷ XLS é um formato de arquivo do *Excel*, que armazena dados de planilhas (Adobe Acrobat, 2025b).

⁸ É um formato de arquivo de texto que armazena dados tabulares (como uma planilha) onde cada linha representa um registro e os valores de cada linha são separados por vírgulas.

⁹ Permite definir e armazenar dados de maneira compartilhável (Amazon, 2024).

necessidade de adquirir licenças de software para processá-los” (Penteado, 2020, p. 90). Nesse nível, a publicação dos dados é realizada em formato aberto, o que facilita a manipulação dos dados e elimina as restrições de *softwares* específicos (Isotani; Bittencourt, 2015).

O nível quatro (quatro estrelas) está relacionado com a utilização de URI que são utilizados para identificar recursos e torná-los localizáveis na internet. Ou seja, os *URIs* permitem o compartilhamento e a identificação individual de cada dado publicado (Carossi; Teixeira Filho, 2020; Isotani; Bittencourt, 2015).

Diversos formatos podem ser empregados para representar dados, como *Atom* e *Really Simple Syndication* (RSS)¹¹, porém, o mais utilizado é o *Resource Description Framework* (RDF)¹². O RDF “[...] permite que os metadados de um determinado conjunto de dados possam ser relacionados com outros conjuntos de dados também disponibilizados nesse padrão” (Carossi; Teixeira Filho, 2020, p. 20; Isotani; Bittencourt, 2015).

Ao empregar a linguagem RDF, há a possibilidade de utilizar dados provenientes de esquemas diferentes, os quais podem evoluir ao longo do tempo sem a necessidade dos usuários realizarem alterações (Rautenberg *et al.*, 2018).

No quinto nível (cinco estrelas), todos os níveis anteriores são incorporados e é acrescentado um elemento central: “[...] para pertencer a esse padrão os dados devem estar ligados a outros dados, promovendo o cruzamento para geração de novos conhecimentos” (Carossi; Teixeira Filho, 2020, p. 21). Essa interligação entre conjuntos de dados favorece análises mais complexas e aprofundadas, além de estimular a criação de soluções inovadoras com base em dados, como os públicos.

No entanto, o Esquema Cinco Estrelas proposto por Tim Berners-Lee, se refere apenas ao formato ou a codificação utilizada para publicar os dados. Dessa forma, é possível que um conjunto de dados possa atingir o nível máximo (cinco estrelas) e, ao mesmo tempo, apresentar baixa qualidade em relação ao conteúdo disponibilizado (Vetrò *et al.*, 2016).

¹⁰ Acrônimo para *JavaScript Object Notation*, padrão aberto de estruturação de dados baseado em texto e legível por humano (Carossi; Teixeira Filho, 2020, p. 17).

¹¹ É um formato de texto para distribuição de informações em tempo real pela internet; é uma forma simplificada e resumida de apresentar o conteúdo do *site* usando linguagem XML.

¹² Uma estrutura padrão para descrever e trocar dados na *web*.

Berners-Lee (2006) propôs ainda, um conjunto de princípios de dados conectados para promover a publicação e interligação de dados estruturados na *web*, a saber:

- (I) uso de URIs como nome para as coisas;
- (II) uso de HTTP para que as pessoas possam pesquisar esses nomes;
- (III) ao procurar por um URI, fornecer informações úteis, utilizando os padrões RDF ou *Protocol and RDF Query Language* (SPARQL), por exemplo, e incluir *hiperlinks* para outros URIs para que possam descobrir outros itens.

Os dados abertos têm como um dos principais objetivos “[...] possibilitar a criação de uma rede de informações na qual as pessoas possam realizar várias análises e comparações por meio das informações públicas disponibilizadas pelas organizações” (Carossi; Teixeira Filho, 2020, p. 21).

Apesar dos avanços na abertura de dados, muitos desafios ainda precisam ser superados. A prática de disponibilizar dados em formato aberto continua em expansão e tem se consolidado como uma tendência adotada por organizações públicas e privadas em nível mundial.

4.3 Tipos de dados

A metodologia da QD tem como principal finalidade analisar e garantir a confiabilidade, precisão e consistência dos dados utilizados em diferentes contextos. Segundo Wang, Ziad e Lee (2002, p. 6), os “[...] dados representam objetos do mundo real, em um formato que pode ser armazenado, recuperado e processado por um procedimento de software e comunicado por uma rede”.

Um amplo conjunto de fenômenos pode ser representado por meio de dados, dada sua versatilidade na forma de representação e aplicabilidade em diversas situações e formatos (Wang; Ziad; Lee, 2002). Esses dados podem estar disponíveis em diferentes formatos e contextos, variando entre dados não estruturados, como aqueles presentes em sistemas de arquivos, repositório de documentos ou portais da *web*, ou mesmo em formatos altamente estruturados, como os sistemas de gestão de bases de dados (Batini *et al.*, 2011).

Conforme Kitchin (2014, p. 4), os dados possuem formas materiais diversas, incluindo números, texto, símbolos, imagens, som, ondas eletromagnéticas e mesmo um espaço vazio. Essa variedade permite classificá-los em duas categorias:

- dados qualitativos: não são representados por números, mas por elementos como textos, imagens e sons. Embora não sejam numéricos, esses dados podem ser analisados por meio de visualizações, uma variedade de estatísticas descritivas e inferências. Além disso, podem servir como entrada para modelos preditivos e de simulação.

- dados quantitativos: esses dados geralmente são extensos e se relacionam com as propriedades físicas dos fenômenos - como comprimento, altura e distância - ou são representativos e se relacionam com características não físicas dos fenômenos, como classe social, nível educacional, entre outros.

Quanto à fonte de dados, a literatura os classifica com base no nível de estruturação que elas apresentam, resultando em três categorias principais:

1) Dados estruturados: são aqueles em que cada elemento de dados possui uma estrutura fixa associada, com um esquema formal definido e explícito (Batini *et al.*, 2011; Wang; Ziad; Lee, 2002). Nessa linha, Kitchin (2014, p. 5) define dados estruturados como “[...] aqueles que podem ser facilmente organizados, armazenados e transferidos, permitindo que esses dados possam ser processados, pesquisados, consultados, combinados e analisados de forma relativamente direta”.

Batini *et al.* (2009) apresentam diferentes formas de representação de um mesmo objeto do mundo real em modelos práticos. Por exemplo, um registro de informações pessoais como: nome, sobrenome, cidade e país de nascimento podem ser representados de maneira estruturada. O Quadro 7 ilustra essa representação com os dados de Patrick Metzisi, nascido na cidade de Masai Mara, no Quênia, demonstrando como essas informações podem ser organizadas em um formato estruturado.

Quadro 7 – Representação de dados estruturados

Patrick	Metzisi	Masai Mara	Quênia

Fonte: Batini *et al.* (2009, p. 10).

Por meio do Quadro 7 é possível observar que os dados estruturados apresentam um formato definido, com estrutura fixa e um esquema formal explicitamente especificado. Além disso, nota-se que os dados estruturados consistem em um conjunto de componentes que possuem relação entre si, geralmente armazenados em forma de tabelas, na qual as colunas e as linhas representam os campos dos registros. As tabelas relacionais e os dados estatísticos representam o tipo mais comum de dados estruturados (Batini *et al.*, 2009).

2) Dados não estruturados: são aqueles que se apresentam como se fossem apenas sequências de símbolos, frequentemente expressos em linguagem natural e sem nenhuma especificação, tipos de estrutura, domínios de tipos definidos e restrições formais definidas (Batini *et al.*, 2011; Wang; Ziad; Lee, 2002). Segundo Kitchin (2014, p. 6) esses dados, “[...] não têm um modelo de dados definido ou estrutura identificável comum [...]”. Dessa forma, cada elemento individual pode ter uma estrutura ou formato específico, mas nem todos os dados pertencentes ao mesmo conjunto de dados compartilham a mesma estrutura.

Para ilustrar o conceito, utilizando o mesmo objeto do mundo real empregado no exemplo dos dados estruturados, o Quadro 8 apresenta sua representação em formato de dados não estruturados:

Quadro 8 – Representação de dados não estruturados

Patrick Metzisi, nascido na cidade de Masai Mara, no Quênia

Fonte: Batini *et al.* (2009, p. 10).

Os dados não estruturados apresentam texto livre e não seguem uma estrutura pré-definida - como, por exemplo, respostas a perguntas abertas - o que torna sua consulta e análise, um desafio (Batini *et al.*, 2009). Em geral, são considerados “[...] dados qualitativos por natureza, mas podem ser frequentemente convertidos em dados estruturados por meio de classificação e categorização” (Kitchin, 2014, p. 6).

3) Dados semiestruturados: possuem características intermediárias entre os dados estruturados e os não estruturados. Frequentemente são representados por linguagem XML e apresentam uma estrutura mais flexível, ou seja, possuem alguma

estrutura, mas não tão rígida como as que existem nas bases de dados (Pagaiame, 2018).

Nesse sentido, Kitchin (2014, p. 5) afirma que os dados semiestruturados “[...] são fracamente estruturados e não possuem modelo/esquema de dados predefinido, e, portanto, não podem ser mantidos em um banco de dados relacional”. A estrutura desses dados é considerada irregular, implícita, flexível e geralmente organizada numa estrutura hierárquica, permitindo certa padronização, mas sem impor esquemas fixos e uniformes (Kitchin, 2014).

Para exemplificar os dados semiestruturados, utiliza-se o mesmo objeto do mundo real aplicado nos exemplos anteriores - de dados estruturados e não estruturados, conforme demonstrado nos Quadros 7 e 8. No Quadro 9 é demonstrada a representação desses dados utilizando o XML. Assim, é possível evidenciar como uma estrutura pode ser organizada de forma mais flexível e sem comprometer a integridade dos dados.

Quadro 9 – Representação de dados semiestruturados

```
<DadosPessoais>

  <nome>Patrick Metzisi</nome>

  <naturalidade>Masai Mara, Quênia</naturalidade>

</DadosPessoais>
```

Fonte: Batini *et al.* (2009, p. 10).

Os dados semiestruturados apresentam algumas características em comum que os diferenciam dos dados estruturados e não estruturados. Segundo Wang, Ziad e Lee (2002), essas características incluem:

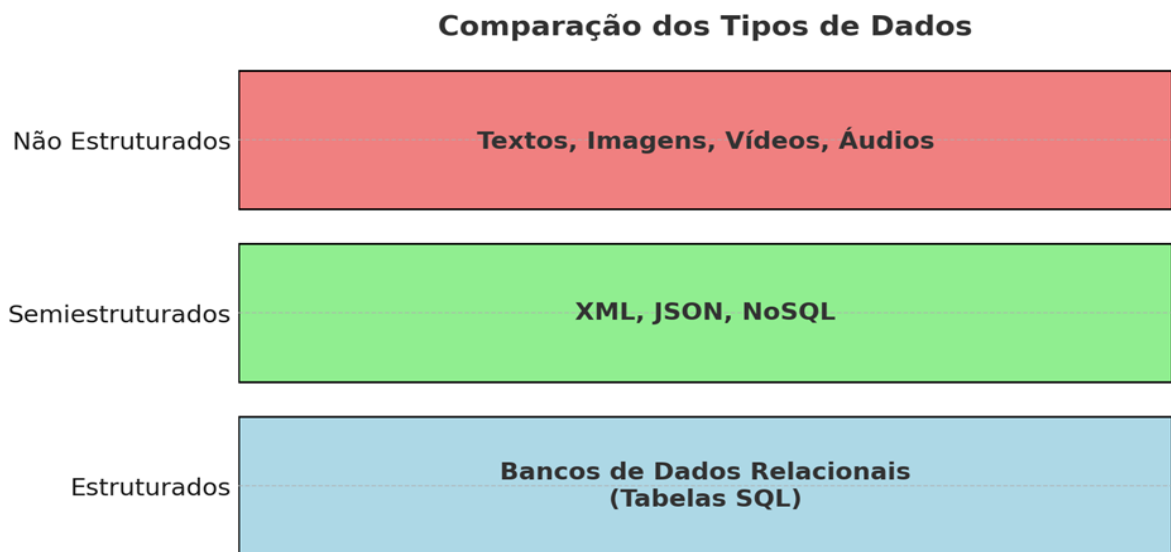
I – os dados podem conter campos que não são conhecidos no momento da definição da estrutura. Por exemplo, um arquivo XML não possui um arquivo de esquema XML associado;

II – o mesmo tipo de dado pode ser representado de várias maneiras. Uma data, por exemplo, pode ser registrada como um único campo ou dividida em vários campos, mesmo inserida em um mesmo conjunto de dados;

III – muitos campos podem não conter valores, o que resulta em estruturas com dados ausentes ou incompletos.

Para facilitar a compreensão das diferenças entre os tipos de dados apresentados, a Figura 3 ilustra de forma comparativa as principais distinções entre dados estruturados, semiestruturados e não estruturados.

Figura 3 – Comparação entre as diferenças dos dados estruturados, semiestruturados e não estruturados



Fonte: Open AI (2025).

De forma complementar, a Figura 4 demonstra uma representação gráfica dos tipos de dados, permitindo visualizar suas diferenças estruturais. Cada tipo de dado possui características específicas e pode ser mais adequado para um determinado objetivo. Por esse motivo, é necessário compreender a estrutura e as particularidades de cada categoria.

Figura 4 – Representação gráfica dos tipos de dados



Fonte: Dataside (2023).

Compreender as diferenças entre os tipos de dados permite que a organização escolha as ferramentas mais adequadas para a construção, gestão, compartilhamento e análise de seus dados.

Nos dados estruturados e semiestruturados são utilizadas dimensões de qualidade, como: completude, precisão, compreensibilidade, entre outras. Essas avaliações são amplamente aplicáveis e independem do contexto, pois estão associadas a algoritmos e métodos de avaliação bem estabelecidos (Batini; Scannapierca, 2006).

Por outro lado, a medição de dados não estruturados apresenta desafios adicionais, uma vez que as técnicas voltadas à sua avaliação são menos consolidadas. De acordo com Batini e Scannapierca (2006), dimensões como precisão e completude são mais difíceis de avaliar em dados não estruturados, como em um texto. No entanto, outras dimensões como a atualidade podem ser aplicadas de forma mais direta nesse tipo de conteúdo.

Diante do exposto, observa-se que é essencial a adoção de estratégias mais eficientes e de ferramentas tecnológicas mais eficazes para compreender a respeito os diferentes tipos de dados e suas características. Com a expansão dos volumes de dados gerados pelas organizações e a evolução constante das TICs, é fundamental assegurar a confiabilidade, precisão e utilidade dos dados.

4.4 Evolução histórica da política de governo digital no Brasil

Antes de abordar a evolução dos DGA, é necessário compreender o processo de transformação digital ocorrido no setor público, que possibilitou a abertura dos dados do governo.

As transformações tecnológicas são vivenciadas pela sociedade desde o século XVIII, na primeira fase da Revolução Industrial. Desde então, a sociedade experimentou a transmissão de dados e informações pelo telégrafo, televisão, telefone fixo, computadores, telefones móveis e internet (Mitkiewicz, 2024).

A partir da segunda metade do século XX, “[...] a protagonista da inovação já não é mais a internet. Vivemos na era da hiperconexão, em que grande parte da população está continuamente conectada por meio de *smartphones* e outros dispositivos. O destaque atual recai sobre a internet das coisas [...]”, sendo os grandes volumes de dados as novas fronteiras a serem exploradas (Avelino; Pompeu; Fonseca, 2021, p. 7).

Essa evolução tecnológica, associada a novos modelos de inovação, tem superado barreiras, inclusive na Administração Pública. O fácil acesso à informação, o crescimento das redes sociais e a comunicação instantânea, foram elementos determinantes nesse processo (Mitkiewicz, 2024).

A transformação digital da economia e da sociedade modificou as expectativas em relação aos governos. Esse processo gerou pressões por maior transparência, abertura e pela criação de espaços e mecanismos que permitam aos cidadãos e empresas expressarem suas necessidades (Organização para Cooperação e o Desenvolvimento Econômico - OCDE, 2018b).

Segundo Gomes (2018), a construção do governo digital vem sendo desenvolvida há aproximadamente 45 anos. Nesse período, diversas transformações ocorreram, tanto na forma de transmissão das informações quanto na denominação dos fenômenos. O primeiro termo utilizado para o governo digital foi assembleia eletrônica (1972), seguido por teledemocracia (1987) e depois, democracia eletrônica (1991). Alguns anos depois, surgiu o termo ciberdemocracia (1994), depois democracia virtual (1996), em seguida democracia informática (1997), que foi superada por democracia digital (1999) e alterada por governo eletrônico (2000). Com a automatização dos processos internos de trabalho do setor público, o

conceito evoluiu para governo digital, em 2015, além de outras denominações que surgiram até o termo ser consolidado (Brasil, 2019; Gomes, 2018).

Inicialmente, o setor público adotou tecnologias digitais com o objetivo de implementar centros de processamento de dados, automatizando alguns serviços internos. Posteriormente, com a internet, possibilitou um estreitamento nas relações entre sociedade e Estado. Dados e informações foram disponibilizados de forma ativa, e os cidadãos passaram a realizar suas solicitações por meios eletrônicos, “[...] dando início à abordagem de prestação de serviços públicos centrados no usuário” (Mitkiewicz, 2024, p. 257).

De acordo com a OCDE, a evolução do governo analógico para o digital é uma construção de três fases, conforme ilustrado na Figura 5.

Figura 5 – Evolução das tecnologias digitais no governo



Fonte: OCDE (2018a, p. 33).

De acordo com a OCDE (2018a), a transformação do governo digital foi um processo gradual, com condução centrada no cidadão. O modelo anterior, caracterizado por operações fechadas e pouca participação do usuário, foi transformado por um governo aberto, sempre visando o interesse do público.

Nessa abordagem mais aberta, e também mais colaborativa e inovadora do governo, os serviços são desenhados considerando as necessidades e dores do cidadão. As decisões em relação às políticas públicas são orientadas por dados, tendo a participação social e a transparência como premissas (Mitkiewicz, 2024, p. 257).

Com base nessas premissas, Leite e Rezende (2015, p. 552) definem governo digital como:

[...] a utilização das TIC no setor público de forma ampla. O governo eletrônico abrange a melhoria na gestão interna, a transparência e accountability, a disponibilização dos serviços públicos por meios eletrônicos diversos e a aproximação e participação dos cidadãos nos processos decisórios, e ainda inclui as TIC que o setor público pode disponibilizar ou fomentar para a melhoria da qualidade de vida dos cidadãos.

Para a OCDE (2020, p. 5), governo digital se refere ao uso de “[...] tecnologias digitais, como parte integrada das estratégias de modernização dos governos, para criar valor público”. No que se refere à dimensão de QD, foco desta pesquisa, é necessário que o governo reconheça os seus dados como um “[...] ativo estratégico fundamental para a geração de valor público [...]” (Mitkiewicz, 2024, p. 260), além de adotar medidas que assegurem a construção de dados com qualidade e acessíveis aos usuários.

Segundo a OCDE (2020, p. 15), em um governo orientado por dados, três atividades contribuem para a geração de valor público:

- **antecipação de planejamento:** utilização de dados na concepção de políticas, no planejamento de intervenções, na antecipação de possíveis mudanças e na previsão de necessidades;
- **entrega:** utilizar dados para informar e melhorar a implementação de políticas, a capacidade de resposta e a prestação de serviços públicos;
- **avaliação e monitoramento:** os dados podem ser utilizados para mensurar impactos, auditar decisões e monitorar desempenhos.

Com base nas atividades apresentadas, a OCDE (2020, p. 15) apresenta 12 facetas necessárias para um setor público fundamentado em dados (Figura 6).

Figuras 6 – Facetas de um setor público fundamentado em dados



Fonte: Mitkiewicz (2024, p. 261).

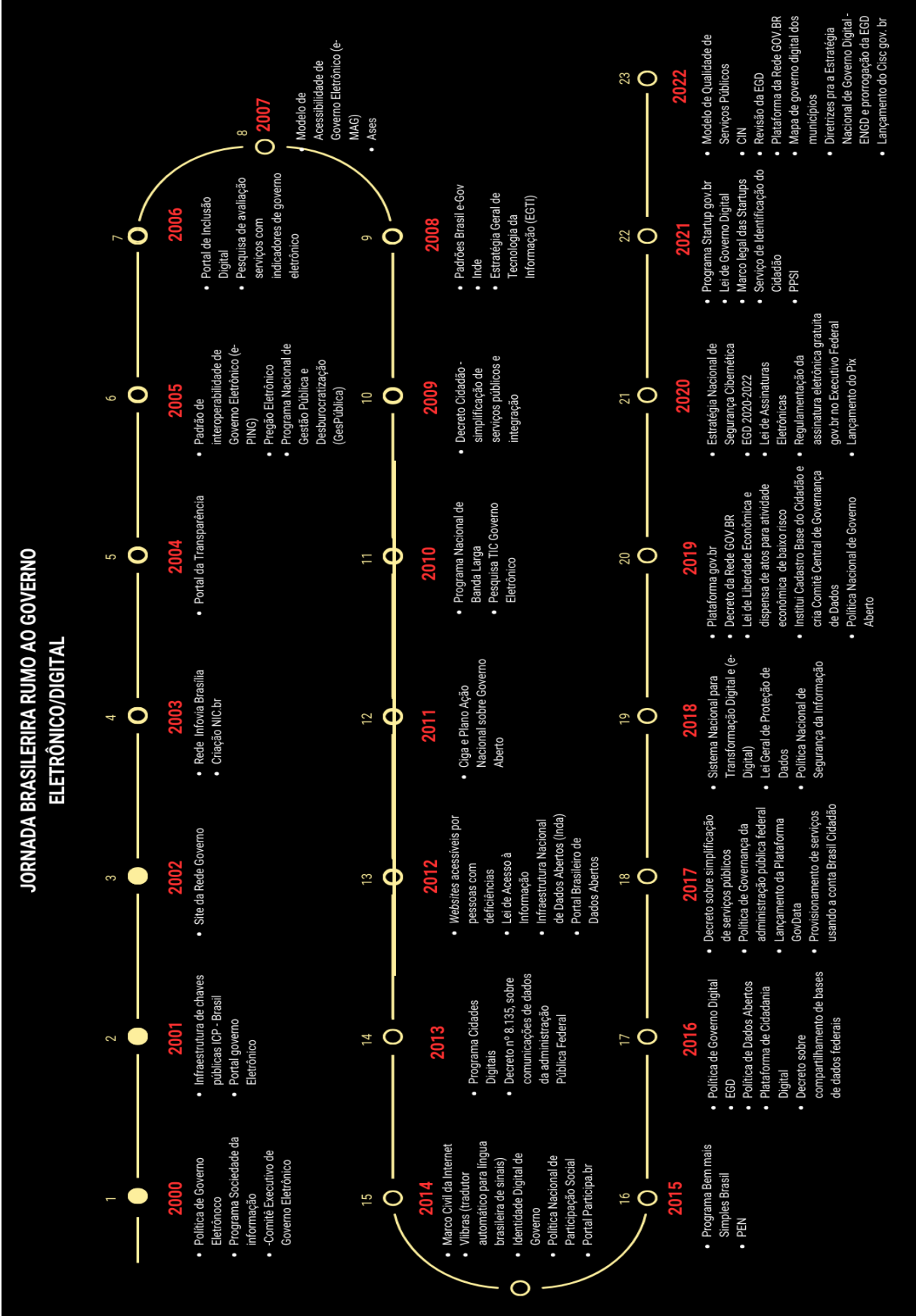
A partir da Figura 5, é possível perceber que a política de um governo orientado por dados é estruturada em três grandes eixos centrais: governança, valor público e confiança, e apoiada por componentes operacionais e princípios fundamentais com o objetivo de promover eficiência, transparência e confiabilidade no uso dos dados públicos.

No Brasil, o Programa de Governo Eletrônico foi criado no ano de 2000, com o desenvolvimento de um grupo de trabalho interministerial em tecnologia da informação. Esse grupo tinha como finalidade examinar e elaborar políticas, diretrizes e normas que aproximassem o governo e o cidadão por meio da utilização das TICs (Mitkiewicz, 2024).

A crise econômica de 2008, que atingiu a confiança em governos relevantes financeiramente, atingindo especialmente os Estados Unidos, diversos países da União Europeia, além de economias da Ásia e da América Latina, impulsionou a transição do governo eletrônico para o governo digital (Arruda, 2020). Com a nova estratégia, o governo brasileiro buscava “[...] um governo mais eficaz, acessível e transparente, possibilitando a aproximação da sociedade aos serviços públicos disponíveis” (Arruda, 2020, p. 314).

A Figura 7 apresenta a evolução histórica do governo eletrônico para o governo digital no Brasil.

Figura 7 – Trajetória histórica do governo digital no Brasil



Fonte: Mitkiewicz (2024, p. 272).

Como se pode observar, entre os anos de 2000 a 2022, diversas medidas foram adotadas para estruturar a abertura dos dados do governo brasileiro como: elaboração dos fundamentos do governo eletrônico, criação da agenda de acessibilidade, transparência e dados abertos, consolidação jurídica da abertura de dados, estruturação da governança de dados, fortalecimento da abertura e a consolidação do governo digital e maturidade da abertura.

Como estratégia do governo digital no Brasil, o país contou com a cooperação da OCDE, que elaborou a Revisão do Governo Digital do Brasil, fundamentada na recomendação sobre estratégias de governo digital. A cooperação teve como principal objetivo, a transição do governo eletrônico para o governo digital. O processo focou nas seis dimensões identificadas pela OCDE como sendo os principais atributos de governos digitais.

O Quadro 10 apresenta as seis dimensões para um governo digital na concepção da OCDE (2018c).

Quadro 10 – Seis dimensões para um governo digital

1) De uma administração centrada no usuário para uma administração conduzida pelo usuário:	Um governo que adota abordagens e toma medidas para que os cidadãos e as empresas possam definir e comunicar suas próprias necessidades e assim conduzir a elaboração de políticas e serviços públicos.
2) De reativo a <u>proativo</u> na elaboração de políticas públicas e prestação de serviços:	Um governo que formula políticas públicas e serviços em antecipação aos processos de desenvolvimento econômico e social, em respeito às necessidades dos usuários e que proporciona serviços antes de serem solicitados. Também os dados devem ser liberados sob a forma de dados abertos (de forma proativa) ao invés de ser reativo aos pedidos de acesso à informação pública.
3) De um governo centrado na informação para um setor público <u>orientado por dados</u>:	Um governo que é capaz de antecipar as tendências sociais e compreender as necessidades dos usuários, transformar a formulação, além de propor o fornecimento e monitoramento de políticas públicas e serviços por meio da gestão e da utilização de dados.
4) Da digitalização de processos existentes ao processo por <u>concepção digital (digital by design)</u>:	Um governo que considera todo o potencial das tecnologias digitais e dos dados, desde o início do processo de formulação de políticas e de serviços, a fim de mobilizar novas tecnologias para repensar, reestruturar e simplificar processos internos e procedimentos. Um governo que fornece setores públicos eficientes, sustentáveis e orientado para os cidadãos, independente do canal utilizado para interagir com as autoridades públicas.

5) De um governo provedor de serviços para um governo como plataforma para cocriação de valor público:	Um governo que utiliza tecnologias digitais e dados para permitir a colaboração com e entre as partes interessadas da sociedade, além de aproveitar sua criatividade e suas capacidades para enfrentar os desafios de um país.
6) Do acesso à informação para a <u>abertura como padrão (open by default)</u>:	Um governo que se compromete de forma proativa na divulgação de dados em formatos abertos e a tornar seus processos acessíveis e apoiados por tecnologias digitais, a menos que haja uma justificativa legítima para não fazer.

Fonte: OCDE (2018b, p. 1).

A implementação das seis dimensões para a construção de um governo digital, permite que o governo federal brasileiro avance em suas políticas, programas e projetos fornecendo orientações e práticas de outros países (OCDE, 2018b).

Segundo Motta e Valle (2022, p. 60) “[...] optar por um governo digital implica em transpor para o ambiente telemático a operação do conjunto de influências, interações e condições que há de se estabelecer entre Administração pública e cidadania”. Os autores ainda complementam, “[...] em tempos de sociedade e cidadania interconectadas, na medida em que o governo vai ficando desconectado, perde sua capacidade de governar” (Motta; Valle, 2022, p. 60).

Diante da realidade social marcada pela era digital, a administração pública ou administração digital (termo preferido), desempenha um papel de protagonismo na atualidade. Para tanto, investir e incorporar novas tecnologias, não é suficiente para o Estado modernizar a administração. É necessário também, que o Governo possa garantir que todos os cidadãos tenham acesso à informação, em cumprimento a “[...] difusão e circulação da informação com maior eficiência, qualidade e rapidez [...]” (Arruda, 2020, p. 294), fornecendo uma participação social mais abrangente e efetiva. Do contrário, o Estado tende a “[...] tornar-se uma administração opaca e hermética, apenas para os poucos que dominam os meios digitais” (Arruda, 2020, p. 294).

Governos ao redor do mundo, como por exemplo, aqueles que integram iniciativas de governo aberto e dados abertos, estão empenhados em oferecer serviços e adotar ferramentas que beneficiem o cidadão na transparência e controle social, melhor qualidade dos serviços públicos, facilitação do acesso à informação, criação de inovação e soluções digitais, tomada de decisões informada, inclusão e acessibilidade. O governo digital visa não somente a transparência pública, mas

envolve todos os serviços públicos, com ênfase especial naqueles mais próximos do cidadão.

4.5 Evolução dos DGA

A abertura dos dados e de informações públicas pelos órgãos governamentais acompanha uma tendência mundial em constante crescimento, abrangendo tanto a administração pública direta quanto a indireta.

Governos ao redor do mundo, comprometidos “[...] com a modernização do Estado e os anseios da sociedade, passaram a promover a abertura de seus dados apoiando-se em ecossistemas digitais avançados [...]”, impulsionados por computação em nuvem e inteligência artificial (Macedo; Lemos, 2021, p. 14). Em um curto intervalo de tempo, os dados abertos se estabeleceram como recurso fundamental na transparência pública (Gray, 2014).

Esse movimento é impulsionado, em grande parte, pelo fato da economia mundial ter se tornado cada vez mais centrada nos dados. Assim, aqueles com capacidade para extrair o máximo de benefícios de seus dados, terão o poder a nível político, social, cultural e, especialmente, econômico (Máchová; Lnenicka, 2017).

Com a intensificação da demanda por transparência e acesso à informação, “[...] governos ao redor do mundo têm dedicado esforços para implementar programas de governo aberto e iniciativas de dados governamentais abertos para os mais diversos fins” (Pinho; Silva, 2019, p. 4).

O primeiro passo rumo à abertura de dados governamentais ocorreu em 2007 nos Estados Unidos da América (EUA) quando um grupo, conhecido como **advogados do governo aberto**, se reuniu com o objetivo de contribuir para iniciativas de transparência governamental, a partir da criação de uma definição para o termo dados abertos (Macedo; Lemos, 2021).

A política de abertura do governo (ou governo aberto) surgiu efetivamente em dezembro de 2009, também nos EUA, quando o então presidente Barack Obama publicou o memorando *Open Government Directive* (OGP). O documento estabeleceu três pilares fundamentais que orientariam os programas e ações governamentais: **transparência, participação e colaboração**. Ainda em 2009, os EUA e a Inglaterra iniciaram a implementação de seus portais de dados

governamentais abertos impulsionando a política de governo aberto em diversos países (Berners-Lee, 2024).

Em 2011, foi criada a organização internacional OGP, decisiva para o estabelecimento da tendência mundial. A OGP passou a estimular iniciativas de governo aberto em todo o mundo. Desde sua fundação, são 75 países e 150 jurisdições que aderiram à parceria e, juntamente com a sociedade civil, estabeleceram parceria para a criação de planos de ação colaborativos com representantes de todos os segmentos da sociedade (OGP, 2025).

A OGP propaga o conceito de governo aberto com base nos princípios da transparência, combate à corrupção, empoderamento cidadão e uso de novas tecnologias. A aplicação destes princípios colabora para “[...] aumentar a disponibilidade de informações; fomentar a participação social; implementar padrões de integridade e ampliar o acesso a novas tecnologias” (Bellix; Guimarães; Machado, 2016, p. 6).

Com a atuação constante da OGP, diversas organizações internacionais passaram a incorporar e difundir a agenda do governo aberto, cada uma delas apresentando sua própria definição de governo aberto. O Brasil define governo aberto como:

[...] uma cultura de governança que promove a colaboração entre governo e sociedade, por meio da transparência na gestão, participação social e da responsabilidade e responsividade dos agentes públicos. Tem como objetivo aprimorar políticas e serviços e solucionar questões de interesse público. Governo aberto se refere a uma nova visão da Administração Pública que promove projetos e ações pautadas nos princípios da transparência na gestão, responsabilidade e responsividade e participação dos cidadãos (Brasil, 2025a).

A parceria entre a OGP com países membros para o Governo Aberto segue ativa e dinâmica. Segundo o relatório anual do OGP o progresso é significativo, a manutenção de estratégias está consolidada, ocorre a integração de novos membros e a realização de encontros regionais e globais com participação ativa da sociedade civil (OGP, 2025).

De acordo com o Relatório Anual de 2024-2025 do OGP, 95% dos países membros para Governo Aberto implementaram seus planos de ação em 2024,

sendo que 11 países apresentaram novos compromissos e três novos países ingressaram como membros (OGP, 2025).

No Brasil, destacam-se avanços concretos na abertura do governo, como o 6º Plano de Ação Nacional de Governo Aberto em 2024, cuja etapa de oficinas colaborativas de cocriação foi concluída. O documento formalizou compromissos da Ciência Aberta e Inovação. Também em 2024, aconteceu o primeiro seminário de Ciência Aberta, juntamente com o lançamento do Centro de Ciência Aberta da Universidade Federal de Goiás (UFG). Além disso, o Instituto Brasileiro de Auditoria de Obras Públicas (IBRAOP) recebeu certificação por sua atuação na elaboração de compromissos sobre infraestrutura e transparência no 6º Plano Nacional de Governo Aberto (OGP, 2025).

A priorização de uma colisão política, espaço cívico, disseminação de inovações e avanços concretos em transparência e participação em relação aos Governo Aberto, seguem como estratégias do OGP para o período de 2023/2028 (OGP, 2025).

O Quadro 11 apresenta os principais conceitos de Governo Aberto formulados em cada organismo internacional.

Quadro 11 – Conceito de Governo Aberto por organismos internacionais

Organismo Internacional	Conceito	Enfoque
Parceria para Governo Aberto por meio da OGP	Governo que promove transparência, empoderamento da sociedade, combate a corrupção e incorpora novas tecnologias para melhoria da governança (OGP, 2011).	Transparência, empoderamento da sociedade, combate à corrupção e incorporação das novas tecnologias para melhoria da governança.
Organização dos Estados Americanos (OEA)	O Governo Aberto se baseia nos princípios de: Transparência: um governo transparente tem informações a respeito de suas responsabilidades, atos, decisões, dados, planos, fontes e tudo relacionado à sociedade. Colaboração: um governo colaborativo envolve e convoca os cidadãos e outros agentes sociais para resolverem os problemas públicos coletivamente,	Transparência, colaboração e participação.

	aproveitando o potencial e a energia de vários setores e atores sociais e privados. Participação: um governo participativo incentiva o direito dos cidadãos de intervir ativamente na tomada de decisões políticas, promovendo a interação entre o Estado e a sociedade para que ambos possam se beneficiar dos conhecimentos, ideias e experiências com o objetivo de melhorar a qualidade e eficácia da ação pública (OEA, 2016).	
Banco Interamericano de Desenvolvimento (BID)	“Governo aberto se insere na área de modernização/reforma do estado, cujo foco é na abertura e accountability” (BID, 2016).	Transparência, <i>accountability</i> .
Comissão Econômica para a América Latina e o Caribe (CEPAL)	“Um governo aberto se caracteriza por ser transparente e promover participação e esquemas de colaboração, com a finalidade de melhor a qualidade dos serviços públicos” (CEPAL, 2016).	Transparência, colaboração e participação para melhoria dos serviços públicos.
Organização para a Cooperação e Desenvolvimento Econômico (OCDE)	Governo aberto - a abertura de processos governamentais, procedimentos, documentos e dados para escrutínio público e envolvimento [...]. Tanto maior transparência e a participação pública não só pode levar a políticas e serviços melhores, como também podem promover a integridade do setor público, que é essencial para recuperar a confiança dos cidadãos, a neutralidade e a fiabilidade das informações públicas administradas (OCDE, 2016).	Transparência, engajamento cívico, <i>accountability</i> e integridade.
Omidyar Network Fundação Avina	Governo que promove transparência, empoderamento da sociedade, combate à corrupção e incorpora novas tecnologias para melhoria da governança (OGP, 2011).	Transparência, empoderamento da sociedade, combate à corrupção e incorporação das novas tecnologias para melhoria da governança.
Red Gealc	Políticas de governo eletrônico focadas nos cidadãos para promover a transparência e participação cidadã (Red Gealc, 2016).	Transparência, e participação, a partir da perspectiva do desenvolvimento do governo eletrônico.

CLAD (2016)	Conjunto de mecanismos e estratégias que contribuem para a governança pública e para o bem governo, baseado nos pilares da transparência, participação cidadã, responsabilização, colaboração e inovação, focando e incluindo a cidadania no processo de tomada de decisões, bem como na formulação e implementação de políticas públicas, para fortalecer a democracia, a legitimidade da ação pública e bem-estar coletivo.	Transparência, participação, colaboração e inovação.

Fonte: Bellix *et al.* (2016, p. 10).

Por meio do Quadro 11, é possível observar que os conceitos apresentados pelos organismos internacionais se baseiam em pilares como transparência, responsabilidade, participação cidadã, inovação e colaboração.

A temática Governo Aberto obteve destaque na administração pública, especialmente porque, com a abertura do governo, benefícios como eficiência, redução da corrupção e aumento da legitimidade do governo passaram a estar presentes e tornaram-se reconhecidos (Bellix *et al.*, 2016).

Em 2013, governos integrantes da cúpula do *Grupo dos Oito*¹³ (G8), por meio da Carta Aberta de Dados, reafirmaram o compromisso de tornar os dados públicos abertos. O documento estabelece princípios fundamentais, como dados abertos por meio de padrões, observação da legislação existente sobre privacidade e propriedade intelectual. Nesse mesmo ano, os EUA tornaram seus dados abertos e legíveis por máquina, adotando esse novo padrão para as informações de seu governo (Macedo; Lemos, 2021).

Vale destacar que a CF de 1988, já reconhecia o direito ao acesso à informação como direito fundamental a todo cidadão. No entanto, a política de acesso à informação pública no Brasil conseguiu maior visibilidade em 2011, com dois importantes marcos regulatórios: O primeiro foi a adesão do Brasil à iniciativa internacional OGP, em parceria com os EUA, para a criação do Governo Aberto, cujo objetivo é o incentivo de práticas de transparência, acesso público e participação social. O segundo marco, foi a promulgação da LAI, que regulamenta o direito ao acesso à informação, estabelecendo diretrizes para a divulgação das informações públicas.

¹³ Grupo formado por sete países mais industrializados (Alemanha, Canadá, Estados Unidos, França, Itália Japão, Reino Unido) e a Rússia.

A publicação da LAI representou um marco decisivo para o avanço acerca do acesso e transparência dos dados abertos no Brasil, pois regulamentou a disponibilização dos dados governamentais abertos.

Outro ponto relevante para a abertura de dados no Brasil, ocorreu em 2015, com o objetivo de impulsionar a disponibilização dos dados e promover a transparência pública. Nesse sentido, o governo brasileiro publicou cinco motivos que comprovam a necessidade da abertura dos dados públicos:

- I - maior transparência da gestão pública;
- II - contribuição da própria sociedade com serviços inovadores ao cidadão;
- III - aprimorar a QD;
- IV - viabilizar novos negócios;
- V - obrigatoriedade legal (Brasil, 2015).

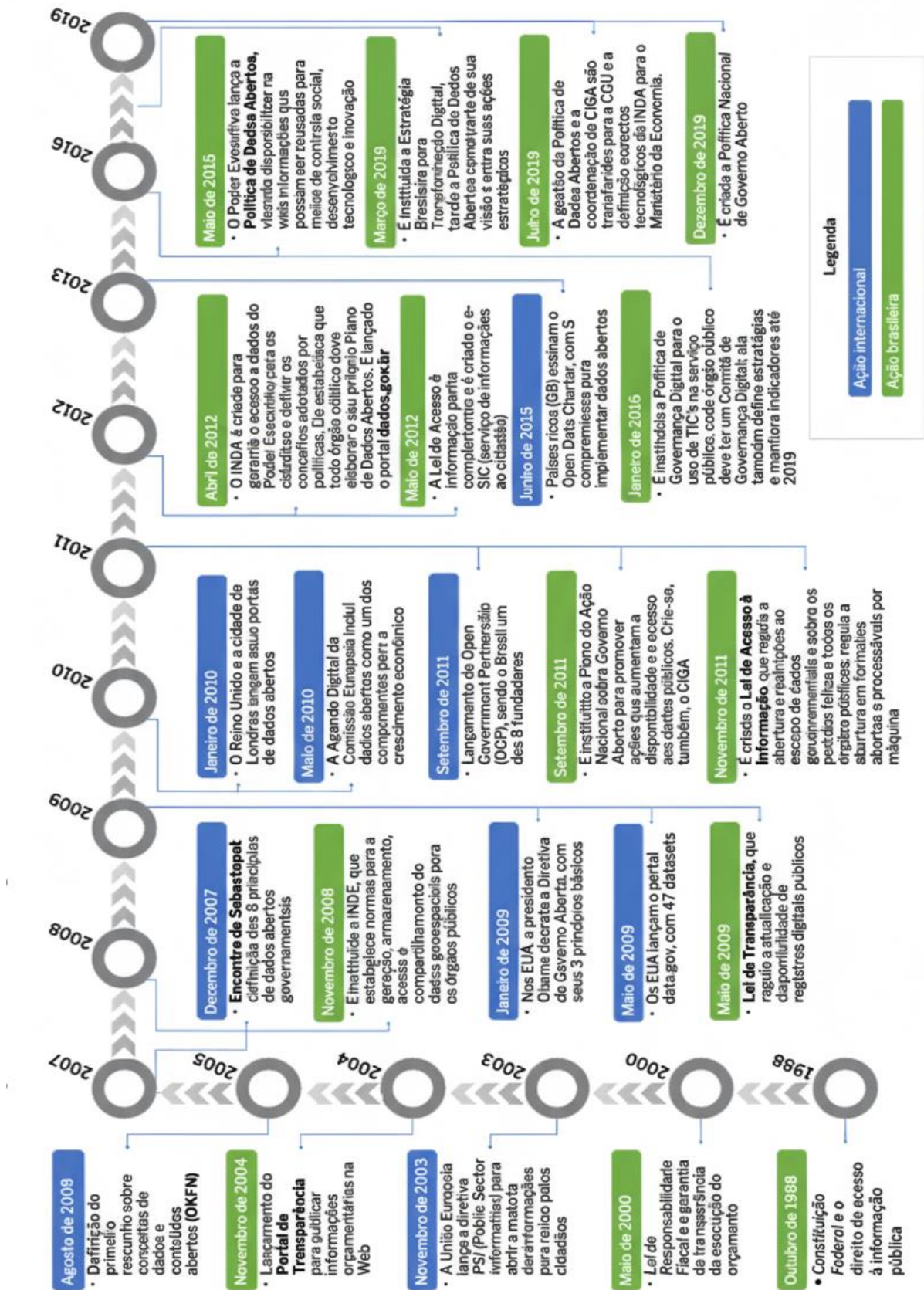
Com a necessidade de operacionalizar os dados governamentais abertos, o Governo Federal instituiu, em 2016, a Política de Dados Abertos do Poder Executivo Federal, por meio do Decreto nº 8.777/2016. Esse decreto criou a Infraestrutura Nacional de Dados Abertos (INDA)¹⁴ e o Plano de Dados Abertos (PDA)¹⁵, impulsionando e incentivando o uso de tecnologias para a disponibilização das informações públicas aos usuários. Essas ações marcaram um importante avanço rumo à transparência e à democratização do acesso à informação.

A Figura 8 apresenta uma contextualização histórica dos marcos mais relevantes o das iniciativas brasileiras e internacionais de DGA, ilustrando a evolução e os principais acontecimentos que contribuíram para o desenvolvimento do movimento de dados abertos.

¹⁴ Disponível em: <https://www.gov.br/governodigital/pt-br/dados-abertos/infraestrutura-nacional-de-dados-abertos>. Acesso em: 05 dez. 2025.

¹⁵ Disponível em: <https://www.gov.br/governodigital/pt-br/estrategias-e-governanca-digital/sisp/guia-do-gestor/pda>. Acesso em: 05 dez. 2025.

Figura 8 – Contextualização histórica dos marcos relevantes dos DGA



Fonte: Penteado (2020, p. 67).

Por meio da Figura 8 é possível visualizar as ações brasileiras e internacionais para a abertura dos dados governamentais. Cada conquista representa um avanço significativo para o cidadão na busca por mais transparência pública e acesso à informação.

Antes da implementação dos dados abertos, a Administração Pública brasileira divulgava suas informações por meio de documentos físicos afixados em locais públicos. Para acessar essas informações, o cidadão precisava se deslocar até o local para consultar as informações desejadas ou solicitá-las por meio de requerimento ao governo.

A adoção de um conjunto de normativas que estabelece diretrizes para a abertura de dados governamentais como a lei n.º 12.527/2011 e o decreto n.º 8.777/2016 possibilitou que uma quantidade superior de informações seja disponibilizada, sem a necessidade de obrigações formais, podendo ser consultado em qualquer lugar com acesso à internet.

Embora a Administração Pública não disponha dos mesmos recursos tecnológicos das organizações privadas para disponibilizar seus dados, tais como arquiteturas de dados em nuvens altamente escaláveis, pipelines automatizados de tratamento e governança de dados, soluções robustas de interoperabilidade e ferramentas de inteligência artificial, tem avançado velozmente para acompanhar as novas tendências digitais. Esse progresso do governo digital não beneficia apenas a sociedade, pois ao promover mais transparência e acessibilidade, aprimora também a Administração Pública.

4.6 Definição de DGA

O conceito de DGA disponível na literatura científica apresenta uma convergência conceitual. Para que um governo seja efetivamente considerado aberto, suas ações devem ser pautadas em três princípios básicos: transparência, participação e colaboração. O objetivo de um governo aberto é a ampliação da participação da sociedade nas políticas públicas e viabilização de meios para que a sociedade participe direta e indiretamente na administração pública (Bandeira *et al.*, 2012, p. 10).

Segundo Brasil (2016), dados governamentais abertos são:

[...] dados acessíveis ao público, representados em meio digital, estruturados em formato aberto, processáveis por máquina, referenciados na internet e disponibilizados sob licença aberta que permita sua livre utilização, consumo ou cruzamento, limitando-se a creditar a autoria ou a fonte.

Essa definição dialoga diretamente com as diretrizes internacionais que orientam o movimento de dados abertos e reforça a necessidade de que os dados sejam disponibilizados de forma livre, reutilizável e interoperável, permitindo seu uso por qualquer usuário.

O Quadro 12 apresenta conceitos de DGA disponíveis na literatura científica:

Quadro 12 – Conceitos de DGA

Autor	Ano	Conceito
Kalampokis <i>et al.</i>	2011	São dados do setor público disponíveis livremente em formato aberto.
Albano e Araújo	2013	É uma solução adotada pelos governos para cumprir exigências de acesso por parte da sociedade. Estas exigências se sustentam em conceitos de transparência, colaboração e participação.
Barbalho	2014	São os dados gerados nos processos governamentais que passam a ser liberados para acesso pela sociedade.
Yannoukakou e Araka	2014	Dados produzidos ou encomendados pelo governo ou entidades controladas pelo governo, disponíveis livremente para uso e redistribuição.
Solar <i>et al.</i>	2014	É uma filosofia de trabalho que busca capacitar os cidadãos a acessar os dados produzidos por entidades públicas permitindo acessar, armazenar em bases de dados próprias, redistribuir e integrá-los com outras fontes de dados.
World Wide Web Consortium (W3C)	2016	Refere-se à publicação e disseminação dos dados do setor público na <i>web</i> , em formato bruto e aberto, compreensíveis logicamente, permitindo sua reutilização em aplicações digitais desenvolvidas pela sociedade.

Fonte: Pinto (2024, p. 40).

Observa-se que os conceitos de dados abertos e a transparência pública estão interligados, pois ambos têm como objetivo tornar os dados governamentais acessíveis a todos os cidadãos, eliminando restrições e garantindo seu pleno acesso

e utilização (Oliveira; Fonseca, 2021). Entre os conceitos analisados, observa-se que a definição apresentada por Solar *et al.* (2014) se aproxima mais diretamente dos princípios do governo aberto, ao enfatizar o empoderamento do cidadão por meio do acesso, armazenamento, redistribuição e integração dos dados públicos. Essa perspectiva destaca o papel ativo da sociedade na reutilização das informações governamentais, reforçando a participação e a colaboração, elementos centrais de um governo orientado pela transparência.

No Brasil, os principais objetivos relacionados aos dados abertos estão descritos no Decreto n.º 8.777/2016, que institui a Política de Dados Abertos do Poder Executivo Federal (Brasil, 2016a). São eles:

- promover a publicação de dados contidos em bases de dados de órgãos e entidades da administração pública federal direta, autárquica e fundacional sob a forma de dados abertos;
- aprimorar a cultura de transparência pública;
- franquear aos cidadãos o acesso, de forma aberta, aos dados produzidos ou acumulados pelo Poder Executivo federal, sobre os quais não recaia vedação expressa de acesso;
- facilitar o intercâmbio de dados entre órgãos e entidades da administração pública federal e as diferentes esferas da federação;
- fomentar o controle social e o desenvolvimento de novas tecnologias destinadas à construção de ambiente de gestão pública participativa e democrática e à melhor oferta de serviços públicos para o cidadão;
- fomentar a pesquisa científica de base empírica sobre a gestão pública;
- promover o desenvolvimento tecnológico e a inovação nos setores público e privado e fomentar novos negócios;
- promover o compartilhamento de recursos de tecnologia da informação, de maneira a evitar a duplicidade de ações e o desperdício de recursos na disseminação de dados e informações;
- promover a oferta de serviços públicos digitais de forma integrada.

A OGP e suas organizações parceiras composta por representantes do governo, sociedade civil, empresas e organizações sem fins lucrativos, comprometeram-se com a defesa dos princípios de governo aberto e transparente.

Nesse contexto, em 2007, foram estabelecidos os “Oito Princípios de Dados Governamentais Abertos” também conhecidos como “Oito princípios de *Sebastopol*”. Esses princípios foram elaborados por defensores e especialistas do governo aberto em *Sebastopol*, Califórnia, EUA, com o propósito de orientar a compreensão de como os dados públicos abertos são essenciais para a democracia (Silva; Pinheiro, 2019).

O Quadro 13 apresenta os Oito Princípios de *Sebastopol*, conforme detalhado a seguir:

Quadro 13 – Oito Princípios de *Sebastopol*

N.º	Princípios	Descrição
1	Completo	Quaisquer dados públicos que não estejam sujeitos a limitações válidas de privacidade, segurança ou privilégio devem estar disponíveis.
2	Primário	Os dados devem ser disponibilizados da mesma forma que foram coletados na fonte, sem estarem sujeitos a processos de agregação ou alteração.
3	Atualizado	Os dados devem ser disponibilizados ao público o mais rápido possível, a partir o momento de sua criação para que seu valor seja o mais elevado possível.
4	Acessível	Os dados são disponibilizados para o maior número possível de usuários e para múltiplas finalidades.
5	Processável por máquina	Os dados são publicados de forma que permitam o processamento automatizado.
6	Não discriminatório	Os dados devem estar disponíveis a todos os interessados, sem exigência de requerimento ou cadastro.
7	Não proprietário	Os dados devem ser disponibilizados em um formato que não seja controlado exclusivamente por uma única entidade.
8	Livre de licença	Os dados não devem estar sujeitos a qualquer regulamento de direitos autorais, patentes, marcas registradas ou segredos comerciais, exceto por restrições razoáveis de privacidade, segurança e privilégios.

Fonte: Dorobat e Posea (2021, p. 2).

Em 2010, a *Sunlight Foundation*¹⁶ expandiu as diretrizes originais dos Oito Princípios de *Sebastopol* ao adicionar mais dois princípios, totalizando dez

¹⁶ Organização apartidária e sem fins lucrativos que utiliza ferramentas de tecnologia cívica, dados abertos, análise de políticas públicas e jornalismo para tornar nosso governo e a política mais responsáveis e transparentes para todos.

princípios. A fundação oferece orientações ainda mais abrangentes para a publicação e o compartilhamento de dados abertos.

O Quadro 14 apresenta os dois princípios complementares da *Sunlight Foundation*, tornando o conjunto de princípios ainda mais fortes e robustos para a implementação das políticas de dados abertos.

Quadro 14 – Princípios de *Sunlight Foundation*

Princípio	Descrição
Permanência	Os dados disponibilizados devem manter-se permanentemente disponíveis, com rastreamento de versão e arquivamento apropriados ao longo do tempo.
Custos de uso	O acesso aos dados não deve envolver custos aos interessados.

Fonte: Dorobat e Posea (2021, p. 2).

Os DGA podem ser produzidos pelo governo em três diferentes contextos:

(I) com o objetivo explícito de divulgação externa, visando a benefícios econômicos ou sociais mais amplos ou, então; (II) para uso interno ou com subprodutos de suas próprias atividades, e; (III) para transparência e responsabilização, permitindo aos cidadãos entenderem como o governo trabalha (Reinhard; Germano, 2016, p. 11).

A partir desses contextos, é possível identificar e classificar os principais atores na execução e utilização dos dados abertos. Com base nos três diferentes contextos, os atores podem ainda, ser agrupados em quatro ecossistemas distintos, conforme apresentado no Quadro 15.

Quadro 15 – Os Quatro Ecossistemas dos DGA

Ecossistemas	Atividades	Atores
Geração	Coletar, produzir e disponibilizar.	Governo.
Consumo	Buscar, acessar e visualizar.	Sociedade civil, empresas, universidades, organizações, governo, entre outros.
Tratamento e análise	Enriquecer, combinar e analisar.	Sociedade civil, empresas, academia, organizações, governo, entre outros.
Discussão e <i>feedback</i>	Melhorar a compreensão do dado.	Sociedade civil, empresas, academia, organizações, governo, entre outros.

Fonte: Adaptado de Brasil (2022).

O Governo é o único ator que integra todos os ecossistemas de dados abertos, sendo ao mesmo tempo produtor, intermediário e usuário dos DGA. As organizações públicas, por sua vez, são as únicas a compor o ecossistema de geração de DGA.

Desde o início da iniciativa de governo aberto, uma grande quantidade de dados foi disponibilizada ao público. No entanto, isso não significa que o objetivo de promover a transparência e facilitar a prestação de contas tenha sido alcançado. Muitos dados governamentais são disponibilizados apenas para cumprimento de exigências legais, sem observar os requisitos mínimos necessários para sua acessibilidade (Attard *et al.*, 2015).

Alcançar todo o potencial que os DGA têm a oferecer ainda é um imenso desafio. Um dos principais obstáculos é a pluralidade de dados e a variedade de formatos utilizados, o que acaba formando uma barreira para o usuário, principalmente aqueles com menor familiaridade com os recursos digitais.

4.7 Legislações que contribuíram para a abertura dos dados governamentais

Com os avanços promovidos pelo governo aberto, especialmente após o lançamento do Portal Brasileiro de Dados Abertos, em 2012, surgiu a necessidade de o Governo Federal formalizar as políticas relacionadas à disponibilização dos dados governamentais. Para isso, foram elaboradas leis, decretos, resoluções, acórdãos e manuais técnicos, com o intuito de regular o assunto e orientar os órgãos públicos.

A primeira referência normativa brasileira que contempla a interação entre governo e sociedade, assegurando o acesso à informação a todos os cidadãos foi estabelecida na CF de 1988. O direito fundamental ao acesso à informação está contemplado em seu artigo 5º, XXXIII, em que garante a todos o direito a receber dos órgãos públicos informações de seu interesse particular, coletivo ou geral. O dispositivo estabelece ainda, que as informações requeridas devem ser prestadas em prazo determinado em lei, sob pena de responsabilização, observadas aquelas cujo sigilo seja imprescindível para a segurança nacional.

Da mesma forma, o artigo 37 da CF, destaca a participação do usuário na administração pública direta e indireta, garantindo o acesso a registros administrativos e as informações sobre atos do governo (Brasil, 1988).

Ainda na CF, o parágrafo 2º do artigo 216, estabelece que cabe à administração pública, observada a legislação, a gestão da documentação pública, bem como, providências para franquear sua consulta a quem necessitar (Brasil, 1988).

Apesar do reconhecimento do acesso à informação, a CF não aborda regras para a disponibilização dos dados públicos, sendo necessária a criação de leis para preencher essa lacuna. Alguns anos após a promulgação da CF, o Governo Federal Brasileiro passou a elaborar legislações voltadas à transparência pública (Carossi; Teixeira Filho, 2020).

O Quadro 16 elenca os principais dispositivos legais que compõem a base legal sobre a política de dados abertos no Brasil.

Quadro 16 – Principais legislações da Política de dados abertos no Brasil

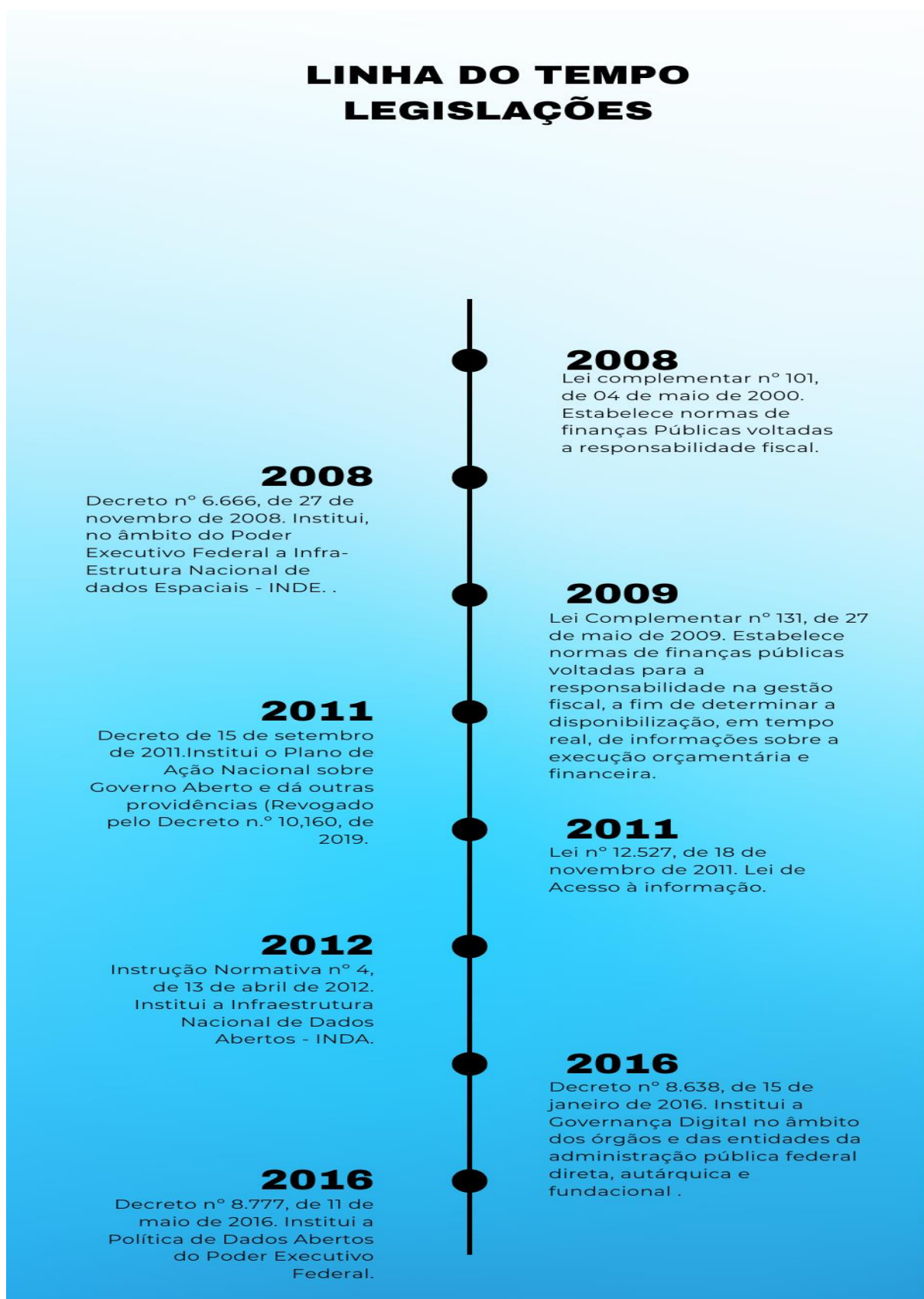
Instrumento	Descrição
Lei complementar n.º 101, de 4 de maio de 2000.	Estabelece normas de finanças públicas voltadas para a responsabilidade na gestão fiscal e dá outras providências (Brasil, 2000).
Decreto n.º 6.666, de 27 de novembro de 2008.	Institui, no âmbito do Poder Executivo Federal, a Infraestrutura Nacional de Dados Espaciais (INDE) - (Brasil, 2008).
Lei complementar n.º 131, de 27 de maio de 2009.	Acrescenta dispositivos à Lei Complementar n.º 101, de 4 de maio de 2000, que estabelece normas de finanças públicas voltadas para a responsabilidade na gestão fiscal e dá outras providências, a fim de determinar a disponibilização, em tempo real, de informações pormenorizadas sobre a execução orçamentária e financeira da União, dos Estados, do Distrito Federal e dos Municípios (Brasil, 2009).
Decreto de 15 de setembro de 2011.	Institui o Plano de Ação Nacional sobre Governo Aberto e dá outras providências (Revogado pelo Decreto n.º 10,160, de 2019).
Lei n.º 12.527, de 18 de novembro de 2011.	Regula o acesso a informações previsto no inciso XXXIII do art. 5º, no inciso II do § 3º do art. 37 e no § 2º do art. 216 da Constituição Federal; altera a Lei n.º 8.112, de 11 de dezembro de 1990; revoga a Lei n.º 11.111, de 5 de maio de 2005, e dispositivos da Lei n.º 8.159, de 8 de janeiro de 1991; e dá outras providências (Brasil, 2011).
Instrução Normativa n.º 4, de 13 de abril de 2012.	Institui a Infraestrutura Nacional de Dados Abertos (INDA) - (Brasil, 2012).

Decreto n.º 8.638, de 15 de janeiro de 2016.	Institui a Governança Digital no âmbito dos órgãos e das entidades da administração pública federal direta, autárquica e fundacional (revogado pelo decreto n.º 10.332, de 2020) (Brasil, 2016).
Decreto n.º 8.777, de 11 de maio de 2016.	Institui a Política de Dados Abertos do Poder Executivo Federal (2016).

Fonte: Adaptado de Silva e Pinheiro (2019).

Para visualização do caminho percorrido, a Figura 9 apresenta a trajetória histórica da legislação brasileira sobre Política de Dados Abertos no Brasil.

Figura 9 – Contextualização histórica da legislação acerca da Política de Dados Abertos no Brasil



Fonte: Adaptado de Silva e Pinheiro (2019).

Segundo Possamai (2016), da promulgação da CF em 1988 a regulamentação efetiva do direito ao acesso à informação pública em 2011, com a LAI, até 2016 da publicação da Política de Dados Abertos do Poder Executivo Federal, diferentes normas objetivam romper com a opacidade histórica da administração pública, tornando visíveis práticas e informações antes restritas. Dessa forma, os dispositivos legais buscam reafirmar, regular e orientar o direito ao acesso, o dever da publicidade e a divulgação das informações públicas (Possamai, 2016, p. 126).

Não existe uma lei única que regule a política de dados abertos no Brasil. Ao contrário disso, a política é constituída por leis dispersas, decretos, resoluções e outros dispositivos legais que são criados para preencher lacunas existentes. A partir necessidades visualizadas, várias outras leis surgiram no ordenamento jurídico para direcionar os DGA. As principais legislações estão reunidas no Quadro 17.

Quadro 17 – Legislação brasileira sobre DGA

Legislação	Ementa
Decreto n.º 9.903, de 8 de julho de 2019.	Altera o Decreto n.º 8.777, de 11 de maio de 2016, que institui a Política de Dados Abertos do Poder Executivo Federal, para dispor sobre a gestão e os direitos de uso de dados abertos (2019).
Resolução n.º 3, de 13 de outubro de 2017.	Aprova as normas sobre a elaboração e publicação de Planos de Dados Abertos, conforme disposto no Decreto n.º 8.777, de 11 de maio de 2016 (Brasil, 2017).
Decreto n.º 10.160, de 09 de dezembro de 2019.	Institui a Política Nacional de Governo Aberto e o Comitê Interministerial de Governo Aberto (2019).
Lei n.º 14.129, de 29 de março de 2021.	Dispõe sobre princípios, regras e instrumentos para o Governo Digital e para o aumento da eficiência pública (2021).
Decreto n.º 10.332, de 28 de abril de 2020.	Institui a Estratégia de Governo Digital para o período de 2020 a 2022, no âmbito dos órgãos e das entidades da administração pública federal direta, autárquica e fundacional e dá outras providências.
Lei n.º 14.129, de 29 de março de 2021.	Dispõe dos princípios, regras e instrumentos para o Governo Digital e para o aumento da eficiência pública e altera a Lei n.º 7.116, de 29 de agosto de 1983, a Lei nº 12.527, de 18 de novembro de 2011 (Lei de acesso à informação), a Lei n.º 12.682, de 09 de julho de 2012, e a Lei n.º 13.460, de 26 de junho de 2017.

Decreto n.º 12.308, de 11 de dezembro de 2024.	Institui o Comitê Interministerial para a Transformação Digital (2024).
PDA	Operacionaliza a Política de Dados Abertos do Poder Executivo Federal, pois planeja as ações que visam a abertura e sustentação de dados abertos nas organizações públicas.

Fonte: Adaptado de Brasil (2017) e Silva (2021).

Cada um dos dispositivos desenvolvidos para fortalecer a temática de Governo Aberto contribuiu significativamente para a criação do Portal Brasileiro de Dados Abertos¹⁷. O Portal foi idealizado para ser o ponto central de acesso aos dados públicos governamentais no Brasil, devido a necessidade de disponibilizar, em um único ambiente, dados provenientes de diversos órgãos governamentais. Além disso, o Portal visa “[...] estabelecer padrões, tecnologias, procedimentos e mecanismos de controle necessários para atender às condições de disseminação e compartilhamento de dados e informações públicas” (Oliveira; Fonseca, 2021, p. 730).

No Portal Brasileiro de Dados Abertos, os dados estão disponíveis para *download* em massa, ou seja, por meio de conjuntos de dados completos que podem ser reutilizados para as mais diversas finalidades. No entanto, apesar de diversos órgãos públicos já disponibilizarem seus dados, muitos ainda não elaboraram seu PDA.

O PDA consiste em um instrumento que operacionaliza a Política de Dados Abertos. Por meio do PDA, os órgãos governamentais elaboram o planejamento das ações de implementação e promoção da abertura de dados de interesse da sociedade (Brasil, 2020b).

Conforme o Decreto n.º 8.777/2016, todos os órgãos e entidades da administração pública federal direta, autárquica e fundacional são obrigados a elaborar seus PDAs e devem publicá-los a cada dois anos (Brasil, 2020b).

Para que a sociedade fiscalize quais órgãos governamentais estão cumprindo com a Política de Dados Abertos, a Controladoria-Geral da União (CGU) criou o Painel de Monitoramento de Dados Abertos¹⁸ (Brasil, 2025). Na ferramenta é

¹⁷ Disponível em: <https://dados.gov.br/home>. Acesso em: 05 dez. 2025.

¹⁸ Disponível em: <https://www.gov.br/cgu/pt-br/governo-aberto/noticias/2017/cgu-lanca-painel-de-monitoramento-de-dados-abertos-do-governo-federal>. Acesso em: 05 dez. 2025.

possível ao usuário: verificar quais órgãos públicos já publicaram seu PDA; quais estão com seu PDA em construção e quais ainda não possuem o PDA.

Segundo o Painel (Brasil, 2025), apenas 93 órgãos governamentais ou 34,83%, possuem seus PDAs publicados e disponíveis. Outros 172 órgãos, que equivalem a 64,42% dos existentes, ainda não elaboraram seu PDA até a última atualização, em fevereiro de 2025. Entretanto 2 ou 0,75% dos órgãos governamentais encontram-se com os seus PDAs em análise pela CGU.

Passados 11 anos desde a publicação do Decreto n.º 8.777/2016, ainda há muito para avançar a respeito da consolidação da política dos DGA. Apesar de exigências previstas no Decreto n.º 8.777/2016 englobar todos os órgãos e entidades da administração pública federal direta¹⁹, autárquica e fundacional²⁰, mais da metade dos órgãos públicos ainda não elaboraram seu PDA.

Segundo Oliveira e Fonseca (2021), algumas instituições justificam que não publicaram seu PDA por falta de entendimento sobre o que são dados abertos, além de alegarem falta de recursos financeiros e humanos para a gestão dos dados. Além desses fatores, outro aspecto importante que colabora com a resistência das instituições a abrirem seus dados é a cultura do sigilo que ainda persiste em parte da administração pública.

¹⁹ É composta pelos órgãos do governo central (União, Estados, DF e municípios) que atuam de forma descentralizada e sem personalidade jurídica própria.

²⁰ A administração indireta é formada por entidades com personalidade jurídica própria e autonomia (como autarquias, empresas públicas e fundações), que atuam na execução de serviços de forma descentralizada, embora vinculadas ao controle estatal.

5 A QUALIDADE DOS DADOS GOVERNAMENTAIS ABERTOS (DGA)

Diversos esforços têm sido empreendidos pelos governos para assegurar um nível adequado de qualidade nos dados disponibilizados ao público. A implementação de um conjunto abrangente e robusto de dimensões de qualidade é essencial para que os órgãos governamentais possam monitorar e aprimorar continuamente seus processos de publicação de dados.

Com o objetivo de efetivar a abertura dos dados governamentais, muitos governos têm investido na criação e manutenção de portais e banco de dados que possibilitam à sociedade civil localizar e utilizar informações de seu interesse de forma ágil e sem barreiras técnicas. No entanto, para que esses portais sejam eficientes, processos de avaliação sistemática são necessários para facilitar a compreensão e o valor que geram (Máchová; Lnenicka, 2017).

Por se tratar de dados oficiais, os dados disponibilizados pelo setor público devem ser precisos e confiáveis. A questão da QD se tornou um fator crucial para o projeto de dados abertos, já que a sua efetividade afeta a tomada de decisão e as receitas dos órgãos que os utilizam. A falta de observância na QD, pode também desacreditar os órgãos responsáveis por sua disponibilização.

A qualidade é definida por Firmani *et al.* (2016, p. 124) como a “[...] totalidade de características de um produto que influenciam sua capacidade de satisfazer as necessidades implícitas, como adequação ao uso, conformidade com os requisitos e a satisfação do usuário”. Assim, para que um conjunto de dados apresente a qualidade necessária, deve-se incorporar uma combinação de dimensões específicas que os tornem aptos para o uso eficiente e eficaz.

No mesmo sentido, Mahanti (2019) define QD como a capacidade dos dados em satisfazerem as necessidades dos usuários, tanto profissional como pessoal. Um dado é considerado de qualidade quando atende a sua finalidade, isto é, quando atende a necessidade informacional do usuário. Para que a qualidade do dado seja alcançada, este deve ser completo, relevante e oportuno.

Este capítulo tem como foco a qualidade dos DGA. Inicialmente, será apresentada uma contextualização sobre o conceito de qualidade dos dados abertos, seguida pela descrição das principais dimensões e métricas a serem aplicadas para garantir que um conjunto de dados atenda aos critérios mínimos estabelecidos.

5.1 A qualidade dos dados abertos (QD)

A preocupação com a QD não surgiu com a disponibilização dos conjuntos de dados pela administração pública. Segundo Silva (2019), o estudo da QD remonta à Antiguidade, sendo mencionado pela primeira vez pelo filósofo Ateniense Tucídides no século IV a.C.

Séculos depois, a estatística foi uma das primeiras disciplinas a investigar a QD, no final da década de 1960, que resultou em uma teoria matemática para considerar duplicatas como conjunto de dados (Batini; Scannapieco, 2006, p. 4). Cerca de vinte anos depois, os pesquisadores de gestão buscavam controlar sistemas de fabricação de dados para detectar e eliminar problemas de sua qualidade (Batini; Scannapieco, 2006, p. 4). Na década de 1990, os pesquisadores da computação iniciaram os estudos para a melhoria da qualidade dos dados eletrônicos armazenados em banco de dados (Batini; Scannapieco, 2006, p. 4).

Entretanto, a pesquisa que marcou a era contemporânea dos estudos da QD foi conduzida por Wang e Strong (1996). Os autores definem a QD como “[...] dados adequados para uso de consumidores de dados” (Wang; Strong, 1996, p. 6). Eles propõem uma abordagem que trata dado como um produto, enfatizando que, “[...] um sistema de informação pode ser visto como um sistema de fabricação de dados que atua na entrada de dados brutos para produzir dados de saída ou produtos de dados” (Wang; Strong, 1996, p. 8). Nessa perspectiva, os consumidores de dados, embora não estejam comprando dados, estão escolhendo usar os dados disponibilizados.

De forma semelhante, Vetrò *et al.* (2016, p. 4) entendem que a QD se refere “[...] à capacidade de uma coleta de dados atender aos requisitos dos usuários”. Já Pagaime (2018, p. 21) define a QD “[...] como um grau em que um conjunto de características dos dados cumpre com determinados requisitos”.

Zaveri *et al.* (2015, p. 6) reforçam que a QD é amplamente conceituada como um construto multidimensional, sendo comumente definida com uma **adequação para uso**, e pode depender de fatores como o contexto do usuário e sua finalidade.

Segundo Jesus, Barité e Santarém Segundo (2024), a partir de uma pesquisa intitulada “O termo qualidade de dados na Ciência da Informação brasileira: uma análise de terminologia pontual” foi constatada que o conceito de qualidade de

dados na CI pode assumir múltiplas interpretações. A QD pode ser compreendida como:

[...] um conceito em si, como um tema de investigação; como um campo consolidado do conhecimento; como um processo; como parte integrante de outros domínios; como um parâmetro para a tomada de decisão, ou ainda como um nível que pode ou não ser atingido (Jesus; Barité; Santarém Segundo, 2024, p. 9).

Em outras palavras, sua definição pode variar conforme os objetivos e abordagens adotadas em cada estudo.

Com o aumento exponencial, em escala mundial, de dados e informações manipulados por computador, a QD se tornou um “[...] componente essencial da qualidade e utilidade das informações derivadas desses dados, onde a maioria dos negócios depende da qualidade dos dados” (International Organization for Standardization - ISO, 2008).

A crescente demanda por dados de qualidade levou a ISO, a elaborar um conjunto de normas internacionais padronizadas (ISO/IEC 25012) para os dados compartilhados, processados e utilizados entre os sistemas de computadores e os usuários (ISO, 2008). Segundo a QD é definida “[...] como o grau em que as características dos dados satisfazem as necessidades declaradas e implícitas quando usados sob condições especificadas” (ISO, 2008, p. 4).

Por outro lado, e com visão nos usuários, Guedes, Fonseca e Strauhs (2021, p. 2), ressaltam que tais conceitos de dados são subjetivos, uma vez que, “[...] um dado pode ser considerado de qualidade e adequado para um indivíduo em um dado momento e inadequado e dispensável em outras tarefas”. Dessa forma, a QD é, muitas vezes, “[...] estabelecida a partir do ponto de vista do usuário” (Guedes; Fonseca; Strauhs, 2021, p. 2).

Considerando a importância do conhecimento para as organizações e indivíduos, e na constatação que o conhecimento é resultado de apropriações de dados e informações, há de ponderar que a QD utilizada neste processo pode influenciar diretamente no conhecimento a ser produzido (Bandeira *et al.*, 2012).

Dados com baixa qualidade tornam-se caros aos usuários. Ao ignorar o fato de que grande parte dos dados disponibilizados apresentam problemas de compreensibilidade e acessibilidade, o usuário, muitas vezes em silêncio, acaba aceitando o problema e não manifesta sua insatisfação. Em uma sociedade em que

os dados são considerados o ativo mais valioso, deve-se priorizar sempre dados de qualidade (Olson, 2003).

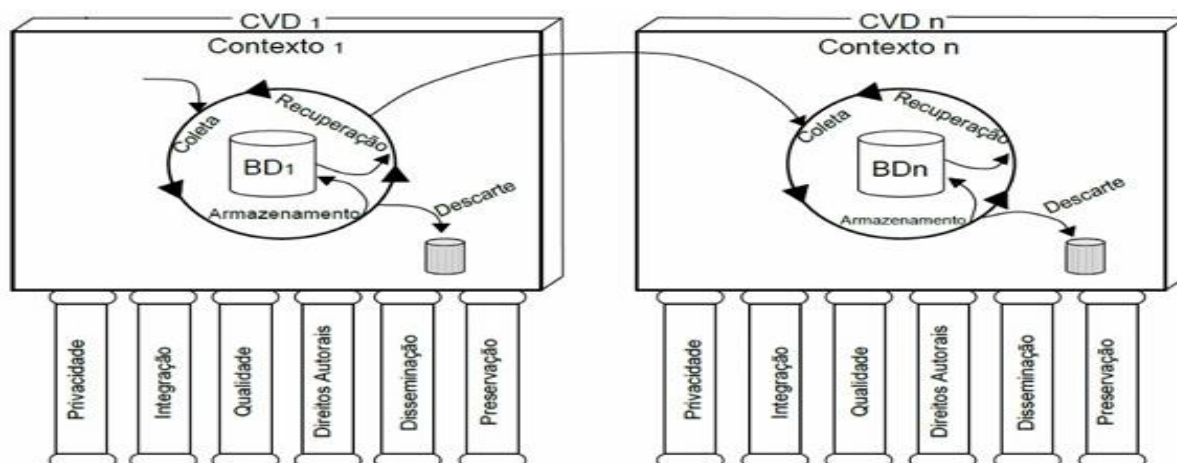
As organizações em geral vêm percebendo que problemas com a QD causam prejuízos como prejuízos financeiros diretos, tomada de decisões comprometidas, impactos operacionais, entre outros. Nesse sentido, Olson (2003, p. 1) afirma que o “[...] custo da má qualidade geralmente está oculto e não é óbvio para aqueles que não o procuram”.

De acordo com Scarisbrick-Hauser e Rouse (2007, p. 170), “[...] os problemas na qualidade dos dados podem ser encontrados em todos os níveis e em todos os componentes da informação, incluindo definição de conteúdo e apresentação”. Assim, a QD pode ser comprometida em todas as fases do Ciclo de Vida dos Dados (CVD), ou seja, seus problemas podem surgir do estágio inicial - coleta/aquisição, ao processamento, transferência, armazenamento ao arquivamento ou descarte (Mahanti, 2019).

No mesmo sentido, Moses, Gavish e Vorwerck (2024, p. 24) afirmam que a QD pode sofrer impacto em qualquer estágio do *pipeline* de dados, antes da ingestão, na produção ou mesmo durante a análise.

O CVD está representado na Figura 10:

Figura 10 – O CVD



Fonte: Sant'Ana (2016, p. 123).

Desde a criação até a preservação, os dados são continuamente transferidos entre sistemas e interfaces, e, à medida que avançam, suas qualidades podem ser afetadas de diferentes formas.

No CVD de Sant'Ana (2016, p. 123) são identificadas quatro fases: coleta, armazenamento, recuperação e descarte. Em todas as fases, os seguintes fatores devem estar presentes: privacidade, integridade, **qualidade**, direitos autorais, disseminação e preservação.

Na fase da coleta, a observação por dados que atendam aos requisitos de qualidade garante a qualidade desde o início de seu CVD. A observação desses fatores evita problemas futuros e contribui para dados confiáveis. Dessa forma, a QD impacta diretamente na confiança que possuem.

A coleta de dados realizada incorretamente pode afetar de forma negativa em toda a cadeia de dados. Se os dados já possuírem erros desde o momento da coleta, toda análise, mesmo que bem executada, estará comprometida. Dessa forma, “[...] melhorar a qualidade dos dados requer a capacidade de avaliar o relacionamento entre entradas e saídas para garantir que as entradas atendam aos requisitos do processo e as saídas estejam em conformidade com as expectativas” (ENAP, 2019, p. 25).

Segundo explica Sant'Ana (2016), a QD na fase de armazenamento refere-se às ações sobre a garantia que os dados devem manter no tempo e no espaço a integridade física e lógica.

Na fase da recuperação de dados, a qualidade é um fator chave no CVD. Os recursos de dados utilizados na recuperação devem ter “[...] especial atenção à interação do usuário” (Sant'Ana, 2016, p. 134).

A etapa de descarte também demanda atenção à qualidade, especialmente no que diz respeito à interação com o usuário e à necessidade de manter registros detalhados sobre o processo. Isso ocorre porque a documentação do descarte assegura rastreabilidade e transparência, permitindo reconstruir o histórico do ciclo de vida da informação e demonstrar conformidade com normas arquivísticas e legais. Além disso, esses registros possibilitam a realização de análises futuras, favorecem a continuidade administrativa, apoiam auditorias e mitigam riscos operacionais e legais relacionados à eliminação incorreta de dados. Assim, manter informações precisas sobre o que foi descartado, quando, como e por qual justificativa, constitui uma prática essencial para a boa governança da informação e para o fortalecimento da confiança dos usuários nos processos institucionais.

Porém, segundo a Escola Nacional de Administração Pública (ENAP, 2019), o processo de criação de dados pode consistir apenas na **coleta de dados** ou ainda, envolver várias etapas:

- **coleta de dados**: captura dos dados;
- **integração em uma *data warehouse***²¹: organização dos dados coletados em um repositório centralizado;
- **agregação em um *data mart***²²: organização de um subconjunto de um *data warehouse* para finalidades específicas.

Olson (2003) destaca que um dos principais problemas relacionados à QD está associado à entrada manual realizada por seres humanos, estes naturalmente suscetíveis a erro. As pessoas podem inserir informações incorretas, fornecer informações incompletas, escolher opções erradas, o que compromete a QD disponibilizadas.

Dessa forma, para que seja possível mensurar a QD, é necessário avaliar uma ou mais de suas dimensões, em determinados contextos, situação e tarefa para as quais os dados serão utilizados (Mahanti, 2019, p. 129).

A avaliação da QD pode ser realizada por meio de diferentes abordagens, que variam conforme o contexto de uso, os objetivos da análise e os tipos de dados envolvidos. Segundo Wang e Strong (1996, p. 7), a literatura científica apresenta três principais abordagens para o estudo da QD:

1) Abordagem Intuitiva: essa abordagem é adotada quando a seleção dos atributos da QD é orientada pela experiência ou pela percepção intuitiva dos pesquisadores, que decidem o que é “importante”. Ou seja, a seleção ocorre com base no julgamento intuitivo, sem recorrer a fundamentos teóricos. As diferentes seleções de atributos realizadas ao longo dos estudos formam um conjunto de características de qualidade comumente utilizadas para avaliar a QD. Geralmente, cada disciplina adota atributos específicos para a avaliação da QD, o que resulta na identificação de múltiplas dimensões. Na Abordagem Intuitiva cada estudo pode selecionar os atributos mais relevantes para os objetivos específicos.

²¹ Repositório centralizado que coleta e armazena dados de várias fontes para análise e geração de relatórios.

²² Um subconjunto de um *data warehouse* focado em uma área de negócios específica, como *marketing* ou finanças, para fornecer dados direcionados a um grupo de usuários.

2) Abordagem Teórica: essa abordagem se concentra nas deficiências que podem surgir durante o processo de produção dos dados. Embora seja frequentemente recomendada, a literatura científica oferece poucos exemplos práticos de aplicação. Um exemplo identificado considera que os atributos da QD decorrem de suas deficiências, entendidas como as inconsistências entre a visão de um sistema do mundo real e a que seria possível obter por meio da observação direta desse mesmo mundo. A abordagem teórica pode, assim, oferecer um conjunto abrangente de atributos da QD que são intrínsecos ao próprio produtor de dados.

3) Abordagem Empírica: nessa abordagem, a QD é avaliada a partir de sua análise quando coletada diretamente com os usuários. O objetivo é identificar quais atributos os usuários consideram relevantes para avaliar se os dados são adequados para suas tarefas. Diferentemente das abordagens intuitiva e teórica, as características identificadas na abordagem empírica não podem ser determinadas previamente ou selecionadas intuitivamente pelos pesquisadores ou por fundamentações teóricas. Ao contrário, reflete a percepção dos usuários no uso dos dados, o que pode resultar na identificação de atributos que os pesquisadores não consideram para avaliação da qualidade.

Partindo das múltiplas definições atribuídas ao termo QD, Jesus, Barité e Santarém Segundo (2024) propuseram uma especificação conceitual em três eixos principais: **(I) - enquanto um campo do conhecimento, um domínio ou ainda um tema de investigação; (II) - enquanto um processo e (III) - enquanto uma medida ou uma ponderação.**

No primeiro eixo, **enquanto um campo do conhecimento, um domínio ou ainda um tema de investigação**, a QD é reconhecida por seu caráter interdisciplinar, despertando o interesse de diversas áreas. Nesse contexto, formas de avaliar e aprimorar os níveis de qualidade dos conjuntos de dados são discutidas, investigadas e propostas, bem como artefatos e metodologias voltadas à sua melhoria (Jesus; Barité; Santarém Segundo, 2024).

No segundo eixo, **enquanto um processo**, a QD é frequentemente abordada como sinônimo da própria avaliação de qualidade. Nessa perspectiva, o foco recai sobre os procedimentos e etapas, utilizados para verificar se os dados atendem a determinados padrões ou requisitos. A avaliação é conduzida com base na escolha de dimensões, critérios e métricas que possibilitam analisar diversos aspectos da QD (Jesus; Barité; Santarém Segundo, 2024).

No terceiro eixo, **enquanto uma medida ou ponderação**, a QD é compreendida como sinônimo dos conceitos de dados de qualidade, qualidade de dados e de qualidade aplicada de dados. A literatura científica frequentemente representa essa ideia por meio da expressão “*fitness for use*”, ou adequado ao uso, destacando que a QD está diretamente relacionada à sua capacidade de atender às necessidades dos usuários em contextos específicos (Jesus; Barité; Santarém Segundo, 2024).

Segundo Jesus, Barité e Santarém Segundo (2024), o termo QD pode ser abordado a partir de diferentes perspectivas ou dimensões, que influenciam diretamente tanto o processo de mensuração da qualidade quanto a identificação de conjuntos de dados considerados adequados.

As dimensões de qualidade referem-se às características dos dados, as quais podem ser tanto quantificadas quanto qualificadas por meio de métricas. Esse processo de mensuração permite que o usuário avalie o quão adequados os dados estão para realização de suas atividades (Jesus; Barité; Santarém Segundo, 2023).

Wang e Strong (1996, p. 11), ao analisarem os atributos de qualidade identificados na literatura científica e a forma como esses atributos são percebidos pelos usuários, bem como a importância atribuída por eles no uso dos dados, desenvolveram um conjunto de dimensões de QD que servem de referência para sua avaliação.

A utilização de uma estrutura robusta de dimensões de QD possibilita aos PIs, compreender e atender fidedignamente às necessidades dos usuários quanto aos dados disponibilizados (Wang; Strong, 1996).

Portanto, essa pesquisa apresentará um conjunto de dimensões de qualidade que permitem mensurar, avaliar e aprimorar a QD disponibilizados por meio dos portais de DGA. As dimensões que serão apresentadas avaliam diferentes variáveis para avaliação dos dados. Essas dimensões possibilitam aos PI, criarem conjuntos de dados e ao mesmo tempo, avaliarem quais dimensões de QD podem ser aplicadas.

5.2 Dimensões para avaliar a qualidade dos dados abertos

As dimensões de QD constituem o núcleo central para a avaliação dos dados disponibilizados. Dentre as metodologias existentes para avaliação da QD, muitas se

fundamentam na identificação e mensuração de dimensões que impactam diretamente a QD.

Ainda que a QD possa ser subjetiva, ou seja, variar conforme o perfil de cada usuário é fundamental, segundo Mahanti (2019, p. 14), “[...] traduzir as necessidades subjetivas do usuário em dimensões objetivas na qual é possível medir é essencial”. Nesse sentido, é papel do PI transformar as necessidades informacionais subjetivas dos usuários em objetivas e passíveis de recuperação.

Para tanto, Zaveri *et al.* (2015, p. 6), explicam que “[...] a avaliação da qualidade de dados envolve medição de dimensões ou critérios de qualidade que são relevantes para o consumidor”. Complementando essa definição, Penteado (2020, p. 87) afirma que uma dimensão ou elemento de qualidade é um ou o “[...] aspecto principal pelo qual a qualidade pode ser observada, comportando diversos critérios de qualidade para caracterizá-lo”.

No mesmo sentido, Mahanti (2019, p. 43) afirma que os dados devem estar em conformidade com as várias dimensões de QD para serem adequados para uso operacional e analítico. Cichy e Rass (2019) conceituam as dimensões de dados como “[...] atributos de qualidade de dados que podem, quando medidos corretamente, indicar o nível geral de qualidade dos dados”. Para autores como Wang e Strong (1996, p. 6), a dimensão de qualidades dos dados é definida “[...] como um conjunto de atributos de qualidade dos dados que representam um único aspecto ou construção da qualidade dos dados”.

A partir das definições, é possível inferir que “[...] cada dimensão captura um aspecto específico incluído sob o guarda-chuva geral da qualidade dos dados” (Batini; Scannapieco, 2006, p. 19). Isso significa que não há consensos acerca das dimensões necessárias para avaliar dados. Por isso, ao definir medidas de QD, deve-se tentar focar nas dimensões que são significativas e pertinentes ao usuário (Mahanti, 2019).

As dimensões utilizadas para avaliar a QD podem ser agrupadas em categorias gerais, conforme proposto por Liu e Chi (2002), Mahanti (2019) e Strong, Lee e Wang (1997):

- **qualidade intrínseca dos dados:** essa categoria indica que os dados possuem qualidade por si só, independentemente do contexto ou da ação a ser realizada na utilização. Isso significa que atributos essenciais, como exatidão,

consistência, completude e integridade, estão presentes nos dados em sua própria natureza;

- **qualidade de dados contextuais:** a qualidade de dados contextuais indica que a QD deve ser considerada no contexto da tarefa a ser executada. Assim, a qualidade de dados contextuais pode ser compreendida como o nível de relevância dos dados nas tarefas dos usuários. Essa categoria tem ênfase nas tarefas dos usuários de dados;

- **qualidade de dados representacionais:** essa categoria representa as dimensões responsáveis por avaliar/medir a qualidade do formato dos dados e o grau em que significados podem ser derivados. A qualidade dos dados representacionais baseia-se na usabilidade direta dos dados.

- **qualidade dos dados de acessibilidade:** refere-se aos requisitos de acesso aos dados, destacando a importância das permissões necessárias para o acesso aos sistemas.

Dessa forma, Wang e Strong (1996, p. 22) compreendem que “[...] dados de alta qualidade devem ser intrinsecamente bons, contextualmente apropriados para a tarefa, claramente representados e acessíveis ao consumidor de dados”.

As dimensões de QD podem ser avaliadas sob três perspectivas diferentes, conforme esclarece Mahanti (2019, p. 133):

- **modelagem de dados:** corresponde ao esquema conceitual ou dimensões de esquema;

- **valor dos dados:** refere-se explicitamente a valores, independentemente da representação interna dos dados;

- **apresentação dos dados:** está relacionada ao formato dos dados.

Mahanti (2019) destaca que os modelos de dados devem ser construídos de maneira consciente e coerente, de modo que, além de atenderem aos requisitos funcionais, também permitam que os usuários os compreendam.

No contexto da QD, também estão inseridas as métricas. Segundo Zaveri *et al.* (2015, p. 6), uma métrica de “[...] avaliação da qualidade dos dados é um procedimento utilizado para medir uma determinada dimensão da qualidade da informação”. Complementando a ideia, Batini *et al.* (2009, p. 3), afirmam que o “[...] termo medição é usado para abordar a questão de medir o valor de um conjunto de

dimensões e qualidade de dados”. Portanto, as métricas são necessárias para avaliar as dimensões de QD.

Dessa forma, “[...] a medição representa a primeira fase para se afastarem do mundo das percepções, mitos, equívocos, suposições, política e entrar no mundo dos fatos e números” (Mahanti, 2019, p. 214). O autor acrescenta ainda que, por meio da medição, é possível revelar verdades ocultas nos dados, reforçando sua importância como etapa inicial na avaliação da QD (Mahanti, 2019).

A literatura científica a respeito da QD fornece uma variedade de dimensões que podem ser aplicadas à sua medição. No entanto, Batini *et al.* (2009, p. 6) ressaltam que não há consenso quanto ao conjunto ideal de dimensões que definem a QD, tampouco sobre o significado exato de cada uma delas. Nesta constante, caberá ao PI selecionar as dimensões que se aplicam ao contexto e necessidades da sua comunidade.

De acordo com Sebastian-Coleman (2013), para as medições de QD serem eficazes, devem ser compreensíveis e interpretáveis. O autor observa que “[...] se os usuários não entenderem qual a dimensão ou métrica que está sendo medida, a medição não ajudará a reduzir a incerteza ou será útil, mesmo que o objeto que está sendo medido seja importante” (Sebastian-Coleman, 2013, p. 128).

Além disso, Sebastian-Coleman (2013) destaca que as medições devem ser reproduzíveis, ou seja, funcionar como uma metodologia que pode ser refeita. Isso significa que, ao utilizar os mesmos instrumentos e condições de medição, os resultados obtidos devem ser consistentes e, ao menos, muito próximos. Nesse sentido, Mahanti (2019) complementa afirmando que “[...] para medir a qualidade dos dados, é preciso medir uma ou mais dimensões da qualidade de dados, dependendo do contexto, da situação e tarefa para a qual os dados serão usados”.

As definições das dimensões de QD representam uma etapa de caracterização conceitual e de discussão teórica que fundamenta o desenvolvimento metodológico proposto na pesquisa. Essa etapa teve como base a RSL descrita na seção dois, por meio da qual foram identificadas 41 dimensões de qualidade utilizadas por diferentes autores em estudos voltados à avaliação da QD. A Análise comparativa dessas dimensões revelou a existência de sobreposições terminológicas, redundâncias conceituais e variações semânticas entre distintas nomenclaturas que, embora divergentes em denominação, apresentavam funções equivalentes no processo de avaliação da QD. Diante disso, foi realizada a

compilação e consolidação de um conjunto de dados composto por 15 dimensões, selecionados por sua abrangência, representatividade e capacidade de caracterizar e mensurar com clareza os conjuntos de DGA. Essas 15 dimensões, configuram a base conceitual de referência para a análise e discussão da QD, atuando como fundamento teórico e estrutural para posterior desenvolvimento do instrumento metodológico de avaliação.

Com o objetivo de demonstrar de que forma as 41 dimensões foram agrupadas e sistematizadas em 15 outras, apresenta-se o Quadro 18.

Quadro 18 – Consolidação das 15 dimensões de QD selecionadas

N.º	Dimensões selecionadas	Dimensões secundárias	Definição
1	Acessibilidade	Inclusão	Capacidade dos usuários com alguma deficiência, seja física, sensorial, intelectual e psicossocial em acessar, compreender e utilizar os dados de forma autônoma.
2	Completeness	Integridade Granularidade	Garante que as informações essenciais estejam presentes e sejam suficientes para o objetivo a que se destinam.
3	Atualidade	Velocidade	Se refere ao tempo de coleta ou geração dos dados e sua disponibilização.
4	Formato do Arquivo	<i>Download</i> de dados	Relaciona-se ao formato em que os dados são disponibilizados e a sua condição de serem manipulados por ferramentas de armazenamento e processamento.
5	Disponibilidade	Oportuno Expiração Oportunidade Segurança Transparência Relevância	Assegura que os conjuntos de dados sejam publicados e permaneçam disponíveis ao longo do tempo. Refere-se ao quanto a informação está presente, pode ser obtida e está pronta para uso.
6	Precisão	Objetividade Exatidão Redundância	Refere-se à correção e a exatidão dos dados, e em como a sua disponibilização reflete a realidade.
7	Recuperabilidade	Metadados Descoberta Categorização	Refere-se à facilidade com que os dados são localizados, acessados e obtidos pelos usuários.

8	Inconsistência		Refere-se ao grau de coerência lógica de determinada variável em função de outra variável complementar. Os dados não apresentam contradições ou discrepâncias.
9	Compreensibilidade	Adequação	Refere-se à clareza e a facilidade em que os dados são apresentados, permitindo que qualquer usuário consiga entendê-los e utilizá-los.
10	Conformidade	Credibilidade	Os dados estão disponibilizados de acordo com a legislação e os regulamentos vigentes.
11	Licença aberta	Não proprietário	Os dados não podem estar sujeitos a qualquer regulamento de direitos autorais, patentes, marcas registradas ou segredos comerciais.
12	Usabilidade	Personalização Navegabilidade Visualização <i>Design</i> Responsividade	Garante que o usuário, mesmo com poucas habilidades técnicas ou conhecimentos específicos, possam explorar, manipular e analisar os dados.
13	Não discricionário	Gratuidade	Refere-se a imparcialidade no acesso ao uso de dados. Permite que todos os usuários tenham acesso irrestrito aos conjuntos de dados, sem que haja distinção ou privilégios baseados em seu perfil.
14	Primário	Rastreabilidade	Refere-se à origem dos dados. São dados disponibilizados conforme estão disponíveis na fonte original.
15	Interoperabilidade		Permite a troca eficiente de dados entre diversos sistemas, aplicações e usuários, sem barreiras técnicas ou semânticas.

Fonte: Elaborado pela autora (2026).

Por meio do Quadro da consolidação das 15 dimensões de qualidade propostas é possível obter uma visão mais clara e organizada das dimensões, permitindo identificar relação de complementaridade e evitar sobreposições conceituais.

5.2.1 Acessibilidade

A acessibilidade não está relacionada apenas à disponibilidade dos dados, mas também com a capacidade dos usuários com alguma deficiência em utilizá-los de forma autônoma e sem barreiras. Dados acessíveis apresentam descrições claras, detalhadas e são comunicados em uma linguagem adequada a diferentes perfis de usuário, incluindo aqueles com limitações técnicas, cognitivas ou sensoriais. Dessa forma, a acessibilidade envolve mais que acesso técnico aos dados.

Pessoas com alguma deficiência podem ter dificuldades para interagir e acessar o conteúdo na *web*; não conseguir concluir um cadastro por não existir um botão com rótulos adequados (Brasil, 2023). Quando os conteúdos disponibilizados digitalmente são acessíveis, beneficiam todos os indivíduos, especialmente as que possuem algum tipo de deficiência (Brasil, 2023).

De acordo com Batini e Scannapieco (2006), a acessibilidade “[...] mede a capacidade do usuário de acessar os dados de sua própria cultura, status/funções físicas e tecnologias disponíveis”. Nessa perspectiva, Mahanti (2019, p. 200) complementa ao definir acessibilidade como a “[...] facilidade com que a existência de dados e/ou metadados pode ser determinada, bem como a adequação da forma ou meio através do qual esses dados podem ser acessados e recuperados com rapidez e simplicidade”.

Dorobat e Posea (2021) reforçam que disponibilizar dados a todos os usuários, sem restringir seu campo de utilização, é uma forma efetiva de garantir a acessibilidade dos conjuntos de dados. A falta de acessibilidade, “[...] tanto no mundo físico quanto no mundo digital, impacta principalmente as pessoas com deficiências permanentes, temporárias ou aquelas que dependem de acessibilidade por qualquer outra condição” (Brasil, 2023).

De acordo com Zhang e Xiao (2020, p. 216), a acessibilidade também integra os “[...] princípios dos dados governamentais abertos [...]” e é definida como “[...] o grau em que os dados podem ser acessados em um contexto específico de uso”. A acessibilidade dos DGA é considerada uma “[...] condição essencial para o uso desses dados governamentais. Se o usuário não conseguir acessar o dado, ele não será capaz de encontrá-lo e fazer uso dele” (Zhang; Xiao, 2020, p. 217).

Ao disponibilizar um conjunto de dados, é imperativo que sejam acessíveis a todos os usuários que desejam utilizá-lo, incluindo aqueles com alguma deficiência. Os dados devem ser acessíveis sem barreiras, garantindo a participação plena e eficaz de toda a sociedade. Para tanto, os dados devem estar em conformidade com normas e leis que assegurem a acessibilidade. Mahanti (2019, p. 200) apresenta questionamentos que devem ser realizados para avaliar a acessibilidade dos DGA:

- I - os dados estão disponíveis para acesso e uso?
- II - se os dados estiverem disponíveis, como eles podem ser acessados?
- III - os dados são facilmente acessíveis?

No que diz respeito à disponibilização de conjuntos de dados sem barreiras e acessíveis a todos os usuários, Batini e Scannapieco (2006) propõem um conjunto de diretrizes fundamentais para garantir a acessibilidade:

- a primeira diretriz refere-se ao fornecimento de alternativas equivalentes ao conteúdo auditivo visual, denominado conteúdo equivalente de texto. O equivalente de texto torna-se uma imagem acessível quando o conteúdo do texto for apresentado ao usuário, como fala sintetizada, *braille* e texto exibido visualmente, tornando a informação acessível a grupos com diferentes deficiências;
- evitar o uso exclusivo de cores para transmitir informações;
- uso de linguagem natural e clara;
- uso de termos simples;
- planejar um *site* considerando as necessidades de todos os usuários;
- fornecer informações e orientações sobre a página, para os usuários.

Mahanti (2019, p. 201) destaca que a acessibilidade, por si só, não é suficiente para garantir a QD. É fundamental que os dados estejam completos, em conformidade com padrões estabelecidos e livres de problemas relacionados a outras dimensões.

Nesse sentido, para que um conjunto de dados seja acessível é necessário a combinação de outras dimensões para assegurar que os dados sejam de qualidade. A adoção de ações que combinem dimensões distintas é essencial para garantir que um conjunto de dados esteja acessível ao maior número possível de usuários, promovendo inclusão e permitindo que todos, independentemente de suas habilidades ou limitações, possam utilizá-lo de maneira plena.

5.2.2 Completude

Outra dimensão de destaque na avaliação da QD é a completude. Enquanto a dimensão acessibilidade é considerada como uma qualidade do serviço e pode ser avaliada pelas necessidades do usuário, a dimensão completude é uma qualidade do produto que pode ser avaliado por especificações (Cichy; Rass, 2019).

A acessibilidade garante que todos os dados estejam acessíveis a todos os usuários, enquanto a completude assegura que os dados sejam completos e sem falhas. Desta forma, a acessibilidade e a completude agrupadas conferem qualidade aos dados, além de tornar os dados acessíveis, completos e adequados para uso.

A princípio, a completude significa o dado possuir todas as partes necessárias ou apropriadas para ser disponibilizado: ser inteiro, acabado e total. Assim, a primeira condição de completude é a existência dos dados, necessidade óbvia, mas esquecida por muitos (Sebastian-Coleman, 2013).

Um conjunto de dados completos significa que todos os dados necessários estão presentes e que nenhuma informação essencial está ausente. Quando um conjunto de dados está incompleto, pode gerar interpretações incorretas ou resultados enviesados, impactando diretamente o seu potencial de uso e reúso dos dados.

De acordo com Moreira *et al.* (2020, p. 121), a completude está relacionada ao grau de preenchimento dos campos analisados, mensurado pela proporção entre a quantidade de campos preenchidos e não preenchidos. Em relação aos DGA, Dorobat e Posea (2021, p. 02) consideram outro ponto de vista, pois - são dados completos “[...] quando não estejam sujeitos a limitações válidas de privacidade, segurança ou privilégio”.

Os dados completos em termos de qualidade “[...] não apresentam espaços vazios ou nulos e possuem um número significativo de registros que permitem a definição de tendências ou comportamentos quando analisados” (Herrera-Melo; González-Sanabria, 2020). Espaços ou valores vazios ou nulos referem-se a dados ausentes, ou seja, a valores que embora existam no mundo real, não estão presentes no conjunto de dados (Batini *et al.*, 2009, p. 7). No que se refere aos metadados, devem fornecer “[...] informações suficientes para compreender o conteúdo, o escopo e a finalidade dos dados, além de possuírem informações que permitam o contato com a fonte” (Herrera-Melo; González-Sanabria, 2020).

Zhang e Xiao (2020, p. 216) reforçam que a completude “[...] é um princípio fundamental dos Dados Governamentais Abertos e um atributo contextual de qualidade dos dados, sendo definida como o número de campos preenchidos em um registro de dados”. Além disso, Zaveri *et al.* (2015) classificam a completude dos dados nas seguintes categorias:

a) **completude do esquema**: grau em que entidades e atributos não estão ausentes em um esquema. O esquema se refere a estrutura formal que define como as informações são organizadas, ou seja, é o modelo de um conjunto de dados.

b) **completude da coluna**: é uma função dos valores ausentes para uma propriedade ou coluna específica;

c) **completude de população**: se refere à proporção de entidades representadas em um sistema de informação para a população completa.

Desta forma, “[...] um conjunto de dados é completo na medida em que contém atributos necessários e um número suficiente de registros, e na medida em que os atributos são preenchidos de acordo com as expectativas do consumidor de dados” (Sebastian-Coleman, 2013).

Para avaliar a dimensão de completude, Mahanti (2019) sugere considerar as seguintes questões:

- I - todas as informações necessárias estão disponíveis?
- II - valores e dados críticos estão ausentes nos registros de dados?
- III - todos os conjuntos de dados estão registrados?
- IV - todos os itens dos dados obrigatórios estão registrados?

É importante observar que a ausência de determinados critérios nos dados pode gerar algumas implicações: em alguns casos, os dados ausentes podem ser irrelevantes, enquanto, em outros, sua ausência é crucial para um processo específico (Mahanti, 2005). Como por exemplo, em um conjunto de dados sobre servidores públicos. Se o atributo número do ramal telefônico estiver ausente, essa ausência pode ser irrelevante para a análise da remuneração dos servidores ou a distribuição de cargos. Por outro lado, a ausência de informações sobre a remuneração mensal dos servidores compromete a análise, pois o elemento essencial para a mensuração está incompleto.

Portanto, a completude representa a condição em que todas as informações necessárias aos dados estão presentes, sem lacunas significativas. A presença de falhas nesse aspecto pode levar o usuário a obter informações inadequadas ou a alcançar resultados incorretos.

5.2.3 Atualidade

A atualidade, assim como a acessibilidade e a completude, contribui diretamente para a QD. Um conjunto de dados acessível e completo não possui utilidade se estiver defasado no tempo e não representar a realidade. A qualidade atualidade relaciona-se ao tempo de coleta ou geração dos dados e sua disponibilização.

A crescente necessidade por dados e informações em diversos setores da sociedade leva os usuários a buscarem dados sempre atualizados. Na tomada de decisões, por exemplo, o uso de informações desatualizadas pode perder valor analítico e resultar em escolhas equivocadas, principalmente em situações que exijam respostas rápidas.

A definição mais geral de atualidade ou pontualidade em relação à QD é entendida como “[...] o grau em que os dados representam a realidade do ponto de tempo necessário”, ou ainda, à capacidade dos dados disponibilizados refletirem o momento presente (Sebastian-Coleman, 2013, p. 165).

Nesse sentido, a atualidade dos dados está relacionada à frequência com que são disponibilizados e atualizados em portais ou sistemas de informação. Pode ser definida como o lapso de tempo que ocorre entre uma mudança de um estado real e seu registro no sistema de informação ou base de dados (Moreira *et al.*, 2020, p. 121).

Liu e Chi (2002, p. 302) ressaltam que a atualidade deve ser avaliada de acordo com a tarefa a ser realizada, pois algumas atividades exigem dados mais recentes, enquanto outras podem ter seus objetivos alcançados mesmo com informações menos atualizadas.

A atualidade possui dois componentes fundamentais: **a idade** e a **volatilidade**. A idade refere-se a uma medida de quão antigo o dado é com base no tempo que levou para ser registrado. Dessa forma, quanto mais recente os dados forem coletados, mais relevantes tendem a ser (Bovve; Srivastava; Mak, 2003). Por

outro lado, a volatilidade é uma medida de instabilidade dos dados. Os dados podem ser alterados a qualquer momento por interesse da organização que a disponibiliza, o que os levam a se tornarem desatualizados mais rapidamente, em uma relação volátil. Os dados não voláteis não sofrem ações como essa e são considerados como estáveis (Bovve; Srivastava; Mak, 2003).

Para ilustrar a volatilidade, Batini e Scannapieco (2006) apresentam exemplos claros: dados estáveis como datas de nascimento são considerados não voláteis, pois não sofrem variações; já as cotações de ações, por variarem frequentemente, possuem um alto grau de volatilidade pois permanecem válidas por intervalos pequenos de tempo.

Os dados são considerados oportunos quando são atualizados com a frequência necessária para atender os objetivos do usuário. Caso contrário, podem tornar-se irrelevantes (Bovve; Srivastava; Mak, 2003). Assim, cabe ao usuário avaliar, de acordo com suas necessidades, se os dados disponíveis são atuais o suficiente para seu propósito específico.

Para garantir a atualidade dos dados, é fundamental estabelecer processos sistemáticos de atualização. A lentidão na atualização dos dados compromete não somente o uso imediato e eficaz, como também pode gerar desconfiança no usuário.

5.2.4 Formato de Arquivo

Outra dimensão relevante da QD se refere ao formato de arquivo em que os dados são disponibilizados. O principal objetivo da disponibilização de dados em portais governamentais é promover seu uso, reúso e redistribuição, garantindo aos usuários acesso irrestrito e efetivo à informação. Nesse contexto, o formato em que os dados são apresentados exerce influência direta sobre a acessibilidade dos dados.

Trata-se de uma dimensão de qualidade que impacta o “[...] arranjo dos elementos informacionais que constituem os objetos do ambiente informacional, suas relações e representações [...]” (Moreira *et al.*, 2020, p. 122). A dimensão formato do arquivo está relacionada à tecnologia que sustenta os conjuntos de dados disponibilizados, bem como à sua capacidade de serem manipulados por ferramentas de armazenamento e processamento.

Para contemplar satisfatoriamente o formato do arquivo, recomenda-se o uso de dados no “[...] formato CSV, pois ele permite uma flexibilidade maior na transferência de dados entre aplicações [...]”, facilitando seu reuso (Moreira *et al.* (2020, p. 125). Além disso, os dados devem ser disponibilizados em formatos processáveis por máquina e não proprietário, de modo que ampliem seu alcance e acessibilidade (Máchová; Lnénicka, 2017).

O formato dos arquivos também se alinha a princípios internacionalmente reconhecidos, como os Oito Princípios de *Sebastopol* e integra o esquema de cinco estrelas criado para dados abertos, proposto por Tim Berners-Lee. Contudo, apesar da relevância, a disponibilização de conjuntos de dados em formato processável por máquinas e não proprietário não é exigida por muitas iniciativas de governo aberto, sendo os dados frequentemente publicados apenas para atender a requisitos formais (Attard *et al.*, 2015).

A adoção de dados em formato padronizados contribui para o funcionamento de outras dimensões de qualidade, como a interoperabilidade entre plataformas, promovendo a integração eficiente entre elas.

Em síntese, a escolha adequada do formato de arquivo é um elemento crítico para assegurar a eficácia no uso dos dados. A adoção de formatos abertos e compatíveis com diversas ferramentas de processamento não apenas amplia a acessibilidade e o potencial de reutilização dos dados, mas também fortalece a transparência e a eficiência na Gestão da Informação (GI).

5.2.5 Disponibilidade

A dimensão da disponibilidade é essencial para a QD, pois está intimamente relacionada à acessibilidade e atualidade dos dados, assegurando o acesso, uso e reuso dos dados sempre que necessário. Em relação aos DGA, a disponibilidade assegura que os conjuntos de dados sejam publicados e permaneçam disponíveis ao longo do tempo.

De acordo com Zaveri *et al.* (2015, p. 17), disponibilidade “[...] refere-se à extensão em que a informação está disponível, ou é facilmente recuperável [...], ou seja, disponibilidade de um conjunto de dados é a extensão em que a informação está presente, pode ser obtida e está pronta para uso”. Para que essa dimensão

seja efetiva, é imprescindível que todos os métodos de acesso²³ aos dados funcionem de forma adequada (Zaveri *et al.*, 2015).

Na mesma linha de pensamento, Batini e Scannapieco (2006, p. 30) afirmam que a disponibilidade dos dados se refere ao “[...] período de tempo em que os dados permanecem válidos”. Assim, a disponibilidade está estreitamente relacionada com a atualidade dos dados: enquanto a atualidade exige que os dados sejam atualizados constantemente, a disponibilidade assegura que os dados estejam disponíveis no tempo para eventos que correspondem ao seu uso.

Segundo Herrera-Melo e González-Sanabria (2020), a disponibilidade está assegurada quando os dados estão “[...] disponíveis para visualização, download, uso e reutilização”. Complementarmente, essa dimensão garante que os usuários tenham acesso tanto às informações principais quanto às informações complementares, no momento e local adequados (Maestre-Góngora; Rangel-Carrillo; Osorio-Sanabria, 2023, p. 10).

Jetzek (2016, p. 91) reforça que a disponibilidade envolve “[...] dados que não estão sujeitos a restrições de privacidade ou limitações de segurança, estando abertos a todos por padrão”. Para Bovee, Srivastava e Mak (2003) “[...] à medida que a disponibilidade dos dados se torna comum, a capacidade de definir e avaliar rapidamente a qualidade dos dados para tomada de decisão fornece uma vantagem estratégica potencial”.

Entretanto, apenas a disponibilidade dos dados não é suficiente para gerar a vantagem estratégica. É necessário considerar outras dimensões de QD, de forma integrada, para que quando disponibilizados, realmente atendam aos critérios de qualidade e, dessa forma, possibilitarem vantagem estratégica.

A disponibilidade dos dados desempenha, portanto, um papel central na eficiência de processos decisórios. Em setores como saúde, finanças e administração pública, o acesso a informações críticas em tempo real pode significar a diferença entre respostas rápidas a emergências e a ineficiência em processos de planejamento e intervenção.

Em síntese, a disponibilidade é um dos pilares fundamentais na gestão da QD, influenciando diretamente a eficácia e a agilidade dos processos decisórios. Investir em tecnologias e estratégias que garantam a disponibilidade contínua dos

²³ São os canais que possibilitam o uso dos dados.

dados é essencial para promover a transparência, inovação e competitividade em diversos setores da sociedade, o que contribuiu para um ambiente de informação mais acessível e dinâmico.

5.2.6 Precisão

A dimensão precisão está diretamente relacionada à correção e à exatidão das informações em relação à realidade que elas representam. No contexto dos DGA, é imprescindível que a precisão seja cuidadosamente avaliada, pois a utilização de dados imprecisos pode levar a interpretações equivocadas e comprometer a tomada de decisão.

As dimensões de qualidade, precisão e completude, embora muitas vezes se confundam, apresentam características que as diferenciam. Enquanto a completude se refere à ausência de dados completos, em que todos os campos obrigatórios estão preenchidos; a precisão diz respeito aos dados disponibilizados refletirem corretamente a realidade.

Para ilustrar essa diferença, a Tabela 2 apresenta um exemplo fictício de cadastro do cidadão Adão de Souza, residente na Av. Brasil, nº 123, em Barbacena, Minas Gerais (MG), nascido em 12/12/1978 e inscrito no Cadastro de Pessoa Física (CPF) n.º 000.000.000-00.

Tabela 2 – Diferença entre completude e precisão

Dimensão	Nome		Endereço	Data de nascimento	CPF
Completude	Adão Souza	de	Av. Brasil nº 123, Barbacena, MG	12/12/1978	
Precisão	Adão Souza	de	Av. Brasil nº 123, Barbacena, MG	12/12/1087	000.000.000-00

Fonte: Elaborado pela autora (2026).

Observa-se que, na linha da completude, os dados não estão completos, uma vez que a informação sobre o CPF do cidadão não foi preenchida em seu cadastro. Além disso, na linha da precisão, os dados inseridos na Tabela se mostram imprecisos, considerando que a data de nascimento do cidadão foi registrada de forma incorreta. Esse exemplo evidencia a diferença entre as dimensões completude e precisão, que pode ser resumida da seguinte forma:

- completude: presença de todos os dados necessários para compreensão do dado;
- precisão: veracidade dos dados que não representam fidedignidade à realidade.

De acordo com Mahanti (2019, p. 180), “[...] a precisão se refere a quão bem os dados armazenados em um sistema refletem a realidade [...]”, ou seja, “[...] é o grau em que os dados descrevem corretamente o objeto, entidade, situação, fenômeno ou evento do mundo real que está sendo descrito” (Mahanti, 2019, p. 180).

Na avaliação da QD, a precisão se destaca como uma das dimensões mais relevantes. Para o usuário que trabalha com dados é primordial, pois quando são imprecisos, não apenas deixam de refletir a realidade, como também não possuem significado (Silva, 2019). Nesse contexto, Mahanti (2019, p. 181) propõe algumas questões a serem realizadas na avaliação da precisão:

- I - os objetos dos dados caracterizam precisamente os valores do mundo real que eles devem representar e as entidades da vida real que eles modelam?
- II - em que medidas os dados capturam corretamente o que eles foram projetados para capturar?
- III - Os dados representam com precisão a realidade ou uma fonte verificável?

De acordo com Moreira *et al.* (2020, p. 120) “[...] a precisão impacta principalmente os elementos estruturais dos conjuntos de dados e pode ser definida como sendo o cuidado, exatidão, grau de conformidade do dado com o padrão estabelecido como exatidão”. Para Silva, Monteiro e Reis (2020, p. 42), a precisão ou acuracidade significa “[...] que o conteúdo dos arquivos de dados da página é correto e confiável. A informação é precisa e próxima da realidade sendo possível navegar por vários níveis para validar as informações”.

Adicionalmente, a precisão é considerada um dos atributos intrínsecos mais importantes no que se refere à QD. Zhang e Xiao (2020, p. 2016) afirmam que a precisão se refere “[...] à medida que um registro de dados descreve corretamente a respectiva informação [...]”. Ela influencia diretamente na descoberta e uso dos dados. Quanto mais preciso os dados, melhor eles se ajustam à descoberta e ao uso dos usuários”.

A precisão dos dados está diretamente relacionada à consistência das medições. Para um dado ser considerado preciso, resultados semelhantes devem ser obtidos em medições refeitas com os mesmos critérios. A precisão está vinculada à capacidade de um método produzir resultados consistentes e confiáveis (Granucci, 2023), quantas vezes for realizado.

A literatura científica distingue dois tipos de precisão:

- precisão sintática: refere-se à proximidade de um valor “V” aos elementos do domínio de definição correspondente “D”;
- precisão semântica: diz respeito à proximidade do valor “V” ao valor verdadeiro de “V”.

Segundo Batini e Scannapieco (2006), essa distinção é fundamental para compreender que a precisão não se limita apenas à conformidade com regras ou padrões formais (precisão sintática), mas também à relevância e acurácia dos dados em refletir a realidade (precisão semântica).

A precisão impacta diretamente na confiabilidade, na utilidade e na relevância dos dados. Garantir que os dados sejam precisos é crucial para que as decisões baseadas neles sejam eficazes, contribuindo para a transparência, a integridade das informações e o suporte adequado à tomada de decisão em diferentes contextos.

No contexto da *web*, a relevância de dados confiáveis e precisos torna-se ainda mais evidente, uma vez que a internet amplia o alcance, a velocidade de circulação e o impacto das informações. Dados incorretos e desatualizados ou imprecisos podem se propagar rapidamente, comprometendo análises, induzindo a erros decisórios e gerando desinformação em larga escala.

5.2.7 Recuperabilidade

Encontrar, acessar e utilizar os DGA de forma eficiente é um dos principais desafios encontrados pelos usuários de dados. A dimensão recuperabilidade se refere à facilidade com que esses podem ser localizados, acessados e obtidos pelos usuários. Nesse contexto, a recuperabilidade configura-se como uma dimensão essencial da construção dos conjuntos de dados para disponibilização.

Segundo Souto (2003, p. 73), entre os principais obstáculos na obtenção ou recuperação de dados e informações em bases de dados, está a filtragem das

informações. Nesse sentido, o uso de linguagens documentárias na organização dos conjuntos de dados por PIs é fundamental para possibilitar e facilitar a recuperação pelos usuários.

Para assegurar uma recuperabilidade satisfatória, é fundamental que os dados estejam organizados e classificados de maneira lógica, padronizados e consistentes, fazendo uso de metadados adequados que facilitem sua busca.

Para tanto, o gerenciamento de metadados é uma atividade imprescindível para a recuperação dos dados. A ENAP (2019, p. 5) assemelha a importância dos metadados para os dados com o catálogo de uma biblioteca. O catálogo além de fornecer informações necessárias acerca dos acervos existentes na biblioteca e como podem ser encontrados, também auxilia os usuários a recuperarem os materiais usando diferentes pontos de partida, como por palavras-chave. Sem um catálogo, seria difícil para os usuários recuperarem o material desejado. Da mesma forma, um conjunto de dados sem metadados bem definidos é como uma biblioteca sem um catálogo.

A recuperação de dados é definida como a “[...] operação pela qual se seleciona documentos, a partir do acervo, em função da demanda do usuário” (Ferneda, 2003, p. 14). O objetivo desta operação é garantir a “[...] recuperação de todos os objetos ou itens que satisfazem precisamente às condições formuladas através de uma expressão de busca” (Ferneda, 2003, p. 15). Para esse propósito, é necessário que o conjunto de dados disponha de mecanismos de busca eficientes, além de descrições claras e acessíveis de seu conteúdo, tal qual ocorrem a partir do uso de linguagens documentárias.

A recuperabilidade também é compreendida como um “[...] atributo que permite manter e preservar um determinado nível de operações e qualidade” (Maestre-Góngora; Rangel-Carrillo; Osorio-Sanabria, 2023, p. 10). A aplicação de ferramentas e técnicas adequadas de recuperabilidade permitem a recuperação eficiente dos conjuntos de dados, de acordo com os critérios da pesquisa (Herrera-Melo; González-Sanabria, 2020).

Além disso, Slibar, Oreski e Begicevic (2021) destacam que a recuperabilidade se refere ao sucesso na busca dos dados e dos conteúdos associados, como metadados e recursos complementares. Dessa forma, investir em estratégias que aprimorem a recuperabilidade dos dados não só melhoram a

experiência de acesso, como também asseguram que as informações estejam disponíveis e sejam de fácil utilização.

5.2.8 Inconsistência

A inconsistência é uma métrica fundamental na avaliação da QD, especialmente no que diz respeito ao conteúdo das informações. Essa dimensão manifesta-se nos elementos estruturais dos conjuntos de dados e pode ser definida como o “[...] grau de coerência lógica de determinada variável em função de outra variável complementar” (Moreira *et al.*, 2020, p. 120).

Frequentemente, os dados são disponibilizados em formatos variados e sem padronização adequada, o que dificulta o “[...] acesso e o tratamento em análises comparativas de dados de órgãos diferentes” (Silva; Monteiro; Reis, 2020, p. 42).

De acordo com Zhang e Xiao (2020, p. 216), a inconsistência é um atributo de qualidade de representação dos dados, caracterizado pelo fato de que “[...] os dados seguem um formato sintático consistente, sem contradição ou discrepância dentro de todo o catálogo de metadados”. Se os dados não possuem esses requisitos, se tornam inconsistentes.

Por outro lado, para Mahanti (2019), a dimensão consistência implica “[...] que os valores dos dados são idênticos para todas as instâncias de um aplicativo, e os dados devem estar sincronizados entre si”. Dessa maneira o formato e apresentação dos dados devem ser consistentes em todos o conjunto de dados relacionados a sua entidade.

Dados inconsistentes são aqueles que apresentam contradições. Abath *et al.* (2014, p. 133) definem consistência como “[...] o grau de coerência do dado”. Batini e Scannapieco (2006) ilustram essa ideia por meio de um exemplo claro de inconsistência:

- estado civil = casado;
- idade = 5 anos.

Essa discrepância entre o estado civil do indivíduo e sua idade revela a falta de coerência lógica entre os valores apresentados. Assim, ao avaliar a consistência, Mahanti (2019, p. 165), propõe os seguintes questionamentos:

I - diferentes fontes de dados ou múltiplos ambientes de dados fornecem informações contraditórias sobre o mesmo objeto de dados subjacente?

II - quais valores de dados fornecem informações contraditórias?

III - os dados são armazenados no mesmo formato em diferentes tabelas e sistemas de dados?

IV - os valores são uniformes em todos os conjuntos de dados e fontes de dados?

V - atributos mutuamente dependentes refletem adequadamente sua consistência estimada?

VI - existem informações diferentes sobre o mesmo objeto de dados subjacente em vários ambientes?

Segundo Wand e Wang (1996, p. 93), a consistência abrange diversos aspectos dos dados, em especial, seus valores e sua representação. Sebastian-Coleman (2013, p. 166), acrescenta que as medidas de consistência dos dados podem envolver “[...] conformidade interna ou podem revelar padrões lógicos criados dentro dos dados que refletem as conexões entre as situações do mundo real que os dados representam”.

Portanto a consistência pode ser avaliada por três perspectivas principais: 1) em relação a um padrão ou regra previamente estabelecida; 2) em relação a outros conjuntos de dados da mesma organização; e 3) em relação a outros conjuntos de dados obtidos de outras fontes.

Garantir a consistência é essencial para assegurar a confiabilidade dos dados e a integridade dos processos decisórios, contribuindo significativamente para a qualidade global dos sistemas de informação e para a eficácia das análises realizadas.

5.2.9 Compreensibilidade

A compreensibilidade dos dados refere-se à clareza e a facilidade em que os dados são apresentados, permitindo que qualquer usuário, independentemente de seu nível de conhecimento, consiga compreendê-los e utilizá-los. Em outras palavras, os dados devem ser inteligíveis, possibilitando ao usuário captar seu significado de forma precisa (Bovee; Srivastava; Mak, 2003).

As dimensões de qualidade, compreensibilidade e acessibilidade dos dados são interdependentes e se complementam. A compreensibilidade dos dados está diretamente relacionada com a facilidade do usuário de compreender os dados apresentados. A utilização de uma linguagem clara e a presença de ferramentas como dicionários de dados, são algumas características que possibilitam a compreensibilidade.

Por outro lado, a acessibilidade tem foco na facilidade em que diversos usuários conseguem acessar os dados, incluindo características como a gratuidade e o acesso e cumprimento de normas de acessibilidade digital.

Moreira *et al.* (2020, p. 125) definem a compreensibilidade como “[...] uma dimensão de qualidade que envolve os elementos estruturais dos conjuntos de dados, seus registros e recursos de visualização”. De acordo com Zaveri *et al.* (2015), esta qualidade está diretamente relacionada ao “[...] nível de compreensão e clareza que o dado é apresentado para o usuário dos dados”.

Para Moreira *et al.* (2020), a compreensibilidade envolve uma reunião de características que permite entender e assimilar os dados, atendendo a uma necessidade informacional específica.

Silva, Monteiro e Reis (2020, p. 43) reforçam que os “[...] conjuntos de dados disponibilizados devem ser facilmente entendidos e compreendidos por todos especialistas e acadêmicos até cidadãos mais leigos”. Para isso, a compreensibilidade demanda “[...] maior detalhamento e maior instrução sobre os dados para a promoção de acesso e transparência” (Silva; Monteiro; Reis, 2020, p. 43).

Em resumo, a compreensibilidade é a dimensão da QD que assegura que todos os usuários, independentemente de seu grau de instrução, entendam, compreendam e interpretem os dados disponibilizados.

5.2.10 Conformidade

A conformidade dos dados é uma dimensão substancial de sua qualidade pois assegura que os dados estejam disponibilizados de acordo com a legislação e os regulamentos vigentes, especialmente no que diz respeito à segurança e à privacidade do usuário. Quando os dados seguem essas diretrizes, tornam-se mais confiáveis para uso, reúso e acesso por parte dos usuários.

Segundo Mahanti (2019, p. 153) “[...] conformidade significa que os dados estão em conformidade com um conjunto de padrões ou diretrizes internas ou externas ou definições de dados padrão, incluindo definições de metadados”. De modo semelhante, Sebastian-Coleman (2013, p. 166) define a conformidade como “[...] o grau em que os dados estão em conformidade com um conjunto de regras de negócios, às vezes expressas como um padrão ou representadas dentro de um domínio de dados definidos”.

O principal objetivo da dimensão conformidade consiste em assegurar que as organizações observem rigorosamente as leis, regulamentações e políticas vigentes, de modo a garantir a proteção das informações sensíveis sob sua responsabilidade, como descrito na Lei n.º 13.709/2018, proteger os direitos fundamentais de liberdade e de privacidade e o livre desenvolvimento da personalidade da pessoa natural. De acordo com Vetrò *et al.* (2016), o conjunto de dados está em conformidade quando houver uma norma específica para o caso e esta for respeitada.

Zhang, Indulska e Sadiq (2019, p. 123) identificam dois aspectos fundamentais que caracterizam a conformidade:

- **conformidade de metadados:** os valores dos dados devem ser consistentes com a definição do atributo, de acordo com os metadados empregados;
- **conformidade com regras de negócios:** os valores dos dados estão em conformidade com regras de negócios específicas.

Além disso, Maestre-Gongora, Rangel-Carrillo e Osorio-Sanabria (2023, p. 10) afirmam que os dados estão em conformidade quando observam “[...] diretrizes e padrões atuais como DCAT (Data Catalog Vocabulary), que padroniza a descrição dos metadados”. Essa padronização facilita a integração e o acesso às informações, reforçando a confiança dos usuários e assegurando o cumprimento dos requisitos legais e operacionais.

A comparação entre itens de dados e metadados permite mensurar o grau de conformidade. Nesse processo, Mahanti (2019, p. 153) propõe alguns questionamentos na conformidade entre dados e metadados:

I - os valores de dados devem estar em conformidade com formatos específicos?

II - se sim, todos os valores precisam estar em conformidade com esses formatos?

III - os dados correspondem às regras ou sintaxe de sua definição, chamadas de metadados?

IV - quais dados não são armazenados em um formato padrão?

V - quais valores de dados estão fora do intervalo?

Além de observar aspectos técnicos, a conformidade também envolve o respeito às leis, regulamentações e políticas estabelecidas. Nesse sentido, dois aspectos devem ser considerados:

- **proteção de dados pessoais conforme legislação:** avalia se os dados atendem a normas legais ou regulatórias como a Lei nº 13.709/2018 - Lei Geral de Proteção de Dados Pessoais (LGPD) (Brasil, 2018);

- **responsabilidade em caso de violação de direitos ou uso inadequado de dados:** avalia a existência de um responsável (controlador e operador de dados) como descrito na LGPD.

Ao aderir a normas, padrões e diretrizes estabelecidas, os conjuntos de dados se tornam não apenas mais úteis e confiáveis, mas também em consonância com as exigências do ambiente legal e regulatório, contribuindo para uma GI mais eficiente e sustentável.

5.2.11 Licença Aberta/Não Proprietária

A ausência de licenças abertas pode restringir significativamente o acesso e a reutilização dos dados públicos. Para superar esse desafio, é fundamental que as organizações unam esforços para publicar seus dados sob licenças abertas e assegurem que esses dados não sejam de propriedade exclusiva de nenhuma entidade.

Segundo Dorobat e Posea (2021, p. 8) “[...] dados livres de licença significam que não podem estar sujeitos a qualquer regulamento de direitos autorais, patentes, marcas registradas ou segredos comerciais, exceto por restrições razoáveis de privacidade, segurança e privilégios”. Moradi, Mazoochi e Ahmadi (2022) reforçam

que um conjunto de dados livres de licença possibilita que estejam sempre abertos para reutilização.

Além disso, conforme salientam Royo-Montañés e Benítez-Gómez (2019, p. 3), os dados “[...] devem estar disponíveis em formatos sobre os quais nenhuma entidade tenha controle exclusivo”. Essa característica assegura que qualquer usuário ou máquina acesse, processe e reutilize os dados, sem depender de ferramentas ou licenças específicas.

A publicação de dados em formato proprietário pode resultar em obstáculos técnicos que exigirão do usuário esforço significativo antes de serem utilizados (Attard *et al.*, 2015, p. 17), como, por exemplo: a necessidade de adquirir *softwares* específicos pagos para abrir, editar ou converter o arquivo; dificuldade de interoperabilidade, entre outros. Portanto, a adoção de formatos não proprietários é indispensável para assegurar que os dados sejam abertos, interoperáveis e facilmente reutilizáveis por uma ampla gama de usuários e sistemas.

A padronização das licenças abertas específicas utilizadas na disponibilização de dados é igualmente essencial. Ela evita a incompatibilidade entre os diferentes conjuntos de dados. Embora determinadas licenças sejam classificadas como abertas, ainda podem conter restrições que impedem que dados com licenças diferentes sejam mesclados para uso específico (Attard *et al.*, 2015).

No contexto da qualidade dos dados abertos, tanto a adoção de uma licença aberta quanto o uso de formatos não proprietários são dimensões extremamente inevitáveis. Estas afetam diretamente a qualidade da disponibilidade, da acessibilidade, da recuperabilidade, da usabilidade, da não discricionariedade, da visualização e da interoperabilidade dos dados. Assim, garantir que os dados sejam publicados sob licenças compatíveis e sem restrições desnecessárias é fundamental para promover a transparência, o compartilhamento e a colaboração entre os diversos usuários e entidades.

5.2.12 Usabilidade

A usabilidade dos conjuntos de dados é uma dimensão indispensável da QD, pois garante que mesmo usuários com poucas habilidades técnicas ou conhecimentos específicos possam explorar, manipular e analisar os dados. Em

outras palavras, os dados devem ser facilmente manipuláveis e operacionais, permitindo que o público-alvo usufrua ao máximo seu conteúdo.

A usabilidade está diretamente relacionada à facilidade de navegação e interação do usuário com o ambiente digital, permitindo que alcance os objetivos pretendidos (Bevan, 1995). Nesse sentido, a usabilidade conecta-se diretamente com a acessibilidade “[...] que visa melhorar a interface gráfica com o usuário” (Rocha; Barbosa; Aganette, 2021, p. 202).

Portanto, enquanto a acessibilidade se preocupa com a facilidade de acesso aos dados pelos usuários com alguma deficiência, a compreensibilidade refere-se a clareza com que os dados são apresentados e compreendidos, a usabilidade está relacionada à praticidade de manipulação dos dados – é a funcionalidade do sistema.

Segundo Silva (2019, p. 76) “[...] dados que não são facilmente operacionalizáveis são tão inúteis quanto os dados não compreensíveis, uma vez que sem operacionalização, a informação fica presa atrás de um método ineficiente de obtenção de dados [...]”. Dessa forma, a dimensão usabilidade, além do aspecto do acesso aos dados, envolve a facilidade com que os usuários podem compreender, processar e fazer uso dos dados.

Usabilidade diz respeito à habilidade do *software* em permitir que o usuário alcance facilmente suas metas de interação com o sistema. Desta forma, problemas de usabilidade estão relacionados com o diálogo da interface.

A usabilidade, assim, garante a própria continuidade e afirmação competitiva de um *site*, de um *software* ou de um sistema de informação na perspectiva da interação com o usuário. É pela interação com o usuário, a partir do seu desempenho e da sua satisfação, que se evidencia a sobrevivência de um sistema de informação.

A usabilidade é um fator crucial para maximizar o valor dos dados abertos, pois facilita o acesso e a aplicação das informações por um público diversificado, contribuindo para a democratização, a transparência e a eficiência na gestão dos dados.

5.2.13 Não Discrecionário

A dimensão de qualidade não discricionária refere-se à imparcialidade no acesso ao uso de dados. Essa dimensão permite que todos os usuários tenham acesso irrestrito aos conjuntos de dados, sem que haja distinção ou privilégios baseados no perfil do usuário. Portanto, todos os usuários possuem igualdade para acessar e utilizar os dados, sem favorecimentos, exigências ou barreiras.

Segundo Dorobat e Posea (2021, p. 8), dados não discricionários significam que “[...] devem estar disponíveis a todos os interessados, e o acesso aos mesmos não deve envolver qualquer processo de autenticação”. Essa característica assegura que o acesso e a reutilização dos dados sejam equivalentes para todas as pessoas físicas e jurídicas, independentemente da origem, motivação ou perfil técnico (Moradi; Mazoochi; Ahmadi, 2022).

Nesse sentido, os conjuntos de dados classificados como não discricionários devem ser acessíveis sem necessidade de cadastro, *login*, códigos de acesso, senhas ou licenças (Royo-Montañés; Benítez-Gómez, 2019, p. 3). Essa prática contribui significativamente para a construção de um ambiente de dados mais transparente, acessível e democrático, essencial para a inovação e a tomada de decisões fundamentadas.

5.2.14 Primário

A primariedade dos dados é uma dimensão que impacta diretamente na confiabilidade dos dados obtidos pelo usuário. Essa dimensão se refere à origem dos dados. Para alcançar o nível máximo de detalhamento e precisão, os conjuntos de dados “[...] devem ser publicados sem processamento ou acréscimos”. (Royo-Montañés; Benítez-Gómez, 2019, p. 3). Dito de outra forma, trata-se dos dados em estado bruto e não manipulados.

Dados primários são aqueles dispostos conforme constam na fonte original – o dado genuíno. Esses dados não estão sujeitos a processos de agregação ou modificação, pois foram coletados ou produzidos pela organização, antes de qualquer processamento, agregação ou interpretação (Attard *et al.*, 2015; Dorobat; Posea, 2021; Fan; Zhao, 2017). Essa abordagem garante que os usuários tenham acesso aos dados na sua forma mais autêntica e detalhada, fortalecendo a confiança e a integridade dos dados utilizados para tomada de decisão e análises.

5.2.15 Interoperabilidade

A interoperabilidade é uma dimensão que permite a troca eficiente de informações entre diferentes sistemas, aplicações e usuários, sem barreiras técnicas ou semânticas. Essa capacidade de integração é essencial para que diferentes sistemas, organizações e usuários possam interagir de forma eficaz, garantindo a comunicação e a colaboração necessárias para a construção de um ecossistema de dados robusto.

Conforme definido por Brasil (2020a), a interoperabilidade é uma qualidade que se refere à “[...] capacidade de diversos sistemas e organizações trabalharem em conjunto (interoperar) de modo a garantir que pessoas, organizações e sistemas computacionais interagem para trocar informações de maneira eficaz e eficiente”. Para isso, é necessário definir conjuntos mínimos de premissas, políticas e especificações técnicas para os conjuntos de dados. Essas capacidades dos dados são: (I) - um conjunto de dados precisa ser interpretado por diferentes sistemas sem perda de significado; (II) - ser combinado com outros conjuntos de dados; e (III) - ser reutilizado por diferentes ferramentas sem a necessidade de conversões complexas (Brasil, 2020a).

Jetzek (2016) destaca que a clareza na semântica e na sintaxe dos dados é fundamental, e que os dados devem ser apresentados utilizando modelos abertos, metadados padronizados e identificadores padrão. Complementando essa visão, Cámara *et al.* (2021, p. 150) enfatizam que na interoperabilidade:

[...] verifica-se a conformidade dos metadados com a especificação DCAT-AP²⁴, a disponibilidade de informações sobre formatos de distribuição, a utilização de formatos bem conhecidos, a utilização de formatos não proprietários e a utilização de formatos processáveis.

Segundo a ENAP (2019, p. 11), os metadados técnicos são essenciais para a interoperabilidade dos dados. Para isso, a ISO/IEC, editou normas com o objetivo permitir a troca de dados entre “[...] ambientes heterogêneos, com base e definições exatas de dados” (ISO/IEC 11179, 2023).

Entretanto, a disponibilização de conjuntos de dados pelas organizações governamentais, por exemplo, ocorre de forma não padronizada, tornando-os

²⁴ *Application Profile for Data Portals in Europe (DCAT-AP).*

heterogêneos, levando a problemas como a falta de interoperabilidade (Attard *et al.*, 2015). Superar esses desafios é fundamental para promover a integração e o compartilhamento de dados, ampliando o potencial de uso e a eficácia dos dados.

É importante ressaltar que as dimensões de QD estão conectadas entre si, existindo dependência e correlação entre elas. Ao considerar uma dimensão mais significativa que outra, pode-se favorecer negativamente as demais (Mahanti, 2019). Como exemplo pode-se citar a dimensão “segurança” que, ao apresentar maior relevância, acabou por comprometer a acessibilidade. Isso ocorre porque recursos de segurança mais complexos, como permissões de acesso, mecanismo de autorização, uso de *login* e senhas de acesso, tendem a dificultar o acesso aos dados.

Como apresentado, cada dimensão contribui para um aspecto específico da QD. Segundo Mahanti (2019, p. 675), “[...] as dimensões de qualidade de dados são os elementos fundamentais de mensuração no domínio da QD e, portanto, cruciais para a gestão da qualidade de dados”. Assim, o termo dimensão é utilizado para identificar os aspectos dos dados que podem ser avaliados e por meio dos quais sua qualidade pode ser descrita e avaliada (Sebastian-Coleman, 2013).

Na sociedade contemporânea, a necessidade das instituições não reside apenas na demanda por dados, mas de quanto e para quais aplicações de uso a organização carece. Portanto, a coleta de dados demasiada não é necessária, mas sim que sejam confiáveis.

Segundo Mahanti (2019, p. 630), “[...] os dados só têm valor quando são usados e são usados apenas quando estão aptos para uso”. Nesse sentido, a aplicação das dimensões de qualidade na construção de um conjunto de dados contribui para seu valor de uso, ao garantir que os dados estejam adequados para atender às demandas de todos os tipos de usuário.

6 DEFINIÇÃO DO INSTRUMENTO PARA MENSURAÇÃO DA QUALIDADE DOS DADOS GOVERNAMENTAIS ABERTOS (DGA)

As dimensões de qualidade representam aspectos conceituais que orientam a mensuração da QD. Cada uma das dimensões de qualidade pode ser associada a uma ou mais métricas, de acordo com a natureza da característica a ser avaliada (Batini *et al.*, 2009). A avaliação sistemática da QD requer considerar múltiplos componentes da estrutura do sistema de dados abertos, incluindo: mecanismos de acesso disponibilizados aos usuários aos portais governamentais, os elementos que compõem o conteúdo informacional presente nos portais, e a estrutura dos conjuntos de dados (Moreira *et al.*, 2020).

As métricas de qualidade, também denominadas como atributos ou critérios “[...] são características particulares dos dados que podem ser tanto objetiva (Ex.: acurácia do dado), quanto subjetiva (Ex.: reputação da fonte de dados)” (Penteado, 2020, p. 87). No mesmo sentido, Mahanti (2019, p. 2017) classifica as métricas avaliando-as objetivamente: aquelas que podem ser quantificadas ou para qual um valor concreto pode ser calculado; e, subjetivamente: aquelas que não podem ser quantificadas, pois dependem da perspectiva do usuário.

Assim, foi construído um instrumento de avaliação da qualidade dos DGA com base nas dimensões e métricas identificadas ao longo da pesquisa. Esse instrumento foi elaborado de forma a integrar fundamentos teóricos provenientes da RSL, normas e legislações voltadas à qualidade da informação pública. A estrutura do instrumento contemplou um conjunto de 15 dimensões de qualidade associadas a métricas específicas, cada uma com critérios e parâmetros definidos para a mensuração objetiva da QD.

Posteriormente, o instrumento aplicado a um conjunto de dados governamental real, de modo a permitir a verificação da sua aplicabilidade e a análise dos resultados. O intuito foi validar sua capacidade de identificar pontos fortes e fracos na gestão e disponibilização dos DGA, contribuindo para o aprimoramento da transparência e do acesso à informação.

6.1 Dimensão: acessibilidade

A dimensão acessibilidade refere-se à garantia de inclusão das pessoas portadoras de deficiência. A falta de acessibilidade, tanto no mundo físico quanto no mundo digital, impacta principalmente pessoas com deficiências permanentes, temporárias ou aquelas que dependem da acessibilidade por qualquer outra condição (Brasil, 2023, p. 6).

A métrica associada à essa dimensão é a **Conformidade com normas de acessibilidade**, a qual avalia se os dados e as interfaces utilizadas para seu acesso estão em conformidade com as normas, diretrizes e legislações que asseguram a acessibilidade para todas as pessoas, inclusive aquelas com alguma deficiência.

A análise considera referências legais e técnicas como:

- Lei n.º 10.098/2000 (Lei da Acessibilidade): estabelece normas gerais e critérios básicos para a promoção da acessibilidade das pessoas portadoras de deficiência ou com mobilidade reduzida;
- Lei n.º 14.129/2021 (Lei do Governo Digital): dispõe sobre princípios, regras e instrumentos para o Governo Digital e para o aumento da eficiência pública;
- Norma Técnica da Associação Brasileira de Normas Técnicas (ABNT), ABNT NBR 17225:2025 (Acessibilidade em conteúdos e aplicações *web* - requisitos): fornece diretrizes de acessibilidade para elaboração de conteúdo *web*;
- Cartilha de Acessibilidade Digital do Governo Federal (2023): fornece diretrizes e orientações para a construção e a adequação de conteúdos digitais, seguindo as melhores práticas de acessibilidade digital e o Modelo de Acessibilidade em Governo Eletrônico.

Dessa forma, a avaliação da acessibilidade busca identificar se o Portal de Dados Abertos garante um acesso efetivamente inclusivo. A análise contempla aspectos técnicos de usabilidade, verificando a adoção de recursos que assegurem a conformidade com as normas vigentes. Os critérios e indicadores definidos para essa avaliação estão organizados no Quadro 19.

Quadro 19 – Recursos de acessibilidade

Categoria/ Recurso	Descrição	Avaliação		
		Excelente	Aceitável	Crítico
Texto alternativo	Todas as imagens, gráficos e ícones possuem descrição textual significativa.	Todas as imagens possuem texto alternativo adequado.	Algumas imagens possuem texto alternativo.	Imagens sem texto alternativo ou inadequado.
Navegação por teclado	O <i>site</i> pode ser totalmente navegável sem <i>mouse pad</i> .	Navegação totalmente funcional por teclado.	Navegação funcional, mas com dificuldades em algumas ações.	Navegação impossível ou muito limitada sem <i>mouse pad</i> .
Contraste de cores	Texto e elementos gráficos têm contraste adequado para leitura.	Contraste atende aos padrões (WCAG ²⁵ AA/AAA). ²⁶	Contraste marginal, leitura possível, mas desconfortável.	Contraste insuficiente, leitura dificultada.
Legendas e Língua Brasileira de Sinais (Libras)	Vídeos e conteúdos multimídia possuem legendas, e os conteúdos possuem interpretação em Libras.	Todos os vídeos com legendas e conteúdos com interpretação em Libras.	Alguns vídeos com legendas e poucos conteúdos com interpretação de Libras.	Nenhum vídeo com legenda ou conteúdo com interpretação com Libras.
Formulários acessíveis	Rótulos claros, instruções, mensagens de erro compreensíveis.	Todos os formulários totalmente acessíveis.	Alguns formulários apresentam pequenas falhas.	Formulários inacessíveis ou confusos.

²⁵ *Web Content Accessibility Guidelines.*

²⁶ *Authentication, Authorization, and Accounting.* Em cada versão das WCAG, o Nível A representa a base para acessibilidade na *web*, com cada nível subsequente incluindo e se baseando no nível anterior. Isso significa que a conformidade com o Nível AA das WCAG inclui a conformidade com o Nível A, e a conformidade com o Nível AAA inclui a conformidade com os Níveis A e AA (Level Access, 2024, tradução nossa).

Estrutura semântica/ Títulos	Uso correto de títulos, listas, menus e hierarquia de conteúdos.	Estrutura ideal e consistente em todo o <i>site</i> .	Estrutura parcialmente correta.	Estrutura confusa ou ausente.
Conteúdo multimídia seguro	Evita conteúdos piscantes ou que causem desconforto.	Não há conteúdo que cause risco de epilepsia ou mal-estar.	Pequenos <i>flashes</i> ou animações ocasionais.	Conteúdos piscantes ou perigosos presentes.
Linguagem e clareza	Texto claro e compreensível para pessoas com deficiência intelectual.	Linguagem clara, direta e objetiva.	Algumas partes com linguagem complexa.	Texto difícil de compreensão confuso ou muito técnico.

Fonte: Elaborado pela autora (2026).

O objetivo é garantir que o acesso aos dados seja pleno e inclusivo. Essa avaliação pode ser utilizada para coletar a percepção dos usuários sobre os diferentes aspectos da acessibilidade. Para interpretar a métrica, serão atribuídas pontuações para o cálculo do índice de acessibilidade, em que cada critério possui um valor correspondente:

- excelente: 3 pontos;
- aceitável: 2 pontos;
- crítico: 1 ponto.

Após a atribuição de pontos para cada recurso avaliado do Quadro 19, será calculada a pontuação obtida. Para o cálculo da qualidade acessibilidade a fórmula da Tabela 3 será aplicada.

Tabela 3 – Cálculo e fórmula da qualidade acessibilidade

Índice de qualidade (%) =	<u>Pontuação obtida x 100</u> Pontuação máxima

Ex.: suponha-se que, dos oito recursos avaliados, sejam obtidos 20 pontos. Tem-se:

Fórmula: $\frac{20}{24} \times 100 = 83\%$

Fonte: Elaborado pela autora (2026).

Para a classificação final da qualidade dos recursos de acessibilidade, serão aplicados os critérios de referências do Quadro 20.

Quadro 20 – Critérios de referência para avaliação da qualidade acessibilidade

Nível de qualidade	Cobertura	Descrição
Excelente	85 a 100%	Os recursos atendem aos critérios de acessibilidade.
Aceitável	75 a 84%	O recurso apresenta implementação parcial dos critérios de acessibilidade.
Crítico	Abaixo de 74	O recurso não atende aos critérios mínimos de acessibilidade.

Fonte: Elaborado pela autora (2026).

Ao aplicar os critérios de referência da qualidade acessibilidade, será possível mensurar se o conjunto de dados está em conformidade com as normas legais e diretrizes internacionais. Desse modo, a qualidade acessibilidade pode permitir que o usuário com alguma deficiência utilize os dados sem barreiras significativas.

6.2 Dimensão: completude

A dimensão completude assegura que informações essenciais estejam presentes e sejam suficientes para o objetivo a que se destinam. Em outras palavras, avalia-se o grau em que os dados estão íntegros, sem lacunas que possam comprometer sua interpretação, análise ou reutilização.

Nesta seção, são exploradas métricas de completude aplicadas na PoC, na qual detalha-se os critérios utilizados para mensuração de cada uma delas.

A partir dos resultados obtidos, sugere-se as seguintes métricas para a avaliação dessa dimensão: **valores ausentes; cobertura de dados coletados; campos obrigatórios; comparabilidade; metadados e distribuição temporal.**

1) Valores ausentes: esta métrica refere-se à proporção de células nulas ou vazias em um conjunto de dados, sendo fundamental para identificar lacunas que

podem comprometer a precisão e a qualidade da análise. As células de uma planilha são unidades básicas em que se pode inserir e armazenar dados (ENAP, 2022).

Nesse caso, a fórmula da Tabela 4 é utilizada para calcular a proporção de células vazias.

Tabela 4 – Cálculo e fórmula da qualidade valores ausentes

Proporção de células vazias =	<u>N.º de células nulas ou vazias x 100</u>
	<u>N.º total de células no conjunto de dados</u>

Ex.: a planilha ilustrada na Figura 10 possui um total de 81 células, das quais 26 são nulas ou vazias. Então:

$$\text{Fórmula: } \frac{26}{81} = 0.3209 \times 100 = 32\%$$

Fonte: Elaborado pela autora (2026).

Figura 11 – Valores ausentes

2,4	7,1	7,5	10,7	8,9	12,2	14,9	12,5	19,0
3,1	5,1	19,3	43,1	63,8	48,6	67,7	76,0	78,7
--	--	--	--	--	--	--	--	--
--	--	--	--	--	--	--	--	--
2,9	3,4	13,3	33,6	35,0	22,9	28,7	27,4	30,5
2,7	0,0	4,0	4,3	4,1	12,9	15,4	12,9	19,3
3,1	5,1	19,3	43,1	63,8	48,6	67,7	76,0	78,7
2,5	9,0	8,5	13,3	11,2	--	--	--	--
2,5	9,0	8,5	13,3	11,2	--	--	--	--

Fonte: Elaborado pela autora (2026).

O resultado de 32% indica uma proporção significativa de células vazias no conjunto de dados, o que pode prejudicar sua análise. Para interpretar essa métrica, serão utilizados os critérios de referência apresentados no Quadro 21.

Quadro 21 – Critérios de referência para avaliação dos valores ausentes

Nível de qualidade	Cobertura	Descrição
Excelente	0-5%	Dados praticamente completos.
Aceitável	6-20%	Apesar de aceitável, compromete a confiabilidade dos dados.

Crítico	> 20%	O conjunto de dados possui alto nível de células vazias causando riscos à análise ou tomada de decisão do usuário.

Fonte: Elaborado pela autora (2026).

A classificação dos percentuais de nível de qualidade aplicados na avaliação dos **valores ausentes** fundamenta-se no princípio de que, quanto menor o nível de células ausentes presentes no conjunto de dados, melhor é sua avaliação em termos de qualidade.

Ao aplicar os critérios de referência na avaliação da métrica **valores ausentes**, é possível garantir que os dados disponibilizados sejam capazes de fornecer informações mais abrangentes e confiáveis.

2) Cobertura de dados coletados: refere-se à proporção de valores preenchidos em relação ao total de dados, servindo como um indicador do nível de completude do conjunto de dados. Calcula-se a quantidade de dados efetivamente preenchidos em comparação com o total.

A fórmula utilizada para o cálculo está apresentada na Tabela 5.

Tabela 5 – Cálculo e fórmula da qualidade cobertura de dados coletados

Cobertura de dados coletados =	<u>N.º de células preenchidas x 100</u> N.º total de células esperadas

Ex.: considerando a planilha mencionada anteriormente, em que o conjunto de dados apresenta o total de 81 células, das quais 55 estão preenchidas e 26 vazias, tem-se:

$$\text{Fórmula: } \frac{55}{81} = 0.680 \times 100 = 68\%$$

Fonte: Elaborado pela autora (2026).

O resultado demonstra que 68% dos dados foram efetivamente coletados, o que fornece uma medida objetiva sobre a cobertura dos dados coletados.

Para interpretar essa métrica, serão utilizados os critérios de referência apresentados no Quadro 22.

Quadro 22 – Critérios de referência da cobertura dos dados coletados

Nível de qualidade	Cobertura	Descrição
Excelente	95-100%	Dados praticamente completos.
Aceitável	80-94%	Apesar de aceitável, compromete a confiabilidade dos dados.
Crítico	> 80%	O conjunto de dados possui alto nível de células vazias causando riscos à análise ou decisão do usuário.

Fonte: Elaborado pela autora (2026).

A aplicação da métrica **Cobertura dos dados coletados** permite avaliar de claramente o grau de completude dos dados, permitindo identificar lacunas que possam comprometer análises ou a confiabilidade das informações. Ao identificar níveis críticos de cobertura, é possível a adoção de ações corretivas.

3) Campos obrigatórios: esta métrica refere-se à verificação da presença de dados em campos considerados essenciais para análise, interpretação ou conformidade de um conjunto de dados. Esses campos exigem informações específicas e indispensáveis para a análise ou para a conformidade do conjunto de dados.

Exemplos comuns de campos obrigatórios:

- CPF;
- data: necessário para ordenação cronológica ou validade dos registros;
- valor monetário: fundamental em análises orçamentárias e fiscais.

Informações como essas são fundamentais para garantir a integridade e a completude dos dados coletados.

Para a avaliação dessa métrica, alguns pontos devem ser analisados, tais como:

- identificar os campos obrigatórios de acordo com o objetivo da base de dados;
- verificar a taxa de preenchimento de cada campo essencial;
- calcular a proporção de registros completos nos campos obrigatórios;
- classificar o nível de conformidade (excelente, aceitável ou crítico).

A fórmula utilizada para o cálculo da métrica de campos obrigatórios será apresentada na Tabela 6:

Tabela 6 – Cálculo e fórmula da qualidade campos obrigatórios

Campos obrigatórios =	<u>N.º de registros com campo</u> <u>preenchido x 100</u> <u>N.º total de registros</u>

Ex.: uma planilha com 20 campos de preenchimento obrigatório, apresenta um total de 15 campos preenchidos, assim:

$$\text{Fórmula: } \frac{15}{20} = 0.750 \times 100 = 75\%$$

Fonte: Elaborado pela autora (2026).

O resultado da fórmula aplicada demonstra que 75% dos **campos obrigatórios** estão preenchidos. Para interpretar essa métrica, serão utilizados os critérios de referência apresentados no Quadro 23.

Quadro 23 – Critérios de referência dos campos obrigatórios

Nível de qualidade	Cobertura	Descrição
Excelente	95-100%	Dados praticamente completos.
Aceitável	80-94%	Apesar de aceitável, compromete a confiabilidade dos dados sendo necessário o preenchimento dos campos.
Crítico	<80%	O conjunto de dados precisa de correção urgente.

Fonte: Elaborado pela autora (2026).

A verificação do preenchimento dos **campos obrigatórios** garante que as informações-chave estejam disponíveis, possibilitando análises e decisões precisas e bem fundamentadas. A ausência do preenchimento dos campos obrigatórios pode comprometer significativamente o conjunto de dados disponibilizados.

4) Comparabilidade: é avaliado se os conjuntos de dados podem ser confrontados entre diferentes períodos de tempo, entidades, regiões ou categorias,

permitindo que os usuários identifiquem semelhanças, diferenças, tendências e padrões.

Para a análise da métrica comparabilidade alguns aspectos devem ser avaliados:

- **estrutura dos conjuntos de dados:** os conjuntos de dados devem ter os mesmos campos (colunas) com significados consistentes. Ex.: uma coluna “valor total” deve representar o mesmo tipo de valor em diferentes bases de dados;
- **unidade de medida padronizada:** os dados devem estar na mesma escala e unidade (como em reais, dólares, porcentagem, entre outros);
- **períodos padronizados:** as datas ou períodos das coletas devem ser equivalentes (mensal, anual e outras). Ex.: comparar um dado trimestral com um dado anual exige que estejam padronizados;
- **classificações consistentes:** as categorias ou códigos utilizados devem seguir a mesma nomenclatura. Ex.: um município identificado como “401236” em um conjunto de dados, e como “1203” em outro, pode gerar conflito.

A avaliação da métrica comparabilidade seguirá os critérios apresentados no Quadro 24.

Quadro 24 – Critérios de avaliação da comparabilidade

Nível de qualidade	Descrição
Excelente (Um ponto)	Os conjuntos de dados apresentam consistência em estrutura, unidades, períodos e classificações, permitindo comparações diretas, sem necessidade de ajustes.
Aceitável (Dois pontos)	Os conjuntos de dados apresentam compatibilidade parcial, alguns elementos permitem comparações, mas são necessários ajustes.
Crítico (Três pontos)	Os conjuntos de dados não apresentam condições de comparabilidade, há divergências na estrutura, unidades, períodos ou classificações dos dados.

Fonte: Elaborado pela autora (2026).

A **comparabilidade** é fundamental para que dimensões como a interoperabilidade entre conjunto de dados seja possível. A comparabilidade garante a padronização na estrutura, nas unidades, nos períodos ou nas classificações garantindo que diferentes conjuntos de dados dialoguem entre si.

5) Metadados: refere-se à presença de informações estruturadas que descrevem, explicam e contextualizam um conjunto de dados permitindo sua compreensão, uso correto e rastreabilidade.

Quando completos, os metadados auxiliam os usuários a entenderem:

- o que o dado representa;
- de onde se origina;
- com que frequência é utilizado;
- como pode ser utilizado legalmente.

Metadados completos incluem elementos essenciais para garantir a clareza, a rastreabilidade e a utilização adequada dos dados em análises, relatórios e tomadas de decisão. Metadados completos devem incluir:

- título: nome claro e preciso do conjunto de dados;
- descrição: resumo detalhado sobre o conteúdo, finalidade e abrangência dos dados;
- fonte: instituição ou entidade responsável pela coleta e disponibilização dos dados;
- periodicidade: frequência com que os dados são atualizados.

Assegurar a presença de metadados padronizados e bem estruturados contribui para a qualidade dos DGA.

Para avaliar a métrica “metadados” serão utilizados os critérios apresentados no Quadro 25.

Quadro 25 – Critérios de avaliação de metadados

Nível de qualidade	Descrição
Excelente (Um ponto)	Os metadados estão completos, padronizados e acessíveis, contemplando todos os elementos essenciais.
Aceitável (Dois pontos)	Os metadados estão presentes, mas de forma incompleta, ou sem padronização.
Crítico (Três pontos)	Ausência de metadados, ou presença de informações superficiais que não garantem a compreensão e o uso adequado dos dados.

Fonte: Elaborado pela autora (2026).

Os **metadados** garantem clareza, rastreabilidade e contextualização, permitindo que públicos diversos utilizem a informação de forma correta. Além disso, a presença de metadados completos e padronizados aumenta a interoperabilidade dos conjuntos de dados.

6) Distribuição temporal: esta métrica refere-se à análise da cobertura dos dados em relação ao tempo.

Aspectos a serem verificados:

- cobertura cronológica: os dados contemplam períodos recentes e relevantes, como os últimos cinco anos;
- periodicidade: a atualização dos dados ocorre de forma regular e consistente, permitindo comparações ao longo do tempo;
- lacunas temporais: existem falhas ou intervalos sem registro de dados que podem comprometer análises longitudinais ou séries históricas.

A análise de qualidade da métrica **distribuição temporal** seguirá os critérios apresentados no Quadro 26.

Quadro 26 – Critérios de avaliação da distribuição temporal

Nível de qualidade	Descrição
Excelente (Um ponto)	A cobertura inclui o período dos últimos cinco anos ou mais. A periodicidade é regular e consistente e não existe lacuna temporal.
Aceitável (Dois pontos)	A cobertura contempla somente uma parte dos períodos recentes (Ex.: três anos). A periodicidade apresenta algumas inconsistências e as lacunas temporais existentes não comprometem totalmente a análise.
Crítico (Três pontos)	A cobertura temporal é desatualizada, sem incluir períodos recentes. A periodicidade é irregular ou inexistente e existem lacunas temporais consideráveis que inviabilizam análises consistentes.

Fonte: Elaborado pela autora (2026).

Analisar a **distribuição temporal** é essencial para avaliar a representatividade e a abrangência dos dados disponibilizados. Conjuntos de dados com uma longa cobertura temporal permitem análises mais completas, tendências históricas e padrões que influenciam a formulação de políticas públicas compatíveis com cada região.

6.3 Dimensão: atualidade

A dimensão atualidade está relacionada à capacidade dos dados refletirem de forma fidedigna os eventos ou fenômenos mais recentes, o que garante sua relevância e utilidade. Dados atualizados permitem análises mais precisas, decisões oportunas e o acompanhamento contínuo de políticas públicas ou fenômenos sociais.

Para avaliar essa dimensão, foram definidas métricas específicas que analisam tanto o tempo decorrido entre o evento e sua disponibilização, quanto à frequência de atualização e a indicação clara da data da coleta ou última modificação. A aplicação dessa métrica permite verificar se os conjuntos de dados estão alinhados ao princípio da atualidade e se são continuamente revisados ou disponibilizados de forma tempestiva.

A seguir, são apresentadas as métricas que compõem essa dimensão, acompanhadas dos critérios utilizados para sua mensuração: **idade dos dados; frequência de atualização; tempo entre a ocorrência do evento e o do registro no sistema e presença de data e hora da coleta dos dados e atualização dos metadados.**

1) Idade dos dados: essa métrica avalia o tempo decorrido entre a coleta e a última atualização dos dados. Além de auxiliar a determinar o quão recentes ou desatualizados estão os dados disponibilizados.

Para se chegar à idade atual dos dados é necessário que alguns pontos sejam identificados, como:

- a data da última atualização dos dados disponibilizados;
- a data atual.

A fórmula utilizada para o cálculo da métrica de idade dos dados será a demonstrada na Tabela 7.

Tabela 7 – Fórmula da qualidade idade dos dados

Idade dos dados =	<u>Data da última atualização x 100</u>
	Data atual

Fonte: Elaborado pela autora (2026).

Exemplo: para um conjunto com periodicidade anual e o intervalo for menor que 365 dias se conclui que o conjunto de dados se encontra dentro do limite. O mesmo critério aplica-se às demais periodicidades: diária (limite de 24 horas), mensal (30 dias), bimestral (60 dias), entre outras.

Com base nessa diferença temporal, os dados serão classificados conforme os critérios dos Quadro 27.

Quadro 27 – Frequência da atualização dos dados abertos

Nível de qualidade	Descrição	Enquadramento
Excelente	Os dados são atualizados no período de tempo proposto.	O período de atualização indicado é diário e qualquer atualização que ocorra em 24 horas ou menos, se enquadra nesta categoria. Essa ação se qualifica nos casos de atualizações diárias, mensais, bimestrais, anuais e outras.
Aceitável	Apesar de atualizados, os dados merecem atenção.	O período de tolerância para as atualizações é de 10%. Assim, se o período indicado de atualização é de 24h e ocorreu após e entre as 24h e 26h, se enquadra na categoria aceitável.
Crítico	Os dados estão muito desatualizados.	Acima de 11% de tolerância a situação de atualização do dado é considerada crítica. Neste caso, se o período indicado para atualização é de 24h e ocorreu após decorridas 26h, a situação é crítica. Por meio das informações fornecidas pelo conjunto de dados não é possível identificar a idade dos dados.

Fonte: Elaborado pela autora (2026).

A avaliação da **atualização dos dados abertos** pode ser adaptada conforme sua frequência (dia, mês ou ano).

2) Frequência de atualização: avalia o intervalo regular entre as atualizações de um conjunto de dados, comparando a periodicidade prevista (ou declarada) com a frequência real observada nas publicações. Essa métrica verifica se os dados estão sendo atualizados conforme o planejado.

Para a análise da frequência de atualização, alguns aspectos devem ser observados:

- a frequência de atualização prevista (mensal, trimestral, anual e outras);
- as datas reais de atualização disponíveis no histórico do conjunto de dados.

Para avaliar a qualidade da frequência de atualização dos dados, os critérios a serem utilizados estão apresentados no Quadro 28.

Quadro 28 – Critérios de avaliação da frequência de atualização dos dados abertos

Nível de qualidade	Descrição
Excelente (Um ponto)	As atualizações ocorrem no prazo previsto e de forma consistente. O intervalo entre as datas reais de atualização corresponde ao ciclo declarado (diário, mensal, semestral ou anual).
Aceitável (Dois pontos)	As atualizações ocorrem, mas com pequenos atrasos em relação a periodicidade prevista, contudo, permite análises confiáveis.
Crítico (Três pontos)	As atualizações não seguem a periodicidade prevista. Os atrasos são recorrentes e os intervalos sem atualizações são longos, comprometendo a análise. O conjunto de dados não fornece informações confiáveis para avaliação.

Fonte: Elaborado pela autora (2026).

A métrica da **frequência de atualização** dos dados abertos garante que os dados estejam sempre atualizados no período esperado.

3) Tempo entre a ocorrência do evento e o do registro no sistema: essa métrica mede o intervalo de tempo entre o momento em que um evento ou fato ocorre, e o momento em que é registrado ou disponibilizado no sistema.

Para calcular a métrica é necessário:

- identificar a data/hora do evento (momento em que o fato ocorreu);
- identificar a data/hora do registro no sistema (momento da inserção ou disponibilização).

Após essas identificações, calcula-se o tempo conforme o Quadro 29.

Quadro 29 – Cálculo da qualidade do tempo de registro

Tempo de registro =	Data do registro	Data do evento

Fonte: Elaborado pela autora (2026).

Com base nesse intervalo, o desempenho será classificado conforme demonstrado no Quadro 30.

Quadro 30 – Tempo entre a ocorrência do evento e do registro no sistema

Tempo de registro	Nível de qualidade	Descrição
No dia da ocorrência	Excelente	Atualização praticamente em tempo real.
Um a cinco dias	Aceitável	A atualização dos dados pode estar comprometida.
Acima de cinco dias	Crítico	Atualização dos dados está com uma defasagem considerável.

Fonte: Elaborado pela autora (2026).

Portanto, se o fato ocorreu, por exemplo, no dia 10 de maio de 2025, e somente foi registrado no sistema em 19/05/2025, o tempo de registro foi de nove dias. Nesse caso, sua avaliação de tempo de ocorrência é ruim, pois, os dados já se encontram defasados para alguns usuários.

4) Presença de data e horário da coleta dos dados: a métrica analisa se a entrada de dados possui campos com as informações sobre data e hora da coleta, permitindo assim, identificar o momento da ocorrência do evento e avaliar a atualidade dos dados. Os critérios estabelecidos para essa métrica podem ser visualizados no Quadro 31.

Quadro 31 – Critérios para avaliação da atualidade dos dados abertos

Nível de qualidade	Descrição
Excelente (Um ponto)	O conjunto de dados apresenta informações sobre a data e o horário da coleta dos dados.
Aceitável (Dois pontos)	O conjunto de dados apresenta somente a data da coleta dos dados.

Crítico (Três pontos)	O conjunto de dados não apresenta nenhuma informação sobre a coleta de dados.

Fonte: Elaborado pela autora (2026).

A manutenção e a disponibilização de dados constantemente atualizados são indispensáveis. Dados recentes transmitem aos usuários maior credibilidade e reforça a confiança na qualidade dos dados governamentais publicados.

5) Atualização dos metadados: é verificado se os metadados associados ao conjunto de dados são atualizados de forma consistente juntamente com os dados principais. Isso inclui campos como data de atualização, identificação da fonte, contato da fonte responsável e versão dos dados. Para cada item presente, atribui-se a pontuação um. A soma da pontuação será a nota final. Para avaliar essa métrica os seguintes critérios serão utilizados:

Quadro 32 – Critérios para avaliação da atualização de metadados

Nível de qualidade	Descrição	Enquadramento
Excelente (Quatro pontos)	Todos os metadados do conjunto de dados estão sendo atualizados.	Data da última atualização; Identificação da fonte responsável; Contato da fonte responsável; Versão dos dados.
Aceitável (Dois a três pontos)	Somente alguns metadados do conjunto de dados estão presentes.	Data da última atualização; Identificação da fonte responsável.
Crítico (Um ponto)	Somente um ou nenhum metadado está presente.	Área técnica responsável.

Fonte: Elaborado pela autora (2026).

Um conjunto de dados é atualizado anualmente, e seus metadados refletem essa atualização, indicando a nova data de publicação, a versão do arquivo e a instituição responsável pela divulgação.

6.4 Dimensão: formato de arquivo

A dimensão de qualidade formato de arquivo está relacionada à forma como os dados são disponibilizados abrangendo aspectos técnicos que determinam sua compatibilidade com diferentes usos e ferramentas.

A avaliação dessa dimensão busca verificar se os dados são oferecidos em formatos abertos, legíveis por máquina, bem como visa identificar a existência de documentação que descreva sua estrutura e especificações técnicas.

As métricas utilizadas para avaliar essa dimensão na PoC são: **formato dos dados acessíveis, multiplicidade de formatos, dados disponíveis em formato compactado, e documentação do formato.**

1) Formato dos dados acessíveis: verifica se os dados estão disponíveis em formato de dados abertos, como JSON, XML, CSV, Planilha *OpenDocument* (ODS), RDF, que não exigem o uso de *softwares* proprietários para sua leitura ou processamento. Para a análise dessa métrica, são analisados os formatos em que os conjuntos de dados são disponibilizados.

Para a avaliação dessa métrica serão utilizados os critérios conforme Quadro 33.

Quadro 33 – Critérios de avaliação para formato dos dados acessíveis

Nível de qualidade	Descrição	Observações
Excelente	O formato é de padrão aberto e amplamente aceito.	Utiliza formatos como: CSV, JSON e XML.
Aceitável	Permite o acesso e a manipulação dos dados sem dificuldades.	Utiliza dados nos formatos: XLS e <i>Document</i> (DOC)
Crítico	O formato dos dados impede ou dificulta severamente o acesso, uso e reúso dos dados. Requer <i>software</i> proprietário e exige esforços consideráveis para conversão.	Utiliza formato de dados PDF ou arquivos de <i>software</i> específicos (DWG ²⁷ ou <i>Photoshop Document</i> ²⁸ (PSD), por exemplo).

Fonte: Elaborado pela autora (2026).

²⁷ Extensão de um arquivo de desenho técnico Duas Dimensões (2D) e Três Dimensões (3D), criado originalmente para *software* AutoCAD.

²⁸ Formato de arquivo nativo do *Adobe Photoshop*, utilizado para salvar projetos que mantêm todas as suas camadas, máscaras e outras configurações de edição.

A aplicação desses critérios auxilia na construção de conjuntos de dados com formatos mais acessíveis e potencializar a qualidade dos DGA.

2) Multiplicidade de formatos: examina se o conjunto de dados está disponível em múltiplos formatos e abertos, ampliando as possibilidades de acesso, análise e reutilização por diferentes perfis de usuários.

Esta métrica analisa se o conjunto de dados está disponibilizado em mais de um formato, como CSV, JSON, XML, ODS ou RDF, permitindo que seja facilmente utilizado por ferramentas distintas.

Para avaliar a multiplicidade de formatos, os critérios descritos no Quadro 34 serão aplicados.

Quadro 34 – Critérios de avaliação da multiplicidade dos formatos

Nível de qualidade	Descrição	Observações
Excelente	O conjunto de dados é considerado excelente quando oferece uma ampla variedade de formatos, atendendo a diferentes necessidades de uso.	Exemplo: quando disponibilizado em CSV, JSON e XML
Aceitável	O conjunto de dados é considerado aceitável quando oferece dois formatos.	Exemplo: disponibilizado em CSV e JSON
Crítico	A multiplicidade de formatos é crítica quando o conjunto de dados está disponível em apenas um formato.	Exemplo: disponibilizado somente em CSV.

Fonte: Elaborado pela autora (2026).

A aplicação desses critérios permite uma avaliação ágil e eficaz da métrica multiplicidade dos formatos. A disponibilização dos dados em múltiplos formatos amplia a utilização dos dados.

3) Dados disponíveis em formato compactado: essa métrica analisa se o conjunto de dados é disponibilizado em formatos compactados, como em .zip²⁹ ou

²⁹ *Zip File Format*: refere-se à estrutura técnica ou ao tipo de arquivo

.tar.gz³⁰, para facilitar *downloads*, especialmente em casos de arquivos volumosos ou composto por múltiplos documentos.

Para a análise da métrica de dados em formato compactado, serão utilizados os critérios apresentados no Quadro 35.

Quadro 35 – Critérios de avaliação de dados abertos em formato compactado

Nível de qualidade	Descrição	Observações
Excelente (Um ponto)	O conjunto de dados está disponível em um formato compactado (zip ou .tar) e a compactação reduz o tamanho do arquivo, otimizando o <i>download</i> . O arquivo compactado também agrupa múltiplos arquivos.	Exemplo: um conjunto de dados com 5 arquivos CSV com 100 <i>Megabyte</i> (MB) cada é compactado em um único arquivo de 50MB.
Aceitável (Dois pontos)	A compactação é utilizada, mas a redução de tamanho é pouco significativa ou o arquivo compactado contém apenas um único arquivo extenso.	Exemplo: um arquivo CSV de 500MB é compactado em um .zip que também tem 500MB.
Crítico (Três pontos)	O conjunto de dados não está compactado, ou quando estão, apresentam um formato obsoleto ou de difícil acesso. Pode ocorrer ainda que a compactação é a única forma de acesso a um arquivo extenso, sem a opção de <i>download</i> individual.	Arquivos de textos sucintos são compactados, criando uma etapa desnecessária para o uso dos dados. Outro obstáculo é o uso de formatos de compactação que exigem <i>softwares</i> específicos, como proprietários.

Fonte: Elaborado pela autora (2026).

A observação dos critérios apresentados no Quadro 34 garante eficiência maior no acesso aos dados, acelerando sua obtenção pelos usuários.

4) Documentação do formato: investiga se o formato ao qual os dados são disponibilizados está claramente documentado. Essa métrica avalia a presença de documentação, notas explicativas ou manuais que detalham o formato dos dados. O Quadro 36 demonstra como utilizar a métrica da documentação do formato.

³⁰ É um formato de arquivo que combina o *Tape ARchiver*(.tar) e a compactação GZIP(.gz) para agrupar múltiplos arquivos e diretórios em um único e compactado.

Quadro 36 – Critérios de avaliação da documentação do formato utilizado

Nível de qualidade	Descrição	Observações
Excelente (Um ponto)	O formato do conjunto de dados está claramente descrito e documentado, incluindo especificações técnicas e instruções de uso.	
Aceitável (Dois pontos)	O formato do conjunto de dados está indicado, mas a documentação existente é parcial ou insuficiente.	Exemplo: o portal informa que o dado está em CSV, mas não apresenta estrutura de campos, dicionário de variáveis ou orientações de leitura.
Crítico (Três pontos)	O formato está indicado sem documentação, impossibilitando que o usuário compreenda a estrutura dos dados. Nesse caso, o uso e acesso aos dados é limitado.	

Fonte: Elaborado pela autora (2026).

Quando o formato do conjunto de dados está bem descrito, possibilita que o usuário identifique como estão estruturados e quais ferramentas pode utilizar para acessá-los.

6.5 Dimensão: disponibilidade

A dimensão da qualidade “disponibilidade” refere-se ao período em que os DGA permanecem acessíveis de forma confiável, contínua e sem barreiras técnicas ou administrativas aos usuários. A dimensão disponibilidade é essencial para assegurar o acesso aos dados públicos, garantindo que os conjuntos de dados possam ser consultados e utilizados a qualquer momento.

Assim, foram delimitadas as seguintes métricas para avaliar a disponibilidade dos dados: **frequência dos períodos indisponíveis**; **hiperlink de download funcional**; **disponibilidade de canais de comunicação com o usuário e múltiplos idiomas**.

1) Frequência dos períodos indisponíveis: analisa com que frequência ocorrem interrupções no acesso no Portal onde os dados são disponibilizados em

um determinado intervalo de tempo. É necessário estipular um período de observação.

A métrica será avaliada seguindo os critérios estabelecidos no Quadro 37.

Quadro 37 – Frequência dos períodos disponíveis

Nível de qualidade	Descrição
Excelente	Nenhuma interrupção identificada.
Aceitável	Breves interrupções identificadas.
Crítico	Dados frequentemente inacessíveis .

Fonte: Elaborado pela autora (2026).

Essa métrica é essencial para conferir a estabilidade do acesso aos DGA. As interrupções frequentes no acesso aos dados apresentam um obstáculo aos usuários.

2) Hiperlink de download funcional: avalia a funcionalidade dos *hiperlinks* de *downloads* nos conjuntos de dados disponibilizados. O objetivo é verificar se os *hiperlinks* são operacionais, se conduzem corretamente ao recurso informado, se o arquivo pode ser acessado ou se o *download* pode ser realizado sem falhas.

Para avaliação dessa métrica serão aplicados os critérios do Quadro 38.

Quadro 38 – Avaliação dos *hiperlinks* de *downloads* funcionais

Nível de qualidade	Descrição	Exemplo
Excelente (Um ponto)	Todos os <i>hiperlinks</i> testados funcionam corretamente e conduzem ao recurso esperado.	
Aceitável (Dois pontos)	O <i>hiperlink</i> de <i>download</i> funciona parcialmente.	O arquivo é acessado, mas apresenta instabilidades, demora excessiva, mensagens de erro ou inconsistência no conteúdo (arquivo diferente do descrito).
Crítico (Três pontos)	Todos os <i>hiperlinks</i> apresentam erros e falhas.	

Fonte: Elaborado pela autora (2026).

A métrica **hiperlink de downloads funcional** garante aos usuários o acesso de forma simplificada e sem barreiras técnicas. *Hiperlinks* sem funcionalidades geram insatisfações nos usuários fazendo com que desistam de acessar os dados.

3) Disponibilidade de canais de comunicação com o usuário: analisa se há canais de comunicação claros, acessíveis e funcionais para que os usuários possam reportar problemas, sanar dúvidas ou fazer sugestões acerca dos dados disponibilizados.

Na disponibilidade de canais de comunicação com o usuário, são considerados aspectos como:

- presença de *e-mail* para contato, formulário *online*, telefone ou sistema de atendimento específico;
- facilidade de acesso às informações de contato;
- indicação de responsável ou setor específico para o atendimento.

Essa métrica será avaliada conforme critérios apresentados no Quadro 39.

Quadro 39 – Critérios de avaliação da disponibilidade de canais de comunicação

Nível de qualidade	Descrição
Excelente	Os Portais disponibilizam mais de um canal de comunicação de forma clara, acessível e funcional. Também há indicação de um responsável de referência, os canais são visíveis e de fácil acesso.
Aceitável	Os Portais disponibilizam apenas um canal de comunicação que funciona, mas com algumas limitações. Também não há indicação evidente a respeito do responsável, mas o usuário consegue estabelecer contato.
Crítico	Os Portais não disponibilizam canais de comunicação, ou os que estão indicados não funcionam. Não há indicação de um responsável.

Fonte: Elaborado pela autora (2026).

A métrica disponibilidade de canais de comunicação com o usuário é essencial para a comunicação direta entre o gerador dos dados e o usuário. Além disso, transmite confiança e demonstra a existência de um setor ou responsável para atendimento quando houver dúvidas.

4) Múltiplos idiomas: essa métrica analisa se os dados e suas respectivas documentações estão disponíveis em mais de um idioma, além da língua vernácula oficial do país.

A presença de opções de traduções amplia o alcance aos dados, promovendo o acesso aos:

- usuários estrangeiros: pessoas que não dominam o idioma local podem acessar e utilizar os dados;
- pesquisadores internacionais: promove a colaboração e o uso dos dados em contextos globais;
- comunidades multilíngues: atende à diversidade linguística no mesmo país ou região.

Aspectos a serem considerados na métrica **múltiplos idiomas**:

- disponibilidade da base de dados em diferentes idiomas;
- tradução das descrições, metadados, legendas e demais conteúdos informativos;
- clareza e qualidade das traduções apresentadas.

A análise da métrica verifica se a oferta de múltiplos idiomas está presente e se é suficiente para garantir o acesso pleno aos dados de forma internacional. Os critérios de avaliação a serem analisados serão os descritos no Quadro 40.

Quadro 40 – Critérios de avaliação de múltiplos idiomas

Nível de qualidade	Descrição
Excelente	Os dados e toda sua documentação estão traduzidos ao menos em um idioma adicional, com qualidade e coerência adequadas.
Aceitável	Apenas uma parte da documentação está traduzida ou, a tradução apresenta limitações de objetividade, mas ainda assim permite o uso.
Crítico	Não há qualquer tradução disponível além do idioma oficial.

Fonte: Elaborado pela autora (2026).

A métrica de múltiplos idiomas é fundamental para ampliar o alcance dos DGA, possibilitando o acesso por usuários que não dominam o idioma do país.

6.6 Dimensão: precisão

A dimensão de qualidade precisão refere-se à fidelidade dos dados em relação à realidade que representam, bem como à consistência e a exatidão das informações disponibilizadas no conjunto de dados. Essa dimensão busca assegurar que os valores, classificações e registros sejam corretos e reflitam os fenômenos observados, evitando distorções que possam vir a comprometer a análise dos dados.

As métricas associadas à dimensão precisão têm como objetivo avaliar aspectos como a conformidade entre os dados divulgados e suas fontes originais, a ausência de inconsistências internas e a clareza em sua estruturação.

Para o contexto da PoC, foram definidas as seguintes métricas acompanhadas dos seus respectivos critérios de avaliação: **taxa de dados corrompidos; dados duplicados; grau de correspondência dos dados com a realidade e ausência de ambiguidades.**

1) Taxa de dados corrompidos: refere-se à proporção de dados ou células que estão danificados, invalidadas ou ilegíveis, comprometendo de alguma maneira a confiabilidade e usabilidade do conjunto de dados. No Quadro 41 são apresentados alguns exemplos de taxa de dados corrompidos.

Quadro 41 – Exemplos de taxas de dados corrompidos

Campo	Valor corrompido
CPF	@@##
Data	00-00-0000
Valor	Nan, ###, #DIV/0!
Texto	ÃŁÃoC

Fonte: Elaborado pela autora (2026).

Para a avaliação dos dados corrompidos será aplicada a fórmula proposta na Tabela 8.

Tabela 8 – Cálculo e fórmula da qualidade taxa de dados corrompidos

Taxa de dados corrompidos =	<u>N.º de células corrompidas x 100</u> N.º total de células

Ex.: a Figura 11 possui um total de 145 células, dentre as quais oito estão corrompidas. Para calcular o percentual de células corrompidas, aplica-se a fórmula:

$$\text{Fórmula: } \frac{08}{145} = 0.055 \times 100 = 5,5\%$$

Fonte: Elaborado pela autora (2026).

A Figura 12 demonstra exemplo de células corrompidas.

Figura 12 – Células corrompidas

Ensino F ^{aq~e##}											
Total	Anos Iniciais	Anos Finais	1º Ano	2º Ano	3º Ano	4º Ano	5º Ano	6º Ano	7º Ano	8º Ano	9º Ano
10,9	7,1	15,7	2,1	4,2	8,1	9,9	11,1	15,0	15,9	16,2	15,6
10,1	6,7	14,4	2,0	4,1	7,7	9,3	10,0	13,7	14,7	15,0	14,3
16,3	10,1	25,9	2,5	4,9	10,6	13,9	18,1	25,1	25,8	26,3	26,4
9,8	4,4	12,1	1,8	3,1	5,0	6,2	5,3	7,7	12,9	13,3	13,8
9,8	4,4	12,1	1,8	3,1	5,0	6,2	5,3	7,7	12,9	13,3	13,8
--	--	--	--	####	###	--	--	###	--	--	--
12,5	6,7	14,5	2,1	4,0	7,8	8,6	9,9	12,6	14,1	15,4	15,8
12,0	5,9	14,0	1,8	3,5	7,0	Nan	8,7	12,1	13,7	14,9	15,3
20,9	16,0	23,3	6,3	11,0	####	19,1	24,3	21,7	22,6	24,1	25,0
12,3	8,0	20,9	2,0	4,5	9,0	11,3	13,0	20,7	21,6	21,4	19,8
11,5	7,6	19,4	@@	###	8,8	10,8	12,0	19,4	20,3	20,0	17,9

Fonte: Adaptado de INEP (2025).

O resultado da Figura 12 demonstra que 5,5% das células estão corrompidas. Para avaliar o nível de qualidade dessa métrica serão aplicadas as diretrizes apresentadas no Quadro 42.

Quadro 42 – Taxa de dados corrompidos

Nível de qualidade	Dados Corrompidos	Descrição
Excelente	0%	Nenhum dado corrompido, ótimo nível de qualidade.
Aceitável	Até 1%	A qualidade do conjunto de dados ainda é boa.

Crítico	Acima de 2%	A qualidade dos dados está comprometida. O conjunto de dados precisa ser corrigido com urgência.

Fonte: Elaborado pela autora (2026).

Aplicando o critério de avaliação, a porcentagem de dados corrompidos do conjunto de dados (Figura 12) está crítica, impactando negativamente em sua utilização pelo usuário.

2) Dados duplicados: esta métrica mensura a proporção de registros duplicados em relação ao total de registros presentes do conjunto de dados. Registros duplicados ocorrem quando há repetição exata das informações em todos os campos considerados relevantes.

A taxa de dados duplicados é calculada por meio da fórmula disponível na Tabela 9.

Tabela 9 – Cálculo e fórmula da qualidade de dados duplicados

Taxa de duplicação =	<u>N.º de registros duplicados x 100</u> Total de registros

Ex.: em um conjunto de dados composto por 148 registros, foram identificados 42 registros com informações duplicadas. Ao aplicar a fórmula, o resultado é:

$$\text{Fórmula: } \frac{42}{148} = 0.2837 \times 100 = 28,38 \%$$

Fonte: Elaborado pela autora (2026).

Portanto, 28,38% dos registros estão duplicados, o que é inviável para a consulta do usuário.

A avaliação dos resultados é realizada com base nos critérios apresentados no Quadro 43.

Quadro 43 – Avaliação dos dados duplicados

Taxa de duplicação	Nível de qualidade	Descrição
0%	Excelente	Não foram encontrados registros duplicados.

>0% a ≤10%	Aceitável	Foram encontrados registros duplicados em proporção moderada. Apesar de tolerável, esse índice pode causar impacto na análise dos conjuntos de dados.
>10%	Crítico	A qualidade dos dados está comprometida. A reestruturação dos dados é necessária.

Fonte: Elaborado pela autora (2026).

Para a avaliação de duplicatas, deve-se considerar os campos principais que identificam o evento ou a unidade de informação (Ex.: Identidade (ID), data, local, nome, código).

Essa métrica assegura a precisão dos DGA, uma vez que, a presença de registros redundantes pode distorcer as informações do conjunto de dados.

3) Grau de correspondência dos dados com a realidade: essa métrica avalia o grau de correspondência entre os dados registrados e a realidade observada ou aferida por meio de verificação cruzada com fontes externas confiáveis, como por exemplo, com o(a):

- IBGE;
- Receita Federal do Brasil;
- Cadastro Único para Programas Sociais (CadÚnico).

Para a avaliação do grau de correspondência será utilizada a fórmula da Tabela 10.

Tabela 10 – Cálculo e fórmula da qualidade grau de correspondência

Grau de correspondência =	<u>N.º de registros como corretos x 100</u> Total de registros verificados

Ex.: em um conjunto de dados com a população estimada das 27 unidades federativas do Brasil, é realizada uma verificação cruzada com os dados oficiais do IBGE para avaliar sua precisão:

- total de registros verificados: 27
- registros corretos (conforme o IBGE): 22

$$\text{Fórmula: } \frac{22}{27} = 81,5 = 81,5 \%$$

Fonte: Elaborado pela autora (2026).

Isso significa que 81,5% dos registros estão em conformidade com a fonte oficial. A avaliação dos resultados é realizada com base nos critérios apresentados no Quadro 44.

Quadro 44 – Avaliação do grau de correspondência dos dados abertos com a realidade

Grau de correspondência	Nível de qualidade	Descrição
100%	Excelente	Todos os dados conferem com as fontes oficiais.
≥ 90% e < 100%	Aceitável	São identificadas algumas divergências em relação às fontes oficiais. Apesar de não comprometerem significativamente as análises, recomenda-se revisão pontual.
< 90%	Crítico	As divergências são relevantes ou numerosas, comprometendo a confiabilidade do conjunto de dados. É necessária realizar a verificação e a correção urgente das inconsistências.

Fonte: Elaborado pela autora (2026).

Ao aplicar o critério de avaliação do grau de correspondência dos dados com a realidade, o conjunto de dados do exemplo do IBGE tem uma avaliação excelente quanto ao grau de correspondência dos dados.

4) Ausência de ambiguidades: avalia se os dados estão apresentados de forma clara e livres de ambiguidades que possam comprometer sua correta interpretação e uso.

Os aspectos a serem analisados são:

- presença de valores ou categorias com significados vagos, imprecisos ou duplos;
- uso de terminologia imprecisa ou inconsistências terminológicas;
- ausência de explicações complementares quando necessário para entendimento dos dados;
- aplicação inconsistente de padrões, nomenclaturas ou convenções que possam gerar múltiplas interpretações.

A análise é realizada com base nas informações contidas nas células da planilha de dados, observando se os elementos apresentados mantêm um padrão claro, uniforme e inequívoco de representação.

Para mensurar a ausência de ambiguidades existentes no conjunto de dados serão considerados os critérios de avaliação do Quadro 45.

Quadro 45 – Critérios de avaliação de ausência de ambiguidades

Nível de qualidade	Descrição
Excelente (Um ponto)	Os dados estão apresentados de forma clara, com terminologia consistente e categorias bem definidas. Não há múltiplas interpretações.
Aceitável (Dois pontos)	Existem ambiguidades ínfimas ou inconsistências terminológicas, mas que não comprometem de forma significativa a compreensão geral dos dados.
Crítico (Três pontos)	Os dados apresentam ambiguidades relevantes, com categorias vagas e com terminologias imprecisas que dificultam ou inviabilizam o uso.

Fonte: Elaborado pela autora (2026).

A importância dessa métrica está em assegurar a clareza, confiabilidade e precisão dos DGA.

6.7 Dimensão: recuperabilidade

A dimensão de qualidade recuperabilidade está relacionada à facilidade com que os DGA podem ser localizados, acessados e extraídos pelos usuários. A dimensão recuperabilidade mensura a eficiência dos mecanismos de busca, a filtragem e organização dos dados disponibilizados, de modo a garantir que sejam efetivamente encontrados e utilizados pelos usuários.

Apresenta-se o conjunto de métricas propostas para avaliar essa dimensão: **função de recuperação de dados; grau de eficiência dos processos e sistemas de recuperação; navegação; taxa de falhas na recuperação de dados; metadados recuperáveis e vocabulários controlados.**

1) Função de recuperação de dados: essa métrica avalia a existência, acessibilidade e eficiência dos mecanismos de busca e filtros disponíveis nos portais de dados abertos. O objetivo é selecionar registros com base em parâmetros como: retorno de resultados relevantes, corretos e acessíveis e verificar se indivíduos com diferentes níveis de habilidades técnicas são capazes de utilizar os mecanismos de busca disponíveis.

Para avaliação serão analisados os seguintes critérios:

- se o Portal possui mecanismo de busca;
- se é possível buscar por palavras-chave;
- se possui filtros para refinamento das buscas, como: ano, município, temática, órgão, idioma e outros;
- se retorna resultados relevantes, corretos e acessíveis.

A análise desses elementos permitirá comensurar o grau de facilidade de navegação e recuperação dos dados, conforme o Quadro 46.

Quadro 46 – Critérios de avaliação da função de recuperação de dados

Nível de qualidade	Descrição
Excelente (Um ponto)	O portal de dados abertos dispõe de ferramentas de busca eficientes, com possibilidade de pesquisa por palavras-chave e múltiplos filtros, retornando resultados corretos e relevantes. Os recursos de busca são de uso facilitado e intuitivo, mesmo para pessoas sem

	conhecimento técnico avançado.
Aceitável (Dois pontos)	O portal possui ferramenta de busca, mas com recursos limitados. Permite apenas pesquisas por palavras-chave ou oferece poucos filtros, retornando resultados relevantes, ainda que com restrições de precisão, acessibilidade e usabilidade.
Crítico (Três pontos)	O portal não possui ferramenta de busca ou está inoperante. Não há filtros disponíveis ou os resultados retornados são inconsistentes ou incorretos.

Fonte: Elaborado pela autora (2026).

Mecanismos de busca e filtros permitem que os usuários localizem informações específicas de forma ágil, reduzindo barreiras no acesso e promovendo aproveitamento satisfatório dos dados disponibilizados.

2) Eficiência de busca: navegação do Portal de Dados Abertos ou outro repositório público de dados. Essa métrica analisa quanto tempo um usuário leva, em média, para localizar um conjunto de dados específico.

Ex.: localizar o conjunto de dados “Taxas de Distorção Idade-Série na Educação Básica - 2024” no Portal de Dados Abertos. O tempo decorrido desde o início da busca até o acesso aos dados será cronometrado.

Para a avaliação desta métrica será utilizado os critérios apresentados no Quadro 47.

Quadro 47 – Grau de eficiência dos processos e sistemas de recuperação

Tempo	Nível de qualidade	Descrição
Até um minuto	Excelente	Os dados são localizados de forma rápida, com interface intuitiva e ferramentas de buscas eficientes.
De um a três minutos	Aceitável	Apesar da busca ser relativamente rápida, existem etapas intermediárias. O tempo de busca indica que o sistema exige certa familiarização técnica para manipulação.
Acima de três minutos	Crítico	Os dados são de difícil localização. Exige que o usuário navegue por vários campos para encontrar os dados que necessita.

Fonte: Elaborado pela autora (2026).

Ao aplicar os critérios da métrica **grau de eficiência dos processos e sistemas de recuperação**, é possível garantir o acesso aos dados de forma instantânea e intuitiva, o que diminui o tempo das buscas.

3) Navegação: essa métrica analisa a quantidade de cliques necessários para que um usuário consiga acessar efetivamente um conjunto de dados a partir da página inicial do Portal de Dados Abertos ou outro repositório público de dados.

Para avaliar essa métrica será contabilizado um clique por interação do usuário. O total de cliques é contado a partir da página inicial até a visualização ou acesso do conjunto de dados.

Para a avaliação da métrica, sugere-se a interpretação do Quadro 48.

Quadro 48 – Avaliação da métrica navegação

N.º de cliques	Nível de qualidade	Descrição
Um - três	Excelente	O conjunto de dados é muito acessível.
Quatro - seis	Aceitável	O conjunto de dados não é muito acessível.
Acima de sete	Crítico	O usuário tem problemas para acessar os dados. Assim, o usuário tem barreiras de usabilidade e sua experiência está comprometida. O usuário pode abandonar a pesquisa.

Fonte: Elaborado pela autora (2026).

O número de cliques necessários para acessar um conjunto de dados reflete diretamente a eficiência da interface e a facilidade de sua localização.

4) Taxa de falhas na recuperação de dados: esta métrica avalia a quantidade de vezes a busca por um termo retorna resultados irrelevantes ou inadequados. Para a avaliação será considerada a porcentagem de buscas que não retornaram resultados corretos ou úteis em relação ao número de tentativas. Para a avaliação do grau de correspondência será utilizada a fórmula da Tabela 11.

Tabela 11 – Cálculo e fórmula da qualidade taxa de falhas na recuperação de dados

Taxa de falhas =	$\frac{\text{N.º de buscas com falhas} \times 100}{\text{Total de buscas realizadas}}$

Ex.: foram realizadas 20 buscas por diferentes conjuntos de dados no Portal de Dados Abertos. Dessas, sete buscas retornaram resultados relevantes, e três retornaram resultados irrelevantes. Ao aplicar a fórmula, têm-se:

$$\text{Fórmula: } \frac{03}{20} = 0,15 \times 100 = 15 \%$$

Fonte: Elaborado pela autora (2026).

O resultado demonstra que a taxa de falhas na recuperação dos dados foi de 15%. Para a avaliação do resultado obtido serão aplicados os critérios apresentados no Quadro 49.

Quadro 49 – Taxa de falhas na recuperação de dados abertos

Taxa de falhas	Nível de qualidade	Descrição
Até 5%	Excelente	A recuperação de dados não apresenta falhas significativas.
Seis a 20%	Aceitável	Presença de algumas falhas na recuperação dos dados. Recuperação de resultados mais amplos que exige do usuário filtragem para chegar ao desejado.
Acima de 20%	Crítico	A recuperação de dados apresenta muitas falhas prejudicando o acesso aos dados. Não retorna o conjunto de dados desejado.

Fonte: Elaborado pela autora (2026).

Como é possível observar no exemplo citado, 15% da taxa de falhas na recuperação dos dados é considerada como “Aceitável”, segundo os parâmetros do Quadro 48.

Essa métrica evidencia a eficiência da recuperação de dados, permitindo avaliar se o sistema está conseguindo atender às necessidades informacionais dos usuários. Por meio da métrica é possível também o aprimoramento contínuo para identificar problemas nos metadados ou falhas na indexação.

5) Metadados recuperáveis: essa métrica avalia se os metadados associados a um conjunto de dados (Ex.: títulos, descrições e palavras-chave) são

indexados e utilizados pelo mecanismo de busca na recuperação e no Portal de Dados Abertos ou repositório de dados.

Para a avaliação serão utilizados os seguintes passos:

- escolha por busca realizada por título, descrição ou palavras-chave;
- buscar por termos livres ou controlados isoladamente;
- verificar se o conjunto de dados pretendido é recuperado entre os resultados;
- classificar o metadado como recuperável ou não recuperável.

Ex.: um conjunto de dados apresenta os seguintes metadados:

Título: “Taxas de Distorção Idade-Série na Educação Básica”

Descrição: Taxa de Distorção idade-série por localização e dependência administrativa, nos níveis de ensino fundamental e médio - Brasil, regiões geográficas e Unidades da Federação - 2024.

Palavras-chave (implícitas): taxa de distorção idade-série, ensino fundamental e médio no Brasil, taxa de distorção educação básica no Brasil.

A avaliação dessa métrica será realizada conforme o Quadro 50.

Quadro 50 – Metadados recuperáveis

Termo de busca	Metadado de origem	Conjunto de dados encontrados	Recuperável
Taxa de distorção	Título	Sim	Sim
Ensino fundamental e médio	Descrição	Sim	Sim
Idade-série	Título	Sim	Sim
Distorção idade-série	Palavras-chave	Sim	Sim

Fonte: Elaborado pela autora (2026).

Para cada recurso (Título, descrição e palavras-chave) recuperável será atribuído **sim** e **não** para os não recuperados. Para avaliar serão utilizados os critérios: total, parcial e não.

Após sistematizados os dados conforme Quadro 49, a avaliação da qualidade dos metadados recuperáveis seguirá os critérios apresentados no Quadro 51.

Quadro 51 – Critérios de avaliação dos metadados recuperáveis

Nível de qualidade	Descrição
Excelente (Total)	Todos os elementos de metadados testados são recuperáveis individualmente pelo mecanismo de busca. O conjunto de dados aparece consistente nos primeiros resultados de busca.
Aceitável (Parcial)	Apenas alguns elementos de metadados são recuperáveis, ou os conjunto de dados aparecem em posições menos relevantes nos resultados de busca.
Crítico (Não)	Nenhum dos elementos de metadados é recuperável isoladamente, ou o conjunto de dados não aparece nos resultados de busca. A descoberta é significativamente prejudicada, comprometendo o acesso aos dados.

Fonte: Elaborado pela autora (2026).

A métrica “metadados recuperáveis” garante que o usuário encontre os conjuntos de dados por meio de busca textual. Metadados bem estruturados reduzem o tempo de recuperação e aumentam a precisão dos resultados.

6) Vocabulários controlados: essa métrica avalia a uniformidade dos dados indexados com utilização de vocabulários controlados, como por ex.: tesauros, taxonomias ou ontologias. Os vocabulários controlados amplamente adotados visam assegurar a interoperabilidade semântica, a integridade conceitual e o reúso em diferentes contextos de aplicação.

A adoção de vocabulários controlados promove a consistência terminológica e integra os dados com outras fontes, sistemas, plataformas e usuários.

Os aspectos a seguir devem ser considerados na avaliação:

- utilização de vocabulários controlados nacionais ou internacionais específicos à demanda;
- referência a vocabulários controlados reconhecidos;
- aderência a convenções utilizadas por órgãos oficiais.

A avaliação da métrica de vocabulários controlados seguirá a classificação de qualidade apresentada no Quadro 52.

Quadro 52 – Critérios para avaliação dos vocabulários controlados

Nível de qualidade	Descrição
Excelente (Um ponto)	Todos os elementos de dados empregam vocabulários controlados reconhecidos. Possuem aderência completa e convenções oficiais e modelos aceitos no domínio.
Aceitável (Dois pontos)	Apenas alguns elementos dos dados empregam vocabulários controlados. Aderência parcial às convenções oficiais ou modelos reconhecidos.
Crítico (Três pontos)	Nenhum elemento dos dados utiliza vocabulários controlados ou seguem padrões reconhecidos. A aderência a convenções e modelos oficiais é inexistente.

Fonte: Elaborado pela autora (2026).

O uso de vocabulários controlados é fundamental para garantir a consistência lógica e a clareza semântica dos DGA, assim como para padronizar a terminologia de um domínio específico.

6.8 Dimensão: inconsistência

A dimensão da avaliação da qualidade “inconsistência” refere-se à presença de contradições ou informações conflitantes em um conjunto de dados ou entre conjuntos de dados distintos, provenientes da mesma fonte. A consistência dos dados é essencial para garantir a confiabilidade, integridade e utilidade dos dados.

A verificação da inconsistência permite identificar falhas estruturais ou semânticas que comprometem a coerência dos dados, com divergências em valores, classificações, períodos de referência ou categorias utilizadas. Assim, apresenta-se as métricas elaboradas a partir do PoC para avaliar o grau de inconsistência dos DGA.

Nesta seção, são apresentadas as métricas elaboradas voltadas à avaliação do grau de consistência dos DGA. As métricas propostas são: **coerência lógica; consistência entre conjuntos de dados; coerência em dados agregados e uso de classificações.**

1) Coerência lógica: verifica se os dados presentes em um conjunto de dados respeitam regras lógicas, temporais, matemáticas ou de integridade

estabelecidas pelo domínio da aplicação. Avalia se um ou mais campos fazem sentido quando analisados em conjunto, ainda que individualmente pareçam válidos.

Exemplos de incoerência lógica:

- a data de término de um contrato deve ser posterior à data de seu início;
- um município, mesmo com nome homônimo, só pode pertencer a um único estado;
- a soma de subgrupos deve representar o total do grupo;
- uma escola rural não pode estar localizada na área urbana.

Para a avaliação do grau de coerência lógica será utilizada a fórmula da Tabela 12.

Tabela 12 – Fórmula da qualidade coerência lógica

Coerência lógica =	<u>N.º de registros incoerentes x 100</u> Total de registros verificados

Fonte: Elaborado pela autora (2026).

Para a avaliação do resultado obtido serão aplicados os critérios apresentados no Quadro 53.

Quadro 53 – Critérios de avaliação da coerência lógica

Taxa de incoerência	Nível de qualidade	Descrição
Até 1%	Excelente	Os registros, mesmo com algumas falhas, atendem plenamente as regras lógicas, temporais, matemáticas e de integridade definidas. O banco de dados pode ser utilizado diretamente em análises sem necessidade de correções.
>1 a 5%	Aceitável	Existem erros pontuais, mas que não inviabilizam o uso do conjunto de dados.
Acima de 5%	Crítico	Existem falhas significativas na consistência do conjunto de dados. O uso dos dados gera alto risco de interpretações equivocadas.

Fonte: Elaborado pela autora (2026).

Garantir a coerência dos dados é fortalecer a confiança dos usuários nos sistemas de dados. A coerência lógica assegura que os dados reflitam a realidade de forma consistente.

2) Consistência entre conjuntos de dados: está relacionada com a uniformidade e não contradição de informações que aparecem em diferentes bases de dados, sistemas ou conjuntos relacionados. Avalia a padronização entre os dados quando replicados, ou compartilhados entre conjuntos, e se mantêm os mesmos valores e significados.

Ex.: um contrato de obras no Portal da Transparência apresenta o valor de R\$1.200.000,00, mas em outra base, o valor está em R\$1.300.000,00.

Para a avaliação do grau de correspondência será utilizada a fórmula da Tabela 13

Tabela 13 – Fórmula da qualidade coerência entre conjunto de dados

Coerência entre conjunto de dados	<u>N.º de registros inconsistentes x 100</u>
=	Total de registros verificados

Fonte: Elaborado pela autora (2026).

Para a avaliação do resultado obtido na consistência entre conjunto de dados serão aplicados os critérios apresentados no Quadro 54.

Quadro 54 – Critérios de avaliação da consistência entre conjuntos de dados

Taxa de incoerência	Nível de qualidade	Descrição
Zero	Excelente	Nenhuma divergência entre os conjuntos de dados são encontrados.
Até 5%	Aceitável	Divergências pouco significativas identificadas, mas sem impacto significativo na análise.
Acima de 5%	Crítico	Divergências significativas, comprometendo a confiabilidade, o uso e reúso dos dados.

Fonte: Elaborado pela autora (2026).

A métrica consistência entre os conjuntos de dados evita que contradições ocorram entre os conjuntos de DGA.

3) Coerência em dados agregados: essa métrica avalia se os valores totais divulgados em um conjunto de dados correspondem corretamente à soma ou cálculo das partes que os compõem.

Ex.: a soma dos gastos do período de janeiro a dezembro, por exemplo, deve ser igual com o total publicado.

Para a avaliação do grau de coerência será utilizada a fórmula apresentada na Tabela 14.

Tabela 14 – Fórmula da qualidade coerência em dados agregados

Coerência em dados agregados =	$\frac{\text{N.º de registros agregados incoerentes} \times 100}{\text{Total de registros verificados}}$

Fonte: Elaborado pela autora (2026).

Para a avaliação do resultado obtido na coerência dos dados agregados serão aplicados os critérios apresentados no Quadro 55.

Quadro 55 – Critérios de avaliação da coerência em dados agregados

Taxa de incoerência	Nível de qualidade	Descrição
Zero	Excelente	Os totais correspondem exatamente à soma ou cálculo das partes.
Até 2%	Aceitável	Algumas inconsistências matemáticas.
Acima de 2%	Crítico	Divergências frequentes comprometendo a confiabilidade do conjunto de dados.

Fonte: Elaborado pela autora (2026).

Essa métrica assegura a integridade matemática dos dados, garantindo que valores sejam atribuídos corretamente.

4) Uso de classificações: essa métrica avalia a adoção de sistemas de classificações padronizados (como códigos IBGE, Classificação Nacional de Atividades Econômicas (CNAE), entre outros) para garantir a uniformidade, padronização e consistência dos conjuntos de dados em registros públicos.

Ex.: um conjunto de dados utiliza o código IBGE: 3550308 que corresponde ao município de São Paulo.

Para a avaliação do grau de correspondência será utilizada a fórmula da Tabela 15.

Tabela 15 – Fórmula da qualidade uso de classificações

Uso de classificações =	<u>N.º das classificações padronizadas x 100</u> Total de registros verificados

Fonte: Elaborado pela autora (2026).

Para a avaliação do resultado obtido no uso de classificações serão aplicados os critérios apresentados no Quadro 56.

Quadro 56 – Critérios de avaliação do uso de classificações padronizadas

Uso de classificações	Nível de qualidade	Descrição
100%	Excelente	Todos os registros utilizam classificações padronizadas.
80% a 99%	Aceitável	Parte significativa dos registros utiliza classificações padronizadas, mas há lacunas com preenchimento livre.
Menos de 80%	Crítico	Muitos registros não utilizam classificações padronizadas, dificultando a integração dos dados.

Fonte: Elaborado pela autora (2026).

O **uso de classificações** controladas garante a conformidade semântica evitando ambiguidades e duplicidades, além de facilitar a integração e a interoperabilidade entre diferentes conjuntos de dados.

6.9 Dimensão: compreensibilidade

A dimensão de qualidade “compreensibilidade” relaciona-se com a facilidade com que os usuários conseguem compreender os dados disponibilizados, seus significados, estruturas e contextos. Essa dimensão busca assegurar que os dados públicos sejam apresentados de forma clara e acessível, permitindo que diferentes perfis de usuários consigam interpretar e utilizar adequadamente os DGA.

A clareza na apresentação, o uso de vocabulários compreensíveis e controlados, a existência de glossários e notas explicativas são métricas essenciais para garantir que diferentes perfis de usuários possam interpretar e utilizar os DGA.

Nesta seção, são apresentadas as métricas elaboradas para medir o nível de compreensão dos DGA: **presença de glossários, notas explicativas ou legendas; consistência na linguagem; descrições claras e detalhadas dos dados e linguagem adequada a diferentes públicos e valores padronizados.**

1) Presença de dicionários, notas explicativas ou legendas: esta métrica analisa se os conjuntos de dados oferecem recursos como dicionários, notas explicativas, legendas ou glossários, que ajudam a auxiliar o usuário a compreender termos técnicos, códigos, abreviações ou conceitos que não são claros ou comuns.

Ex.:

- glossário de dados: tabela com significado dos termos, por exemplo:
- UF = Unidade Federativa;
- legenda: NA = Não aplicável;
- SC = Sem classificação;
- nota explicativa: valores ajustados pelo Índice Nacional de Preços ao Consumidor Amplo (IPCA) de 2024.

Para a avaliação do resultado serão aplicados os critérios apresentados no Quadro 57.

Quadro 57 – Critérios de avaliação da presença de dicionários, notas ou legendas

Nível de qualidade	Descrição
Excelente (Total)	O conjunto de dados apresenta dicionário completo e atualizado, além de notas explicativas ou legendas apropriadas para todos os termos técnicos.
Aceitável (Parcial)	O conjunto de dados apresenta alguns recursos explicativos parciais que auxiliam, mas são de difícil acesso e não eliminam todas as dúvidas.
Crítico (Não)	O conjunto de dados não apresenta recursos explicativos, dificultando a compreensão dos campos, códigos ou conceitos utilizados.

Fonte: Elaborado pela autora (2026).

Essa métrica auxilia na interpretação correta dos dados, reduzindo ambiguidades, permitindo também que todos os usuários compreendam os significados do conjunto de dados.

2) Consistência na linguagem: avalia o uso padronizado de termos, nomes de campos em uma tabela, abreviações, formatos e estruturas entre diferentes conjuntos de dados.

Ex.: se em um conjunto de dados de uma coluna é denominado por “UF” e, em outro, que deveria representar a mesma informação, a coluna é intitulada de “estado”.

Para a avaliação do grau de correspondência será utilizada a fórmula da Tabela 16.

Tabela 16 – Fórmula da qualidade consistência na linguagem

Consistência na linguagem =	<u>N.º de campos/termos consistentes x 100</u> Total de registros comparáveis entre os conjuntos de dados

Fonte: Elaborado pela autora (2026).

Para a avaliação dos resultados obtidos na consistência da linguagem serão aplicados os critérios apresentados no Quadro 58.

Quadro 58 – Critérios de avaliação da consistência na linguagem

Uso de classificações	Nível de qualidade	Descrição
100%	Excelente	Todos os conjuntos de dados possuem consistência na linguagem utilizada.
Até 10%	Aceitável	Existem pequenas variações de linguagem entre os conjuntos que não comprometem a compreensão geral.
Mais de 10%	Crítico	O conjunto de dados apresenta divergências significativas, comprometendo a padronização e dificultando a integração.

Fonte: Elaborado pela autora (2026).

A métrica de consistência na linguagem garante que termos iguais tenham o mesmo significado em diferentes bases de dados.

3) Descrições claras e detalhadas dos dados: esse critério avalia se os dados disponibilizados são acompanhados de descrições claras e acessíveis, de modo que qualquer usuário, independentemente do seu nível técnico, consiga interpretar as informações apresentadas.

A análise considera aspectos como:

- definição clara dos campos (variáveis): cada coluna ou dado precisa ter um nome que não apresente dúvidas sobre o que representa;
- explicação das unidades de medida utilizadas: se houver números, é fundamental informar as unidades (por ex.: R\$³¹ para valores monetários e Kg de quilograma para peso);
- indicação e descrição das codificações aplicadas (se houver): caso os dados usem códigos (por ex.: “um” para “ativo” e “zero” para “inativo”), devem ser explicados de forma completa.

A verificação é realizada a partir da análise do conteúdo informativo presente nas células da planilha, observando se as informações são suficientemente claras e detalhadas para orientar o usuário na leitura e compreensão dos dados.

Para a avaliação do grau de correspondência será utilizada a fórmula disposta na Tabela 17.

Tabela 17 – Fórmula da qualidade descrições claras e detalhadas dos dados

Descrições claras e detalhadas dos dados =	$\frac{\text{N.º de registros com divergências} \times 100}{\text{Total de registros verificados}}$

Fonte: Elaborado pela autora (2026).

Para a avaliação dos resultados obtidos nas descrições claras e detalhados dos dados serão aplicados os critérios apresentados no Quadro 59.

³¹ É o símbolo da moeda brasileira, o Real.

Quadro 59 – Critérios de avaliação das descrições claras e detalhadas dos dados abertos

Descrições claras e detalhadas	Nível de qualidade	Descrição
Até 5%	Excelente	As descrições dos conjuntos de dados são claras e detalhadas.
6 a 20%	Aceitável	Algumas descrições apresentam problemas, mas não afetam a análise do conjunto de dados.
> 20%	Crítico	O conjunto de dados não apresenta dados claros e detalhados, comprometendo a padronização e dificultando a integração.

Fonte: Elaborado pela autora (2026).

A aplicação desta métrica é fundamental para garantir a acessibilidade e a usabilidade dos dados.

4) Linguagem adequada a diferentes públicos: essa métrica avalia se a linguagem utilizada na apresentação e documentação dos dados é inteligível, objetiva e acessível. São considerados elementos como:

- uso de termos simples e de fácil compreensão: evita-se vocabulários complicados, optando por palavras corriqueiras, também conhecidas por linguagem natural. Entretanto, os vocabulários controlados são aplicados em domínios do conhecimento específicos, quando for o caso;
- construção de frases diretas e concisas: as informações devem ser apresentadas em sentenças diretas, de fácil leitura e assimilação;
- evitar o uso de jargões técnicos ou, quando necessário, incluir explicações adequadas: evitar termos específicos da área. Se for necessário utilizá-los, devem estar acompanhados de explicações específicas e de fácil entendimento.

O objetivo é garantir que os dados possam ser compreendidos por um público amplo, garantindo que instituições públicas, privadas e sociedade civil possam acessar, interpretar e utilizar os dados de forma plena e inclusiva.

A análise dessa métrica contempla todo o conjunto de dados, incluindo:

- títulos: devem ser informativos e descomplicados;
- descrições: precisam ser explicativas sem ser excessivamente complexas;

- legendas: devem decifrar símbolos ou representações de forma simples;
- qualquer outra informação em texto: é recomendado que todo conteúdo escrito seja revisado para garantir clareza;
- linguagem natural e controlada: definir a linguagem natural para públicos não especializados, e a controlada para públicos especializados.

Nesta métrica, são avaliados recursos e práticas como os apresentados no Quadro 60.

Quadro 60 – Recursos de linguagem adequada a diferentes públicos

Categoria/Recurso	Excelente	Aceitável	Crítico
Uso de termos simples e de fácil compreensão	Todos os termos utilizados são adequados e de fácil compreensão.	Alguns termos utilizados são complexos e de difícil compreensão.	Todos os termos utilizados são complexos e de difícil compreensão.
Construção de frases diretas e concisas	As frases são breves e objetivas. Cada oração transmite uma ideia principal, sem informações desnecessárias.	A maioria das frases é direta, mas algumas podem ser um pouco longas ou complexas. As frases transmitem a ideia principal, mas poderiam ser mais concisas.	As frases são longas e o leitor tem dificuldade para identificar a ideia central.
Evitar o uso de jargões técnicos ou, quando necessário, incluir explicações adequadas	Não há uso de jargões técnicos. O vocabulário é acessível e qualquer pessoa compreende o conteúdo sem dificuldade.	Há uso de alguns jargões, mas são imediatamente explicados de forma clara. O texto não é 100% acessível, mas o leitor consegue compreender o significado dos termos técnicos sem necessitar de pesquisa externa.	O texto está repleto de jargões técnicos sem nenhuma explicação. A comunicação é complicada mesmo por especialistas da área.

Fonte: Elaborado pela autora (2026).

Essa avaliação pode ser utilizada para coletar a percepção dos usuários sobre os diversos aspectos da linguagem aplicada. Para interpretar a métrica, será

atribuída uma pontuação para o cálculo do índice de linguagem adequada a diferentes públicos, em que:

- excelente: 3 pontos;
- aceitável: 2 pontos;
- crítico: 1 ponto.

Após a atribuição de pontos para cada recurso avaliado do Quadro 58, será calculada a pontuação obtida. Para o cálculo da métrica, a fórmula aplicada será a apresentada na Tabela 18.

Tabela 18 – Fórmula da qualidade linguagem adequada a diferentes públicos

Linguagem adequada a diferentes públicos =	$\frac{\text{Pontuação obtida} \times 100}{\text{Pontuação máxima}}$

Fonte: Elaborado pela autora (2026).

Para a classificação final da qualidade dos recursos de linguagem adequada a diferentes públicos serão aplicadas as referências do Quadro 61.

Quadro 61 – Critérios de referência para avaliação da linguagem adequada

Nível de qualidade	Cobertura	Descrição
Excelente	90 a 100%	A linguagem utilizada é adequada para os diferentes públicos.
Aceitável	75 a 89%	A linguagem atinge o objetivo, mas poderia ser mais simples.
Crítico	Abaixo de 74	Os dados excluem públicos diversos, comprometendo o uso e reúso dos dados.

Fonte: Elaborado pela autora (2026).

A aplicação dessa métrica é fundamental para tornar os dados abertos e úteis para públicos distintos.

5) Valores padronizados (datas, moedas, medidas e outros): esse critério avalia se os dados seguem formatos padronizados e amplamente reconhecidos para

a representação de elementos como datas, moedas, unidades de medida, números decimais, entre outros.

A utilização de valores padronizados descomplica a leitura, possibilita a comparação e reutilização dos dados por diferentes públicos, sistemas e aplicações.

Para tanto, são considerados aspectos como:

- **uso consistente de formatos internacionais ou nacionais reconhecidos:** datas, por exemplo, devem seguir um padrão único em toda a base de dados, como nas moedas, que devem ser representadas da mesma forma;
- **indicação compreensível de unidade de medida:** se houver valores numéricos, a unidade de medida precisa estar explícita de forma nítida e consistente. Ex.: usar sempre Kg para peso ou Km (quilômetro) para distância;
- **representação uniforme de valores monetários, com símbolo e separadores adequados:** a representação de moeda deve ser padronizada, incluindo o uso de símbolo e dos separadores de décimos, centésimos e milésimos. Ex.: R\$ 1.250,50 (Um mil, duzentos e cinquenta reais e cinquenta centavos);
- **padronização para porcentagem e números decimais:** a forma como os decimais são escritos deve ser a mesma em toda a base de dados. Ex.: 2,5 ou 2.5.

A análise é realizada a partir da observação direta contida nas células da planilha. São avaliados recursos e práticas como os apresentados no Quadro 62.

Quadro 62 – Recursos de avaliação de valores padronizados

Categoria/ Recurso	Excelente	Aceitável	Crítico
Formatos consistentes	Todos os valores (datas, moedas, unidades de medida) seguem padrão único em toda a base de dados.	A maioria dos valores segue um padrão, mas há algumas inconsistências que não interferem na análise do conjunto de dados.	Os formatos de datas, moedas e unidades de medidas são totalmente inconsistentes.
Indicação detalhada de unidade de medida	Todas as unidades de medida estão especificadas. Não há dúvidas sobre valores em metros, quilogramas ou em	A maioria das unidades de medida está indicada, mas há algumas lacunas ou pequenas inconsistências.	Não há indicação evidente das unidades de medida.

	reais.		
Valores monetários uniformes	Todos os valores monetários seguem um único formato.	A maioria dos valores monetários é uniforme, mas existem algumas pequenas inconsistências.	Os valores monetários não possuem um padrão definido.
Padronização para números decimais e porcentagens	Todos os números decimais e porcentagens seguem padrão único e consistente em toda a base de dados.	A maioria dos valores decimais e porcentagens segue um padrão, mas há inconsistências.	Não existe um padrão definido para números decimais e porcentagens.

Fonte: Elaborado pela autora (2026).

Essa métrica de avaliação de valores padronizados pode ser utilizada para coletar a percepção dos usuários sobre seus os diferentes aspectos. Para interpretar a métrica, pontuações serão atribuídas para o cálculo do índice de valores padronizados, aos quais:

- excelente: 3 pontos;
- aceitável: 2 pontos;
- crítico: 1 ponto.

Após a atribuição de pontos para cada recurso avaliado do Quadro 60, será calculada a pontuação obtida. Para o cálculo da métrica, a fórmula será aplicada conforme a Tabela 19.

Tabela 19 – Cálculo e fórmula da qualidade valores padronizados (datas, moedas, medidas e outros)

Valores padronizados =	<u>Pontuação obtida x 100</u> Pontuação máxima

Ex.: suponha-se que, dos quatro recursos avaliados, o total de 10 pontos sejam obtidos. Tem-se:

Fórmula: $\frac{10}{12} \times 100 = 83 \%$

12

Fonte: Elaborado pela autora (2026).

Para a classificação final da qualidade dos valores padronizados (datas, moedas, unidades de medidas e outros) serão aplicadas as referências do Quadro 63.

Quadro 63 – Critérios de referência para avaliação dos valores padronizados (datas, moedas, unidades de medidas e outros)

Nível de qualidade	Cobertura	Descrição
Excelente	90 a 100%	Os valores utilizados são padronizados.
Aceitável	80 a 89%	Apenas alguns valores não estão padronizados, podendo prejudicar a análise dos dados.
Crítico	Abaixo de 79%	A maioria dos dados disponibilizados estão fora dos padrões, ocasionando erros e dificultando a análise.

Fonte: Elaborado pela autora (2026).

A padronização de datas, moedas, unidade de medidas, entre outros, permite que diferentes sistemas processem e disponibilizam os dados sem falhas. Além disso, os usuários conseguem interpretar corretamente valores sem risco de equívocos.

6.10 Dimensão: conformidade

A dimensão de qualidade conformidade refere-se ao grau em que os DGA atendem a normas, padrões, regulamentações e requisitos legais aplicáveis à sua produção, publicação e utilização. Essa dimensão assegura que os conjuntos de dados estejam em consonância com os princípios da transparência pública, da proteção de dados pessoais e sensíveis e da padronização técnica.

A seguir, são apresentadas as métricas desenvolvidas para avaliar a conformidade dos dados com as diretrizes técnicas, legal e normativa: **proteção de dados pessoais e dados sensíveis conforme legislação; origem dos dados e responsável em caso de violação de direitos ou uso inadequado dos dados.**

1) Proteção de dados pessoais e dados sensíveis conforme legislação: a métrica avalia se os conjuntos de dados respeitam as diretrizes da LGPD, garantindo que dados pessoais e dados sensíveis não sejam expostos indevidamente;

Para avaliação dessa métrica serão observados os pontos descritos no Quadro 64.

Quadro 64 – Proteção de dados pessoais e dados sensíveis

Tipo de Informação	Observação
Nome completo de pessoa física.	Deve ser anonimizado ou removido.
CPF, Registro Geral (RG), endereço, telefone, <i>e-mail</i> .	Dados sensíveis, sujeitos à LGPD.
Dados sensíveis (religião, saúde e outros).	Devem ser anonimizados ou tratados com base legal explícita.
Dados agregados e anonimizados.	Permitidos, desde que sem possibilidade de reidentificação.

Fonte: Elaborado pela autora (2026).

Para tanto, nessa métrica serão avaliados os itens apresentados no Quadro 65, e para cada recurso será atribuído uma nota: Excelente ou total (três), Aceitável ou parcial (dois), e Crítico ou não (um).

Quadro 65 – Conformidade com a LGPD

Item avaliado	Excelente (Total)	Aceitável (Parcial)	Crítico (Não)
Dados pessoais identificáveis presentes?			
Há justificativa legal para tratamento de dados?			
Os dados foram anonimizados ou agregados adequadamente?			
Há algum risco de reidentificação com o cruzamento de dados?			

Fonte: Elaborado pela autora (2026).

Após a soma dos pontos de cada item é aplicada a classificação apresentada no Quadro 66.

Quadro 66 – Avaliação da métrica proteção de dados pessoais e sensíveis

Resultado da Avaliação	Avaliação Final
Nenhum dados pessoal ou sensível exposto (total).	Excelente (12 pontos)
Alguns riscos identificados (parcial).	Aceitável (5 a 11 pontos)
Dados pessoais ou sensíveis visíveis ou reidentificáveis (não).	Crítico (0 a 4 pontos)

Fonte: Elaborado pela autora (2026).

A conformidade com a LGPD além de resguardar os direitos fundamentais dos cidadãos, fortalece a confiança dos usuários na disponibilização dos DGA.

2) Origem dos dados: a métrica analisa se é possível identificar visivelmente a origem dos dados disponibilizados em cada conjunto, seja por processos internos de órgãos diversos, por fontes externas (outros órgãos ou instituições) ou por fusão de múltiplas fontes.

Para cada item avaliado será atribuído uma pontuação: total (um), parcial (dois) e não (três). A avaliação se dará pela soma dos itens avaliados. Os itens para avaliação dessa métrica estão apresentados no Quadro 67.

Quadro 67 – Origem dos dados abertos

Item avaliado	Total (1 ponto)	Parcial (2 pontos)	Não (3 pontos)	Observações
A fonte de dados está notoriamente indicada?				Nome do órgão, base ou sistema de origem.
Tipo de origem (interna, externa, mista) é identificável?				Transparência do fluxo de dados.
Existe documentação explicando a procedência?				Hiperlinks, metadados ou notas explicativas.

Fonte: Elaborado pela autora (2026).

A avaliação da métrica origem dos dados abertos será classificada conforme parâmetros apresentados no Quadro 68.

Quadro 68 – Grau de identificação de origem dos dados abertos

Grau de identificação de origem	Avaliação
Totalmente identificada (total)	Excelente
Parcialmente identificada (parcial)	Aceitável
Não identificada (não)	Crítico

Fonte: Elaborado pela autora (2026).

A identificação evidente da origem dos dados é essencial para garantir a confiabilidade e a rastreabilidade dos dados disponibilizados. Essa métrica contribui diretamente para a identificação da estrutura e da legitimidade dos DGA.

3) Responsável em caso de violação de direitos ou uso inadequado dos dados: essa métrica verifica se o conjunto de dados ou sistema em que estão disponibilizados identifica formalmente os agentes responsáveis pelo tratamento dos dados, conforme exigidos LGPD. Esses responsáveis podem ser descritos como controlador, operador e encarregado de dados. Essa métrica tem como objetivo demonstrar quem são os agentes responsáveis no tratamento de dados e quais suas responsabilidades legais, como segue:

- **controlador:** quem delibera a respeito das decisões sobre o tratamento dos dados;
- **operador:** quem realiza o tratamento dos dados sob ordens do controlador;
- **Encarregado de Dados ou Data Protection Officer (DPO):** ponto de contato entre controlador, titulares dos dados e a Autoridade Nacional de Proteção de Dados (ANPD).

Para avaliar a métrica responsável em caso de violação de direitos ou uso inadequado dos dados serão avaliados os itens conforme Quadro 69. Para cada item verificado, atribui-se uma pontuação: **sim** (um ponto), significa se a informação está disponível no conjunto de dados e **não** (dois pontos), significa se a informação

não está presente no conjunto de dados. A avaliação será a soma da pontuação obtida.

Quadro 69 – Responsáveis legais por dados pessoais ou sensíveis disponibilizados

Item verificado	Sim (1 ponto)	Não (2 pontos)
Identificação do controlador dos dados.		
Nomeação do DPO com contato acessível.		
Indicação de responsabilidades, conforme LGPD.		
Canal para denúncia ou solicitação de direitos do cidadão.		

Fonte: Elaborado pela autora (2026).

Após a avaliação dos itens pode-se aplicar a classificação apresentada no Quadro 70.

Quadro 70 – Grau de clareza institucional

Grau de clareza Institucional	Avaliação
Todos os recursos estão claramente identificados com canal de contato.	Excelente (4 pontos)
Apenas alguns recursos estão identificados.	Aceitável (5 a 7 pontos)
Nenhuma informação identificada.	Crítico (8 pontos)

Fonte: Elaborado pela autora (2026).

Informações precisas acerca dos agentes de tratamento dos dados são fundamentais para reforçar a responsabilidade da instituição e segurança jurídica para os usuários e seus dados.

6.11 Dimensão: licença aberta/não proprietária

A dimensão de qualidade licença aberta/não proprietária avalia o grau de clareza e abertura no que se refere ao uso, reúso, modificação e redistribuição dos DGA. A adoção de uma licença explícita e acessível é condição essencial para

assegurar que os dados públicos possam ser livremente utilizados por qualquer cidadão, instituição ou desenvolvedor.

A adoção dessa dimensão busca verificar se os conjuntos de dados disponibilizados especificam de forma coerente os direitos e restrições associados ao seu uso.

As métricas propostas permitem identificar se os conjuntos de dados estão devidamente licenciados, se as permissões estão bem definidas e se os usuários têm garantias quanto à estabilidade e continuidade do uso das informações. As métricas consideradas são: **existência da licença; tipo de licença; permissão para reutilização; período de validade para o uso dos dados; ausência de restrições e declaração de domínio público.**

1) Existência das licenças: a licença empregada deve apresentar se há uma declaração explícita, visível e facilmente identificável que orienta o uso do conjunto de dados.

Para avaliação dessa métrica será avaliado se:

- a licença está precisamente indicada na página a qual o conjunto de dados está publicado;
- a licença é padronizada em relação a ?;
- se o *hyperlink* para acesso ao texto completo da licença está disponível;
- se as informações sobre a licença são encontradas de forma simples no *site*.

Ex.: na *homepage* do conjunto de dados deve constar as informações:

Licença: **Creative Commons (CC)**

Uniform Resource Locator (URL) da licença: _____

Para a avaliação da métrica será serão avaliados os recursos sobre a licença e atribuído uma pontuação para cada recurso: **sim** (um ponto), significa se a informação está disponível no conjunto de dados e **não** (dois pontos), significa se a informação não estiver presente no conjunto de dados. A avaliação será a soma da pontuação obtida, conforme Quadro 71.

Quadro 71 – Existência de licenças

Item verificado	Sim (1 ponto)	Não (2 pontos)
A licença está indicada na <i>homepage</i> do conjunto de dados?		
A licença é padronizada?		
O <i>hiperlink</i> para o texto completo da licença está disponível?		
A informação está em local de fácil acesso?		

Fonte: Elaborado pela autora (2026).

Após a avaliação dos itens aplicar-se-á a classificação apresentada no Quadro 72.

Quadro 72 – Critérios de avaliação da existência de licenças

Nível de qualidade	Descrição
Excelente (4 pontos)	A licença está visivelmente indicada, padronizada, acompanhada do <i>hiperlink</i> para o texto completo e localizada em espaço de fácil acesso na página do conjunto de dados.
Aceitável (5 a 7 pontos)	Apenas parte dos aspectos está contemplada.
Crítico (8 pontos)	Não há indicação de licença.

Fonte: Elaborado pela autora (2026).

A existência de uma licença explícita é fundamental para assegurar a segurança jurídica no uso dos DGA, pois define os direitos e restrições aplicáveis.

2) Tipos de licenças: analisa se a licença aplicada ao conjunto de dados é compatível com os princípios dos dados abertos, ou seja, se permite o uso, redistribuição, reutilização e modificação dos dados por qualquer pessoa e para qualquer finalidade.

Exemplo de licenças compatíveis com dados abertos (recomendadas):

- **Creative Commons Zero (CC0):** é a licença de domínio público. Ela é a licença mais permissiva. Renúncia a todos os direitos autorais, colocando os dados

em domínio público. Não exige atribuição. **(Altamente recomendada para dados abertos)**;

- ***Creative Commons Attribution (CC BY)***: permite qualquer uso, como o comercial ou não, desde que a fonte original seja atribuída. **(Altamente recomendada e amplamente utilizada)**;

- ***Open Data Commons Public Domain Dedication and Licence (PDDL)***: é a licença equivalente de CC0 para bancos de dados;

- ***Open Data Commons Attribution License (ODC-BY)***: licença que equivalente a CC BY para bancos de dados;

- ***Open Government Licence (OGL)*** utilizada pelo governo do Reino Unido que permite livre reutilização com atribuição da fonte. Licenças similares podem existir em outros países;

- ***Massachusetts Institute of Technology (MIT) License, Apache License*** (para *software*, mas princípios semelhantes podem ser aplicados a dados se apropriado): embora sejam licenças de *software*, os princípios de permissividade e atribuições são compatíveis com a filosofia de dados abertos.

Exemplos de licenças incompatíveis com dados abertos:

As licenças apresentadas a seguir não estão em conformidade com a licença aberta:

- ***Creative Commons Attribution NoDerivatives (CC BY – NC, não comercial)***: restringe o uso comercial. Os dados não podem ser utilizados para fins lucrativos;

- ***Creative Commons Attribution (CC BY – ND, sem obras derivadas – não comercial)***: impede modificações, adaptações ou criação de novos dados baseados no original.

Para avaliação da métrica **tipo de licenças** serão utilizados os critérios apresentados no Quadro 73.

Quadro 73 – Critérios de avaliação dos tipos de licenças

Nível de qualidade	Descrição	Observação

Excelente (Total)	Licença plenamente compatível com os princípios de dados abertos. Permite uso, redistribuição, reutilização e modificação por qualquer pessoa e para qualquer finalidade.	CCO, CC BY, PDDL, ODC-BY, OGL, MIT <i>License</i> , Apache <i>License</i>
Aceitável (Parcial)	Licença parcialmente aberta. Permite acesso e uso, mas com restrições que limitam o reuso pleno.	CC BY-NC
Crítico (Não)	Licença inexistente, não informada ou incompatível com dados abertos.	CC BY-ND

Fonte: Elaborado pela autora (2026).

A observação da qualidade dessa métrica é essencial para garantir a abertura efetiva dos DGA, especialmente ao longo do tempo.

3) Permissões para reutilização de licenças: essa métrica avalia se a licença aplicada ao conjunto de dados permite a:

- **reutilização:** o uso dos dados para diferentes finalidades;
- **redistribuição:** o compartilhamento dos dados com terceiros;
- **modificação:** a criação de obras e estudos derivados;
- **limitações dos dados:** como a proibição de uso comercial.

Ao avaliar essa métrica, os critérios apresentados no Quadro 74 serão aplicados.

Quadro 74 – Critérios para avaliação de permissões para reutilização de licenças

Nível de qualidade	Descrição
Excelente (Total)	A licença permite reutilização irrestrita, redistribuição e modificação.
Aceitável (Parcial)	A licença permite reutilização parcial e impõe restrições que limitam o uso dos dados.
Crítico (Não)	A licença não permite reutilização, redistribuição ou modificação.

Fonte: Elaborado pela autora (2026).

A permissão para a reutilização compõe o núcleo central dos dados abertos, pois garante que os conjuntos de dados possam ser utilizados, redistribuídos e modificados sem limitações.

4) Período de validade para o uso dos dados: avalia a existência de permissão com um período de validade para uso dos conjuntos de dados e, se há previsão de revogabilidade.

- **prazo de validade:** a permissão para usar os dados tem uma data de expiração?

Aspectos a serem considerados:

1. a licença tem uma data de expiração definida?
2. após o vencimento da licença adotada, o acesso e uso dos dados ainda são permitidos?

Os conjuntos de dados abertos não devem possuir validade determinada, pois sua validade deve ser indeterminada para promover o reúso contínuo.

- **revogabilidade:** a licença ou permissão pode ser retirada pelo fornecedor dos dados a qualquer momento, e sob quais condições?

Aspectos a serem considerados:

1. a licença pode ser revogada unilateralmente?
2. as condições de revogação estão explícitas e transparentes?

Para garantir a confiança nos conjuntos de dados, as licenças abertas devem ser irrevogáveis. Ao avaliar essa métrica serão observados os critérios apresentados no Quadro 75.

Quadro 75 – Critérios de avaliação do período de validade para acesso e uso dos dados abertos

Nível de qualidade	Descrição
Excelente (Total)	A licença é indeterminada e irrevogável.
Aceitável	A licença é indeterminada, mas contém cláusulas de revogabilidade

(Parcial)	condicionada.
Crítico (Não)	A licença possui prazo de validade definido ou pode ser revogada unilateralmente pelo fornecedor de dados.

Fonte: Elaborado pela autora (2026).

A métrica **período de validade para uso dos dados abertos** garante a estabilidade para o reuso dos dados em projetos de longo prazo. As licenças devem possuir validade indeterminada e serem irrevogáveis para assegurar que qualquer pessoa possa utilizá-los de forma contínua e previsível.

5) Ausência de restrições (Domínio Público): essa métrica analisa se o conjunto de dados está livre de restrições legais, como direitos autorais, patentes ou outras formas de propriedade intelectual.

Ex.: análise de um conjunto de dados disponibilizado no Portal de Dados Abertos com licença *Creative Commons Zero* (CC0). Essa licença declara que o conjunto de dados foi inserido no domínio público. Assim, o conjunto de dados está livre de direitos autorais e restrições de uso, atendendo ao critério de ausência de restrições.

A avaliação será classificada conforme parâmetros apresentados no Quadro 76.

Quadro 76 – Critérios de avaliação da ausência de restrições (Domínio Público)

Nível de qualidade	Descrição
Excelente (Total)	O conjunto de dados está explicitamente incorporado no Domínio Público ou é disponibilizado sob uma licença equivalente.
Aceitável (Parcial)	O conjunto de dados é liberado sob uma licença aberta, mas impõe restrições mínimas que não impedem o reuso de forma significativa.
Crítico (Não)	Existem restrições legais significativas no conjunto de dados que impedem ou limitam seu reuso.

Fonte: Elaborado pela autora (2026).

A métrica de ausência de restrições, como os dados em Domínio Público, é fundamental para a abertura dos dados governamentais, permitindo que qualquer pessoa faça o uso e reuso dos dados para qualquer finalidade.

6) Declaração de Domínio Público: analisa a existência de declaração exata, explícita e acessível que explique que os dados estão no domínio público, além da existência de políticas institucionais que garantam essa condição.

Para avaliação dessa métrica serão observados os níveis descritos no Quadro 77.

Quadro 77 – Critérios de avaliação da declaração de Domínio Público

Nível de qualidade	Descrição
Excelente (Total)	A declaração de Domínio Público é explícita e facilmente acessível.
Aceitável (Parcial)	A declaração de Domínio Público não é clara ou inacessível.
Crítico (Não)	Não há declaração de Domínio Público.

Fonte: Elaborado pela autora (2026).

A declaração de Domínio Público garante ao usuário usar e reusar dos dados sem se preocupar com consequências legais que podem ser aplicadas.

6.12 Dimensão: usabilidade

A dimensão de qualidade usabilidade está relacionada com a facilidade em que o usuário encontra, compreende e utiliza os conjuntos de dados disponibilizados. Um conjunto de dados completo e atualizado, perde seu valor se os usuários não conseguem encontrá-lo, entendê-lo ou utilizá-lo de forma eficaz e eficiente.

A dimensão de qualidade usabilidade concentra-se em garantir que o usuário tenha uma experiência positiva na interação com o sistema. É caracterizada também pela clareza na navegação, na disponibilidade dos dados em formatos acessíveis e na ausência de barreiras técnicas.

Para tanto, foram definidas as seguintes métricas para avaliação dessa dimensão: **conjunto de dados disponíveis para *download* sem necessidade de autenticação e interface amigável.**

1) Conjunto de dados disponíveis para *download* sem necessidade de autenticação: analisa se os conjuntos de dados podem ser acessados, visualizados e se o *download* pode ser realizado por qualquer indivíduo, a qualquer momento, sem a exigência de um *login*, senha, registro ou qualquer forma de identificação.

Requisitos para avaliação:

- sem necessidade de *login* obrigatório;
- sem necessidade de cadastro prévio;
- sem necessidade de autenticação por *token* ou chave de autenticação;
- disponível 24 horas por dia, sete dias na semana.

Para a métrica, são avaliados recursos e práticas conforme apresentados no Quadro 78.

Quadro 78 – Requisitos para acesso aos dados disponíveis para *download* direto

Categoria/Recurso	Excelente	Aceitável	Crítico
Acesso anônimo	Não existe nenhuma barreira para o usuário acessar.	Algumas barreiras ínfimas estão presentes, mas não impedem o reuso de forma significativa.	Existem barreiras significativas que tornam o acesso aos dados difícil ou inviável.
Sem formulários ou solicitações	O processo é direto, não existe etapa intermediária entre o usuário e os dados.	Existe a necessidade de preenchimento de formulário, mas é simples, rápido e não seletivo.	A existência de formulários ou solicitações causa barreiras significativas para o usuário.
Sem barreiras de verificação humana ou <i>scripts</i> de bloqueio	Nenhuma barreira impede o acesso automatizado.	Existem barreiras mínimas e esporádicas não representando impedimento para acesso.	Existem barreiras tecnológicas que impedem ou dificultam severamente o acesso automatizado.
Hiperlink direto e estável	O <i>hyperlink</i> é direto e estável.	O <i>hyperlink</i> , apesar de não ser direto, é funcional e relativamente estável.	O <i>hyperlink</i> não é direto e é instável.

Fonte: Elaborado pela autora (2026).

Para interpretar a métrica, será atribuída uma pontuação para o cálculo do índice de *download* disponível, em que:

- excelente: 3 pontos;
- aceitável: 2 pontos;
- crítico: 1 ponto.

Após a atribuição de pontos para cada recurso avaliado no Quadro 76, será calculada a pontuação obtida. Para o cálculo da qualidade da métrica **conjunto de dados disponíveis para *download* sem necessidade de autenticação**, será aplicada a fórmula demonstrada na Tabela 20.

Tabela 20 – Cálculo e fórmula da qualidade conjunto de dados disponíveis para *download* sem necessidade de autenticação

Dados disponíveis para <i>download</i> sem necessidade de autenticação =	<u>Pontuação obtida x 100</u> Pontuação máxima

Ex.: suponha-se que, dos quatro recursos avaliados, 11 pontos sejam obtidos. Têm-se:

$$\text{Fórmula: } \frac{11}{12} \times 100 = 92 \%$$

Fonte: Elaborado pela autora (2026).

Para a classificação final da qualidade da métrica, será aplicada as referências do Quadro 79.

Quadro 79 – Critérios de avaliação para dados abertos disponíveis por *download*

Nível de qualidade	Cobertura	Descrição
Excelente	90% a 100%	Os aspectos estão em conformidade com praticamente todos os requisitos necessários.
Aceitável	80% a 89%	O conjunto de dados cumpre a maioria dos requisitos de qualidade, porém, apresenta falhas notáveis.
Crítico	Abaixo de 79%	Os aspectos estão em

		desconformidade com uma parte significativa dos requisitos, inviabilizando o <i>download</i> .

Fonte: Elaborado pela autora (2026).

Essa métrica garante que os dados além de disponíveis, estejam acessíveis e que possibilitam que os dados governamentais sejam utilizáveis.

2) Interface amigável: avalia a qualidade da experiência do usuário ao navegar pelo *site* em que o conjunto de dados está publicado. A interface amigável deve possuir de navegação e interação simplificada, intuitiva e eficiente. Elementos como botões, barras de ferramentas, menus e formulários devem estar dispostos de maneira lógica e acessível. O *layout* deve ser claro, sem excessos de informações visuais e informacionais.

Exemplo de interface amigável:

- presença de um menu simples, com categorias bem definidas na página inicial;
- botões para baixar realizar *download* de conjunto de dados grandes, nitidamente identificados e agrupados por tema ou ano;
- utilização de cores que garantam contraste e leitura eficientes;
- títulos apresentados em negrito;
- sem incidência de anúncios, *pop-ups* ou outros elementos que interrompam a navegação.

Para a avaliação dessa métrica serão avaliados os recursos: menu, *downloads*, cores, títulos e navegação. Para cada recurso adequadamente disponibilizados no conjunto de dados será atribuída a pontuação (sim – um) e se o recurso não estiver presente ou adequado (não – dois), a avaliação final é a soma dessas pontuações. Serão aplicados os critérios apresentados no Quadro 80.

Quadro 80 – Grau de interface amigável

Grau de disponibilização de dados	Nível de qualidade
Interface amigável.	Excelente (5 pontos)

Interface pouco amigável.	Aceitável (6 a 9 pontos)
Interface com obstáculos que dificultam a navegação.	Crítico (10 pontos)

Fonte: Elaborado pela autora (2026).

A avaliação dessa métrica permite identificar o grau de maturidade dos portais de dados abertos no que se refere à experiência do usuário e, à eficiência da interação com o sistema. Interfaces bem projetadas favorecem o acesso equitativo aos dados públicos, reduzem barreiras de uso e estimulam o reúso dos dados por diferentes públicos.

6.13 Dimensão: não discricionário

A dimensão de qualidade de não discricionário refere-se à capacidade de acesso igualitário entre os usuários dos DGA. A avaliação dessa dimensão considera a ausência de barreiras técnicas, administrativas ou econômicas que possam limitar o acesso e uso dos dados, como exigências de cadastros, autenticações, aprovações prévias ou quaisquer outras formas de controle para acesso.

As métricas propostas para mensurar o grau de não discricionariedade dos dados são: **dados disponíveis diretamente ao usuário e custo de acesso.**

1) Dados abertos disponíveis diretamente ao usuário: essa métrica verifica se os conjuntos de dados estão disponíveis diretamente ao público, sem a exigência de solicitação formal via Sistema de Informação ao Cidadão (SIC) para obter os conjuntos de dados que deveriam ser públicos. Os dados classificados como abertos devem estar disponíveis para consulta, *download* e reúso por qualquer usuário, sem barreiras adicionais.

A avaliação dessa métrica será realizada conforme critérios exibidos no Quadro 81.

Quadro 81 – Critérios de avaliação de dados abertos disponíveis diretamente aos usuários

Nível de qualidade	Descrição
Excelente (Total)	Todos os dados do conjunto de dados estão disponíveis diretamente ao usuário, sem necessidade de solicitação formal.
Aceitável (Parcial)	Parte dos dados do conjunto de dados estão disponíveis diretamente, mas existem outros que só podem ser acessados mediante solicitação formal.
Crítico (Não)	Os dados do conjunto de dados não estão disponíveis diretamente, sendo necessário recorrer ao SIC ou outro procedimento formal.

Fonte: Elaborado pela autora (2026).

A observação dessa métrica garante que o acesso aos dados seja realizado de forma imediata, sem necessidade de solicitações formais, tornando o processo ágil, inclusivo e democrático.

2) Custo de acesso aos dados abertos: é examinado se os dados estão disponíveis de forma gratuita ou com baixo custo de acesso, de modo a garantir o acesso universal e equitativo às informações.

Para avaliação dessa métrica, são considerados aspectos como:

- disponibilidade gratuita para visualização ou *download*;
- ausência de barreiras financeiras para o acesso completo ao conteúdo;
- existência de taxas simbólicas justificadas e compatíveis com o princípio da ampla acessibilidade.

O objetivo da métrica é assegurar que qualquer usuário, independentemente de sua condição socioeconômica, possa acessar os dados sem restrições.

Os critérios para avaliação dessa métrica serão apresentados no Quadro 82.

Quadro 82 – Critérios de avaliação dos custos de acesso aos dados abertos

Nível de qualidade	Descrição
Excelente (Total)	Os dados estão integralmente disponíveis de forma gratuita e sem cobrança de qualquer tipo de taxa.

Aceitável (Parcial)	Os dados são disponibilizados com taxas simbólicas ou justificadas, que não inviabilizam o acesso ao conteúdo.
Crítico (Não)	Existem barreiras financeiras significativas, restringindo o uso e reúso dos dados pelos usuários.

Fonte: Elaborado pela autora (2026).

Os conjuntos de DGA devem estar disponíveis sem restrições econômicas, de modo que qualquer usuário, independentemente de sua condição financeira, possa consultar, utilizar e reutilizar os dados.

6.14 Dimensão: primário

A dimensão de qualidade primário refere-se à origem e autenticidade dos dados disponibilizados em portais governamentais abertos, com o objetivo de verificar se os conjuntos de dados foram publicados por entidades responsáveis diretamente por sua coleta, produção e manutenção.

A análise dessa dimensão permite identificar se os dados mantêm um vínculo direto com a instituição produtora, garantindo que o conteúdo publicado seja original e fiel ao processo de geração dos dados públicos. Também, possibilita a verificação da existência de documentação que descreve o fluxo de produção, tratamento e publicação dos dados.

As métricas utilizadas para avaliar essa dimensão são: **fonte de dados; processos dos dados documentais e rastreabilidade dos dados.**

1) Fontes dos dados abertos: essa métrica avalia a origem dos dados e os processos utilizados na sua coleta e publicação, conforme segue:

- **Fonte primária** – considera-se fonte primária quando os dados foram coletados diretamente por uma entidade pública ou pelo próprio responsável pela publicação; dados que não sofreram intermediações ou derivação de outras fontes.

Ex.: o conjunto de dados “Indicadores Educacionais da Educação Básica” foi coletado por meio do Censo Escolar da Educação Básica, coordenado pelo INEP, órgão oficial do governo federal. O Censo Escolar é realizado anualmente, e suas informações auditadas e validadas por meio do sistema Educacenso.

• **Fonte secundária** – são dados de fonte secundária quando são derivados, agregados ou compilados de fontes intermediárias ou de terceiros; envolvem transformações realizadas por terceiros; podem levar a restrições de uso, licenciamento ou direitos autorais.

Ex.: um relatório elaborado por uma empresa privada que utiliza dados públicos, mas o agrupa de forma a satisfazer as necessidades da sua empresa, o que constitui uma fonte secundária.

Para a avaliação serão considerados os critérios descritos no Quadro 83.

Quadro 83 – Fonte de dados abertos

Nível de qualidade	Critério
Excelente (Total)	A fonte de dados é visivelmente identificada.
Aceitável (Parcial)	A fonte dos dados é mencionada de forma genérica.
Crítico (Não)	Não há qualquer identificação sobre a fonte dos dados.

Fonte: Elaborado pela autora (2026).

Essa métrica garante a autenticidade e confiabilidade dos dados. Dados de fonte primária apresentam maior fidedignidade e rastreabilidade.

2) Processo dos dados documentais: essa métrica avalia se existe documentação objetiva, completa e acessível sobre as etapas que envolvem a produção, tratamento e disponibilização dos dados públicos.

Para o processamento dos dados, os critérios avaliados serão os apresentados no Quadro 84.

Quadro 84 – Processo dos dados documentais

Nível de qualidade	Critério
Excelente (Total)	A documentação é detalhada e de fácil acesso. A documentação descreve todas as etapas do CVD.
Aceitável	A documentação é parcial ou não cobre todas as etapas do processo,

(Parcial)	limitando a compreensão plena da origem e do tratamento dos dados. A documentação é de difícil localização, tornando-se um obstáculo para o usuário.
Crítico (Não)	Não há qualquer identificação sobre a origem dos dados ou é insuficiente.

Fonte: Elaborado pela autora (2026).

A métrica **processo dos dados documentais** é fundamental para que o usuário avalie a qualidade, a confiabilidade ou a adequação ao propósito de uso.

3) Rastreabilidade dos dados: avalia se o conjunto de dados possui um histórico documentado, acessível e transparente acerca de todas as etapas pelas quais os dados passaram, desde sua origem até a versão atual publicada.

As informações verificadas nessa métrica são:

- modificações: alterações realizadas nos dados ao longo do tempo (correções, atualizações e outros);
- integrações: combinações com outras fontes de dados ou fusões de bases distintas;
- processamento: transformações aplicadas, como limpeza, agregações, cálculos, entre outros.

Para avaliar a métrica será atribuído um ponto para cada recurso presente no conjunto de dados e dois se o recurso não estiver presente no conjunto de dados. A soma da pontuação determinará a avaliação da métrica.

Os critérios para avaliação da métrica são os apresentados no Quadro 85.

Quadro 85 – Rastreabilidade dos dados abertos

Níveis de qualidade	Descrição
Excelente (3 pontos)	A documentação é completa, acessível e atualizada, cobrindo todas as modificações, integrações e etapas de processamento dos dados.
Aceitável (4 a 5 pontos)	A documentação existente é parcial, contemplando apenas algumas etapas, pois não cobre todo o histórico dos dados.
Crítico (6 pontos)	Não existe documentação ou é insuficiente, pois não permite a rastreabilidade do conjunto de dados.

Fonte: Elaborado pela autora (2026).

Com base nas métricas apresentadas para a dimensão de qualidade primário é possível aprimorar a prática de disponibilização dos conjuntos de DGA.

6.15 Dimensão: interoperabilidade

A dimensão de qualidade interoperabilidade está relacionada à capacidade dos DGA serem integrados, compartilhados e reutilizados por diferentes sistemas, plataformas e aplicações, de forma automatizada, eficiente e sem perda de informação.

A interoperabilidade é uma dimensão fundamental para a consolidação do ecossistema do governo digital, pois permite a integração entre bases de dados públicas, o cruzamento de informações e o desenvolvimento de soluções inovadoras a partir dos DGA.

As métricas aplicadas para avaliar essa dimensão são: **capacidade de reutilização e integração em sistemas.**

1) Capacidade de reutilização dos dados abertos: essa métrica analisa se o conjunto de dados disponibiliza informações relacionadas a reutilização do conjunto de dados, permitindo compreender o grau de interesse e aproveitamento pela sociedade. A disponibilização de estatísticas de reuso contribui para mensurar o impacto social e científico dos dados abertos. Os aspectos a serem avaliados são:

- o número de reusos que o conjunto de dados possui;
- a quantidade de *downloads* realizados do conjunto de dados.

Para mensurar a qualidade dessa métrica, serão aplicados os critérios apresentados no Quadro 86.

Quadro 86 – Critérios de avaliação da capacidade de reutilização dos dados abertos

Níveis de qualidade	Descrição
Excelente (Total)	As informações sobre a quantidade de reusos e <i>downloads</i> são externalizadas, atualizadas e acessíveis.
Aceitável (Parcial)	As informações são parciais e contemplam apenas um dos aspectos do reuso ou <i>download</i> .

Crítico (Não)	Não existe qualquer informação disponível sobre reúsos ou <i>downloads</i> dos conjuntos de dados.

Fonte: Elaborado pela autora (2026).

Essa métrica tem capacidade de demonstrar o alcance e a demanda dos DGA na sociedade.

2) Integração em sistemas: essa métrica analisa a capacidade dos dados disponibilizados serem facilmente conectados, consumidos e utilizados por outras plataformas ou sistemas externos, como ferramentas *Business Intelligence* (BI), *Enterprise Resource Planning* (ERP), *Customer Relationship Management* (CRM), ou outros bancos de dados.

Entre os recursos de integração, têm-se:

- *Application Programming Interface (API) RESTful*³² com padrão JSON: a API segue o padrão *Representational State Transfer* (REST) e retorna dados em JSON, facilitando a integração com sistemas e aplicações (Brasil, 2024);
- documentação interativa: a documentação da API está disponível no *site*, permitindo explorar os *endpoints* diretamente no navegador;
- autenticação via *token*: para acessar a API, é necessário gerar um *token* de acesso;
- conectores e bibliotecas: algumas ferramentas de terceiros, como um pacote *Python* facilita o consumo de dados em ambiente de análise.

Os critérios para avaliação da métrica integração em sistemas externos serão descritos no Quadro 87.

Quadro 87 – Critérios de avaliação da integração sem sistemas externos

Níveis de qualidade	Descrição
Excelente (Total)	Os dados oferecem APIs RESTful em padrão aberto (JSON ou Similar), com documentação interativa, autenticação notável e suporte a conectores/bibliotecas que facilitam a integração.

³² O mesmo que API REST. Significa a junção de APT com REST.

Aceitável (Parcial)	Os dados oferecem algum recurso de integração, como uma API simples ou exportação em formato aberto, mas com documentação incompleta ou ausência de bibliotecas auxiliares.
Crítico (Não)	Não existe suporte para integração em sistemas externos, limitando o uso dos dados a <i>downloads</i> manuais ou formatos pouco interoperáveis.

Fonte: Elaborado pela autora (2026).

Essa métrica é essencial para a interoperabilidade e a efetiva utilização dos DGA.

Dessa forma, o modelo de avaliação para as métricas da dimensão de qualidade interoperabilidade permite identificar em que medida os DGA estão aptos a serem utilizados de maneira eficiente em diferentes contextos tecnológicos.

Por meio deste instrumento, será possível realizar a avaliação geral de cada dimensão, bem como do conjunto de dados analisado. Para a mensuração das dimensões, definiu-se um peso por escala, no qual cada métrica integrante de uma dimensão é avaliada com base nos seguintes valores:

- excelente: 10 pontos;
- aceitável: 05 pontos;
- crítico: 00 pontos.

Para o cálculo será utilizada a fórmula apresentada na Tabela 21.

Tabela 21 – Cálculo e fórmula da qualidade das dimensões do conjunto de dados

Avaliação geral da dimensão =	<u>Pesos atribuídos x 100</u> Peso máximo

Ex.: Se uma dimensão possui três métricas:

Peso máximo: 10+10+10= 30

Soma realizada: Se as métricas foram avaliadas como Excelente (10), Aceitável (5) e Crítico (0), a soma é 10 + 5 + 0 = 15.

$$\text{Fórmula: } \frac{15}{30} \times 100 = 50 \%$$

Fonte: Elaborado pela Autora (2026)

Com base no resultado percentual obtido, a dimensão é classificada conforme os critérios de qualidade estabelecidos no Quadro 88.

Quadro 88 – Critérios de avaliação de cada dimensão

Nível de qualidade	Cobertura	Descrição
Excelente	90% a 100%	Quase todas as métricas apresentam os requisitos necessários de qualidade.
Aceitável	50% a 89%	As métricas que compõe a dimensão cumprem com a maioria dos requisitos de qualidade, porém, apresentam falhas notáveis.
Crítico	Abaixo de 49%	As métricas estão em desconformidade com uma parte significativa dos requisitos.

Fonte: Elaborado pela Autora (2023).

Para a mensuração geral do conjunto de dados, foi consolidada por meio da média ponderada de todas as dimensões avaliadas. Soma-se o peso máximo de todas as dimensões (o valor que seria obtido se tudo fosse peso dez). Em seguida soma-se o valor real obtido em cada uma das dimensões.

O cálculo utilizado está representado na Tabela 22.

Tabela 22 – Cálculo e fórmula para avaliação geral do conjunto de dados

Avaliação geral do conjunto de dados =	<u>Soma do peso total das dimensões x 100</u> Soma da avaliação real das dimensões

Ex.: Foram avaliadas 15 dimensões onde a soma do peso máximo foi de: 570

Soma realizada: Após a avaliação, a soma real das dimensões foi de: 265

$$\text{Fórmula: } \frac{265}{570} \times 100 = 46 \%$$

Fonte: Elaborada pela Autora (2026).

A classificação geral do conjunto de dados segue os critérios de avaliação apresentados do Quadro 89.

Quadro 89 – Critérios de avaliação geral do conjunto de dados.

Nível de qualidade	Cobertura	Descrição
Excelente	90% a 100%	O conjunto de dados apresenta os requisitos necessários de qualidade.
Aceitável	50% a 89%	O conjunto de dados apresenta falhas notáveis.
Crítico	Abaixo de 49%	O conjunto de dados está desconformidade com os requisitos de qualidade.

Fonte: Elaborado pela Autora (2026).

O instrumento possibilitará mensurar a contribuição geral de cada dimensão na avaliação. Para o cálculo foi atribuído o mesmo peso para todas as dimensões, considerando que todas as dimensões têm a mesma importância para a qualidade dos dados.

Para o cálculo será utilizado a Tabela 23.

Tabela 23 - Cálculo e fórmula para avaliação do peso efetivo de cada dimensão

Peso efetivo da dimensão =	<u>Peso atribuído x 100</u> Peso efetivo da dimensão

Ex.: À dimensão Completude foi atribuído peso: 1

Desempenho da dimensão: Repete-se o valor encontrado no cálculo da tabela 21. Como exemplo utiliza-se o valor 50.

Peso efetivo da dimensão: O resultado será a divisão do desempenho da dimensão pelo seu peso:

$$\text{Fórmula: } \frac{50}{1 \times 100} = 0,50$$

Fonte: Elaborado pela Autora (2026).

Com o resultado do peso efetivo que a dimensão apresenta na avaliação da qualidade dos dados, é possível encontrar a contribuição de cada dimensão no índice geral da avaliação.

O cálculo utilizado está representado na Tabela 24.

Tabela 24 – Cálculo e fórmula para avaliação da contribuição da dimensão do índice geral

Contribuição da dimensão no índice geral =	<u>Peso efetivo da dimensão x 100</u> Peso máximo das dimensões

Ex.: A dimensão Completude possui peso efetivo de: 0,50

Peso máximo: O peso máximo de todas as dimensões é de: 15

$$\text{Fórmula: } \frac{0,50 \times 100}{15} = 3 \%$$

Fonte: Elaborado pela Autora (2026).

Conforme o cálculo apresentado na Tabela 24, a contribuição geral da dimensão Completude na avaliação geral da qualidade do conjunto dos dados é de 3%. A soma da contribuição de todas as dimensões corresponde ao índice da qualidade dos dados que seguirá os critérios de avaliação do Quadro 88.

Diante das métricas utilizadas para mensurar a qualidade dos conjuntos de DGA, propõe-se o desenvolvimento de um modelo de instrumento de avaliação que reúna as dimensões e métricas apresentadas anteriormente. O instrumento foi idealizado para orientar a análise e mensuração da QD, permitindo identificar pontos fortes e fragilidades dos dados públicos. Sua estrutura apresenta critérios qualitativos e quantitativos, organizados de maneira a possibilitar a avaliação em diferentes contextos institucionais.

A seguir, é apresentada a estrutura do instrumento proposto, que consolida dimensões, métricas, critérios e escalas de qualificação adotadas nesta pesquisa, constituindo-se como base para a aplicação da Prova de Conceito.

Devido à extensão e à complexidade da base de dados, estruturada em planilhas dinâmicas com múltiplas abas, o arquivo completo encontra-se disponível digitalmente para consulta. O material pode ser acessado no repositório *Google Drive* através do *hiperlink*:

<https://docs.google.com/spreadsheets/d/16vB9NmZ7oitTSWNPp8t3HzxlseDcFOaK/edit?usp=sharing&ouid=115836244919798871489&rtpof=true&sd=true>.

A disponibilização neste formato visa garantir a transparência e permitir a navegabilidade entre as diferentes categorias de análise.

7 PROVA DE CONCEITO (PoC)

Esta seção apresenta a Prova de Conceito desenvolvida como etapa prática, aplicada e exploratória para validação do instrumento desenvolvido para avaliar a qualidade dos GDA. O objetivo desta fase é demonstrar a viabilidade da metodologia proposta, bem como das dimensões e métricas de qualidades selecionadas, verificando se eles se comportam de forma consistente quando aplicadas em um ambiente real de dados públicos. Nessa etapa, busca-se confirmar se cada métrica, fundamentada conceitualmente na etapa teórica, é capaz de mensurar na prática aspectos de qualidade a que se propõe.

A PoC consiste em um modelo experimental que permite demonstrar a viabilidade técnica de uma ideia proposta por uma pesquisa. De acordo com Silva (2014), trata-se de uma ferramenta para validar conceitos de forma prática. Na mesma linha, Gaspar (2019, p. 18-9), define PoC como um mecanismo que “[...] permite demonstrar que uma ideia é tecnicamente possível, ou seja, pode ajudar a verificar se uma arquitetura é construível”. No campo da TI, a PoC é amplamente utilizada na criação de protótipos, funcionando como recurso de experimentação para testar a viabilidade técnica de projeto e arquiteturas (Silva, 2014).

No contexto desta pesquisa, a PoC tem como a avaliação a metodologia e o instrumento construído para mensurar a qualidade dos dados abertos governamentais. O uso de métricas na avaliação da qualidade favorece o enquadramento de indicadores que permitem obter uma melhor avaliação.

Com base nos fundamentos apresentados, esta prova de conceito foi aplicada no conjunto de dados do Instituto Nacional de Estudos e Pesquisas Educacionais - “Indicadores Educacionais da Educação Básica - Taxa de Distorção Idade-Série por Localização e Dependência Administrativa, nos Níveis de Ensino Fundamental e Médio - Brasil, Regiões Geográficas e Unidades da Federação - 2024”, disponível no Portal Dados Abertos, com o objetivo de testar a aplicabilidade prática das dimensões e métricas propostas para a avaliação da qualidade dos dados governamentais abertos.

A escolha pelo conjunto de dados disponibilizado pelo Instituto Nacional de Estudos e Pesquisas Educacionais - “Indicadores Educacionais da Educação Básica”, justifica-se por diversos fatores. Em primeiro lugar, trata-se de um órgão que confere credibilidade, autenticidade e confiabilidade às informações.

Além disso, o conjunto de dados apresenta ampla relevância social e acadêmica, uma vez que reúne informações fundamentais para o acompanhamento e a formulação de políticas públicas na área educacional. O que evidencia a importância de se assegurar a qualidade, acessibilidade e reuso desses dados.

Outro aspecto importante que motivou a escolha foi a abrangência e regularidade do conjunto de dados, atualizado a partir do Censo Escolar da Educação Básica, um levantamento completo da educação no país. Característica que possibilita uma avaliação consistente das dimensões e métricas de qualidade propostas nesta pesquisa, já que oferece dados volumosos e de interesse público.

A escolha do conjunto de dados do INEP permite testar a aplicabilidade das dimensões e métricas em um cenário real, complexo e representativo, garantindo que a PoC se desenvolva sobre um objeto de análise de relevância nacional e de acordo com os princípios de transparência pública e de dados governamentais abertos.

Dessa forma, será possível verificar a utilidade e a consistência do modelo proposto quando confrontado com dados concretos disponibilizados por portais de dados governamentais abertos. Para avaliação dos resultados, as métricas foram categorizadas nas classificações: **Excelente, Aceitável e Crítico**.

A aplicação da PoC no conjunto de dados do Instituto Nacional de Estudos e Pesquisas Educacionais - “Indicadores Educacionais da Educação Básica - Taxa de Distorção Idade-Série por Localização e Dependência Administrativa, nos Níveis de Ensino Fundamental e Médio - Brasil, Regiões Geográficas e Unidades da Federação - 2024”, disponibilizado no Portal de Dados Abertos e em seu próprio portal institucional, permitiu verificar, em ambiente real, a operacionalidade das métricas propostas. A íntegra dos dados resultados da PoC pode ser consultado no repositório Google Drive disponível em : https://docs.google.com/spreadsheets/d/1zZgaVpn6oVsSdYLDw6mi1zmgY_IRPluV/edit?usp=sharing&oid=115836244919798871489&rtpof=true&sd=true.

Com base nos resultados obtidos por meio da aplicação da PoC, constatou-se que, das 57 métricas avaliadas, 30 (52,63%) foram classificadas como “excelente”, 17 (29,82%) como “aceitável” e 10 (17,54%) como “crítico”. Esse resultado de distribuição evidencia um desempenho predominantemente positivo dos conjuntos de dados analisados, contudo, existem aspectos importantes que demandam aprimoramentos.

7.1 Análise dos resultados da avaliação

A aplicação da PoC permitiu verificar que o conjunto de dados “Indicadores Educacionais da Educação Básica - Taxa de Distorção Idade-Série por Localização e Dependência Administrativa, nos Níveis de Ensino Fundamental e Médio - Brasil, Regiões Geográficas e Unidades da Federação - 2024”, demonstrou desempenho satisfatório nas dimensões estruturais e técnicas. Observou-se elevada conformidade nas métricas associadas às dimensões precisão, estabilidade, não discricionariedade e baixos índices de inconsistência.

Em contrapartida, foram identificadas fragilidades significativas nas dimensões de acessibilidade, atualidade e compreensibilidade. Estes resultados evidenciam lacunas na interface com o usuário final e sinalizam a necessidade de ajustes e melhorias nos processos de documentação e atualização para garantir a efetiva qualidade e uso dos dados.

O Quadro 90 apresenta, de forma detalhada, as métricas que obtiveram classificação excelente, representando os melhores desempenhos na avaliação.

Quadro 90 – Síntese das métricas com qualificação excelente

Dimensões	Métricas
Compleitude	Campos obrigatórios. Comparabilidade. Metadados. Distribuição temporal.
Atualidade	Idade dos dados; Atualização de metadados.
Formato de arquivo	Formato dos dados acessíveis.
Disponibilidade	Frequência dos períodos indisponíveis.
Precisão	Taxa de dados corrompidos; Dados duplicados; Grau de correspondência dos dados com a realidade; Ausência de ambiguidades.
Recuperabilidade	Grau de eficiência dos processos e sistemas de recuperação; Metadados recuperáveis.
Inconsistência	Coerência lógica; Consistência entre conjuntos de dados; Coerência entre dados agregados; Uso de classificações.

Compreensibilidade	Consistência na linguagem.
Conformidade	Proteção de dados pessoais conforme legislação vigente.
Licença aberta/não proprietária	Tipo de licença; Permissão para reutilização; Período de validade para uso dos dados; Ausência de restrições.
Usabilidade	Dados disponíveis para acesso e <i>download</i> direto; Interface amigável.
Não discricionário	Dados disponíveis diretamente ao usuário; Custo de Acesso.
Primário	Fonte de dados.
Interoperabilidade	Capacidade de reutilização.

Fonte: Elaborado pela Autora (2026)

As métricas classificadas como **Excelente** evidenciam que o conjunto de dados do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP) apresenta elevado grau de maturidade informacional e técnica. O resultado reflete a adoção de boas práticas de gestão, governança e padronização dos dados públicos, denotando a existência de um ambiente institucional comprometido com a qualidade, a confiabilidade e a transparência dos DGA.

Dentre as métricas com melhor desempenho destacam-se: campos obrigatórios, metadados, formato dos dados acessíveis, multiplicidade de formatos, taxa de dados corrompidos, dados duplicados, coerência lógica, consistência entre os conjuntos de dados, valores padronizados e capacidade de reutilização. A excelente avaliação dessas métricas demonstra que o conjunto de dados possui integridade estrutural, padronização semântica, assegurando o correto funcionamento e a interoperabilidade entre sistemas. Além disso, a ausência de duplicidade ou erros lógicos e de dados corrompidos, confirma o compromisso do órgão com a acurácia informacional, facilitando a reprodutibilidade e a efetiva reutilização das informações pela sociedade.

Do ponto de vista jurídico-normativo, as métricas referentes à proteção de dados pessoais, tipo de licença, permissão para reutilização, período de validade para uso dos dados, ausência de restrições e declaração de domínio reforçam a conformidade do conjunto de dados com os princípios da abertura e não

discricionariedade. Estes elementos estão previstos tanto na legislação nacional quanto nas diretrizes internacionais de dados abertos. A adoção de licença aberta, pública e irrevogável, garante segurança jurídica aos usuários, promovendo a previsibilidade e a continuidade no uso, reúso e redistribuição das informações públicas.

No que se refere à eficiência informacional e a usabilidade dos dados, métricas como: comparabilidade, distribuição temporal, metadados recuperáveis, interface amigável, dados disponíveis para download direto, não discricionário, custo de acesso, indicam que o Portal de Dados Abertos proporciona ao usuário uma experiência de navegação eficiente. As métricas coerência lógica, uso de classificações e consistência entre conjuntos de dados reforçam a existência de uma governança institucional consolidada.

No entanto, ainda que os resultados sejam amplamente positivos, determinadas métricas demandam aprimoramento contínuo, de modo a garantir acesso equitativo e compreensível aos dados públicos. O Quadro 91 apresenta as métricas que obtiveram qualificação aceitável, as quais, embora satisfatórias, indicam potencial de melhoria para o fortalecimento da qualidade do conjunto de dados analisados.

Quadro 91 – Síntese das métricas com qualificação aceitável

Dimensões	Métricas
Compleitude	Valores ausentes; Cobertura dos dados coletados.
Formato de arquivo	Multiplicidade de formatos; Dados disponíveis em formato compactado; Documentação do formato.
Disponibilidade	<i>Hiperlink</i> de <i>download</i> funcional; Disponibilidade de canal de comunicação com o usuário.
Recuperabilidade	Função de recuperação de dados; Taxa de falhas na recuperação de dados.
Compreensibilidade	Presença de dicionários, notas explicativas ou legendas; Descrição dos recursos.
Conformidade	Origem dos dados.
Licença aberta/não proprietária	Existência de licença Declaração de domínio público

Primário	Processo dos dados documentais; Rastreabilidade dos dados .
Interoperabilidade	Integração em sistemas externos.

Fonte: Elaborado pela Autora (2026).

As métricas classificadas como aceitáveis representam um nível intermediário de qualidade no contexto do Portal de Dados Abertos. O resultado indicando que, embora 52,63% das métricas avaliadas cumprem com os requisitos essenciais de abertura e acesso, o conjunto de dados do INEP ainda apresenta limitações técnicas que podem comprometer o reuso dos dados. Esse resultado indica a necessidade de aperfeiçoamentos estruturais e documentais para que o conjunto alcance níveis de excelência.

As métricas, valores ausentes e cobertura dos dados coletados, pertencentes à dimensão completude, revelam que, embora a maior parte dos campos estejam devidamente preenchidos, existem lacunas que podem comprometer a análise e a confiabilidade dos dados. A ausência parcial de dados reduz a precisão das inferências possíveis, o que impacta diretamente o potencial de uso e reuso dos dados. Segundo Bolla e Vuyyuru (2024, p. 8), “[...] quando algumas partes dos dados estão faltando, podem existir lacunas que podem levar a informações inadequadas ou a resultados incorretos”.

Em relação à dimensão formato de arquivo, observou-se que, o conjunto de dados é disponibilizado em formato aberto e acessível, em conformidade com os princípios dos dados governamentais abertos, porém, a documentação técnica é limitada, com descrições pouco detalhadas sobre os processos de compactação e estruturação dos dados. Essa limitação dificulta o entendimento pleno da organização interna dos dados e pode comprometer a análise por terceiros.

Na dimensão disponibilidade, foram constatadas falhas nas métricas *hiperlink* de *download* funcional e disponibilização de canal de comunicação com o usuário, devido à presença de *hiperlinks* inoperantes e limitações no suporte ao usuário. Essas fragilidades reduzem não só a eficiência do acesso como podem gerar barreiras no relacionamento entre o portal e os usuários, contrariando o princípio da acessibilidade da informação pública.

As métricas de recuperação de dados, eficiência de busca e taxa de falhas na recuperação de dados demonstram que os mecanismos de busca do portal são

operacionais, porém limitados, exigindo múltiplas etapas para localização e para o acesso ao conjunto de dados. Esse cenário implica uma redução na eficiência da recuperação da informação. Este cenário, segundo Moreira *et al.* (2016, p. 75), compromete a correspondência entre a necessidade do usuário e os resultados retornados, uma vez que “[...] na recuperação de dados busca-se por uma correspondência exata à necessidade”.

A dimensão compreensibilidade também apresentou fragilidades, especialmente quanto à ausência de dicionários, notas explicativas ou legendas. A inexistência desses recursos reduz a capacidade de compreensão do conjunto por diferentes perfis de usuários e limita o potencial de reuso dos dados.

A métrica “rastreabilidade” mostrou-se parcialmente atendida. Embora a origem dos dados seja identificável, o processo de coleta, transformação e publicação não está claramente documentado. A ausência de documentação detalhada impede o rastreamento completo do fluxo de produção da informação, comprometendo a transparência e a confiabilidade institucional. Conforme Mahanti (2019), a rastreabilidade exerce impacto positivo na qualidade dos dados, pois permite aos usuários rastrear os dados até sua origem e fornece informações sobre a introdução e modificação dos registros dos dados, incentivando os usuários a confiar neles.

A integração com sistemas externos revela que o conjunto de dados não dispõe de uma API pública que viabilize o acesso automatizado aos indicadores educacionais, restringindo a interoperabilidade. A integração com outras plataformas depende do *download* manual dos arquivos, o que limita o reuso dos dados e evidencia a necessidade de aprimoramentos técnicos para permitir integração automatizada e em tempo real.

De forma geral, as métricas avaliadas como **Aceitáveis** demonstram que o conjunto de dados do INEP apresenta estrutura funcional consistente, ainda apresenta falhas nos requisitos de documentação, rastreabilidade e comunicação com o usuário. As lacunas informacionais e processuais restringem a experiência do usuário e o acesso total aos dados.

A análise demonstrou que o conjunto de dados do INEP – Taxa de Distorção Idade-Série apresenta métricas avaliadas como Críticas, evidenciando fragilidades significativas em dimensões essenciais da qualidade dos dados. O Quadro 92

sintetiza as dimensões prejudicadas e suas respectivas métricas avaliadas com nível de qualidade Crítico.

Quadro 92 – Síntese das métricas com qualificação crítico

Dimensões	Métricas
Acessibilidade	Conformidade com normas de acessibilidade.
Atualidade	Frequência de atualização; Tempo entre a ocorrência do evento e o do registro no sistema; Presença de data e hora da coleta dos dados.
Disponibilidade	Múltiplos idiomas.
Recuperabilidade	Navegação; Vocabulários controlados.
Compreensibilidade	Descrições claras e detalhadas dos dados. Linguagem adequada a diferentes públicos.
Conformidade	Responsável em caso de violação de direitos ou uso inadequado dos dados.

Fonte: Elaborado pela Autora (2026).

O desempenho classificado como **Crítico** na métrica conformidade com as normas de acessibilidade indicam que o conjunto de dados não atendem integralmente às legislações e normas técnicas vigentes relacionadas à acessibilidade digital. Foram identificadas falhas significativas, tais como: ausência de texto alternativo para imagens e ícones, contraste de cores inadequadas, navegação por teclado ineficiente e ausência de legendas ou tradução em libras. Esses aspectos justificam a classificação crítica atribuída, indicando que o portal apresenta barreiras que comprometem a acessibilidade plena de seus conteúdos, excluindo uma parcela significativa da população.

Essas deficiências restringem o exercício do direito fundamental de acesso à informação, em especial por pessoas com deficiência, contrariando os princípios de acessibilidade universal estabelecidos pela Lei Brasileira de Inclusão (Lei n.º 13.146/2015) e pelas Diretrizes de Acessibilidade para conteúdo da *Web* (WCAG). Conforme observam Batini e Scannapieco (2012, p. 34), um usuário precisa não apenas acessar a rede, mas também compreender a linguagem de navegação e ser capaz de consultar a web e perceber com seus sentidos as informações disponibilizadas.

No campo da dimensão atualidade, as métricas referentes à: frequência de atualização; ao tempo entre a ocorrência do evento e o do registro no sistema e a presença de data e hora da coleta dos dados também foram avaliados como críticos.

A análise demonstrou que o conjunto de dados não apresenta informações claras e consistentes sobre a data de coleta, periodicidade das atualizações e histórico de versões. A ausência dessas informações inviabiliza a verificação da atualidade e a confiabilidade temporal dos dados disponibilizados. Essa lacuna dificulta a rastreabilidade temporal, impedindo que o usuário identifique qual a versão mais recente dos dados.

Quanto às métricas da dimensão compreensibilidade, observou-se o uso de terminologia excessivamente técnica, com ausência de explicações adequadas sobre o significado das variáveis, indicadores e categorias presentes no conjunto de dados. Essa limitação reduz o potencial de uso e reúso dos dados públicos. Como definem Zaveri *et al.* (2012, p. 21), a compreensibilidade é a “[...] extensão em que todos os dados são facilmente compreendidos pelo consumidor de informação”.

Outro ponto crítico refere-se à ausência de informações sobre a responsabilidade institucional em casos de violação de direitos ou uso inadequado dos dados, bem como à falta de canais de denúncia em desconformidade com os preceitos da LGPD. Essa omissão fragiliza os mecanismos de governança e responsabilização, elementos fundamentais para a confiança e a segurança jurídica no ambiente digital.

As métricas múltiplos idiomas e vocabulários controlados também apresentam desempenhos insatisfatórios. Constatou-se que o conjunto de dados é disponibilizado exclusivamente em português e utiliza vocabulário institucional restrito, o que limita a interoperabilidade semântica. Essa limitação dificulta o cruzamento de dados com outros sistemas e o uso dos dados por usuários estrangeiros. A ausência do uso de ontologias comuns restringe o alcance dos dados.

De modo geral, as métricas classificadas como críticas apontam falhas estruturais e funcionais que, embora não comprometam totalmente a abertura dos dados, limitam seu alcance, usabilidade e equidade de acesso.

O desempenho insatisfatório em dimensões como acessibilidade, atualidade, disponibilidade, recuperabilidade, compreensibilidade e conformidade demonstra

que o INEP além do foco nas questões técnicas e documentais, deve observar também as questões centradas no usuário e na equidade de acesso.

Ao comparar os resultados obtidos na PoC com o modelo proposto por Tim Berners-Lee constata-se que, o conjunto de dados do INEP alcança somente três estrelas da escala.

A elevação para os níveis da quarta e quinta estrelas são inviabilizadas pela falta de recursos de *URIs* e *Hiperlinks* para outros dados abertos vinculados. Apesar dos dados serem abertos, sem esses recursos o conjunto de dados permanece isolado em um sítio informacional quando poderia estar se conectando automaticamente a outros portais governamentais.

O Quadro 93 apresenta sugestões para melhorias das dimensões que obtiveram avaliação aceitável e crítico.

Quadro 93 – Sugestões para melhoria da avaliação das métricas

Dimensões	Sugestões
Acessibilidade	Garantir a aplicação da Lei Brasileira de Inclusão e implementação das diretrizes WCAG/e-MAG com aplicação de testes nas interfaces, eliminando as barreiras de acesso por pessoas com deficiência.
Compleitude	Implementar ferramentas de verificação para que o conjunto de dados não apresente células com vazias; Melhorar a cobertura dos dados coletados.
Atualidade	Incluir no conjunto de dados as informações referentes a data da coleta, data da publicação e a data da última atualização, necessárias para a verificação da atualidade dos dados.
Formato de arquivo	Disponibilizar o conjunto de dados em diversos formatos no mínimo três, aumentando a disponibilidade dos dados; Melhorar a compactação dos conjuntos de dados; Disponibilizar a documentação dos formatos disponibilizados em local visível ao usuário.
Disponibilidade	Disponibilizar o conjunto de dados ou no mínimo os metadados essenciais em outra língua que não o vernáculo oficial do país. Realizar checagem mensal da funcionalidade dos <i>hiperlinks</i> para correção; Disponibilização de canais de comunicação com o usuário.
Recuperabilidade	Utilizar vocabulários e ontologias globais ao invés de nomenclaturas internas.
Compreensibilidade	Adicionar glossários, notas explicativas e legendas para explicar o significado de termos técnicos, regionais, estrangeiros ou de difícil

	compreensão facilitando o entendimento; Ao inserir unidades de medidas ou codificações, incluir descrições detalhadas das mesmas.
Conformidade	Indicação e publicação em local visível informação sobre controlador, operador e encarregado responsáveis pelo tratamento dos dados pessoais e um canal de denúncia, conforme Lei n.º 13.709/2018 – LGPD.
Licença aberta/não proprietária	Disponibilizar em local de fácil acesso documentação sobre a licença utilizada e a declaração de domínio público.
Primário	Inserir documentos em local de fácil acesso sobre o processo de coleta, tratamento, validação e publicação dos dados; Tornar os dados rastreáveis para que o usuário consiga acompanhar o histórico dos dados, desde a coleta até a sua publicação.
Interoperabilidade	Utilização ferramentas/recursos para integrar o conjunto de dados a outras plataformas.

Fonte: Elaborado pela Autora (2026).

Ao analisar o quadro, é possível perceber que pequenos ajustes na disponibilização dos conjuntos de dados elevam significativamente sua qualidade. Essas melhorias transformam uma simples exposição de dados, convertendo um conjunto isolado em um ecossistema de conhecimento possibilitando a transparência, o uso e reúso dos dados.

8 CONSIDERAÇÕES FINAIS

As TICs integraram-se definitivamente ao cotidiano da sociedade, gerando um fluxo contínuo de dados acessíveis globalmente. No entanto, como evidenciado nesta pesquisa, a simples publicação dos dados não assegura a transparência, uso ou o reúso efetivo. A qualidade dos dados emerge como o fator crítico, frequentemente subestimado. Disponibilizar dados com qualidade ultrapassa a área técnica, devendo estar alinhada à responsabilidade ética, planejamento cuidadoso e manutenção contínua. Nesse contexto, um dos maiores desafios para os profissionais da informação reside na definição de dimensões e métricas para avaliação da qualidade dos dados, que acompanhem a evolução tecnológica e a crescente dependência informacional da sociedade.

A presente pesquisa identificou a escassez de materiais técnicos e acadêmicos dedicados aos critérios de QD, especificamente no que tange à definição de dimensões e métricas a serem observadas para a construção de um conjunto de dados de qualidade. Observou-se que as referências sobre o tema são fragmentadas, sendo necessário um esforço de síntese para estabelecer um consenso sobre os atributos essenciais para QD.

Diante disso, surgiu o problema da pesquisa: Como estabelecer parâmetros de avaliação da qualidade de DGA, com base em normas e legislações para o acesso, uso e reúso pela sociedade?

A abordagem qualitativa e de natureza aplicada, sustentada pela Revisão de Literatura e Pesquisa Documental, permitiu um aprofundamento na trajetória do governo digital. O referencial teórico possibilitou confirmar que a qualidade dos dados é um atributo fundamental e primário na disponibilização dos dados governamentais abertos.

Complementarmente, a RSL permitiu identificar e sintetizar o estado da arte sobre o tema. Todos os trabalhos científicos selecionados na RSL apresentavam o consenso de que, a mensuração da qualidade de dados, irá depender da definição de dimensões e suas respectivas métricas.

A RSL identificou inicialmente quarenta e uma dimensões distintas utilizadas para avaliar a qualidade dos dados. Por meio de uma análise comparativa dessas dimensões este estudo promoveu a compilação e consolidação de quinze dimensões fundamentais, selecionadas por sua abrangência, representatividade e

capacidade de caracterizar e mensurar com clareza os conjuntos de DGA: acessibilidade, completude, atualidade, formato de arquivo, disponibilidade, precisão, recuperabilidade, inconsistência, compreensibilidade, conformidade, licença aberta, usabilidade, não discricionário, primário e interoperabilidade.

O processo de síntese dessas dimensões exigiu uma leitura crítica, uma vez que a literatura especializada carece de consenso terminológico e conceitual. Observou-se que diferentes autores frequentemente utilizam nomenclaturas distintas para dimensões idênticas, o que demandou esforço para a padronização por parte da pesquisa.

Ademais, constatou-se uma lacuna nos trabalhos selecionados, pois em nenhum deles foram encontradas métricas associadas a essas dimensões. Diante disso, foi necessário um estudo detalhado de cada uma das dimensões, bem como de normas, diretrizes e legislações vigentes, para a seleção das métricas. Esse procedimento viabilizou a derivação de métricas específicas para cada dimensão.

A observância dessas dimensões e métricas é imperativa, pois, dados de baixa qualidade podem gerar modelos tendenciosos e falhas de divulgação, resultando em implicações críticas em aplicações do mundo real.

Contudo, para a mensuração da QD se faz necessário instrumentos capazes de avaliar sistematicamente os conjuntos de dados. Diante da escassez de instrumentos padronizados, esta pesquisa propôs um modelo de mensuração onde métricas foram vinculadas às dimensões, respeitando as suas naturezas técnicas e funcionais.

Diante da lacuna teórica na literatura de QD, foi necessária a elaboração própria das fórmulas para mensuração das métricas, respeitando as especificidades técnicas de cada uma. O desenvolvimento dessas equações demandou a realização de diversos testes até a consolidação dos modelos propostos.

Para a mensuração, optou-se pela escala de avaliação: Excelente, Aceitável e Crítico, por permitir uma classificação clara e intuitiva dos níveis de conformidade identificados. Para cada estágio da escala, estabeleceram-se índices de cobertura específicos que delimitam o enquadramento dos dados, assegurando o rigor metodológico na categorização dos resultados. Tais índices são variáveis, ajustando-se às particularidades de cada dimensão e métrica, o que confere ao instrumento a flexibilidade necessária para avaliar a heterogeneidade dos dados

governamentais abertos. Foram realizadas pesquisas relacionadas aos índices a serem utilizados, que serviram de fundamentação para a estrutura adotada.

A cada índice de cobertura foi atribuída uma descrição, oferecendo ao usuário maior clareza sobre os critérios de avaliação e facilitando a interpretação dos níveis de qualidade alcançados. Para o cálculo de determinadas métricas, aplicaram-se fórmulas específicas para se chegar ao valor de cobertura. Definidos os critérios de avaliação de cada uma das dimensões e métricas, passou-se à construção do instrumento em planilhas eletrônicas, visando à organização e análise dos dados. O arquivo é composto por dezessete abas, sendo uma para cada dimensão, além de um painel consolidado e um *dashoboard*.

O instrumento foi elaborado não somente para os profissionais da área, mas para todos os usuários não especialistas que desejam verificar a qualidade dos conjuntos de dados que utilizam ou pretendem utilizar. As informações essenciais de cada dimensão e métricas apresentadas no capítulo 5 foram integradas ao instrumento, dispensando o usuário da consulta recorrente ao texto da tese durante a realização da verificação. Também, para cada métrica foi elaborado uma questão para auxiliar o usuário em sua respectiva análise.

As dimensões e métricas para a qualidade dos conjuntos de dados abertos não se esgotam nas quinze dimensões apresentadas nesta pesquisa, sua observação assegura, tanto ao órgão provedor quanto ao usuário, que os dados atendem requisitos mínimos de qualidade.

A viabilidade da proposta foi validada por meio de uma PoC aplicada no conjunto de dados Taxa de Distorção Idade-Série na Educação Básica - Indicadores Educacionais da Educação Básica do Instituto Nacional de estudos e Pesquisas Educacionais Anísio Teixeira. Os resultados da PoC demonstraram que a aplicação sistemática dessas quinze dimensões e suas métricas específicas fornece um diagnóstico preciso e replicável sobre a qualidade dos dados governamentais abertos. Desta forma, ao utilizar o instrumento de mensuração proposto, torna-se possível verificar individualmente em quais aspectos cada dimensão e métrica necessita de correções pontuais. Essa capacidade de diagnosticar é fundamental para orientar intervenções precisas, permitindo elevar de forma significativa a qualidade dos dados governamentais abertos de forma eficiente, garantindo que a transparência evolua de uma disponibilidade passiva para uma transparência ativa.

A pesquisa demonstrou a viabilidade de se construir conjuntos de dados que atendam a rigorosos atributos de qualidade. O instrumento de avaliação proposto destaca-se por sua versatilidade e fácil manuseio, ultrapassando o uso exclusivamente técnico, pois foi concebido para ser manuseado tanto por profissionais da área, quanto por usuários comuns. Essa característica é fundamental para fomentar o controle social e garantir que os Dados Governamentais Abertos cumpram seu papel de forma plena e acessível.

Embora o Governo Federal tenha instituído a obrigatoriedade, para as entidades da administração federal direta, autárquica e fundacional, a elaboração do PDA, o foco atual permanece em etapas processuais e administrativas. A literatura destaca a relevância das dimensões de qualidade, nenhuma das etapas para a elaboração do PDA, recomenda-se sua adoção, evidenciando que apesar dos esforços governamentais na abertura dos dados, ainda há um longo caminho a percorrer quanto à qualidade dos dados publicados.

Diante da crescente demanda por dados públicos, a inclusão de novas etapas na elaboração de um PDA torna-se essencial, especialmente no que tange à definição de dimensões e métricas de qualidade. A inclusão dessa etapa permite que as discussões sobre dados governamentais abertos possam avançar para além da simples disponibilização, promovendo a padronização dos *datasets* e garantindo que a informação seja confiável e útil para a sociedade.

Considerando o protagonismo dos dados públicos na sociedade, sugerem-se, como pesquisas futuras, o desenvolvimento de algoritmos e ferramentas de automação que integrem as dimensões e métricas propostas neste estudo. Esse avanço permitiria o monitoramento em tempo real da qualidade de grandes volumes de dados governamentais abertos, reduzindo esforços manuais e agilizando as correções necessárias.

REFERÊNCIAS

ABATH, Marcele de Brito; LIMA, Maria Luiza Lopes Timóteo de; LIMA, Priscila de Souza; SILVA, Maria Carmelita Maia e; LIMA, Maria Luiza Carvalho de. Avaliação da completude, da consistência e da duplicidade de registros de violências do Sinan em Recife, Pernambuco, 2009-2012. **Epidemiologia e Serviços de Saúde**, São Paulo, SP, v. 23, n. 1, p. 131-142, jan. 2014. Disponível em: <https://www.scielo.br/j/ress/a/fyDMwnQ7LvKkKdrd7MqhzfNq/abstract/?lang=pt#https://www.scielo.br/j/ress/a/fyDMwnQ7LvKkKdrd7MqhzfNq/abstract/?lang=pt#>. Acesso em: 25 jul. 2024.

ABELS, Eileen; JONES, Rebecca; LATHAM Jonh; MAGNONI Dee; MARSHALL, Joanne Gard. Competencies for Information Professionals of the 21st Century. Special Libraries Association. 2003. Disponível em: <https://dbiosla.org/Competencies%20for%20Information%20Professionals%20of%20the%2021st%20Century.pdf>. Acesso em: 13 de janeiro de 2025.

ADOBE ACROBAT. **O que é um arquivo XLS e como utilizá-lo**. 2025a. Disponível em: <https://www.adobe.com/br/acrobat/resources/document-files/spreadsheet-files/xls.html#o-que-%C3%A9-um-arquivo-xlshttps://www.adobe.com/br/acrobat/resources/document-files/spreadsheet-files/xls.html#o-que-%C3%A9-um-arquivo>. Acesso em: 27 fev. 2025.

ADOBE ACROBAT. **O que significa PDF?** 2025b. Disponível em: <https://www.adobe.com/br/acrobat/about-adobe-pdf.html>. Acesso em: 28 fev. 2025.

ATTARD, Judie; ORLANDI Fabrizio; SCERRI, Simon; AUER, Sören. A systematic review of open government data initiatives. **Government Information Quarterly**, University de Bonn, Alemanha v. 32, n. 4, p. 399-418, 2015. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S0740624X1500091X>. Acesso em: 27 mar. 2024.

BOLLA, Raja Ashok; VUYYURU, Hari Kiran. *Data quality in generative AI: ensuring reliability, fairness, and governance for AI-driven innovations*. [S. l.]: Independently published, 2024.

AVELINO, Daniel Pitangueira de; POMPEU, João Cláudio; FONSECA, Igor Ferraz da. **Democracia digital**: mapeamento das experiências em dados abertos, governo digital e ouvidorias públicas. Brasília, DF: Instituto de Pesquisa Econômica Aplicada (IPEA), 2021. (Série monográfica: texto para discussão, 2624).

ÁVILA, Thiago José Tavares; ARAÚJO, Anderson Henrique dos Santos; LEVINO, Natally de A. A importância das plataformas de informação para o apoio ao processo de planejamento: o caso do portal Alagoas em dados e informações. In: CONGRESSO BRASILEIRO DE GESTÃO DO CONHECIMENTO (KM BRASIL), 2012, São Paulo. **Anais** [...]. São Paulo, SP: Sociedade Brasileira de Gestão do Conhecimento, 2012. Disponível em: <https://sbgc.org.br/kmbrasil/>. Acesso em: 10 mar. 2025.

BANDEIRA, Judson Melo; ALCÂNTARA, Williams Lourenço de; BARBOSA SOBRINHO, Armando; ÁVILA, Thiago José Tavares; PINTO, Ig Ibert Bittencourt Santana; ISOTANI, Siji. Dados abertos conectados. *In*: ISOTANI, S.; BITTENCOURT, Ig Ibert. **Jornada de Atualização em tecnologia da informação**. Alagoas: EDUFAL, 2012.

BARBALHO, Fernando. **Dados abertos**: a necessidade de alinhamentos significativos. [S.l.]: Médium, 2018. Disponível em: <https://medium.com/tchiluanda/dados-abertos-a-necessidade-de-alinhamentos-de-significados-a00fb5293147https://medium.com/tchiluanda/dados-abertos-a-necessidade-de-alinhamentos-de-significados-a00fb5293147>. Acesso em: 12 out. 2025.

BARDIN, Laurence. **Análise de conteúdo**. Tradução Luís Antero Reto e Augusto Pinheiro. São Paulo, SP: Edições 70, 2016.

BATINI, Carlos; CAPPIELLO, Cinzia; FRANCALANCI, Chiara; MAURINO, Andrea. Methodologies for data quality assessment and improvement. **ACM Computing Surveys (CSUR)**, United States, v. 41, n. 3, 2009. Disponível em: <https://dl.acm.org/doi/10.1145/1541880.1541883>. Acesso em: 03 out. 2024.

BATINI, Carlos; BARONE, Daniele; GREGA, Simone; CABITZA, Federico. A data quality methodology for heterogeneous data. **International Journal of Database Management Systems**, [S.l.], v. 3, n. 1, p. 60-79, 2011. Disponível em: <https://www.airccse.org/journal/ijdms/papers/3111ijdms05.pdf>. Acesso em: 28 out. 2024.

BATINI, Carlos; SCANNAPIECO, Monica. **Data quality**: concepts, methodologies and techniques. Berling: Springer-Verlag, 2006.

BATISTA, P. Uma nova mentalidade de segurança da informação: por que focar na conformidade não é suficiente. **Linkedin**, 2023. Disponível: <https://www.linkedin.com/pulse/uma-nova-mentalidade-de-seguran%C3%A7a-da-informa%C3%A7%C3%A3o-por-que-batista/>. Acesso em: 26 jul. 2024.

BELLIX, Laila; GUIMARÃES, Caroline Burle S.; MACHADO, Jorge. Qual conceito de governo aberto? Uma aproximação aos seus princípios. *In*: CONGRESSO INTERNACIONAL EM GOBIERNO, ADMINISTRACIÓN Y POLÍTICAS PÚBLICAS (GIGAPP), 7., 2016, Madrid, Espanha. **Anales** [...]. Madrid, Espanha: GIGAPP, 2016. Disponível em: <https://www.gigapp.org/ewp/index.php/GIGAPP-EWP/article/view/43>. Acesso em: 25 nov. 2025.

BERNERS-LEE, Tim. **Linked data**. 2006. Disponível em: <https://www.w3.org/DesignIssues/LinkedData.html>. Acesso em: 13 mar. 2025.

BERNERS-LEE, Tim. **Putting Government data online**. 2024. Disponível em: <https://www.w3.org/DesignIssues/GovData.html>. Acesso em: 13 mar. 2025.

BEVAN, Nigel. Usability is quality of use: advances in human factors/ergonomics, **Elsevier**, Reino Unido, v. 20, p. 349-354, 1995. Disponível em:

<https://www.sciencedirect.com/science/article/abs/pii/S092126470680241>. Acesso em: 26 set. 2024.

BOVEE, Matthew; SRIVASTAVA, Rajenda P.; MAK, Brenda. A conceptual framework and belief-function approach to assessing overall information quality. **International Journal of Intelligent Systems**, Espanha, v. 18. n. 1, p. 311-328, 2003. Disponível em: https://www.researchgate.net/publication/220063339_A_Conceptual_Framework_and_Belief-Function_Approach_to_Assessing_Overall_Information_Quality. Acesso em: 29 jan. 2025.

BRASIL. [Constituição (1988)]. **Constituição da República Federativa do Brasil de 1988**. Brasília, DF: Diário Oficial da União, 1988.

BRASIL. Controladoria-Geral da União. **Manual de elaboração de Planos de Dados Abertos (PDAs)**. Brasília, DF: Controladoria-Geral da União, 2020a. Disponível em: https://repositorio.cgu.gov.br/bitstream/1/46702/5/manual_de_elaboracao_de_planos_de_dados_abertos_pdas.pdf. Acesso em: 19 mar. 2025.

BRASIL. Controladoria-Geral da União. **O que é Governo Aberto**. Brasília, DF: Controladoria-Geral da União, 2025a. Disponível em: <https://www.gov.br/cgu/pt-br/governo-aberto/governo-aberto-no-brasil/principios/o-que-e-governo-aberto>. Acesso em: 28 de nov. 2025.

BRASIL. Controladoria-Geral da União. **Painel de monitoramento de dados abertos**. Brasília, DF: Controladoria-Geral da União, 2025. Disponível em: <https://centralpaineis.cgu.gov.br/visualizar/dadosabertos>. Acesso em: 19 mar. 2025.

BRASIL. Controladoria-Geral da União. **Reúso de dados abertos: afinal, para que servem?** Brasília, DF: Controladoria-Geral da União, 15 de dezembro de 2022. Disponível em: <https://www.youtube.com/watch?v=4E3AxGwa5r0>. Acesso em: 28 mar. 2023.

BRASIL. **Decreto de 15 de setembro de 2011**. Institui o Plano de Ação Nacional sobre Governo Aberto e dá outras providências. Brasília, DF: Diário Oficial da União, 2011. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2011-2014/2011/dsn/dsn13117.htm. Acesso em: 19 mar. 2025.

BRASIL. **Decreto nº 10.160, de 09 de dezembro de 2019**. Institui a Política Nacional de Governo Aberto e o Comitê Interministerial de Governo Aberto. Brasília, DF: Diário Oficial da União, 2019a. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2019-2022/2019/decreto/d10160.htm. Acesso em: 19 mar. 2025.

BRASIL. **Decreto nº 10.332, de 28 de abril de 2020**. Institui a Estratégia de Governo Digital para o período de 2020 a 2022, no âmbito dos órgãos e das entidades da administração pública federal direta, autárquica e fundacional e dá outras providências. Brasília, DF: Diário Oficial da União, 2020a. Disponível em:

https://www.planalto.gov.br/ccivil_03/_Ato2019-2022/2020/Decreto/D10332.htm#art14. Acesso em: 19 mar. 2025.

BRASIL. **Decreto nº 12.308, de 11 de dezembro de 2024**. Institui o Comitê Interministerial para a Transformação Digital. Brasília, DF: Diário Oficial da União, 2024. Disponível em: https://www.planalto.gov.br/ccivil_03/_Ato2023-2026/2024/Decreto/D12308.htm. Acesso em: 19 mar. 2025.

BRASIL. **Decreto nº 6.666, de 27 de novembro de 2008**. Institui, no âmbito do Poder Executivo Federal, a Infraestrutura Nacional de Dados Espaciais - INDE, e dá outras providências. Brasília, DF: Diário Oficial da União, 2008. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2007-2010/2008/decreto/d6666.htm. Acesso em: 19 mar. 2025.

BRASIL. **Decreto nº 8.638, de 15 de janeiro de 2016**. Institui a política da Governança Digital no âmbito dos órgãos e das entidades da administração pública federal direta, autárquica e fundacional. Brasília, DF: Diário Oficial da União, 2016a. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2016/decreto/d8638.htm. Acesso em: 19 mar. 2025.

BRASIL. **Decreto nº 8.777, de 11 de maio de 2016**. Institui a Política de Dados Abertos do Poder Executivo Federal. Brasília, DF: Diário Oficial da União, 2016b. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2016/decreto/d8777.htm. Acesso em: 17 mar. 2025.

BRASIL. **Decreto nº 9.903, de 8 de julho de 2019**. Altera o Decreto n.º 8.777, de 11 de maio de 2016, que institui a Política de Dados Abertos do Poder Executivo Federal, para dispor sobre a gestão e os direitos de uso de dados abertos. Brasília, DF: Diário Oficial da União, 2019c. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2019-2022/2019/decreto/d9903.htm. Acesso em: 19 mar. 2025.

BRASIL. **Instrução Normativa nº 4, de 13 de abril de 2012**. Institui a Infraestrutura Nacional de Dados Abertos - INDA. Brasília, DF: Diário Oficial da União, 2012. Disponível em: https://www.gov.br/agricultura/pt-br/acesso-a-informacao/dadosabertos/arquivos-raiz/in04_2012.pdf. Acesso em: 19 mar. 2025.

BRASIL. **Lei Complementar nº 101, de 04 de maio de 2000**. Estabelece normas de finanças públicas voltadas para a responsabilidade na gestão fiscal e dá outras providências. Brasília, DF: Diário Oficial da União, 2000b. Disponível em: https://www.planalto.gov.br/ccivil_03/leis/lcp/lcp101.htm. Acesso em: 19 mar. 2025.

BRASIL. **Lei Complementar nº 131, de 27 de maio de 2009**. Acrescenta dispositivos à Lei Complementar no 101, de 4 de maio de 2000, que estabelece normas de finanças públicas voltadas para a responsabilidade na gestão fiscal e dá outras providências, a fim de determinar a disponibilização, em tempo real, de informações pormenorizadas sobre a execução orçamentária e financeira da União, dos Estados, do Distrito Federal e dos Municípios. Brasília, DF: Diário Oficial da União, 2009. Disponível em: https://www.planalto.gov.br/ccivil_03/leis/lcp/lcp131.htm. Acesso em: 19 mar. 2025.

BRASIL. **Lei nº 12.527, de 18 de novembro de 2011.** Regula o acesso a informações previsto no inciso XXXIII do art. 5º, no inciso II do § 3º do art. 37 e no § 2º do art. 216 da Constituição Federal; altera a Lei nº 8.112, de 11 de dezembro de 1990; revoga a Lei nº 11.111, de 5 de maio de 2005, e dispositivos da Lei nº 8.159, de 8 de janeiro de 1991; e dá outras providências. Brasília, DF: Diário Oficial da União, 2011. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2011-2014/2011/lei/l12527.htm. Acesso em: 19 mar. 2025.

BRASIL. **Lei nº 13.709, de 14 de agosto de 2018.** Lei Geral de Proteção de Dados Pessoais (LGPD). Brasília, DF: Diário Oficial da União, 2018. Disponível: https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm. Acesso em: 31 mar. 2025.

BRASIL. **Lei nº 14.129, de 29 de março de 2021.** Dispõe sobre princípios, regras e instrumentos para o Governo Digital e para o aumento da eficiência pública. Brasília, DF: Diário Oficial da União, 2021. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2019-2022/2021/lei/l14129.htm. Acesso em: 19 mar. 2025.

BRASIL. Ministério da Gestão e da Inovação em Serviços. **Interoperabilidade.** Brasília, DF: Ministério da Gestão e da Inovação em Serviços, 2020c. Disponível em: <https://www.gov.br/governodigital/pt-br/governanca-de-dados/interoperabilidade>. Acesso em: 26 jul. 2024.

BRASIL. **Resolução nº 3, de 13 de outubro de 2017.** Aprova as normas sobre a elaboração e publicação de Planos de Dados Abertos, conforme disposto no Decreto n.º 8.777, de 11 de maio de 2016. Brasília, DF: Diário Oficial da União, 2017. Disponível em: <https://www.gov.br/conarq/pt-br/legislacao-arquivistica/resolucoes/resolucao-n-o-3-de-13-de-outubro-de-2017>. Acesso em: 19 mar. 2025.

BRASIL. Secretaria de Governo Digital do Ministério da Gestão e da Inovação em Serviços Públicos. **Guia de boas práticas para acessibilidade digital.** Brasília, DF: Secretaria de Governo Digital do Ministério da Gestão e da Inovação em Serviços Públicos, 2023. Disponível em: <https://www.gov.br/governodigital/pt-br/acessibilidade-e-usuario/acessibilidade-digital/guiaboaspraaticasparaacessibilidadedigital.pdf>. Acesso em: 08 jul. 2025.

BRASIL. Tribunal de Contas da União. **5 motivos para a abertura de dados na administração pública.** Brasília, DF: Tribunal de Contas da União, 2015. Disponível em: https://portal.tcu.gov.br/data/files/81/55/71/DB/A592C710D79E7EB7F18818A8/5_motivos_abertura_dados_administracao_publica.PDF. Acesso em: 13 mar. 2025.

CAROSSO, Daniel Fernando; TEIXEIRA FILHO, José Gilson de Almeida. **Dados abertos:** o que o cidadão mais busca das instituições públicas. Curitiba, PR: Appris, 2020.

CHIGNARD, Simon. A brief history of open data. **ParisTech Review**, France, 2013. Disponível em: <https://www.paristechreview.com/2013/03/29/brief-history-open-data/>. Acesso em: 10 mar. 2025.

CICHY, Corinna; RASS, Stefan. An overview of data quality frameworks. **IEEE Access**, Estados Unidos, v. 7, 2019. Disponível em: <https://ieeexplore.ieee.org/document/8642813>. Acesso em: 24 out. 2024.

COORDENAÇÃO DE APERFEIÇOAMENTO DE PESSOAL DE NÍVEL SUPERIOR (Capes). **Portal de Periódicos da CAPES lança novo sistema de busca**. Brasília, DF, 2024. Disponível em: <https://www.gov.br/capes/pt-br/assuntos/noticias/portal-de-periodicos-da-capes-lanca-novo-sistema-de-busca>. Acesso em: 03 mar. 2025.

DATASIDE. **Tipos de dados**: estruturados, semiestruturados e não estruturados. 2023. Disponível em: <https://www.dataside.com.br/dataside-community/big-data/tipos-de-dados-estruturados-semi-estruturados-e-nao-estruturados>. Acesso em: 17 jan. 2025.

DAVENPORT, Thomas H; PRUSAK, Laurence. **Conhecimento empresarial: como as organizações gerenciam o seu capital intelectual**. – Rio de Janeiro: Elsevier, 2003. – 10ª reimpressão.

DINIZ, Vagner. Como conseguir dados governamentais abertos. *In*: CONGRESSO CONSAD DE GESTÃO PÚBLICA, 3., 2010, Brasília, DF. **Anais [...]**. Brasília, DF: Conselho Nacional de Secretários de Estado da Administração (CONSAD), 2010. Disponível em: <https://acervo.ceweb.br/acervos/conteudo/b4a5d170-7ff1-41e1-b672-f4ba6e064be6>. Acesso em: 14 abr. 2023.

DOROBAT, Ilie Cristian; POSEA, Vlad. Indicador de dados abertos: uma metodologia acumulativa para medir a qualidade de dados governamentais abertos. *In*: CONFERÊNCIA INTERNACIONAL SOBRE ELETRÔNICA, COMPUTADORES E INTELIGÊNCIA ARTIFICIAL (ECAI), 13, 2021, Pitesti, Romênia. **Anais [...]**. Pitesti, Romênia: ECAI, 2021. p. 1-4. DOI: 10.1109/ECAI52376.2021.9515147. Disponível em: <https://ieeexplore.ieee.org/document/9515147>. Acesso em: 26 fev. 2024.

EAVES, David. **The three laws of open government data**. 2009. Disponível em: <https://eaves.ca/2009/09/30/three-law-of-open-government-data/>. Acesso em: 21 nov. 2023.

ESCOLA NACIONAL DE ADMINISTRAÇÃO PÚBLICA (ENAP). **Gerenciamento de metadados e da qualidade de dados**. Brasília, DF: ENAP, 2019.

ESCOLA NACIONAL DE ADMINISTRAÇÃO PÚBLICA (ENAP). **Introdução ao Excel**. Brasília, DF: ENAP, 2022.

ERMEL, Ana Paula Cardoso; LACERDA, Daniel Pacheco; MORANDI, Maria Isabel W. M.; GAUS, Leandro. **Revisões da literatura**: um método para a geração de conhecimento científico e tecnológico. Rio de Janeiro, RJ: Fólio Digital, 2022.

FAN, Bo; ZHAO, Yupan. The moderating effect of external pressure on the relationship between internal organizational factors and the quality of open government data. **Government Information Quarterly**, United States, v. 34, n. 3, p. 396-405, 2017. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0740624X17300801>. Acesso em: 14 mar. 2024.

FARIA, Paulo M. **Revisão sistemática da literatura**: contribuindo para um novo paradigma investigativo [...]. Portugal: Whitebooks, 2016.

FERNEDA, Edberto. **Recuperação da Informação**: análise sobre a contribuição da Ciência da Computação para a Ciência da Informação. 2003. Tese (Doutorado em Ciências da Comunicação), Escola de Comunicação e Artes, Universidade de São Paulo, São Paulo, 2003.

FIRMANI, Donatella; MECELLA, Massimo; SCANNAPIECO, Mônica; BATINI, Carlo. Sobre a significância da qualidade do Big Data (artigo convidado). **Data Science and Engineering**, Alemanha, v. 1, n. 1, 2016. Disponível em: <https://doi.org/10.1007/s41019-015-0004-7>. Acesso em: 13 jan. 2025.

FLICK, Uwe. **Introdução à pesquisa qualitativa**. 3. ed. Porto Alegre, RS: Artimed, 2009.

GIL, Antônio Carlos. **Como elaborar projetos de pesquisa**. 6. ed. São Paulo, SP: Atlas, 2018.

GIL, Antônio Carlos. **Métodos e técnicas de pesquisa social**. São Paulo, SP: Atlas, 2024.

GRANUCCI, Raphael. Acurácia e precisão: qual a diferença? **Minds Digital**, 2023. Disponível em: <https://minds.digital/digital-transformation/acuracia-e-precisao/>. Acesso em: 27 fev. 2025.

GRAY, Jonathan. Towards a genealogy of open data. *In*: GENERAL CONFERENCE OF THE EUROPEAN CONSORTIUM FOR POLITICAL RESEARCH, 2014, Glasgow. **Anais** [...]. Glasgow: [s.n.], 2014. Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2605828. Acesso em: 11 mar. 2025.

GUEDES, Josefina Soares; FONSECA, Rita de Cássia; STRAUHS, Faimara do Rocio. Uso de indicadores e métricas para avaliação da qualidade da informação. **Brazilian Journal of Information Science: Research trends**, Marília, SP, v. 15, p. e02121, out. 2021. DOI: 10.36311/1981-1640.2021.v15.e02121. Disponível em: <https://revistas.marilia.unesp.br/index.php/bjis/article/view/12189>. Acesso em: 15 mar. 2025.

HERRERA-MELO, Camila Andrea; GONZÁLEZ-SANABRIA, Juan Sebastián. Proposal for the evaluation of open data portals. **Revista Facultad de Ingeniería**, Colômbia, v. 29, n. 54, e10194, 2020. Disponível em:

<https://revistas.uptc.edu.co/index.php/ingenieria/article/view/10194/8410>. Acesso em: 22 fev. 2024.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION (ISO). **Software engineering**: software product quality requirements and evaluation (SquaRE): Data quality model. [ISO/IEC25012:2008]. Londres: Comitê Técnico ISO/IEC JTC 1/SC, 2008.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION; INTERNATIONAL ELECTROTECHNICAL COMMISSION. ISO/IEC 11179:2023. **Information technology**: metadata registries. Geneva: ISO/IEC, 2023.

ISOTANI, Seiji; BITTENCOURT, Ig Ibert. **Dados abertos conectados**. São Paulo, SP: Novatec, 2015.

JANSSEN, Katleen. The influence of the PSI directive on open government data: an overview of recent developments. **Government Information Quarterly**, United States, v. 28, n. 4, p. 446-456, 2011. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S0740624X11000517>. Acesso em: 08 fev. 2024.

JESUS, Ananda Fernanda de; BARITÉ, Mario; SANTAREM SEGUNDO, José Eduardo. O termo qualidade de dados na ciência da informação brasileira: uma análise de terminologia pontual. *In*: ENCONTRO NACIONAL DE PESQUISA E PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO (ENANCIB), 24., 2024, Vitória, ES. **Anais** [...]. Vitória, ES: ANCIB, 2024. Disponível em: <https://enancib.ancib.org/index.php/enancib/xxivenancib/paper/view/2295>. Acesso em: 24 abr. 2025.

JETZEK, Thorhildur; AVITAL, Michel; BJON-ANDERSEN, Niels. Generating sustainable value from open data in a sharing society. *In*: CREATING VALUE FOR ALL THROUGH IT. IFIP ADVANCES IN INFORMATION AND COMMUNICATION TECHNOLOGY, 429., 2014, Berlin, Heidelberg. **Proceedings** [...]. Berlin, Heidelberg: Springer, 2014. Disponível em: https://doi.org/10.1007/978-3-662-43459-8_5. Acesso em: 11 mar. 2025.

JETZEK, Thorhildur. Managing complexity across multiple dimensions of liquid open data: The case of the Danish Basic Data Program. **Government Information Quarterly**, United States, v. 33, n. 1, p. 89-104, 2016. Disponível em: <https://doi.org/10.1016/j.giq.2015.11.003>. Acesso em: 08 mar. 2024.

JOHNSON, Jeffrey Alan. From open data to information justice. **Ethics and Information Technology**, [S.l.], v. 16, p. 263-274, 2014. Disponível em: <https://link.springer.com/article/10.1007/s10676-014-9351-8>. Acesso em: 04 set. 2025.

KITCHENHAM, Barbara. Procedures for performing systematic reviews. **Keele University Technical Report**, United Kingdom, c2004. Disponível em: <http://bibliotecadigital.fgv.br/ocs/index.php/ctd/ctd2018/paper/viewFile/6981/1961>. Acesso em: 31 out. 2025.

KITCHIN, Rob. **The data revolution**: big data, open data, data infrastructures & their consequences. Los Angeles: SAGE, 2014.

LAKE, Peter; CROWTHER, Paul. **Concise guide to databases**: a practical introduction. Berlin: Springer, Verlag London, 2013.

LE COADIC, Yves-François. **A ciência da informação**. Tradução Maria Yeda F. S. Filgueiras Gomes. Brasília, DF: Briquet de Lemos, 1996.

LEVEL ACCESS. **WCAG Conformance Levels**: how high should you aim? 2024. Disponível em: <https://www.levelaccess.com>. Acesso em: 19 nov. 2025.

LEITE, Leonardo de Oliveira; REZENDE, Denis Alcides. **E-gov estratégico**: governo eletrônico para gestão do desempenho da administração pública. Curitiba, PR: Appris, 2015.

LIMA, José Leonardo Oliveira; ALVARES Lillian. Organização e representação da informação e do conhecimento. In: ALVARES (org.). Organização da informação e do conhecimento: conceitos, subsídios interdisciplinares. São Paulo: B4 Ed., 2012.

LIU, Liping; CHI, Lauren N. Evolutional data quality: a theory-specific view. In: MIT INTERNATIONAL CONFERENCE ON INFORMATION QUALITY, 7., 2002, Massachusetts. **Proceedings** [...]. Massachusetts: Massachusetts Institute of Technology (MIT), 2002. Disponível em: <https://www.semanticscholar.org/paper/Evolutional-Data-Quality%3A-A-Theory-Specific-View-Liu-Chi/cd8474a22c554b60ba975af93b99e0c0cdc9893f>. Acesso em: 29 jan. 2025.

LOPES, Ana Lúcia Mendes; FRACOLLI, Lislaine Aparecida. Revisão sistemática de literatura e metasíntese qualitativa: considerações sobre sua aplicação na pesquisa em enfermagem. **Texto & Contexto – Enfermagem**, Florianópolis, SC, v. 17, n. 4, p. 771-778, 2008. Disponível em: <https://doi.org/10.1590/S0104-07072008000400020>. Acesso em: 31 out. 2022.

MACEDO, Dirceu Flávio; LEMOS, Daniela Lucas da Silva. Dados abertos governamentais: iniciativas e desafios na abertura de dados no Brasil e outras esferas internacionais. **AtoZ: Novas Práticas em Informação e Conhecimento**, Curitiba, PR, v. 10, n. 2, p. 14-26, abr. 2021. Disponível em: <https://revistas.ufpr.br/atoz/article/view/77737>. Acesso em: 30 mar. 2023.

MÁCHOVÁ, Renata; LNENICKA, Martin. Evaluating the quality of open data portals on the national level. **Journal of Theoretical and Applied Electronic Commerce Research**, Talca, Chile, v. 12, n. 1, p. 21-41, 2017. Disponível em: https://www.scielo.cl/scielo.php?script=sci_arttext&pid=S0718-18762017000100003. Acesso em: 09 jan. 2025.

MAESTRE-GONGORA, Gina; RANGEL-CARRILLO, Adriana; OSORIO-SANABRIA, Mariutsi. Exploración sobre datos gubernamentales abiertos y big data desde una perspectiva de calidad. **Revista Colombiana de Computación**, Bucamaranga,

Colômbia, v. 24, n. 1, p. 1-17, 2023. Disponível em: <https://revistas.unab.edu.co/index.php/rcc/article/view/4389>. Acesso em: 08 mar. 2024.

MAHANTI, Rupa. **Data quality**: dimensions, measurement, strategy, management, and governance. Milwaukee, Wisconsin: ASQ Quality Press, 2019.

MATHIESEN, Kay. Human rights without cultural imperialism. In: JAEGER, Paul (ed.). **Perspectives on libraries as institutions of human rights and social justice**. [S.l.]: Emerald Publishing, 2016, p. 265-286. Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2943742. Acesso em: 04 set. 2025.

MELO, Diego. **O que é HTML?** TECNOBLOG, 2021. Disponível em: <https://tecnoblog.net/responde/o-que-e-html-guia-para-iniciantes/>. Acesso em: 28 fev. 2025.

MITKIEWICZ, Fernando André Coelho. Transformação digital: análise da implantação da plataforma Gov.br e da evolução da maturidade da política de governo digital no Brasil. In: KUBOTA, Luís Claudio (org.). **Digitalização e tecnologias da informação e comunicação**: oportunidades e desafios para o Brasil. Rio de Janeiro, RJ: IPEA, 2024. p. 255-294.

MORADI, Mohammad; MAZOOCHI, Mojtaba; AHMADI, Mohammad. A comprehensive method for improving the quality of open government data and increasing citizens' willingness to use data by analyzing the complex system of citizens and organizations. **Complexity**, United States, v. 22, n. 1, 2022. Disponível em: <https://onlinelibrary.wiley.com/doi/10.1155/2022/5876035>. Acesso em: 27 mar. 2024.

MOREIRA, Fábio Mosso; BISI, Pedro Henrique Santos; BOTEGA, Leonardo Castro.; SANTAREM SEGUNDO, José Eduardo; SANT'ANA, Ricardo Cesar Gonçalves. A qualidade na recuperação de dados governamentais: um estudo sobre dados de políticas públicas na internet. **Perspectivas em Ciência da Informação**, Belo Horizonte, MG, v. 25, n. 2, p. 103-132, 2020. Disponível em: <https://periodicos.ufmg.br/index.php/pci/article/view/24121>. Acesso em: 06 mar. 2024.

MOSES, Barr; GAVISH, Lior; VORWERCK, Molly. **Fundamentos da qualidade de dados**: guia prático para criar pipelines de dados confiáveis. Tradução Cibelle Ravaglia. Rio de Janeiro, RJ: Alta Books, 2024.

MOTTA, Fabrício; VALLE, Vanice Regina Lírio do. **Governo digital por inovação na administração pública**: a Lei nº 14.129, de 29 de março de 2021. Belo Horizonte, MG: Fórum, 2022.

NIKIFOROVA, Anastasija; MCBRIDE, Keegan. Open government data portal usability: a user-centred usability analysis of 41 open government data portals. **Telematics and Informatics**, Netherlands, v. 58, 2021. Disponível em: <https://doi.org/10.1016/j.tele.2020.101539>. Acesso em: 14 mar. 2024.

OLIVEIRA, Wanessa Queiroz de Souza; FONSECA, Igor Ferraz da. Fatores de sucesso na abertura de dados: o caso do Banco Central do Brasil. **Revista do Serviço Público (RSP)**, Brasília, DF, v. 72, n. 4, out./dez., p. 724-752, 2021. Disponível em: <https://repositorio.enap.gov.br/jspui/bitstream/1/6798/1/4535-Texto%20do%20Artigo-21236-1-10-20211227.pdf>. Acesso em: 17 mar. 2023.

OLSON, Jack Eduard. **Data quality**: the accuracy dimension. San Francisco: Elsevier, 2003.

OPEN AI. **ChatGPT**: representação gráfica das estruturas de dados. San Francisco: Open AI, 19 fev. 2025.

OPEN GOVERNMENT PARTNERSHIP (OGP). **Membros**. 2011. Disponível em: <https://www.opengovpartnership.org/our-members/>. Acesso em: 13 mar. 2025.

OPEN GOVERNMENT PARTNERSHIP (OGP). **Open Government Partnership 2024-2025 annual report**. 2025. Disponível em: <https://www.opengovpartnership.org/2024-2025-annual-report/>. Acesso em: 07 set. 2025.

OPEN KNOWLEDGE BRASIL (OKBR). **Maturidade em dados abertos**: entenda as 5 estrelas. 2013. Disponível: <https://ok.org.br/noticia/maturidade-em-dados-abertos-entenda-as-5-estrelas/>. Acesso em: 22 jan. 2025.

OPEN KNOWLEDGE FOUNDATION (OKF). **O que são dados abertos?** 2010. Disponível em: https://opendatahandbook.org/guide/pt_BR/introduction/. Acesso em: 23 jan. 2025.

ORACLE CLOUD INFRASTRUCTURE (OCI). **O que é NoSQL?** 2025. Disponível em: <https://www.oracle.com/br/database/nosql/what-is-nosql/>. Acesso em: 31 mar. 2025.

ORGANIZAÇÃO PARA A COOPERAÇÃO E DESENVOLVIMENTO ECONÔMICO (OCDE). **Digital government review of Brazil**: rumo à transformação digital do setor público. Paris: OECD Publishing, OECD Digital Government Studies, 2018a. Disponível em: <https://doi.org/10.1787/9789264307636-en>. Acesso em: 21 mar. 2025.

ORGANIZAÇÃO PARA A COOPERAÇÃO E DESENVOLVIMENTO ECONÔMICO (OCDE). **Revisão do governo digital do Brasil rumo à transformação digital do setor público**: principais conclusões. Paris: OECD Digital Government Studies, OECD Publishing, 2018b. Disponível em: <https://repositorio.enap.gov.br/jspui/handle/1/3627>. Acesso em: 23 mar. 2025.

ORGANIZAÇÃO PARA A COOPERAÇÃO E DESENVOLVIMENTO ECONÔMICO (OCDE). **The OCDE digital government policy framework**: six dimensions of a digital government. Paris: OCDE Public Governance Policy Papers [n. 2], OCDE Publishing, 2018c.

ORGANIZAÇÃO PARA A COOPERAÇÃO DESENVOLVIMENTO ECONÔMICO (OCDE). **The OCDE The OCDE digital government policy framework: six dimensions of a digital government**. Paris: OCDE Public Governance Policy Papers [n. 2], OCDE Publishing, 2020.

PAGAIME, Rúben Gonçalo Teixeira. **Gestão da qualidade dos dados em contexto de dados abertos**. 2018. Dissertação (Mestrado em Gestão de Informação) – Instituto Superior de estatística e Gestão de Informação, Universidade Nova de Lisboa, Lisboa, 2018.

PENTEADO, Elias Bruno. **Modelo de Infraestrutura para publicação de dados abertos governamentais conectados de qualidade**. 2020. Tese (Doutorado em Ciências da Computação e Matemática Computacional) - Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Paulo, 2020.

PINHO, Maria Dominguez Costa; SILVA, Maria Paula Almada. Governo aberto e dados abertos governamentais: um mapeamento e sistematização da produção acadêmica. **Comunicação & Inovação, PPGCOM/USCS**, São Caetano do Sul, SP, v. 20, n. 43, p. 3-25, maio/ago. 2019. Disponível em: https://seer.uscs.edu.br/index.php/revista_comunicacao_inovacao/article/view/5711/2787. Acesso em: 04 abr. 2025.

PINTO, Marcus Vinícius. **A abertura de dados no mundo, no Brasil e em Belo Horizonte**. [S.l.]: [s.n.], c2023.

RAUTENBERG, Sandro; SOUZA, Lucélia de; DALL'AGNOL, Josiane Michalak Hauagge; MICHELON, Gisane Aparecida. **Guia prático para publicação de dados abertos conectados na web**. Curitiba, PR: Appris, 2018.

REINHARD, Nicolau; GERMANO, Edson Carlos. **Guia de Incentivo ao (re)uso de dados abertos**. 2016. Disponível em: <https://nic.br/media/docs/publicacoes/13/Guia%20Incentivo%20ao%20Reúso%20de%20Dados%20Abertos.pdf>. Acesso em: 18 mar. 2025.

ROCHA, Vladimir Alexei Rodrigues; BARBOSA, Ricardo Rodrigues; AGANETTE, Elisângela Cristina. Usabilidade: um estudo de prática da arquitetura da informação em website. **Múltiplos Olhares em Ciência da Informação**, Belo Horizonte, MG, n. esp. 2021. Disponível em: <https://periodicos.ufmg.br/index.php/moci/article/view/37261>. Acesso em: 26 jul. 2024.

ROYO-MONTAÑÉS, Sonia; BENÍTEZ-GÓMEZ, Alberto. Portales de datos abiertos: metodología de análisis y aplicación a municipios españoles. **El profesional de la información**, Espanha, v. 28, n. 6, 2019. Disponível em: <https://doi.org/10.3145/epi.2019.nov.09>. Acesso em: 08 mar. 2024.

RUVALCABA GÓMEZ, E. Datos abiertos como política pública dentro del Gobierno abierto. Estado Abierto. **Revista sobre el Estado, la Administración y las Políticas Públicas**, Caba, Argentina, v. 3, n. 2, p. 99-116, 2019. Disponível em:

<https://publicaciones.inap.gob.ar/index.php/EA/article/view/85>. Acesso em: 14 abr. 2025.

SANT'ANA, Ricardo César Gonçalves. Ciclo de vida dos dados: uma perspectiva a partir da Ciência da Informação. **Informação & Informação**, Londrina, PR, v. 21, n. 2, p. 116-142, 2016. DOI: 10.5433/1981-8920.2016v21n2p116. Disponível em: <https://ojs.uel.br/revistas/uel/index.php/informacao/article/view/27940>. Acesso em: 27 mar. 2025.

SANT'ANA, Ricardo César Gonçalves. **Tecnologia e gestão pública municipal: mensuração da interação com a sociedade**. São Paulo: Editora Unesp; São Paulo: Cultura Acadêmica, 2009. 178p. Disponível em: <https://static.scielo.org/scielobooks/8v2y2/pdf/santana-9788579830105.pdf>. Acesso em: 28 abr. 2024.

SCARISBRICK-HAUSER, Anne-Marie; ROUSE, Christina. The whole truth and nothing but the truth? The role of data quality today. **Journal of Research in Interactive Marketing**, United Kingdom, v. 1, n. 3, p. 161-171, 2007. Disponível em: <https://www.emerald.com/insight/content/doi/10.1108/17505930710779333/full/html>. Acesso em: 26 mar. 2025.

SCHULTE, Frederik; CHUNPIR, Hashim Iqbal; VOB, Stefan. Open data evolution in information systems research: considering cases of data-intensive transportation and grid systems. In: MARCUS, A. (ed.). **Design, user experience, and usability: technological contexts**. Cham, Suíça: Springer International Publishing Switzerland, 2016. v. 9748. DOI: https://doi.org/10.1007/978-3-319-40406-6_18. Disponível em: https://link.springer.com/content/pdf/10.1007/978-3-319-40406-6_18.pdf. Acesso em: 09 mar. 2024.

SEBASTIAN-COLEMAN, Laura. **Measuring data quality for ongoing improvement: a data quality assessment framework**. Waltham: Morgan Kaufmann, 2013.

SENA, Priscila Machado Borges. Justiça informacional em ciência, tecnologia e inovação no Brasil: reflexões e ações necessárias em Ciência da Informação. **Encontros Bibli**, Florianópolis, SC, v. 28, n. esp., p.1-18, 2023. Disponível em: <https://doi.org/10.5007/1518-2924.2023.e93046>. Acesso em: 04 set. 2025.

SEVERINO, Antônio Joaquim. **Metodologia do trabalho científico**. 2. ed. São Paulo, SP: Cortez, 2017.

ŠLIBAR, Babara; OREŠKI, Dijana; BEGIČEVIC, Redep, N. Importância da avaliação de dados abertos: uma visão sobre as dimensões de qualidade de (meta)dados. **Sage Open**, United States, v. 11, n. 2, 2021. Disponível em: <https://doi.org/10.1177/21582440211023178>. Acesso em: 27 mar. 2024.

SILVA, Ambrozina de Abreu Pereira; MONTEIRO, Doraliza Auxiliadora Abranches; REIS, Anderson de Oliveira. Qualidade da informação dos dados governamentais abertos: análise do Portal de Dados Abertos Brasileiro. **Revista Gestão em Análise (REGEA)**, Fortaleza, CE, v. 9, n. 1, p. 31-47, jan./abr. 2020. Disponível em:

<https://periodicos.unichristus.edu.br/gestao/article/view/2555>. Acesso em: 06 mar. 2024.

SILVA, Franciéle Carneiro Garcês da; GARCEZ, Dirnéle Carneiro; ROMEIRO, Nathália Lima; FEVRIER, Pricila Rufino; ALVES, Ana Paula Meneses. Justiça para quem? Justiça social, informacional, racial e de gênero em bibliotecas. *In*: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO (ENANCIB), 21., 2021, Rio de Janeiro. **Anais [...]**. Rio de Janeiro, RJ: ANCIB, 2021, p. 1-16, Disponível em: <https://enancib.ancib.org/index.php/enancib/xxienancib/paper/viewFile/619/397>. Acesso em: 05 set. 2025.

SILVA, Leonardo Ferreira da. **Dados abertos governamentais conectados em Big Data**: framework conceitual. São Paulo, SP: Dialética, 2021.

SILVA, Patrícia Nascimento; PINHEIRO, Marta Macedo Kerr. Execução da política de dados abertos no Brasil: uma avaliação dos três anos do decreto n.º 8.777/2016. **Tendências da Pesquisa Brasileira em Ciência da Informação (TPBCI)**, [S.l.], v. 12, n. 2, 2019. Disponível em: <https://revistas.ancib.org/index.php/tpbci/article/view/495>. Acesso em: 19 mar. 2025.

SILVA, Thiago Emídio Esteves da. **Percepções do principal sobre a qualidade dos dados abertos governamentais**: uma análise à luz da teoria principal-agente. 2019. 118f. Dissertação (Mestrado em Administração) - Centro de Ciências Sociais Aplicadas, Universidade Federal da Paraíba (UFPB), João Pessoa, PB, 2019.

SOUTO, L. F. Recuperação de informações em bases de dados: usos de tesouro. **TransInformação**, Campinas, SP, v. 15, n. 1, p. 73-81, jan. 2003. Disponível: <https://www.scielo.br/j/tinf/a/XRW5WHHGpDyddz8Y38mTbKx/?lang=pt>. Acesso em: 22 abr. 2025.

STRONG, Diane M.; LEE, Yang W.; WANG, Richard Y. Data quality context. **Communications of the ACM**, United States, v. 40, n. 5, p. 103-110, 1997. Disponível em: <https://dl.acm.org/doi/10.1145/253769.253804>. Acesso em: 11 fev. 2025.

TAYI, Giri Kumar; BALLOU, Donald P. Examining data quality. **Communications of the ACM**, United States, v. 41, n. 2, p. 54-57, 1998. Disponível em: <https://dl.acm.org/doi/abs/10.1145/269012.269021>. Acesso em: 19 abr. 2025.

TERNOSKI, Simão. **A pesquisa quantitativa e qualitativa nas ciências sociais aplicadas**. Ponta Grossa, PR: Atena, 2022.

UNIVERSIDADE ESTADUAL PAULISTA (Unesp). **Programa de Pós-Graduação em Ciência da Informação**. Marília, 2025. Disponível em: <https://www.marilia.unesp.br/#!/pos-graduacao/mestrado-e-doutorado/ciencia-da-informacao/programa/sobre/>. Acesso em: 03 set. 2025.

URENA-CAMARA, Manuel Antonio; NOGUERAS-ISO, Javier; LACASTA, Javier. Revisión de la calidad de los conjuntos de datos abiertos sobre presupuestos.

Revista Cartografica, Ciudad de México, n. 103, p. 145-163, dic. 2021. DOI: <https://doi.org/10.35424/rcarto.i103.1026>. Disponível em: http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S2663-39812021000200145&lng=es&nrm=iso. Acessado em: 27 nov. 2025.

VALENTIM, Marta Lígia Pomim. Equipes multidisciplinares na gestão da informação e conhecimento. In: BAPTISTA, Sofia Galvão; MULLER, Suzana Pinheiro Machado (Orgs.). *Profissional da Informação e o espaço de trabalho*. Brasília: Thesaurus, 2004, cap. 7, p. 154-176.

VETRÒ, Antônio; CANOVA, Lorenzo; TORCHIANO, Marco; MINOTAS, Camilo Orozco; IEMMA Raimundo; MORANDO, Frederico. Open data quality measurement framework: definition and application to open government data. **Government Information Quarterly**, United States, v. 33, n. 2, p. 325-337, 2016. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S0740624X16300132>. Acesso em: 27 fev. 2024.

WAND, Yair; WANG, Richard Y. Anchoring data quality dimensions in ontological foundations. **Communications of the ACM**, United States, v. 39, n. 11, p. 86-95, 1996. DOI: <https://doi.org/10.1145/240455.240479>. Disponível em: <https://dl.acm.org/doi/10.1145/240455.240479>. Acesso em: 31 jan. 2024.

WANG, Richard, Y; STRONG, Diane M. Beyond accuracy: what data quality means to data consumers. **Journal of Management Information Systems**, United States, v. 12, n. 4, p. 5-33, 1996. Disponível em: <https://www.tandfonline.com/doi/abs/10.1080/07421222.1996.11518099>. Acesso em: 13 jan. 2025.

WANG, Richard Y; ZIAD, Mostapha; LEE Yang W. **Data quality**. New York: Kluwer Academic Publishers, 2002.

WORLD WIDE WEB (WWW). **The Open Data Barometer**. 2017. Disponível em: https://opendatabarometer.org/?_year=2017&indicator=ODB. Acesso em: 26 fev. 2025.

ZAVERI, Amrapali; RULA, Anisa; MAURINO, Andrea; PIETROBON, Ricardo; LEHMANN, Jens; AUER, Soren. Quality Assessment Methodologies for Linked Open Data (2012). **Semantic Web Journal**, [S.l.], v. 1, n. 5, 2015. Disponível em: <https://www.semantic-web-journal.net/content/quality-assessment-methodologies-linked-open-data>. Acesso em: 13 jan. 2025.

ZHANG, Hui; XIAO, Jianying. Quality assessment framework for open government data. **The Electronic Library**, United Kingdom, v. 38, n. 2, p. 209-222, 2020. Disponível em: <https://www.emerald.com/insight/content/doi/10.1108/EL-06-2019-0145/full/html>. Acesso em: 28 fev. 2024.

ZHANG, Roujing; INDULSKA, Marta; SADIQ, Shazia. Discovering data quality problems. **Business & Information Systems Engineering**, Germany, v. 61, n. 5, p. 575-593, 2019. Disponível em: <https://aisel.aisnet.org/bise/vol61/iss5/3>. Acesso em: 08 mar. 2024.

APÊNDICE A – INSTRUMENTO PARA AVALIAÇÃO DA QUALIDADE DOS DADOS ABERTOS

Dimensão:	ACESSIBILIDADE		
Métrica:	Conformidade com as normas de acessibilidade As normas que garantem a acessibilidade dos usuários com deficiência estão sendo respeitadas?	Avaliação:	Crítico
Recursos de avaliação	Descrição dos recursos	Avaliação recursos (3 - excelente/ 2 - aceitável/ 1- Crítico)	
Texto Alternativo	Todas as imagens, gráficos e ícones possuem descrição textual significativa.	3	
Navegação por teclado	O <i>site</i> pode ser totalmente navegável sem <i>mouse</i> .	1	
Contraste de cores	Texto e elementos gráficos têm contraste adequado para leitura	2	
Legendas e libras	Vídeos e conteúdos multimídia têm legendas e os conteúdos possuem interpretação em Libras.	2	
Formulários acessíveis	Rótulos claros, instruções, mensagens de erro compreensíveis.	2	
Estrutura Semântica	Uso correto de títulos, listas, menus e hierarquia de conteúdo.	2	
Conteúdo multimídia seguro	Evita conteúdos piscantes ou que causem desconforto.	3	
Linguagem e clareza	Texto claro e compreensível para pessoas com deficiência intelectual.	1	
		67	
Instruções	a) Para a análise da métrica Conformidade com as Normas de Acessibilidade, verifica-se a presença dos recursos exigidos no conjunto de dados. A cada item analisado é atribuída uma pontuação conforme a escala: 3 (Excelente) , 2 (Aceitável) ou 1 (Crítico) . A avaliação final da métrica é obtida automaticamente pelo somatório dessas notas individuais.		
Nota	Os critérios detalhados para a aplicação de cada nota/pontuação estão disponíveis nas orientações da ferramenta, acessíveis através da célula de resultado. Os campos digitáveis da planilha encontram-se destacados em cinza.		

Dimensão:	COMPLETUDE		
Métrica:	Valores ausentes O conjunto de dados apresenta valores ausentes em proporção que podem prejudicar a análise dos dados?	Avaliação:	Aceitável
Fórmula:	Proporção de células vazias= <u>Número de células nulas ou vazias</u> x100	20	20
	Número total de células no conjunto de dados	100	
Instruções	A análise da métrica 'Valores Ausentes' baseia-se na aplicação de uma fórmula para calcular o percentual de dados faltantes. Para realizar o cálculo, é necessário identificar o número de células vazias e o total de células do conjunto de dados, inserindo esses valores no campo de avaliação. Como exemplo, considera-se um conjunto que possui um total de <u>100</u> células e apresenta <u>20</u> células vazias; ao inserir esses dados, o resultado final da métrica será gerado automaticamente.		
Métrica:	Cobertura dos dados coletados O conjunto de dados apresenta os dados em sua completude?	Avaliação:	Crítico
Fórmula:	Cobertura= <u>Número de células preenchidas</u> x 100	80	80
	Número total de células esperadas	100	
Instruções	A análise da métrica Cobertura dos dados coletados baseia-se na aplicação de uma fórmula para calcular o percentual de preenchimento. Para realizar o cálculo, é necessário identificar o número de células preenchidas e o total de células esperadas no conjunto de dados, inserindo esses valores no campo de avaliação. Por exemplo: caso o conjunto possua <u>80</u> células preenchidas de um total de <u>100</u> , insere-se esses dados e o resultado final da métrica será gerado automaticamente.		
Métrica:	Campos obrigatórios Todos os campos com informações específicas e obrigatórias estão preenchidos?	Avaliação:	Excelente

Fórmula:	Campo obrigatório = $\frac{\text{Número de registros com campo preenchido} \times 100}{\text{Número total de registros}}$	20	100
		20	

Instruções	Analisa o percentual de preenchimento dos dados considerados essenciais para a validade do registro. Diferente da completude geral, esta métrica foca apenas nas colunas indispensáveis (ex: ano, unidade geográfica, localização). O cálculo é realizado identificando o número de campos obrigatórios efetivamente preenchidos em relação ao total de campos obrigatórios exigidos no conjunto de dados. Exemplo: Uma planilha foi configurada com <u>20</u> células destinadas a dados obrigatórios e verificou-se que todas as <u>20</u> células estão devidamente preenchidas. Ao inserir esses valores no campo de avaliação, o resultado final da métrica será gerado automaticamente.		

Métrica:	Comparabilidade É possível realizar a comparação do conjunto de dados em diferentes situações?	Avaliação:	Aceitável
		Total/Parcial/Não	Parcial

Instruções	A análise desta métrica é de natureza qualitativa, fundamentada na verificação técnica quanto à capacidade dos dados de serem confrontados e cruzados com outros conjuntos de dados (temporal ou espacialmente). A avaliação segue critérios predefinidos para mensurar se os recursos de padronização necessários para a comparação estão presentes. Escala de avaliação: <u>Total</u> (Excelente - o conjunto de dados apresenta consistência plena em todos os recursos avaliados, permitindo comparabilidade direta); <u>Parcial</u> (Aceitável - o conjunto de dados apresenta condições limitadas de comparabilidade); <u>Não</u> (Crítico - o conjunto de dados não apresenta condições de comparabilidade).		

Métrica:	Metadados Os metadados estão preenchidos de forma estruturada contextualizando e descrevendo os dados?	Avaliação:	Crítico
		Total/Parcial/Não	Não

Instruções	A avaliação da métrica de Metadados é conduzida por meio de uma abordagem qualitativa, fundamentada no julgamento técnico do avaliador quanto à completude e à precisão das informações descritivas que acompanham o conjunto de dados (como autoria, data de criação, periodicidade e licença). A análise verifica se os metadados fornecidos são suficientes para contextualizar e gerenciar o recurso informacional. A escala de classificação adota os seguintes critérios: <u>Total</u> (Excelente - o conjunto de dados apresenta metadados completos e padronizados, atendendo a todos os requisitos de descrição; <u>Parcial</u> (Aceitável - O conjunto de dados apresenta metadados, porém com lacunas ou inconsistências na descrição; <u>Não</u> (Crítico - O conjunto de dados não apresenta metadados ou estes são insuficientes.

Métrica:	Distribuição temporal O conjunto de dados atende aos requisitos temporais?	Avaliação:	Excelente
		Total/Parcial/Não	1

Instruções	Esta métrica examina a extensão da série histórica disponibilizada, verificando se a cobertura temporal dos dados é suficiente para permitir análises longitudinais e identificar tendências recentes. A avaliação é de natureza qualitativa e fundamenta-se na verificação técnica do intervalo de anos abrangido pelo conjunto de dados em relação ao ano corrente. Escala de avaliação: Excelente (1 ponto): A cobertura temporal é abrangente, incluindo dados dos últimos cinco anos ou mais; Aceitável (2 pontos): A cobertura temporal é restrita, incluindo somente os dados dos últimos quatro anos; Crítico (3 pontos): A cobertura temporal é insuficiente ou desatualizada, não abrangendo períodos recentes.

Nota	Os critérios detalhados para a aplicação de cada nota/pontuação estão disponíveis nas orientações da ferramenta, acessíveis através da célula de resultado. Os campos digitáveis da planilha encontram-se destacados em cinza.

Dimensão:	ATUALIDADE		
Métrica:	Idade dos dados O conjunto de dados apresenta defasagem temporal entre a coleta e a última atualização que compromete a atualidade dos dados?	Avaliação:	Excelente
Fórmula:	Idade dos dados = <u>Data da última atualização</u> Data atual	20/05/2025	189
		25/11/2025	Anual
Instruções	A análise da métrica Idade dos Dados consiste no cálculo do intervalo temporal entre a data da última atualização e a data atual. A verificação de conformidade ocorre confrontando esse resultado com a periodicidade declarada do conjunto. Exemplo: Para um conjunto de periodicidade anual (deve ser preenchida pelo usuário), cuja última atualização ocorreu em <u>20/05/2025</u> e a data de análise é <u>25/11/2025</u> , tem-se um intervalo de <u>189 dias</u> . Neste caso, conclui-se que não há defasagem, pois o valor encontra-se dentro do limite de um ano. O mesmo critério aplica-se às demais periodicidades: diária (limite de 1 dia), mensal (30 dias), bimestral (60 dias), entre outras. Ao inserir esses valores no campo de avaliação, o resultado final da métrica será gerado automaticamente.		
Métrica:	Frequência de atualização O conjunto de dados está sendo atualizado na frequência prevista (mensal, trimestral, anual etc.) ou há discrepâncias entre a periodicidade declarada e a prática observada?	Avaliação:	Excelente
			1
Instruções	A métrica avalia a regularidade e a pontualidade na manutenção dos dados, verificando se as novas informações são disponibilizadas em conformidade com a periodicidade declarada pelo órgão produtor. A análise é de natureza qualitativa e fundamenta-se na verificação técnica da aderência aos prazos estabelecidos. Para avaliação será utilizada as escalas: <u>Excelente</u> (1 ponto) as atualizações ocorrem dentro do prazo; <u>Aceitável</u> (2 pontos) ocorrem pequenos atrasos nas atualizações, mas a periodicidade é mantida; <u>Crítico</u> (3 pontos) as atualizações não seguem a periodicidade previstas.		

Métrica:	Tempo entre a ocorrência do evento e o registro no sistema O intervalo entre a ocorrência do evento e o seu efetivo registro no sistema garante a atualidade necessária?	Avaliação:	Aceitável
Fórmula:	Tempo de registro = $\frac{\text{Data de registro}}{\text{Data do evento}}$	01/09/2025 27/08/2025	5

Instruções	A análise da métrica tempo entre a ocorrência do evento e do registro no sistema consiste no cálculo do intervalo temporal entre a ocorrência do evento e do registro no sistema. Exemplo: O evento ocorreu em <u>27/08/2025</u> mas somente foi disponibilizado no dia <u>27/10/2025</u> . Ao inserir esses valores no campo de avaliação, o resultado final da métrica será gerado automaticamente. Se o conjunto de dados não disponibiliza das datas do evento e do registro, o avaliador deverá identificar com N (não) no local indicado para as datas dos eventos.		

Métrica:	Presença de data e hora da coleta dos dados A data e hora estão indicadas claramente no conjunto de dados?	Avaliação:	Crítico
			3

Instruções	A análise da métrica Presença de data e hora da coleta dos dados é realizada de forma qualitativa e é realizada de forma qualitativa, baseada na verificação técnica do avaliador. Para avaliação atribuir-se-á a escala: Excelente (1 ponto) o conjunto de dados apresenta informações sobre a data e hora da coleta dos dados; Aceitável (2 pontos) somente parte das informações são indicadas no conjunto de dados; Crítico (3 pontos) o conjunto de dados não disponibiliza informações sobre hora e data da coleta.		

Métrica:	Atualização dos metadados Os metadados são atualizados de forma consistente e sincronizada com os conjuntos de dados, garantindo que as informações descritivas reflitam corretamente o conteúdo disponibilizado?	Avaliação:	Aceitável
Recursos avaliados	Descrição dos recursos	Avaliação recursos	

	(cada recurso vale 1 ponto)	
Atualização	Data da última atualização	1
Fonte	Identificação da fonte responsável	1
Responsável	Contato da fonte responsável	0
Dados	Versão dos dados	1
		3

Instruções	Para a análise da métrica atualização dos metadados, verifica-se a presença dos recursos exigidos no conjunto de dados. O avaliador checa uma lista pré-definida de itens obrigatórios. Para cada item presente, atribui-se a pontuação 1. A nota final é a soma simples dessas pontuações. A avaliação final da métrica é obtida automaticamente pelo somatório dessas notas individuais.

Nota	Os critérios detalhados para a aplicação de cada nota/pontuação estão disponíveis nas orientações da ferramenta, acessíveis através da célula de resultado. Os campos digitáveis da planilha encontram-se destacados em cinza.

Dimensão:	FORMATO DE ARQUIVO		
Métrica:	Formato dos dados acessíveis Os dados estão disponíveis em formato aberto e reutilizável?	Avaliação:	Excelente
		Formato disponível:	CSV
Instruções	A análise da métrica Formato dos Dados consiste na identificação da extensão do conjunto de dados disponibilizado. Ao inserir o formato do conjunto no campo específico, o sistema gera o resultado da avaliação automaticamente. Por exemplo: caso o conjunto utilize o formato <u>CSV</u> , a avaliação será classificada como Excelente, por se tratar de um formato aberto e não proprietário.		
Métrica:	Multiplicidade de formatos O conjunto de dados é fornecido em mais de um formato, permitindo maior acessibilidade?	Avaliação:	Aceitável
		Número de formatos:	2
Instruções	A análise da métrica Multiplicidade de Formatos consiste na verificação da diversidade de arquivos oferecidos. Ao inserir o número de formatos distintos no campo específico, a avaliação é gerada automaticamente. Por exemplo: um conjunto disponibilizado em <u>2</u> formatos receberá a classificação Aceitável.		
Métrica:	Dados disponíveis em formato compactado Os dados estão disponíveis em formatos compactados, facilitando o <i>download</i> completo?	Avaliação:	Crítico
			3
Instruções	A análise da métrica Dados em Formato Compactado consiste em verificar se o conjunto de dados é		

	disponibilizado por meio de arquivos comprimidos (ex: .zip, .rar). A avaliação segue a escala: <u>Excelente</u> (1 ponto); <u>Aceitável</u> (2 pontos); <u>Crítico</u> (3 pontos).		
Métrica:	Documentação do formato A documentação sobre o formato dos dados suficientemente clara e detalhada para permitir que diferentes usuários compreendam, utilizem e integrem corretamente as informações disponibilizadas?	Avaliação:	Aceitável
			2
Instruções	A análise desta métrica consiste em verificar se as especificações técnicas do conjunto de dados são disponibilizadas de forma clara e detalhada. A avaliação adota a escala: <u>Excelente</u> (1 ponto), <u>Aceitável</u> (2 pontos) e <u>Crítico</u> (3 pontos).		
Nota	Os critérios detalhados para a aplicação de cada nota/pontuação estão disponíveis nas orientações da ferramenta, acessíveis através da célula de resultado. Os campos digitáveis da planilha encontram-se destacados em cinza.		

Dimensão:	DISPONIBILIDADE		
Métrica:	Frequência dos períodos indisponíveis A frequência e a duração dos períodos de indisponibilidade dos conjuntos de dados comprometem a continuidade do acesso e o uso efetivo das informações pelos usuários?	Avaliação:	Aceitável
		Interrupções:	breves

Instruções	A análise da métrica Frequência dos Períodos Indisponíveis visa monitorar a estabilidade do acesso ao conjunto de dados, verificando a ocorrência e a duração de eventuais quedas. O monitoramento deve ser realizado em horários alternados durante um período pré-determinado pelo avaliador. Para a classificação, utilizam-se os seguintes critérios: <u>Nenhuma</u> (Excelente), <u>Breves</u> (Aceitável); <u>Inacessível</u> (Crítico). Ao inserir a categoria observada no campo específico, a atribuição da nota final será automática.		

Métrica:	Hiperlink de download funcional Os links disponibilizados no conjunto de dados podem ser operacionalizados?	Avaliação:	Excelente
			1

Instruções	Esta métrica verifica a operabilidade dos links de download, identificando se existem falhas que impeçam o acesso ao conjunto de dados ou a partes específicas do conteúdo. A avaliação utiliza a seguinte escala: <u>Excelente</u> (1 ponto), <u>Aceitável</u> (2 pontos) e <u>Crítico</u> (3 pontos).		

Métrica:	Disponibilidade de canal de comunicação com o usuário Há canais claros e acessíveis e funcionais para comunicação entre o usuário e o órgão que disponibiliza os dados?	Avaliação:	Aceitável
		Número de canais:	1

Instruções	A avaliação da métrica Disponibilidade de Canal de Comunicação analisa a existência de meios de contato entre a sociedade e o órgão publicador, bem como a identificação dos responsáveis pelos dados. O usuário deve verificar o conjunto de dados e quantificar os canais disponíveis. Ao inserir esse número no sistema, a classificação será gerada automaticamente seguindo os critérios: <u>Excelente</u> (2 ou mais canais); <u>Aceitável</u> (1 canal); <u>Crítico</u> (nenhum canal).		

Métrica:	Múltiplo idiomas O conjunto de dados está disponibilizado em outros idiomas além do vernáculo oficial?	Avaliação:	Crítico
		Idiomas	Não

Instruções	A métrica 'Múltiplos Idiomas' verifica a internacionalização do conjunto de dados, analisando sua disponibilidade em línguas adicionais ao vernáculo oficial. A avaliação segue os critérios detalhados nas orientações (acessíveis no canto superior direito da célula). Para o preenchimento, deve-se selecionar uma das opções o que gerará a nota automática: <u>Total</u> : Toda a documentação está traduzida (Avaliação: Excelente); <u>Parcial</u> : Apenas parte da documentação está traduzida (Avaliação: Aceitável); <u>Não</u> : Disponível apenas no idioma oficial (Avaliação: Crítico).		

Nota	Os critérios detalhados para a aplicação de cada nota/pontuação estão disponíveis nas orientações da ferramenta, acessíveis através da célula de resultado. Os campos digitáveis da planilha encontram-se destacados em cinza.		

Dimensão:	PRECISÃO		
Métrica:	Taxa de dados corrompidos A ocorrência de dados corrompidos nos conjuntos disponibilizados compromete a integridade, a confiabilidade e a possibilidade de reuso das informações?	Avaliação:	Crítico
Fórmula	Taxa de dados corrompidos = $\frac{\text{Número de células corrompidas} \times 100}{\text{Número total de células}}$	45 500	9
Instruções	A análise da métrica taxa de dados corrompidos consiste no cálculo do percentual de células que contêm erros, caracteres inválidos ou formatações que impedem a análise ou interpretação correta dos dados. Exemplo: Um conjunto de dados com <u>500</u> células foram identificadas <u>45</u> células corrompidas. Ao inserir esses valores no campo de avaliação, o resultado final da métrica será gerado automaticamente.		
Métrica:	Dados duplicados A presença de dados duplicados no conjuntos de dados compromete a consistência, a confiabilidade e a qualidade das análises realizadas pelos usuários?	Avaliação:	Aceitável
Fórmula	Taxa de duplicação = $\frac{\text{Número de registros duplicados} \times 100}{\text{Total de registros}}$	10 500	2
Instruções	A métrica dados duplicados avalia a redundância não intencional no conjunto de dados, calculando a porcentagem de células (ou registros) que aparecem repetidas desnecessariamente. Divide-se o número de células duplicadas pelo número total de células do conjunto de dados. O resultado é multiplicado por 100. Ao inserir esses valores no campo de avaliação, o resultado final da métrica será gerado automaticamente. Exemplo: Um conjunto de dados com <u>500</u> células apresenta <u>10</u> células duplicadas.		
Métrica:	Grau de correspondência dos dados com a realidade Os dados refletem de forma fidedigna com fontes oficiais e confiáveis?	Avaliação:	Aceitável

Fórmula	Grau de Correspondência = $\frac{\text{Número de registros como corretos}}{\text{Total de registros verificados}} \times 100$	180	90
		200	

Instruções	A análise da métrica grau de correspondência dos dados com a realidade avalia se os dados correspondem fielmente à realidade ou às fontes primárias de onde foram extraídos, garantindo a veracidade da informação. Divide-se o número de registros verificados como fidedignos (corretos em relação à fonte) pelo número total de registros analisados. O resultado é multiplicado por 100. Ao inserir esses valores no campo de avaliação, o resultado final da métrica será gerado automaticamente. Exemplo: Um conjunto de dados com 200 registros ao comparar com a documentação oficial, constatou-se que 180 registros são fidedignos e 20 apresentavam divergências.		

Métrica:	Ausência de ambiguidades Os dados são claros objetivos e não geram múltiplas interpretações?	Avaliação:	Excelente
			1

Instruções	A métrica Ausência de Ambiguidades avalia a clareza semântica do conjunto de dados, verificando a consistência terminológica e a definição precisa das categorias, de modo a eliminar margens para múltiplas interpretações. A pontuação segue a escala: Excelente (1 ponto); Aceitável (2 pontos); 3 (Crítico).		

Nota	Os critérios detalhados para a aplicação de cada nota/pontuação estão disponíveis nas orientações da ferramenta, acessíveis através da célula de resultado. Os campos digitáveis da planilha encontram-se destacados em cinza		

Dimensão:	RECUPERABILIDADE		
Métrica:	Função de recuperação de dados O sistema permite localizar e acessar facilmente por meio de ferramentas de busca, filtros ou navegação temática?	Avaliação:	Excelente
			1
Instruções	A métrica função de recuperação de dados analisa a existência e o funcionamento de ferramentas de busca (por palavras-chave), filtros (por data, categoria, formato) e a intuitividade da navegação no conjunto de dados. A avaliação possui natureza qualitativa, sendo fundamentada na experiência do avaliador ao simular a recuperação de informações específicas. Escala para avaliação: Excelente (1 ponto); Aceitável (2 pontos); Crítico (3 pontos).		
Métrica:	Eficiência de busca Os mecanismos de buscas e filtros funcionam com rapidez, retornando resultados completos?	Avaliação:	Aceitável
Fórmula	Tempo decorrido entre o início da busca até o acesso aos dados		3
Instruções	Avalia a eficiência da arquitetura da informação e das ferramentas de pesquisa, quantificando o tempo que o usuário despende para localizar o conjunto de dados desejado. A aferição é realizada cronometrando a interação. A contagem de tempo inicia-se com o primeiro clique realizado para começar a busca (no menu ou campo de pesquisa) e encerra-se quando o conjunto de dados é encontrado. Para avaliar serão utilizados os seguintes critérios: até 1 minuto (Excelente); 1 a 3 minutos (Aceitável); acima de 3 minutos (Crítico).		
Métrica:	Navegação A estrutura de navegação do portal é intuitiva, com menus claros, hierarquia lógica	Avaliação:	Excelente

	e facilidade para localizar conjuntos de dados por diferentes caminhos?		
Fórmula	Para avaliar a métrica será contabilizado quantos cliques o usuário necessitou dar para acessar o conjunto de dados		0

Instruções	A métrica Navegação avalia o caminho percorrido pelo usuário para acessar a informação, quantificando o número de interações (cliques) necessárias para localizar o conjunto de dados. Realiza-se a contagem acumulada de cliques do mouse (ou toques em tela). A contagem inicia-se com o primeiro clique para começar a busca e encerra-se no clique final que abre ou disponibiliza o conjunto de dados. Para avaliar serão utilizados os seguintes critérios: de 1 a 3 cliques (Excelente); 4 a 6 cliques (Aceitável); acima de 7 cliques (Crítico). O avaliador pode consultar as diretrizes detalhadas nos comentários, acessíveis ao clicar no canto superior direito da célula de resultado.		

Métrica:	Taxa de falhas na recuperação de dados O processo de recuperação de dados ocorre normalmente sem interrupções?	Avaliação:	Aceitável
Fórmula	Taxa de falhas = $\frac{\text{Número de buscas com falhas}}{\text{Total de buscar realizadas}} \times 100$	1 5	20

Instruções	Avalia a robustez e a confiabilidade do motor de busca do portal. Mensura a porcentagem de tentativas de pesquisa que resultaram em erro de sistema ou que falharam em retornar resultados para termos sabidamente existentes. O avaliador realiza uma série de testes utilizando diferentes palavras-chave. O cálculo é feito dividindo o número de buscas que apresentaram falha pelo número total de buscas realizadas. O resultado é multiplicado por 100. Ao inserir esses valores no campo de avaliação, o resultado final da métrica será gerado automaticamente. Exemplo: o Avaliador realizou 05 buscas, dessas tentativas, 01 busca não funcionou adequadamente.		

Métrica:	Metadados Recuperáveis Os metadados estão disponíveis, organizados e recuperáveis?	Avaliação:	Excelente
Recursos avaliados	Descrição dos recursos	Avaliação recursos (Sim/Não)	

Título	Termo de busca	Recuperável	Sim
Descrição	Termo de busca	Recuperável	Sim
Palavras-chave	Termo de busca	Recuperável	Sim
		Total/Parcial/Não	Total

Instruções	A métrica metadados recuperáveis verifica a eficácia da indexação dos dados, analisando se os campos descritivos essenciais (Título, Descrição e Palavras-chave) são rastreáveis pela ferramenta de busca do portal. O avaliador deve realizar três testes de busca distintos para o mesmo conjunto de dados: realizar uma busca utilizando os termos exatos do Título; realizar uma busca utilizando trechos significativos da descrição; Realizar uma busca utilizando as Palavras-chave cadastradas. Para avaliar serão aplicados os critérios: Total (O conjunto de dados foi recuperado com sucesso nos três testes (Título, Descrição e Palavras-chave)); Parcial (O conjunto de dados foi recuperado em apenas um ou dois dos testes realizados); Não (O conjunto de dados não foi recuperado em nenhum dos testes). O avaliador pode consultar as diretrizes detalhadas nos comentários, acessíveis ao clicar no canto superior direito da célula de resultado.		

Métrica:	Vocabulário controlados Os vocabulários utilizados no conjunto de dados são controlados, padronizados e consistentes, facilitando a recuperação?	Avaliação:	Aceitável
			2

Instruções	A métrica vocabulário controlados avalia a padronização terminológica do conjunto de dados. Verifica se são utilizadas taxonomias, tesauros, listas de códigos oficiais (ex: Código IBGE, ISO 3166) ou ontologias reconhecidas para descrever os atributos. A análise é qualitativa, dependendo da inspeção do avaliador sobre as colunas categóricas e metadados para identificar o uso de padrões estabelecidos. Escala de avaliação: Excelente (1 ponto); Aceitável (2 pontos); Crítico (3 pontos).		

Nota	Os critérios detalhados para a aplicação de cada nota/pontuação estão disponíveis nas orientações da ferramenta, acessíveis através da célula de resultado. Os campos digitáveis da planilha encontram-se destacados em cinza.		

Dimensão:	INCONSISTENCIA		
Métrica:	Coerência lógica Os dados seguem uma estrutura lógica sem contradição entre os campos relacionados?	Avaliação:	Crítico
Fórmula:	Coerência lógica = $\frac{\text{Número de registros incoerentes}}{\text{Total de registros verificados}} \times 100$	50	10
		500	

Instruções	Analisa a integridade semântica dos dados, verificando se os registros atendem a regras de negócios predefinidas, incluindo lógica temporal (ex: data de fim posterior à de início), matemática (ex: somatórios corretos) e de integridade relacional. O objetivo é garantir que os dados possam ser utilizados diretamente, sem necessidade de saneamento prévio. Divide-se o número de registros identificados como incoerentes pelo número total de registros verificados. O resultado é multiplicado por 100. Ao inserir esses valores no campo de avaliação, o resultado final da métrica será gerado automaticamente. Exemplo: foram analisados <u>500</u> registros e identificou-se <u>50</u> registros que apresentavam inconsistência.		

Métrica:	Consistência entre os conjuntos de dados As informações apresentadas em diferentes conjuntos relacionados mantêm consistência nos valores, nomenclaturas e dimensões?	Avaliação:	Crítico
Fórmula	Coerência entre conjuntos de dados = $\frac{\text{Nº de registros inconsistentes}}{\text{Total de registros verificados}} \times 100$	30	6
		500	

Instruções	A métrica consistência entre os conjuntos de dados Analisa a equivalência e a harmonização entre conjuntos de dados relacionados, verificando a uniformidade de valores, nomenclaturas e dimensões. O objetivo é assegurar que, ao comparar a mesma informação em plataformas ou tabelas diferentes, não haja divergências (ex: o nome de uma cidade deve ser idêntico na tabela de "Vendas" e na tabela de "Logística"). A mensuração é feita dividindo o número de registros identificados como inconsistentes (que divergem da fonte de referência) pelo número total de registros comparados. O resultado é multiplicado por 100. Ao inserir esses valores no campo de avaliação, o resultado final da métrica será gerado automaticamente. Exemplo: foram		

	analisados <u>500</u> registros e identificou-se <u>30</u> registros inconsistência.

Métrica:	Coerência entre dados agregados Os totais, médias e demais dados agregados conferem com a soma ou cálculo realizado a partir de dados individuais?	Avaliação:	Crítico
Fórmula	Coerência em dados agregados = $\frac{\text{Nº de registros agregados incoerentes}}{\text{Total de registros verificados}} \times 100$	3	11
	Total de registros verificados	27	

Instruções	A métrica coerência entre dados agregados verifica a precisão matemática entre os dados detalhados e os valores totais apresentados no conjunto de dados. O objetivo é garantir que a soma das partes corresponda exatamente ao todo informado. Recalcula-se os totais a partir dos dados individuais e compara-se com os valores agregados apresentados. A mensuração é feita dividindo o número de registros agregados incoerentes (onde a soma não bate) pelo número total de registros verificados. O resultado é multiplicado por 100. Ao inserir esses valores no campo de avaliação, o resultado final da métrica será gerado automaticamente. Exemplo: foram apresentados relatórios de vendas para <u>27</u> estados, ao somar as vendas individuais de cada estado, verificou-se que para <u>03</u> estados o valor da soma não conferia com o total apresentado na tabela.		

Métrica:	Uso de Classificações As classificações adotadas seguem padrões oficiais e amplamente reconhecidos?	Avaliação:	Aceitável
Fórmula	Uso de classificações = $\frac{\text{Número classificações padronizadas}}{\text{Total de registros verificados}} \times 100$	190	95
	Total de registros verificados	200	

Instruções	Avalia a conformidade dos dados categóricos em relação a tabelas de referência ou padrões oficiais (ex: CNAE, CBO, ISO de Países). Verifica se os valores inseridos correspondem exatamente aos códigos ou termos previstos na norma adotada. A mensuração é feita dividindo o número de classificações padronizadas pelo número total de registros verificados. O resultado é multiplicado por 100. Ao inserir esses valores no campo de avaliação, o resultado final da métrica será gerado automaticamente. Exemplo: Um conjunto de dados contém uma coluna para "Atividade Econômica", que deve seguir a tabela CNAE (Classificação Nacional de Atividades Econômicas). O total é de 200 empresas. Ao verificar a coluna, constatou-se que 190 registros		

	usavam os códigos CNAE oficiais, enquanto 10 registros usavam termos informais.
Nota	Os critérios detalhados para a aplicação de cada nota/pontuação estão disponíveis nas orientações da ferramenta, acessíveis através da célula de resultado. Os campos digitáveis da planilha encontram-se destacados em cinza.

Dimensão:	COMPREENSIBILIDADE		
Métrica:	Presença de dicionários, notas explicativas ou legendas Os conjuntos de dados apresentam glossários, dicionários, notas explicativas e legendas que auxiliam na compreensão dos termos técnicos?	Avaliação:	Crítico
			Total/Parcial/Não
			Não

Instruções	Avalia se os conjuntos de dados oferecem recursos de apoio, como dicionários de dados, notas explicativas, legendas ou glossários, que auxiliam o usuário a compreender termos técnicos, códigos, abreviações ou conceitos que não são claros. Para a análise, será observado se os conjuntos de dados dispõem desses recursos de documentação. Escala de avaliação : <u>Total</u> (Excelente), o conjunto de dados apresenta todos os recursos; <u>Parcial</u> (Aceitável), o conjunto de dados apresenta alguns recursos; <u>Não</u> (Crítico), não apresenta qualquer recurso explicativo. O avaliador deverá analisar se os recursos estão presentes e indicar no campo específico conforme o modelo.		

Métrica:	Consistência na linguagem O conjunto de dados utiliza nomenclaturas dos campos, termos e formatos padronizados?	Avaliação:	Crítico
Fórmula	Consistência na linguagem = $\frac{\text{Número de campos/termos consistentes}}{\text{Total de registros comparáveis entre os conjuntos}} \times 100$	8	80
		10	

Instruções	A métrica avalia o uso padronizado de termos, nomes de campos, abreviações, formatos e estruturas entre diferentes conjuntos de dados. Para a análise, o avaliador deverá verificar se a linguagem utilizada no conjunto de dados apresenta consistência em relação aos conjuntos de dados de exercícios anteriores e a outros conjuntos de dados do mesmo órgão disponibilizador. Para a avaliação, divide-se o número de termos consistentes pelo número total de termos analisados, multiplicando-se o resultado por 100. Exemplo: se o avaliador analisar 10 campos (nomes de colunas) de um conjunto de dados e constatar que 8 deles mantêm a mesma nomenclatura utilizada no ano anterior (consistentes) e 2 foram alterados, o cálculo será: $(8/10) \times 100 = 80\%$ de consistência.		

Métrica:	Descrições claras e detalhadas dos dados Os dados possuem descrições claras, incluindo definições, unidades de medida e explicações para codificações utilizadas?	Avaliação:	Crítico
Fórmula	Descrições claras e detalhadas = $\frac{\text{Número de registros com divergências}}{\text{Total de registros verificados}} \times 100$	3 10	30

Instruções	A métrica avalia o grau de clareza das informações disponibilizadas, verificando se os dados são acompanhados de descrições acessíveis que permitam a interpretação inequívoca por usuários de diferentes níveis técnicos. O procedimento de análise consiste no análise da clareza descritiva, observando especificamente se o órgão produtor detalha as unidades de medida e as codificações utilizadas. Para a mensuração, calcula-se a razão entre o número de termos que apresentam divergências (ou ausência de detalhamento) e o número total de termos verificados, multiplicando-se o resultado por 100. Exemplo: se o avaliador verificar as descrições de 10 colunas de uma tabela e identificar que 3 delas não informam a unidade de medida ou possuem explicações confusas (divergências), o cálculo será: $(3/10) \times 100 = 30\%$ de descrições com problemas de clareza.		

Métrica:	Linguagem adequada a diferentes públicos A linguagem empregada é simples, direta e acessível, facilitando o entendimento por diversos perfis de usuários?	Avaliação:	Crítico
Fórmula	Índice de qualidade (%) = $\frac{\text{Pontuação obtida}}{\text{Pontuação máxima}} \times 100$	5 9	56
Recursos avaliados	Descrição dos recursos	Avaliação recursos (3 - excelente/ 2 - aceitável/ 1- Crítico)	
Termos	Uso de termos simples e de fácil compreensão	3	
Frases	Construção de frases diretas e concisas	1	
Jargões técnicos	Uso de jargões técnicos ou, quando seu uso for necessário, a presença de explicações ou resoluções adequadas	1	
		5	

Instruções	A métrica mensura o nível de acessibilidade da linguagem utilizada no conjunto de dados, verificando sua adequação a diferentes perfis de usuários, desde o cidadão comum até especialistas. O objetivo é identificar se o vocabulário é restritivo ou excessivamente técnico. A avaliação baseia-se na análise de três recursos fundamentais: termos, estrutura frasal e uso de jargões técnicos. Para cada recurso, atribui-se uma pontuação na escala: <u>Excelente</u> (3 pontos); <u>Aceitável</u> (2 pontos); <u>Crítico</u> (1 ponto). O índice final é calculado pela soma das notas atribuídas dividida pela pontuação máxima possível, multiplicada por 100." Exemplo: na análise realizada, verificou-se que os termos empregados apresentam avaliação Excelente (3 pontos). Por outro lado, as frases carecem de objetividade e concisão (Crítico: 1 ponto) e notou-se o uso de jargões técnicos sem as devidas explicações (Crítico: 1 ponto). O somatório da avaliação resulta em 5 pontos. Aplicando-se a fórmula obtém-se um índice de aproximadamente 56%.

Métrica:	Valores Padronizados Os dados seguem formatos padronizados (ex: datas no padrão ISO, uso correto de separadores decimais, símbolos monetários e unidades), o que favorece a leitura e reutilização dos dados?	Avaliação:	Crítico
Fórmula	$\text{Índice de qualidade (\%)} = \frac{\text{Pontuação obtida}}{\text{Pontuação máxima}} \times 100$	8 12	67
Recursos avaliados	Descrição dos recursos	Avaliação recursos (3 - excelente/ 2 - aceitável/ 1- Crítico)	
Formatos	Os formatos são consistentes	3	
Medida	Indicação clara de unidade de medida	2	
Valores	Valores monetários uniformes	1	
Padronização	Padronização para números decimais e porcentagens	2	
		8	

Instruções	A métrica Valores Padronizados verifica a conformidade dos dados em relação a padrões técnicos estabelecidos e amplamente reconhecidos (como normas ISO ou ABNT) para a representação de datas, moedas, unidades de medida e valores numéricos. A avaliação baseia-se na análise de quatro recursos fundamentais: formatos, medidas, valores e padronização. Para cada recurso, atribui-se uma pontuação na escala: <u>Excelente</u> (3 pontos);

	<u>Aceitável</u> (2 pontos); <u>Crítico</u> (1 ponto). O índice final é calculado pela soma das notas atribuídas dividida pela pontuação máxima possível, multiplicada por 100. Exemplo: na análise realizada foi verificado que os formatos de data são consistentes (ISO 8601), recebendo avaliação Excelente (3). No entanto, as unidades de medida variam dentro da mesma coluna (ex: 'kg' e 'gramas'), sendo consideradas Aceitáveis (2). Os valores decimais apresentam mistura de vírgulas e pontos, dificultando a leitura automática, o que configura estado Crítico (1). Por fim, a padronização geral segue parcialmente as normas do órgão regulador, recebendo nota Aceitável (2). O somatório da avaliação resulta em 8 pontos. Aplicando-se a fórmula obtém-se um índice de aproximadamente 67%.
Nota	Os critérios detalhados para a aplicação de cada nota/pontuação estão disponíveis nas orientações da ferramenta, acessíveis através da célula de resultado. Os campos digitáveis da planilha encontram-se destacados em cinza

Dimensão:	CONFORMIDADE		
Métrica:	Proteção de dados pessoais conforme legislação O conjunto de dados atende aos princípios da LGPD. Não apresentando dados pessoais identificáveis, dados sensíveis individualizados, possui finalidade legítima e clara e não existe risco de reidentificação de indivíduos?	Avaliação:	Aceitável
Recursos avaliados	Descrição dos recursos	Avaliação recursos (3 - Excelente/ 2 - Aceitável/ 1 - Crítico)	
Dados pessoais	Os dados pessoais não estão sendo identificados	3	
Justificativa	Justificativa legal para o tratamento de dados	3	
Dados	Os dados foram anonimizados ou agregados adequadamente	1	
Reidentificação	Existe algum risco de reidentificação com cruzamento de dados	1	
		8	

Instruções	A métrica Proteção de Dados Pessoais verifica a conformidade dos conjuntos de dados com as diretrizes da Lei Geral de Proteção de Dados (LGPD - Lei nº 13.709/2018), assegurando que informações pessoais e sensíveis não sejam expostas indevidamente. A avaliação baseia-se na análise de quatro recursos fundamentais: dados pessoais, justificativa, dados e reidentificação. Para cada recurso, atribui-se uma pontuação na escala: <u>3 (Excelente)</u> , <u>2 (Aceitável)</u> ou <u>1 (Crítico)</u> . As orientações detalhadas para a pontuação podem ser acessadas clicando com o botão direito do mouse no canto superior esquerdo do campo destinado à avaliação. Exemplo: na análise realizada, verificou-se que o tratamento de dados pessoais é Excelente (3), pois não constam identificadores diretos, e a justificativa legal para a publicação está clara, sendo também avaliada como Excelente (3). No entanto, foram encontrados dados sensíveis sem a devida anonimização, o que configura nível Crítico (1), e detectou-se risco de reidentificação através do cruzamento de variáveis indiretas, também avaliado como Crítico (1).		

Métrica:	Origem dos Dados O conjunto de dados indica a procedência dos dados?	Avaliação:	Aceitável

Recursos avaliados	Descrição dos recursos	Avaliação recursos (1 - Total/ 2 - Parcial/ 3 - Não)	
Fonte de dados	A fonte de dados está claramente indicada (Nome do órgão, base ou sistema de origem)	1	
Origem	Tipo de origem(interna, externa ou mista) é identificável. Transparência do fluxo de dados	3	
Documentação	Existe documentação explicando a procedência (links, metadados ou notas explicativas)	3	
		7	

Instruções	Esta métrica examina a rastreabilidade e a transparência quanto à procedência das informações, verificando a possibilidade de identificar inequivocamente se os dados derivam de processos internos do órgão, de fontes externas ou da integração de múltiplas bases. A avaliação fundamenta-se na análise de três componentes estruturantes: a declaração das fontes de dados, a especificação da origem e a documentação de suporte. Para cada componente, atribui-se uma pontuação conforme a escala de conformidade: Total (1 ponto), Parcial (2 pontos); Não (3 pontos). No instrumento de avaliação, as diretrizes detalhadas para a atribuição da pontuação estão dispostas como notas explicativas na interface de preenchimento. Exemplo: O avaliador verificou que a fonte de dados está indicada (1 ponto), porém, o conjunto de dados não informa o tipo de origem (3) e não existe documentação sobre a procedência dos dados (3).		

Métrica:	Responsável em caso de violação de direitos ou uso inadequado dos dados O conjunto de dados apresenta identificação clara do controlador e do encarregado (DPO), bem como canais de contato e informações?	Avaliação:	Aceitável
Recursos avaliados	Descrição dos recursos	Avaliação recursos (1 - Sim/ 2 - Não)	
Controlador	Identificação do controlador dos dados?	2	
Encarregado (DPO)	Nomeação do encarregado (DPO) com contato acessível	2	
Responsabilidades	Indicação de responsabilidades conforme LGPD	2	
Denúncias	Existência de canal de denúncia ou solicitação de direito do cidadão	1	
		7	

Instruções	A métrica verifica se o conjunto de dados ou o portal de disponibilização identifica formalmente os agentes responsáveis pelo tratamento da informação, em conformidade com os requisitos da Lei Geral de Proteção de Dados Pessoais (Lei nº 13.709/2018). O objetivo é assegurar a transparência quanto à identidade e às atribuições legais das figuras do controlador, do operador e do encarregado de dados. A avaliação fundamenta-se na análise da presença de quatro componentes essenciais: identificação do controlador, identificação do encarregado, definição de responsabilidades e existência de canais para denúncias. Para cada componente, atribui-se uma pontuação conforme a escala de conformidade binária: Sim (1 ponto) ou Não (2 pontos). Exemplo: Ao analisar um conjunto de dados é verificado que apenas a identificação do canal de denúncias, que será atribuído pontuação 1, os demais recursos não estão presentes no conjunto de dados, atribui-se pontuação 2.
Nota	Os critérios detalhados para a aplicação de cada nota/pontuação estão disponíveis nas orientações da ferramenta, acessíveis através da célula de resultado. Os campos digitáveis da planilha encontram-se destacados em cinza

Dimensão:	LICENÇA ABERTA		
Métrica:	Existência de licença A licença está claramente indicada na página do conjunto de dados, é padronizada, possui link acessível para o texto completo e está em local de fácil visualização?	Avaliação:	Aceitável
Recursos avaliados	Descrição dos recursos	Avaliação recursos (1 - Sim/ 2 - Não)	
Licença	A licença está indicada na página inicial do conjunto de dados	1	
Licença	A licença é padronizada	2	
Licença	O link para o texto completo da licença está disponível	2	
Licença	A informação sobre a licença está em local de fácil acesso	1	
		6	
Instruções	Esta métrica verifica a presença de uma declaração explícita, visível e facilmente identificável que regule o uso do conjunto de dados. A mensuração é realizada por meio de um <i>checklist</i> composto por quatro critérios de verificação. Para cada critério atendido (Sim), atribui-se 1 ponto; para cada critério não atendido (Não), atribuem-se 2 pontos. A classificação final é dada pelo somatório dessas notas, seguindo a escala: <u>Total</u> (Excelente - 4 pontos), quando todos os critérios de licenciamento estão presentes e adequados; <u>Parcial</u> (Aceitável - 5 a 7 pontos), quando apenas parte dos aspectos regulatórios é atendida; e <u>Não</u> (Crítico - 8 pontos), quando não existe indicação válida sobre a licença ou todos os critérios são inadequados.		
Métrica:	Tipos de licença A licença aplicada é compatível com os princípios dos dados abertos (ex: CC0, CC BY, ODC-BY), permitindo uso, redistribuição e modificação dos dados?	Avaliação:	Aceitável
		Licença	Parcial

Instruções	A métrica tipo de licença avalia se a licença aplicada ao conjunto de dados é compatível com os princípios dos dados abertos, ou seja, se ela permite o uso, redistribuição, reutilização e modificação dos dados por qualquer pessoa, para qualquer finalidade. A mensuração é realizada por meio da verificação dos tipos de licenças aplicadas no conjunto de dados. Escala de avaliação: <u>Total</u> (Excelente: licença plenamente compatível com os princípios de dados abertos permite uso, redistribuição, reutilização e modificação por qualquer pessoa, para qualquer finalidade); <u>Parcial</u> (Aceitável: licença parcialmente aberta. Permite acesso e uso, mas com restrições que limitam o reuso pleno); <u>Não</u> (Crítico: licença inexistente, não informada ou incompatível com dados abertos).		

Métrica:	Permissão para reutilização A licença autoriza expressamente o uso, redistribuição e modificação dos dados, sem limitações como uso não comercial ou proibição de obras derivadas?	Avaliação:	Aceitável
		Total/Parcial/Não	Parcial

Instruções	Essa métrica avalia se a licença aplicada ao conjunto de dados permite a: Reutilização; Redistribuição; Modificação e Limitações dos dados. A mensuração é realizada por meio da verificação das licenças aplicadas no conjunto de dados. Escala de avaliação: <u>Total</u> (Excelente: a licença permite reutilização irrestrita, redistribuição e modificação, sem restrições significativas); <u>Parcial</u> (Aceitável: A licença permite reutilização parcial, mas impõe restrições que limitam o uso dos dados); <u>Não</u> (Crítico: A licença não permite reutilização, redistribuição ou modificação). Exemplo: O conjunto de dados avaliado permite apenas a reutilização de alguns dados, então a avaliação é Parcial (Aceitável).		

Métrica:	Período de validade para o uso dos dados A licença é irrevogável e sem prazo de validade definido, assegurando a continuidade no uso dos dados sem restrições futuras ou expirabilidade?	Avaliação:	Excelente
		Total/Parcial/Não	Total

Instruções	Avalia a segurança jurídica quanto à continuidade do direito de uso dos dados, verificando se há prazos de expiração ou possibilidade de revogação da permissão. O procedimento consiste na análise das cláusulas de vigência e rescisão nos termos de uso. A avaliação seguirá a seguinte escala: <u>Total</u> (Excelente: licença é indeterminada e irrevogável); <u>Parcial</u> (Aceitável: a licença é indeterminada, mas contém cláusulas de		

	revogabilidade condicionada); <u>Não</u> (Crítico: licença possui prazo de validade definido ou pode ser revogada unilateralmente pelo fornecedor de dados). Exemplo: Os dados estão sob a licença <i>Creative Commons Attribution</i> (CC-BY 4.0). O texto legal da licença afirma explicitamente que ela é "irrevogável" (desde que o usuário continue dando os créditos). O desenvolvedor tem certeza de que seu aplicativo poderá usar esses dados para sempre.

Métrica:	Ausência de restrições (domínio público) O conjunto de dados analisado segue a Política de Dados Abertos, ou seja, são disponibilizados sob o domínio público, sem restrições legais?	Avaliação:	Aceitável
		Total/Parcial/Não	Parcial

Instruções	A métrica avalia se o conjunto de dados está livre de embaraços legais decorrentes de direitos autorais, patentes ou outras formas de propriedade intelectual. O procedimento consiste em analisar os termos de uso para verificar se os dados estão em domínio público ou se há incidência de direitos proprietários. A avaliação segue a escala: <u>Total</u> (Excelente - Domínio Público), atribuído quando o conjunto de dados está explicitamente em domínio público ou sob licença equivalente (como CC0), garantindo isenção total de direitos; <u>Parcial</u> (Aceitável - Restrições Mínimas), aplicado quando o conjunto é liberado sob licença aberta que impõe apenas obrigações simples (como a atribuição de autoria - CC-BY), sem impedir o reuso significativo; <u>Não</u> (Crítico - Restrições Totais), quando existem restrições legais severas ou bloqueios proprietários que impedem ou limitam drasticamente a utilização dos dados.

Métrica:	Declaração de domínio público O conjunto de dados está livre de direitos autorais, patentes ou qualquer outro tipo de restrição legal que limite seu uso, indicando domínio público?	Avaliação:	Aceitável
		Total/Parcial/Não	Parcial

Instruções	A métrica declaração de domínio público avalia a formalização do status jurídico dos dados, verificando a existência de uma declaração clara, explícita e acessível de que o conjunto pertence ao domínio público, isento de restrições de direitos autorais. O procedimento consiste em analisar a documentação, os termos de uso e os metadados em busca de manifestações expressas ou selos (como CC0) que garantam essa condição. A avaliação segue a escala: <u>Total</u> (Excelente), atribuído quando a declaração de domínio público é inequívoca, facilmente localizável junto aos dados e amparada por políticas institucionais visíveis; <u>Parcial</u> (Aceitável),

	aplicado quando a declaração existe, mas é ambígua, genérica ou de difícil localização, gerando incerteza sobre a abrangência da renúncia de direitos; <u>Não</u> (Crítico), quando não existe qualquer declaração de domínio público associada ao conjunto de dados.
Nota	Os critérios detalhados para a aplicação de cada nota/pontuação estão disponíveis nas orientações da ferramenta, acessíveis através da célula de resultado. Os campos digitáveis da planilha encontram-se destacados em cinza

Dimensão:	USABILIDADE		
Métrica:	Conjunto de dados disponível para download sem necessidade de autenticação O conjunto de dados está disponível para acesso e download direto, sem barreiras técnicas ou administrativas?	Avaliação:	Aceitável
Fórmula	$\text{Índice de qualidade (\%)} = \frac{\text{Pontuação obtida} \times 100}{\text{Pontuação máxima}}$	10 12	83
Recursos avaliados	Descrição dos recursos	Avaliação recursos (3 - excelente/ 2 - aceitável/ 1- Crítico)	
Acesso	Acesso anônimo	3	
Solicitações	Sem formulários ou solicitações	3	
Verificação humana	Barreiras de verificação humana ou scripts de bloqueio	3	
Link	Link estável e direto	1	
		10	

Instruções	A métrica avalia se o acesso ao conjunto de dados é direto ou se impõe obstáculos ao usuário. Verifica a existência de exigências como <i>login</i> (autenticação), formulários de solicitação prévia, barreiras de verificação humana (CAPTCHA) ou scripts de bloqueio automatizado. A avaliação é baseada em um sistema de pontuação para 4 recursos específicos: Acesso (<i>Login</i>), Solicitações (Formulários), Verificação Humana (CAPTCHA) e Integridade do Link. Para avaliação de cada recurso atribuída a pontuação: 3 (Excelente); 2 (Aceitável); 1 (Crítico) para cada recurso. Soma-se a pontuação obtida nos 4 itens e divide-se pela pontuação máxima possível (12 pontos). O resultado é multiplicado por 100 para obter o percentual. O resultado é multiplicado por 100. Exemplo: um avaliador tenta baixar um arquivo no portal: <i>Login</i> não exigido (3 pontos); Solicitações não exigida (3 pontos); verificação humana exigida (1 ponto); Link funciona mais redireciona para outra página antes de baixar (2 pontos).

Métrica:	Interface amigável O Portal oferece uma interface amigável, acessível e eficiente, onde usuários de diferentes tipos de familiaridade com sites conseguem navegar e baixar as informações?	Avaliação:	Aceitável
Recursos avaliados	Descrição dos recursos	Avaliação recursos (1 - Sim/ 2 - Não)	
Menu	Presença de menu simples, com categorias bem definidas na página inicial	2	
Downloads	Os botões para baixar os conjunto de dados são grandes, claramente identificados e agrupados por tema	1	
Cores	Utilização de cores que garantam bom contraste e fácil leitura	1	
Títulos	Títulos apresentados em negrito	1	
Navegação	Anúncios, pop-ups ou outros elementos que interrompam a navegação	2	
		7	

Instruções	A métrica interface amigável Analisa a usabilidade e a acessibilidade visual do portal onde os dados estão hospedados. O objetivo é verificar se a interface é intuitiva e eficiente, permitindo que diferentes perfis de usuários consigam navegar e baixar as informações sem dificuldades. A mensuração é realizada através de um <i>checklist</i> de 5 recursos essenciais: Menu, Downloads, Cores (Contraste), Títulos e Navegação. Para cada recurso o avaliador atribuirá 1 ponto se o recurso estiver adequado (sim - 1 pontos) se o recurso não estiver adequado (não - 2 ponto). A avaliação final é a soma dessas notas. Exemplo: O avaliador verificou o portal e constatou que o Menu é confuso (Nota 2) e a navegação trava (Nota 2), mas os downloads, cores e títulos estão corretos (Nota 1 cada). Cálculo: $2 + 2 + 1 + 1 + 1 = 7$. Resultado: Como a soma é 7 - Aceitável

Nota	Os critérios detalhados para a aplicação de cada nota/pontuação estão disponíveis nas orientações da ferramenta, acessíveis através da célula de resultado. As células digitáveis da planilha encontram-se destacadas em cinza.

Dimensão:	NÃO DISCRICIONÁRIO		
Métrica:	Dados disponíveis diretamente ao usuário Os dados estão disponíveis de forma imediata, sem depender de solicitação por meio de canais como o Serviço de Informação ao Cidadão (SIC) ou requerimentos adicionais?	Avaliação:	Aceitável
		Total/Parcial/Não	Parcial
Instruções	A métrica Avalia se a disponibilização dos dados é proativa, permitindo o acesso imediato sem que o usuário precise formalizar pedidos prévios via Serviço de Informação ao Cidadão (SIC) ou trâmites burocráticos similares. Para avaliar o recurso o avaliador verifica a disponibilidade direta dos dados. Escala avaliação: <u>Total</u> (todos os dados e recursos do conjunto são acessíveis diretamente, sem necessidade de solicitação); <u>Parcial</u> (penas uma parte dos dados está disponível imediatamente; para acessar o restante é necessário fazer uma solicitação); <u>Não</u> (Restrito): Para obter qualquer parte do conjunto de dados, é obrigatório realizar uma solicitação formal. Exemplo: Um portal exige o conjunto de dados Gastos Públicos. O usuário consegue baixar as planilhas de 2024 e 2025 diretamente. Porém, ao tentar acessar o histórico de 2020 a 2023, é necessário solicitar ao SIC. Como o acesso não é completo para todo o conjunto, a avaliação é Parcial.		
Métrica:	Custo de acesso O acesso aos dados é gratuito, sem cobrança de taxas, pagamentos ou assinatura para obtenção do conteúdo?	Avaliação:	Crítico
		Total/Parcial/Não	Não
Instruções	Esta métrica avalia a existência de barreiras financeiras no acesso aos dados governamentais. Segundo os princípios de Dados Abertos, os conjuntos de dados devem ser disponibilizados ao público preferencialmente de forma gratuita, fomentando o reuso e a transparência. O avaliador deve verificar a disponibilidade direta do recurso, simulando o processo de obtenção do dado para identificar a existência de taxas de assinatura, custos de licença ou taxas de processamento. Escala de avaliação: <u>Total</u> (Excelente: os dados estão integralmente disponíveis de forma gratuita); <u>Parcial</u> (Aceitável: o acesso aos dados é majoritariamente gratuito, mas existem taxas para volumes massivos, taxas administrativas justificadas ou necessidade de cadastro pago para		

	recursos não essenciais; <u>Não</u> (Crítico: existem barreiras financeiras significativas).
Nota	Os critérios detalhados para a aplicação de cada nota/pontuação estão disponíveis nas orientações da ferramenta, acessíveis através da célula de resultado. As células digitáveis da planilha encontram-se destacadas em cinza.

Dimensão:	PRIMÁRIO		
Métrica:	Fonte de dados O conjunto de dados declara de forma explícita a fonte de origem dos dados?	Avaliação:	Crítico
		Total/Parcial/Não	Não
Instruções	Avalia se a origem primária dos dados é declarada de forma explícita e inequívoca. A identificação correta da fonte permite ao usuário verificar a autoridade produtora da informação e garante a atribuição correta de autoria. O avaliador deve examinar o conjunto de dados em busca da identificação do órgão, departamento ou entidade responsável pela criação ou gestão do dado. Para avaliação será utilizada a escala: <u>Total</u> (Excelente: a fonte dos dados é claramente identificada); <u>Parcial</u> (Aceitável: a fonte de dados é mencionada de forma genérica); <u>Não</u> (Crítico: a fonte de dados não é mencionada).		
Métrica:	Processo de dados documentais Há documentação acessível sobre o processo de coleta, tratamento, validação e publicação dos dados?	Avaliação:	Aceitável
		Total/Parcial/Não	Parcial
Instruções	Esta métrica avalia se o órgão disponibiliza documentação que possibilite rastrear a trajetória e o tratamento das informações contidas no conjunto de dados. O procedimento consiste em examinar o recurso para verificar a existência de registros sobre o ciclo de vida dos dados. Escala de avaliação: <u>Total</u> (Excelente: a documentação é detalhada e de fácil acesso descrevendo todas as etapas do ciclo de vida dos dados); <u>Parcial</u> (Aceitável: a documentação é parcial ou não cobre todas as etapas do processo, limitando a compreensão plena da origem e do tratamento dos dados. A documentação é de difícil localização, tornando-se um obstáculo para o usuário); <u>Não</u> (Crítico: não há qualquer identificação sobre a origem dos dados ou ela é insuficiente).		
Métrica:	Rastreabilidade dos dados É possível acompanhar o histórico dos dados desde sua origem até sua publicação?	Avaliação:	Aceitável

Recursos avaliados	Descrição dos recursos	Avaliação recursos (1 - Sim/ 2 -Não)	
Modificações	Alterações realizadas nos dados ao longo do tempo (correções, atualizações)	1	
Integrações	Combinações com outras fontes de dados ou fusões de bases distintas.	2	
Processamento	Transformações aplicadas, como limpeza, agregações, cálculos entre outros.	2	
		5	

Instruções	Esta métrica avalia se o conjunto de dados possui um histórico documentado, acessível e transparente sobre todas as etapas pelas quais os dados passaram, desde sua origem até a versão atual publicada. O procedimento consiste em examinar o conjunto de dados para verificar se as informações podem ser rastreadas em todo o seu ciclo de vida, da origem à publicação. A avaliar da métrica seguirá os seguintes requisitos: Para cada recurso presente no conjunto de dados será atribuído a avaliação <u>1</u>; se o recurso não estiver presente será atribuído a avaliação <u>2</u>. Exemplo: O avaliador verificou o portal e constatou a existência de problemas no processamento (Nota 2) e a integração com outras fontes apresenta problemas (Nota 2), mas as modificações estão corretas (Nota 1 cada). Cálculo: $1+2+2 = 5$. Resultado: Como a soma é 5 - Aceitável.

Nota	Os critérios detalhados para a aplicação de cada nota/pontuação estão disponíveis nas orientações da ferramenta, acessíveis através da célula de resultado. As células digitáveis da planilha encontram-se destacadas em cinza.

Dimensão:	INTEROPERABILIDADE		
Métrica:	Capacidade de reutilização As informações sobre reuso e a quantidade de downloads realizados do conjunto de dados estão disponíveis para o usuário?	Avaliação:	Excelente
		Total/Parcial/Não	Total
Instruções	A métrica capacidade de reutilização avalia se o portal ou conjunto de dados disponibiliza indicadores de impacto, permitindo compreender o grau de interesse e o aproveitamento das informações pela sociedade. O procedimento consiste em verificar a existência de métricas quantitativas (downloads) e qualitativas (exemplos de aplicações ou reuso). Escala de avaliação: <u>Total</u> (Excelente: as informações sobre número de reuso e download são claras, atualizadas e acessíveis); <u>Parcial</u> (Aceitável: as informações são parciais, contemplando apenas um dos aspectos (reuso ou download)); <u>Não</u> (Crítico: não existe qualquer informação disponível sobre reuso ou downloads do conjunto de dados).		
Métrica:	Integração entre sistemas O conjunto de dados apresenta mecanismos técnicos que permitem a integração dos dados a outras plataformas?	Avaliação:	Excelente
Recursos avaliados	Descrição dos recursos	Avaliação recursos (Sim/Não)	
		Total/Parcial/Não	Total
Instruções	A métrica capacidade dos dados serem automaticamente conectados, consumidos e integrados por sistemas externos, como ferramentas de <i>Business Intelligence</i> (BI), ERPs e CRMs. O procedimento consiste na verificação dos mecanismos técnicos de disponibilização para identificar o suporte a interfaces de programação ou padrões de troca de dados Escala de avaliação: <u>Total</u> (Excelente: os dados oferecem <i>APIs RESTful</i> em padrão aberto (JSON ou Similar), com documentação interativa, autenticação clara e suporte a conectores/bibliotecas que facilitam a integração.); <u>Parcial</u> (Aceitável: os dados oferecem algum recurso de integração, como uma API simples ou exportação em formato aberto, mas com documentação incompleta ou		

	ausência de bibliotecas auxiliares; <u>Não</u> (Crítico: Não existe suporte para integração em sistemas externos, limitando o uso dos dados a downloads manuais ou formatos pouco interoperáveis).
Nota	Os critérios detalhados para a aplicação de cada nota/pontuação estão disponíveis nas orientações da ferramenta, acessíveis através da célula de resultado. As células digitáveis da planilha encontram-se destacadas em cinza.

PAINEL CONSOLIDADO DE AVALIAÇÃO DA QUALIDADE DOS DADOS GOVERNAMENTAIS ABERTOS

QUADRO 1

Dimensão	Excelente	Aceitável	Crítico	Total de métricas	Excelente(%)	Aceitável(%)	Crítico (%)
Acessibilidade	0	0	1	1	0,00	0,00	100
Compleitude	2	2	2	6	33,33	33,33	33,33
Atualidade	2	2	1	5	40,00	40,00	20,00
Formato de arquivo	1	2	1	4	25,00	50,00	25,00
Disponibilidade	1	2	1	4	25,00	50,00	25,00
Precisão	1	2	1	4	25,00	50,00	25,00
Recuperabilidade	3	3	0	6	50,00	50,00	0,00
Inconsistência	0	1	3	4	0,00	25,00	75,00
Compreensibilidade	0	0	5	5	0,00	0,00	100,00
Conformidade	0	3	0	3	0,00	100,00	0,00
Licença aberta	1	5	0	6	16,67	83,33	0,00
Usabilidade	0	2	0	2	0,00	100,00	0,00
Não discricionário	0	1	1	2	0,00	50,00	50,00
Primário	0	2	1	3	0,00	66,67	33,33
Interoperabilidade	2	0	0	2	100,00	0,00	0,00

QUADRO 2

	Total métricas	Porcentagem
Excelente	13	22,81
Aceitável	27	47,37
Crítico	17	29,82

QUADRO 3	
Dimensão com maior % de métricas excelentes	Interoperabilidade
Maior % Excelente	100
Dimensão com maior % de métricas críticas	Acessibilidade
Maior % Crítica	100

QUADRO 4		
Dimensão	Métricas	Avaliação
Acessibilidade	Conformidade com as normas de acessibilidade	Crítico
Compleitude	Valores ausentes	Aceitável
	Cobertura de dados coletados	Crítico
	Campos obrigatórios	Excelente
	Comparabilidade	Aceitável
	Metadados	Crítico
	Distribuição temporal	Excelente
Atualidade	Idade dos dados	Excelente
	Frequência de atualização	Excelente
	Tempo entre a ocorrência do evento e do registro no sistema	Aceitável
	Presença de data e hora da coleta de dados	Crítico

QUADRO 5				
AVALIAÇÃO FINAL DA DIMENSÃO				
Dimensão	Pontuação Total	Soma da avaliação real	Desempenho da dimensão	Avaliação Final
Acessibilidade	10	0	0	Crítico
Compleitude	60	30	50	Aceitável
Atualidade	50	30	60	Aceitável
Formato de arquivo	40	20	50	Aceitável
Disponibilidade	40	20	50	Aceitável
Precisão	40	20	50	Aceitável
Recuperabilidade	60	45	75	Aceitável
Inconsistência	40	5	13	Crítico
Compreensibilidade	50	0	0	Crítico
Conformidade	30	15	50	Aceitável
Licença aberta	60	35	58	Aceitável
Usabilidade	20	10	50	Aceitável
Não discricionário	20	5	25	Crítico
Primário	30	10	33	Crítico

	Atualização de metadados	Aceitável
Formato do Arquivo	Formato dos dados acessíveis	Excelente
	Multiplicidade de formatos	Aceitável
	Dados disponíveis em formato compacto	Crítico
	Documentação do formato	Aceitável
Disponibilidade	Frequência dos período indisponíveis	Aceitável
	Hiperlink de downloads funcional	Excelente
	Disponibilidade de canal de comunicação com o usuário	Aceitável
	Múltiplos idiomas	Crítico
Precisão	Taxa de dados corrompidos	Crítico
	Dados duplicados	Aceitável
	Grau de correspondência dos dados com a realidade	Aceitável
	Ausência de ambiguidades	Excelente
Recuperabilidade	Função de recuperação de dados	Excelente
	Eficiência de busca	Aceitável
	Navegação	Excelente

Interoperabilidade	20	20	100	Excelente

QUADRO 6	
AVALIAÇÃO FINAL DO CONJUNTO DE DADOS	
Peso Total do conjunto de dados	570
Soma da avaliação real	265
Cálculo do índice	46
Avaliação final do conjunto de dados	Crítico

QUADRO 7				
CONTRIBUIÇÃO GERAL DE CADA DIMENSÃO PARA A AVALIAÇÃO				
Dimensão	Desempenho da Dimensão	Contribuição da dimensão no índice geral	Peso Atribuído	Peso Determinado
Acessibilidade	0	0	1	0
Compleitude	0,5	0,03	1	0,03
Atualidade	0,6	0,04	1	0,04
Formato de arquivo	0,5	0,03	1	0,03
Disponibilidade	0,5	0,03	1	0,03
Precisão	0,5	0,03	1	0,03

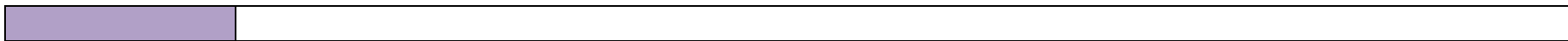
	Taxa de falhas na recuperação de dados	Aceitável
	Metadados recuperáveis	Excelente
	Vocabulários controlados	Aceitável
Inconsistência	Coerência lógica	Crítico
	Consistência entre os conjuntos de dados	Crítico
	Coerência em dados agregados	Crítico
	Uso de classificações	Aceitável
Compreensibilidade	Presença de dicionários notas explicativas ou legendas	Crítico
	Origem dos dados	Aceitável
	Responsável em caso de violação de direitos ou uso inadequado dos dados	Aceitável
Licença aberta	Existência de licença	Aceitável
	Tipo de licença	Aceitável
	Permissão para reutilização	Aceitável
	Período de validade para uso dos dados	Excelente
	Ausência de restrições	Aceitável
	Declaração de domínio público	Aceitável

Recuperabilidade	0,8	0,05	1	0,05
Inconsistência	0,1	0,01	1	0,0083
Compreensibilidade	0,0	0,00	1	0
Conformidade	0,5	0,03	1	0,03
Licença aberta	0,6	0,04	1	0,039
Usabilidade	0,5	0,03	1	0,033
Não discricionário	0,3	0,02	1	0,017
Primário	0,3	0,02	1	0,022
Interoperabilidade	1,0	0,07	1	0,067
		SOMA	15	

Usabilidade	Conjunto de dados disponível para download sem necessidade de autenticação	Aceitável
	Interface amigável	Aceitável
Não discricionário	Dados disponíveis diretamente ao usuário	Aceitável
	Custo de acesso	Crítico
Primário	Fonte de dados	Crítico
	Processo dos dados documentais	Aceitável
	Rastreabilidade dos dados	Aceitável
Interoperabilidade	Capacidade de reutilização	Excelente
	Integração com sistemas externos	Excelente

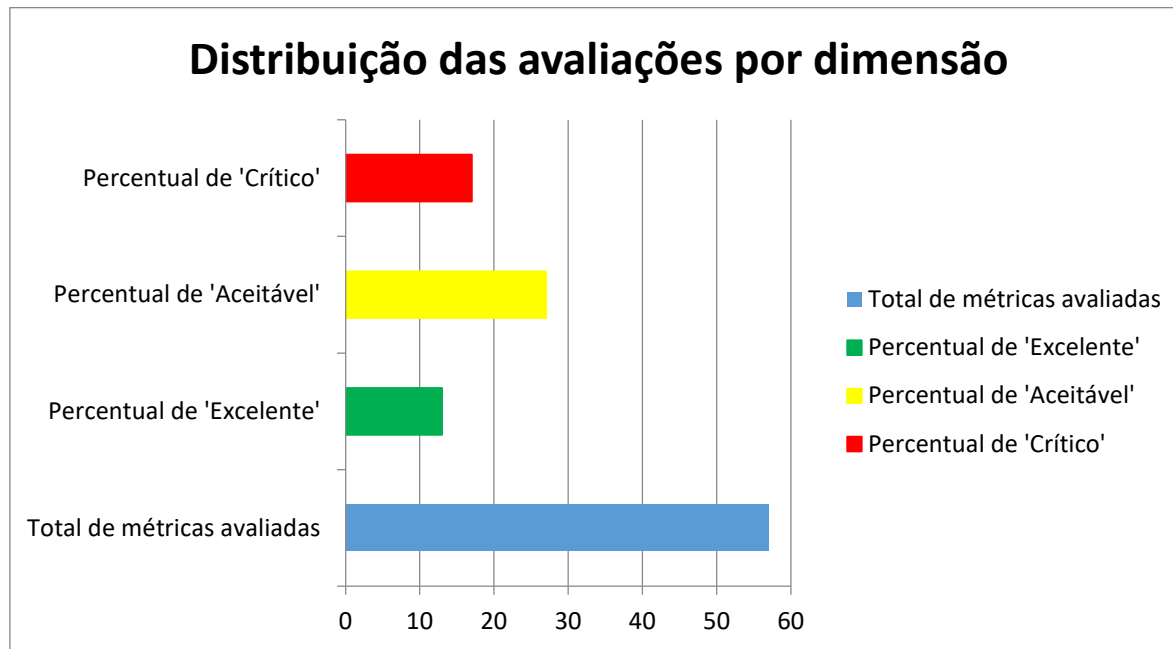
Notas	1 - O Quadro 1 detalha, para cada dimensão, a quantidade de métricas enquadradas nas categorias excelente, aceitável e crítico, juntamente com a porcentagem relativa a cada avaliação.
	2 - O Quadro 2 sintetiza o total de métricas avaliadas como excelentes, aceitáveis e críticas referentes ao conjunto de dados.
	3 - O Quadro 3 apresenta a dimensão com a maior porcentagem de métricas classificadas como excelentes e aquela com a maior porcentagem de métricas avaliadas como críticas. Além disso, evidencia a maior proporção de métricas excelentes e a maior proporção de métricas críticas dentre todas as dimensões analisadas.
	4 - O Quadro 4 apresenta os resultados consolidados da avaliação das métricas por dimensão.
	5 - Para calcular o desempenho de cada dimensão, siga as etapas abaixo utilizando a escala de pontuação: Excelente (10), Aceitável (05) e Crítico (00). Identificação da Pontuação Total: Multiplique a quantidade de métricas da dimensão pela pontuação máximo (10). Este valor representa o potencial máximo (100%).

	Soma da Avaliação Real: Some os pontuações atribuídas a cada métrica conforme o desempenho observado (10, 05 ou 00). Desempenho da dimensão: Aplique a fórmula: Soma da avaliação dividido pela pontuação total multiplicado por 100. Avaliação Final: Excelente: De 90% a 100% Aceitável: De 50% a 89% Crítica: Menor que 49%
	6 - O avaliação final do conjunto de dados é consolidado através da média ponderada de todas as dimensões avaliadas: Soma da Pontuação Total (Potencial): Soma-se o peso máximo de todas as dimensões do conjunto (o valor que seria obtido se tudo fosse nota 10). Soma da Avaliação Real: Soma-se o valor real obtido em cada uma das dimensões. Cálculo do Índice: Aplique a fórmula: Soma da avaliação dividido pelo peso total multiplicado por 100. Classificação do Conjunto: Excelente: 90% a 100% Aceitável: 50% a 89% Crítica: Menor que 49%
	7 - No quadro 7 é calculado a contribuição geral de cada dimensão para a avaliação da qualidade do conjunto de dados: Peso: Foi atribuído peso 1 para todas as dimensões multiplicadas por 100. Desempenho de cada dimensão: Por meio do Quadro 5 (Avaliação final de cada dimensão) foi possível descobrir o desempenho de cada dimensão pelo cálculo das métricas (nota 6). Repete-se os dados encontrados na coluna desempenho da dimensão. Peso efetivo de cada dimensão: para encontrar o peso efetivo de cada dimensão foi dividido o valor do desempenho da dimensão pelo peso total. Contribuição de cada dimensão no índice geral: Para o cálculo da contribuição de cada dimensão na avaliação foi dividido o peso efetivo de cada dimensão pela quantidade total de dimensões (15).
	8 - A contribuição geral de cada dimensão levando em conta o mesmo peso para todas as dimensões é realizada da seguinte forma: Peso total das dimensões: Soma total dos pesos das dimensões. Soma do peso real: Soma do peso total efetivo de cada dimensão multiplicado por 100. Cálculo do índice: Soma da contribuição de cada dimensão no índice geral. Avaliação Final: Excelente: 90% a 100% Aceitável: 50% a 89% Crítica: Menor que 49%

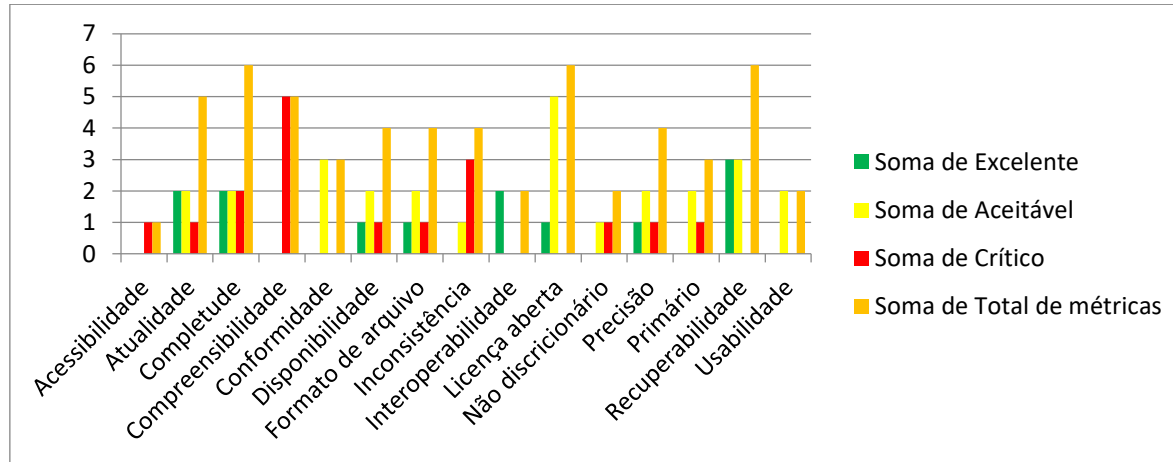


PAINEL CONSOLIDADO

Indicador	Valor
Total de métricas avaliadas	57
Percentual de 'Excelente'	13
Percentual de 'Aceitável'	27
Percentual de 'Crítico'	17



Rótulos de Linha	Soma de Excelente	Soma de Aceitável	Soma de Crítico	Soma de Total de métricas
Acessibilidade	0	0	1	1
Atualidade	2	2	1	5
Compleitude	2	2	2	6
Compreensibilidade	0	0	5	5
Conformidade	0	3	0	3
Disponibilidade	1	2	1	4
Formato de arquivo	1	2	1	4
Inconsistência	0	1	3	4
Interoperabilidade	2	0	0	2
Licença aberta	1	5	0	6
Não discricionário	0	1	1	2
Precisão	1	2	1	4
Primário	0	2	1	3
Recuperabilidade	3	3	0	6
Usabilidade	0	2	0	2
Total Geral	13	27	17	57



AVALIAÇÃO GERAL DAS DIMENSÕES

Soma de Cálculo do Índice				
Rótulos de Linha	Aceitável	Crítico	Excelente	Total Geral
Acessibilidade		0		0
Atualidade	60			60
Completude	50			50
Compreensibilidade		0		0
Conformidade	50			50
Disponibilidade	50			50
Formato de arquivo	50			50
Inconsistência		13		13
Interoperabilidade			100	100
Licença aberta	58			58
Não discricionário		25		25

Precisão	50			50
Primário		33		33
Recuperabilidade	75			75
Usabilidade	50			50

APÊNDICE B – PROVA DE CONCEITO

Dimensão:	ACESSIBILIDADE		
Métrica:	Conformidade com as normas de acessibilidade As normas que garantem a acessibilidade dos usuários com deficiência estão sendo respeitadas?	Avaliação:	Crítico
Recursos de avaliação	Descrição dos recursos	Avaliação recursos (3 - excelente/ 2 - aceitável/ 1- Crítico)	
Texto Alternativo	Todas as imagens, gráficos e ícones possuem descrição textual significativa	1	
Navegação por teclado	O site pode ser totalmente navegável sem mouse	1	
Contraste de cores	Texto e elementos gráficos têm contraste adequado para leitura	2	
Legendas e libras	Vídeos e conteúdos multimídia têm legendas e os conteúdos possuem interpretação em Libras	1	
Formulários acessíveis	Rótulos claros, instruções, mensagens de erro compreensíveis	3	
Estrutura Semantica	Uso correto de títulos, listas, menus e hierarquia de conteúdo	3	
Conteúdo multimídia seguro	Evita conteúdos piscantes ou que causem desconforto	3	
Linguagem e clareza	Texto claro e compreensível para pessoas com deficiência intelectual	2	
		67	
Comentário	O conjunto de dados não apresentam o requisito texto alternativo ou versões acessíveis para pessoas com deficiência visual. Na análise foi observado que alguns menus não abrem pelo teclado, havendo a necessidade de mouse para interagir. Nas abas com os anos (2024, 2023, 2022...), não é possível acessar o conteúdo sem o auxílio do mouse. Existe a dependência do <i>JavaScript</i> para conteúdo de abas. Alguns aspectos do contraste de cores dificultam a leitura. O texto cinza-claro em fundo branco tem baixo contraste, não atendendo ao mínimo exigido. alguns ícones com textos brancos também não atingem		

	contraste suficiente. O uso do cinza sobre cinza possui baixo contraste, dificultando a leitura por pessoas com baixa visão ou daltonismo. O Portal e o conjunto de dados são compostos em sua maioria por textos, não apresentando conteúdo audiovisual para aplicação de legendas. Porém, a aplicação de libras para os textos somente foi observada para alguns conteúdos da página inicial do conjunto de dados do INEP. O Portal de Dados Abertos não possui o recurso de libras em seu conteúdo. Os rótulos e instruções são acessíveis. Ambos possuem estrutura correta e consistente em todo o site. As páginas apresentam conteúdo de multimídia seguro. Tanto o Portal de Dados Abertos quanto o conjunto de dados do INEP utilizam linguagem que podem ser difíceis para pessoas com deficiência intelectual ou alfabetização limitada. No caso do conjunto de dados, é necessário conhecimento prévio em educação e estatística para compreender alguns termos
Nota	Os critérios detalhados para a aplicação de cada nota/pontuação estão disponíveis nas orientações da ferramenta, acessíveis através da célula de resultado. Os campos digitáveis da planilha encontram-se destacados em cinza.

Dimensão:	COMPLETUDE		
Métrica:	Valores ausentes O conjunto de dados apresenta valores ausentes em proporção que podem prejudicar a análise dos dados?	Avaliação:	Aceitável
Fórmula:	Proporção de células vazias= $\frac{\text{Número de células nulas ou vazias}}{\text{Número total de células no conjunto de dados}} \times 100$	1.165 12.327	9
Comentário	A análise do conjunto de dados localizou 1.165 de células vazias num total de 12.327 células. Foi utilizado a fórmula para o cálculo da proporção de células vazias $1.165/12.327 \cdot 100 = 9\%$, o que torna o índice de qualidade da métrica aceitável.		
Métrica:	Cobertura dos dados coletados O conjunto de dados apresenta os dados em sua completeude?	Avaliação:	Aceitável
Fórmula:	Cobertura= $\frac{\text{Número de células preenchidas}}{\text{Número total de células esperadas}} \times 100$	11.162 12.327	91
Comentário	A análise localizou o total de 11.162 de células preenchidas num total de 12.327 células. Para o cálculo do nível de qualidade da cobertura dos dados coletados foi utilizado a $11.162/12.327 \cdot 100 = 91\%$.		
Métrica:	Campos obrigatórios Todos os campos com informações específicas e obrigatórias estão preenchidos?	Avaliação:	Excelente

Fórmula:	Campo obrigatório = $\frac{\text{Número de registros com campo preenchido} \times 100}{\text{Número total de registros}}$	2.368	100
		2.368	

Comentário	Para a análise da métrica, foram identificados os campos obrigatórios do conjunto de dados (Ano, unidade geográfica, localização, dependência administrativa, total, ensino fundamental, ensino médio). Por meio da análise foi possível verificar que os campos considerados essenciais para a análise, interpretação ou conformidade do conjunto de dados foram todos preenchidos.

Métrica:	Comparabilidade É possível realizar a comparação do conjunto de dados em diferentes situações?	Avaliação:	Excelente
		Total/Parcial/Não	Total

Comentário	Para o teste de qualidade da métrica comparabilidade foi utilizado os conjuntos de dados dos anos anteriores, sendo possível a comparabilidade entre os diferentes períodos de tempo, entidades, regiões ou categorias. Para a análise foram avaliadas as estruturas do conjunto de dados, unidades de medidas, períodos padronizados e as classificações. Os conjuntos de dados possuem padronização na estrutura, nas unidades, nos períodos e nas classificações, proporcionando a interação entre eles.

Métrica:	Metadados Os metadados estão preenchidos de forma estruturada contextualizando e descrevendo os dados?	Avaliação:	Excelente
		Total/Parcial/Não	Total

Comentário	Ao analisar os metadados do conjunto de dados foi possível constatar que os mesmos são completos e possuem elementos essenciais que garantem a clareza, a rastreabilidade e a utilização do conjunto de dados.

	Foram analisados elementos como: título, descrição, fonte e periodicidade.		
Métrica:	Distribuição temporal O conjunto de dados atende aos requisitos temporais?	Avaliação:	Excelente
			1
Comentário	A cobertura temporal do conjunto de dados compreende os últimos 18 anos (2006 a 2024). O conjunto de dados possui periodicidade regular e consistente permitindo a comparação ao longo do tempo. Não existem lacunas temporais que prejudiquem a análise dos dados ao longo prazo.		
Nota	Os critérios detalhados para a aplicação de cada nota/pontuação estão disponíveis nas orientações da ferramenta, acessíveis através da célula de resultado. Os campos digitáveis da planilha encontram-se destacados em cinza.		

Dimensão:	ATUALIDADE		
Métrica:	Idade dos dados O conjunto de dados apresenta defasagem temporal entre a coleta e a última atualização que compromete a atualidade dos dados?	Avaliação:	Excelente
Fórmula:	Idade dos dados = <u>Data da última atualização</u> Data atual	20/05/2025 25/11/2025	232 Anual
Comentário	Aplicando as regras para se chegar a idade atual dos dados (data da última atualização - data atual), a diferença temporal demonstra que a qualidade da métrica idade dos dados é excelente. Podendo classificar em dados recentes.		
Métrica:	Frequência de atualização O conjunto de dados está sendo atualizado na frequência prevista (mensal, trimestral, anual etc.) ou há discrepâncias entre a periodicidade declarada e a prática observada?	Avaliação:	Crítico
			3
Comentário	A frequência de atualização do conjunto de dados é anual, porém, os conjuntos de dados dos anos anteriores não apresentam a data de publicação para verificar se as atualizações seguem a periodicidade prevista ou não.		
Métrica:	Tempo entre a ocorrência do evento e o do registro no sistema O intervalo entre a ocorrência do evento e o seu efetivo registro no sistema garante a atualidade necessária?	Avaliação:	Crítico
Fórmula:	Tempo de registro = <u>Data de registro</u>	n	#VALOR!

		Data do evento	n
Comentário	Não é possível medir o intervalo de tempo entre a ocorrência do fato e sua disponibilização no sistema, pois essas informações não estão presentes no conjunto de dados.		
Métrica:	Presença de data e hora da coleta dos dados A data e hora estão indicadas claramente no conjunto de dados?	Avaliação:	Crítico
			3
Comentário	O conjunto de dados não fornece informações com as informações sobre o período em que foi realizada a coleta dos dados, não sendo possível identificar o momento da ocorrência do evento.		
Métrica:	Atualização dos metadados Os metadados são atualizados de forma consistente e sincronizada com os conjuntos de dados, garantindo que as informações descritivas reflitam corretamente o conteúdo disponibilizado?	Avaliação:	Excelente
Recursos avaliados	Descrição dos recursos	Avaliação recursos cada recursos vale 1 ponto	
Atualização	Data da última atualização	1	
Fonte	Identificação da fonte responsável	1	
Responsável	Contato da fonte responsável	1	
Dados	Versão dos dados	1	

Comentário	Ao analisar a métrica atualização de metadados foi possível verificar que os metadados do conjunto de dados estão sendo atualizados. O conjunto de dados apresenta data da última alteração dos metadados (28/03/2025); identificação da fonte responsável (Diretoria de Estatísticas Educacionais - DEED); contato da fonte responsável (deed@inep.gov.br); e, versão dos dados (1.0).
Nota	Os critérios detalhados para a aplicação de cada nota/pontuação estão disponíveis nas orientações da ferramenta, acessíveis através da célula de resultado. Os campos digitáveis da planilha encontram-se destacados em cinza.

Dimensão:	FORMATO DE ARQUIVO		
Métrica:	Formato dos dados acessíveis Os dados estão disponíveis em formato aberto e reutilizável?	Avaliação:	Excelente
Formato disponível:			CSV
Comentário	Analisando a página inicial do conjunto de dados no Portal de Dados Abertos, foi possível identificar que o conjunto de dados está disponível em formato aberto e reutilizável (CSV e ODS).		
Métrica:	Multiplicidade de formatos O conjunto de dados é fornecido em mais de um formato, permitindo maior acessibilidade?	Avaliação:	Aceitável
Número de formatos:			2
Comentário	O conjunto de dados está disponível em dois formatos abertos, ampliando a possibilidade de acesso, análise e reuso dos dados (CSV e ODS).		
Métrica:	Dados disponíveis em formato compactado Os dados estão disponíveis em formatos compactados, facilitando o <i>download</i> completo?	Avaliação:	Aceitável
			2

Comentário	O conjunto de dados está disponibilizado em formato compactado (.zip), porém a redução do tamanho é pequena (tamanho normal 171.884bytes após ser compactado 168.206bytes).		

Métrica:	Documentação do formato A documentação sobre o formato dos dados suficientemente clara e detalhada para permitir que diferentes usuários compreendam, utilizem e integrem corretamente as informações disponibilizadas?	Avaliação:	Aceitável
			2

Comentário	O formato está indicado na página inicial do conjunto de dados, porém, não apresenta documentação sobre o formato.		

Nota	Os critérios detalhados para a aplicação de cada nota/pontuação estão disponíveis nas orientações da ferramenta, acessíveis através da célula de resultado. Os campos digitáveis da planilha encontram-se destacados em cinza.		

Dimensão:	DISPONIBILIDADE		
Métrica:	Frequência dos períodos indisponíveis A frequência e a duração dos períodos de indisponibilidade dos conjuntos de dados comprometem a continuidade do acesso e o uso efetivo das informações pelos usuários?	Avaliação:	Excelente
		Interrupções:	Nenhuma
Comentários	Durante o período de observação (19/09/2025 a 30/09/2025), em horários diversos, foi realizado o monitoramento do conjunto de dados e foi possível constatar que não houve indisponibilidade do conjunto de dados.		
Métrica:	Hiperlink de download funcional Os links disponibilizados no conjunto de dados podem ser operacionalizados?	Avaliação:	Aceitável
			2
Comentários	Na verificação da funcionalidade dos links de download foi constatado apenas um link quebrado na página inicial dos Indicadores do INEP. O link para <i>preview</i> dos Indicadores Educacionais indica erro.		
Métrica:	Disponibilidade de canal de comunicação com o usuário Há canais claros e acessíveis e funcionais para comunicação entre o usuário e o órgão que disponibiliza os dados?	Avaliação:	Aceitável

Número de canais:	1

Comentários	O conjunto de dados disponibiliza uma canal de comunicação com a área responsável mas não apresenta um responsável de referência. O canal está disponível na página inicial do conjunto de dados sendo visível e de fácil acesso.

Métrica:	Múltiplo idiomas O conjunto de dados está disponibilizado em outros idiomas além do vernáculo oficial?	Avaliação:	Crítico
		Idiomas	Não

Comentários	As informações do conjunto de dados não estão disponíveis em outro idioma além do português.

Nota	Os critérios detalhados para a aplicação de cada nota/pontuação estão disponíveis nas orientações da ferramenta, acessíveis através da célula de resultado. Os campos digitáveis da planilha encontram-se destacados em cinza.

Dimensão:	PRECISÃO		
Métrica:	Taxa de dados corrompidos A ocorrência de dados corrompidos nos conjuntos disponibilizados compromete a integridade, a confiabilidade e a possibilidade de reúso das informações??	Avaliação:	Excelente
Fórmula	Taxa de dados corrompidos = $\frac{\text{Número de células corrompidas} \times 100}{\text{Número total de células}}$	0 12.327	0
Comentários	Foram analisadas as 12.327 células do conjunto de dados e não foi verificada a presença de dados corrompidos. Após foi aplicando a fórmula $(0/12.327 = 0 \times 100 = 0\%)$ para o cálculo do nível de qualidade.		
Métrica:	Dados duplicados A presença de dados duplicados no conjuntos de dados compromete a consistência, a confiabilidade e a qualidade das análises realizadas pelos usuários?	Avaliação:	Excelente
Fórmula	Taxa de duplicação = $\frac{\text{Número de registros duplicados} \times 100}{\text{Total de registros}}$	0 12.327	0
Comentários	Na análise das 12.327 células não foram encontrados dados duplicados assegurando a precisão do conjunto de dados.		
Métrica:	Grau de correspondência dos dados com a realidade Os dados refletem de forma fidedigna com fontes oficiais e confiáveis?	Avaliação:	Excelente
Fórmula	Grau de Correspondência = $\frac{\text{Número de registros como corretos} \times 100}{\text{Total de registros verificados}}$	12.327 12.327	100

Comentários	Para a avaliação foram comparados os dados do conjunto de dados com a fonte oficial dos dados (Censo Escolar da Educação Básica). O conjunto de dados apresenta excelente grau de correspondência com a fonte oficial.		
Métrica:	Ausência de ambiguidades Os dados são claros objetivos e não geram múltiplas interpretações?	Avaliação:	Excelente
			1
Comentários	Analisando as informações contidas nas células da planilha de dados foi observado que os dados são apresentados de forma clara, com terminologias consistentes e categorias bem definidas, não dando espaço para múltiplas interpretações.		
Nota	Os critérios detalhados para a aplicação de cada nota/pontuação estão disponíveis nas orientações da ferramenta, acessíveis através da célula de resultado. Os campos digitáveis da planilha encontram-se destacados em cinza.		

Dimensão:	RECUPERABILIDADE		
Métrica:	Função de recuperação de dados O sistema permite localizar e acessar facilmente por meio de ferramentas de busca, filtros ou navegação temática?	Avaliação:	Aceitável
			2

Comentários	O Portal de Dados Abertos possui ferramentas de busca. O Portal permite a busca por palavras-chave “Taxa de Distorção Idade-Série”, que retorna o conjunto correto. A recuperação é possível tanto por meio de busca livre quanto por navegação por tema. Foi verificada a presença de filtros avançados (tema, palavras-chave, formato, reuso, dados de raça/etnia, periodicidade de atualização, licença dados de gênero e objetivo), contudo, o usuário pode apresentar dificuldade em encontrar os filtros, pois estes não estão visíveis. Ao recuperar o conjunto de dados “Taxa de Distorção Idade-Série” no Portal de Dados Abertos o usuário é direcionado para um conjunto de dados mestre do INEP onde contém 16 conjuntos de dados secundários, para ter acesso ao conjunto de dados pretendido é necessário pesquisar entre os 16 conjuntos. Após identificado o conjunto o usuário é redirecionado para o site do INEP onde os dados são disponibilizados. A recuperação exige navegar entre diferentes páginas, sendo a primeira o repositório central do Portal de Dados Abertos e posteriormente a página específica do INEP. Todas essas etapas podem impactar a usabilidade, pois o usuário precisa compreender que a disponibilização é descentralizada. Todos esse percurso aumenta o esforço do usuário podendo fazer com que o mesmo desista da consulta.		

Métrica:	Eficiência de busca Os mecanismos de buscas e filtros funcionam com rapidez, retornando resultados completos?	Avaliação:	Excelente
Fórmula	Tempo decorrido entre o início da busca até o acesso aos dados		1

Comentários	Apesar da busca ser realizada por meio do termo “taxa de distorção idade-série”, a busca retorna todos os

	conjuntos de dados que possuem indicadores da Educacionais da Educação Básica” (16 conjuntos), o que torna a busca confusa e de difícil localização. O usuário com experiência prévia no conjunto de dados leva um tempo médio de 1 minuto para acessar o conjunto de dados. Já o usuário iniciante sem familiaridade com o conjunto de dados pode levar um tempo médio de 5 a 7 minutos.		

Métrica:	Navegação A estrutura de navegação do portal é intuitiva, com menus claros, hierarquia lógica e facilidade para localizar conjuntos de dados por diferentes caminhos?	Avaliação:	Crítico
Fórmula	Para avaliar a métrica será contabilizado quantos cliques o usuário necessitou dar para acessar o conjunto de dados		13

Comentários	Para acessar efetivamente o conjunto de dados a partir da página inicial do Portal de Dados Abertos é necessário que o usuário clique 13 vezes comprometendo a usabilidade e a experiência do usuário.		

Métrica:	Taxa de falhas na recuperação de dados O processo de recuperação de dados ocorre normalmente sem interrupções?	Avaliação:	Aceitável
Fórmula	Taxa de falhas = $\frac{\text{Número de buscas com falhas}}{\text{Total de buscas realizadas}} \times 100$	1 6	17

Comentários	Para mensurar o nível de qualidade da métrica foram realizadas 6 tentativas com palavras-chave diferentes (taxa de distorção idade-série; distorção idade-série; taxa de distorção; indicadores educacionais; defasagem escolar; idade série INEP. Somente uma tentativa apresentou falha na recuperação. Aplicando a fórmula para avaliação da métrica: $\frac{1}{6} \times 100$: 16%, assim o sistema de recuperação funciona adequadamente quando o usuário digita exatamente o título oficial “Taxa de distorção idade-série” ou parte dele, mas apresenta falhas quando são utilizados sinônimos (defasagem escolar).		

Métrica:	Metadados Recuperáveis Os metadados estão disponíveis, organizados e recuperáveis?	Avaliação:	Excelente
Recursos avaliados	Descrição dos recursos	Avaliação recursos (Sim/Não)	
Título	Termo de busca	Recuperável	Sim
Descrição	Termo de busca	Recuperável	Sim
Palavras-chave	Termo de busca	Recuperável	Sim
		Total/Parcial/Não	Total

Comentários	Todos os elementos de metadados testados são recuperáveis individualmente.

Métrica:	Vocabulário controlados Os vocabulários utilizados no conjunto de dados são controlados, padronizados e consistentes, facilitando a recuperação?	Avaliação:	Crítico
			3

Comentários	O conjunto de dados utiliza termo técnico do INEP/MEC, mostrando consistência institucional, mas não suficiente para o usuário geral. Não existe indicação de uso de vocabulário controlado formal (tesauros, ontologias, vocabulários padrão do Governo Federal). Isso pode dificultar a recuperação dos dados do usuário que não conhece previamente o termo “taxa de distorção idade-série”.

Nota	Os critérios detalhados para a aplicação de cada nota/pontuação estão disponíveis nas orientações da ferramenta, acessíveis através da célula de resultado. Os campos digitáveis da planilha encontram-se destacados em cinza.

Dimensão:	INCONSISTENCIA		
Métrica:	Coerência lógica Os dados seguem uma estrutura lógica sem contradição entre os campos relacionados?	Avaliação:	Excelente
Fórmula:	Coerência lógica = $\frac{\text{Número de registros incoerentes}}{\text{Total de registros verificados}} \times 100$	0 12.327	0
Comentários	Por meio da análise do conjunto de dados foi possível observar que os registros atendem plenamente as regras lógicas, temporais, matemáticas e de integridade definidas. Os dados podem ser utilizados diretamente sem a necessidade de correções.		
Métrica:	Consistência entre os conjuntos de dados As informações apresentadas em diferentes conjuntos relacionados mantêm consistência nos valores, nomenclaturas e dimensões?	Avaliação:	Excelente
Fórmula	Coerência entre conjuntos de dados = $\frac{\text{Nº de registros inconsistentes}}{\text{Total de registros verificados}} \times 100$	0 12.327	0
Comentários	Não foi encontrado divergência entre o conjunto de dados do censo escolar de onde os dados foram extraídos.		
Métrica:	Coerência entre dados agregados Os totais, médias e demais dados agregados conferem com a soma ou cálculo realizado a partir de dados individuais?	Avaliação:	Excelente
Fórmula	Coerência em dados agregados = $\frac{\text{Nº de registros agregados incoerentes}}{\text{Total de registros verificados}} \times 100$	0 1718	0

Comentário	A validação de 1.718 células de dados agregados (totais e médias) confirmou a integridade dos cálculos, sem apontar divergências.		

Métrica:	Uso de Classificações As classificações adotadas seguem padrões oficiais e amplamente reconhecidos?	Avaliação:	Excelente
Fórmula	Uso de classificações= $\frac{\text{Número classificações padronizadas} \times 100}{\text{Total de registros verificados}}$	2.369	100
		2.369	

Instruções	O conjunto de dados adota o sistema de classificações de padrões oficiais e reconhecidos. São utilizadas classificações como as do IBGE.		

Nota	Os critérios detalhados para a aplicação de cada nota/pontuação estão disponíveis nas orientações da ferramenta, acessíveis através da célula de resultado. Os campos digitáveis da planilha encontram-se destacados em cinza.		

Dimensão:	COMPREENSIBILIDADE		
Métrica:	Presença de dicionários, notas explicativas ou legendas Os conjuntos de dados apresentam glossários, dicionários, notas explicativas e legendas que auxiliam na compreensão dos termos técnicos?	Avaliação:	Aceitável
		Total/Parcial/ Não	Parcial
Comentários	O conjunto de dados apresenta recursos explicativos parciais que auxiliam na compreensão dos termos técnicos. Contudo, esses recursos explicativos costumam ser centralizados em uma seção específica de notas técnicas do seu portal na qual o usuário precisa passar por várias seções para ter acesso. Todas as notas técnicas do INEP estão agrupadas na mesma seção, o que dificulta ainda mais a pesquisa do usuário.		
Métrica:	Consistência na linguagem O conjunto de dados utiliza nomenclaturas dos campos, termos e formatos padronizados?	Avaliação:	Excelente
Fórmula	Consistência na linguagem = <u>Número de campos/termos consistentes</u> x 100 Total de registros comparáveis entre os conjuntos	12.327 12.327	100
Comentários	O conjunto de dados apresenta consistência na linguagem em relação aos conjuntos de dados dos anos anteriores bem como a outros conjuntos de dados de indicadores do censo escolar. Adotando os critérios de avaliação da qualidade o número de campos consistentes é de 100%.		
Métrica:	Descrições claras e detalhadas dos dados Os dados possuem descrições claras, incluindo definições, unidades de medida e explicações para codificações utilizadas?	Avaliação:	Crítico

Fórmula	Descrições claras e detalhadas = $\frac{\text{Número de registros com divergências}}{\text{Total de registros verificados}} \times 100$	12.327	100
		12.327	

Comentários	Aplicando a verificação das descrições no conjunto de dados foi possível verificar que os campos possuem uma definição clara. Não existe explicação das unidades de medidas, apesar do nome do conjunto de dados “taxa de distorção Idade-Série parecer apresentar percentual, mas não está indicado nas colunas.

Métrica:	Linguagem adequada a diferentes públicos A linguagem empregada é simples, direta e acessível, facilitando o entendimento por diversos perfis de usuários?	Avaliação:	Crítico
Fórmula	Índice de qualidade (%) = $\frac{\text{Pontuação obtida}}{\text{Pontuação máxima}} \times 100$	5	56
		9	
Recursos avaliados	Descrição dos recursos	Avaliação recursos (3 - excelente/ 2 - aceitável/ 1- Crítico)	
Termos	Uso de termos simples e de fácil compreensão	2	
Frases	Construção de frases diretas e concisas	2	
Jargões técnicos	Uso de jargões técnicos ou, quando seu uso for necessário, a presença de explicações ou resoluções adequadas	1	
		5	

Comentários	Os dados excluem um público amplo, comprometendo o uso e reúso dos dados, pois depende de jargões técnicos, falta de legenda para categorias e ausência de descrições explicativas. O conjunto de dados é compreensível apenas para usuários familiarizados com os indicadores educacionais.

Métrica:	Valores Padronizados Os dados seguem formatos padronizados (ex: datas no padrão ISO, uso correto de separadores decimais, símbolos monetários e unidades), o que favorece a leitura e reutilização dos dados?	Avaliação:	Aceitável
Fórmula	Índice de qualidade (%) = $\frac{\text{Pontuação obtida}}{\text{Pontuação máxima}} \times 100$	10 12	83
Recursos avaliados	Descrição dos recursos	Avaliação recursos (3 - excelente/ 2 - aceitável/ 1- Crítico)	
Formatos	Os formatos são consistentes	3	
Medida	Indicação clara de unidade de medida	1	
Valores	Valores monetários uniformes	3	
Padronização	Padronização para números decimais e porcentagens	3	
		10	

Comentários	Os dados do conjunto de dados seguem formatos padronizados.

Nota	Os critérios detalhados para a aplicação de cada nota/pontuação estão disponíveis nas orientações da ferramenta, acessíveis através da célula de resultado. Os campos digitáveis da planilha encontram-se destacados em cinza.

Dimensão:	CONFORMIDADE		
Métrica:	Proteção de dados pessoais conforme legislação O conjunto de dados atende aos princípios da LGPD. Não apresentando dados pessoais identificáveis, dados sensíveis individualizados, possui finalidade legítima e clara e não existe risco de reidentificação de indivíduos?	Avaliação:	Excelente
Recursos avaliados	Descrição dos recursos	Avaliação recursos (3 - Excelente/ 2 - Aceitável/ 1 - Crítico)	
Dados pessoais	Os dados pessoais não estão sendo identificados	3	
Justificativa	Justificativa legal para o tratamento de dados	3	
Dados	Os dados foram anonimizados ou agregados adequadamente	3	
Reidentificação	Existe algum risco de reidentificação com cruzamento de dados	3	
		12	

Comentários	O conjunto de dados analisado atende plenamente aos princípios da LGPD. Não apresenta dados pessoais identificáveis, dados sensíveis individualizados, possui finalidade legítima e clara e não oferece risco de reidentificação dos indivíduos.		

Métrica:	Origem dos Dados O conjunto de dados indica a procedência dos dados?	Avaliação:	Aceitável
Recursos avaliados	Descrição dos recursos	Avaliação recursos (1 - Total/ 2 - Parcial/ 3 - Não)	
Fonte de dados	A fonte de dados está claramente indicada (Nome do órgão, base ou sistema de origem)	3	
Origem	Tipo de origem(interna, externa ou mista) é identificável. Transparência do fluxo de dados	3	

Documentação	Existe documentação explicando a procedência (links, metadados ou notas explicativas)	2
		8

Comentários	A fonte dos dados aparece de forma clara. O tipo de origem é interna, uma vez que os dados são produzidos pelo INEP a partir do Censo Escolar. No arquivo dos dados não existe documentação explicando a procedência, porém, no Portal do INEP existe toda a documentação metodológica. Para ter acesso o usuário precisa pesquisar nas seções do site do INEP.

Métrica:	Responsável em caso de violação de direitos ou uso inadequado dos dados O conjunto de dados apresenta identificação clara do controlador e do encarregado (DPO), bem como canais de contato e informações?	Avaliação:	Crítico
Recursos avaliados	Descrição dos recursos	Avaliação recursos (1 - Sim/ 2 - Não)	
Controlador	Identificação do controlador dos dados?	2	
Encarregado (DPO)	Nomeação do encarregado (DPO) com contato acessível	2	
Responsabilidades	Indicação de responsabilidades conforme LGPD	2	
Denúncias	Existência de canal de denúncia ou solicitação de direito do cidadão	2	
		8	

Comentários	O conjunto de dados não atende a esse requisito da LGPD. Ausência de identificação do controlador, do encarregado, indicação de responsabilidades e canais para solicitações à proteção de dados.

Nota	Os critérios detalhados para a aplicação de cada nota/pontuação estão disponíveis nas orientações da ferramenta, acessíveis através da célula de resultado. Os campos digitáveis da planilha encontram-se destacados em cinza.

Dimensão:	LICENÇA ABERTA		
Métrica:	Existência de licença A licença está claramente indicada na página do conjunto de dados, é padronizada, possui link acessível para o texto completo e está em local de fácil visualização?	Avaliação:	Aceitável
Recursos avaliados	Descrição dos recursos	Avaliação recursos (1 - Sim/ 2 - Não)	
Licença	A licença está indicada na página inicial do conjunto de dados	1	
Licença	A licença é padronizada	1	
Licença	O link para o texto completo da licença está disponível	2	
Licença	A informação sobre a licença está em local de fácil acesso	1	
		5	

Comentários	A licença está indicada na página inicial do conjunto de dados (<i>Creative Commons Attribution</i>) e faz parte do conjunto oficial de licenças <i>Creative Commons</i> . O conjunto de dados não apresenta o link para o texto completo da licença. A informação sobre a licença está na página inicial do conjunto de dados do INEP, estando em um lugar de fácil acesso.		

Métrica:	Tipos de licença A licença aplicada é compatível com os princípios dos dados abertos (ex: CC0, CC BY, ODC-BY), permitindo uso, redistribuição e modificação dos dados?	Avaliação:	Excelente
		Total/Parcial/Não	Total

Comentários	A licença utilizada pelo conjunto de dados do INEP (CC BY), é recomendada para dados abertos. Permite		

	qualquer uso, desde que a fonte original seja atribuída.		

Métrica:	Permissão para reutilização A licença autoriza expressamente o uso, redistribuição e modificação dos dados, sem limitações como uso não comercial ou proibição de obras derivadas?	Avaliação:	Excelente
		Total/Parcial/Não	Total

Comentários	A Creative Commons Attribution (CC BY), autoriza expressamente que qualquer pessoa pode utilizar os dados, inclusive para fins comerciais; permite copiar e compartilhar os dados em qualquer meio ou formato; e, é permitido modificar e criar obras derivadas, sem restrição.		

Métrica:	Período de validade para o uso dos dados A licença é irrevogável e sem prazo de validade definido, assegurando a continuidade no uso dos dados sem restrições futuras ou expirabilidade?	Avaliação:	Excelente
		Total/Parcial/Não	Total

Comentários	A licença CC BY é irrevogável e sem prazo de validade definido. A licença só pode ser suspensa para quem violar os termos.		

Métrica:	Ausência de restrições (domínio público) O conjunto de dados analisado segue a Política de Dados Abertos, ou seja, são disponibilizados sob o domínio público, sem restrições legais?	Avaliação:	Excelente
		Total/Parcial/Não	Total

Comentários	O conjunto de dados analisado segue a Política de Dados Abertos do Governo Federal, ou seja, são disponibilizados sob o domínio público, sem restrições legais.

Métrica:	Declaração de domínio público O conjunto de dados está livre de direitos autorais, patentes ou qualquer outro tipo de restrição legal que limite seu uso, indicando domínio público?	Avaliação:	Aceitável
			Total/Parcial/Não
			Parcial

Comentários	No conjunto de dados “taxas de distorção Idade-série da Educação Básica” não há declaração explícita de que os dados estão no domínio público. No entanto, o INEP, por meio da Portaria n.º 380 de 15 de junho de 2020 define a política de dados abertos do INEP, onde estabelece que dados abertos: são dados acessíveis ao público, representados em meio digital, estruturados em formato aberto, processáveis por máquina, referenciados na internet e disponibilizados sob licença aberta que permita sua livre utilização, consumo ou cruzamento, limitando-se a creditar sua autoria ou a fonte. Apesar de não haver declaração explícita no conjunto de dados, a política de dados abertos do INEP assegura a abertura e acessibilidade dos dados.

Nota	Os critérios detalhados para a aplicação de cada nota/pontuação estão disponíveis nas orientações da ferramenta, acessíveis através da célula de resultado. Os campos digitáveis da planilha encontram-se destacados em cinza.

Dimensão:	USABILIDADE		
Métrica:	Conjunto de dados disponível para download sem necessidade de autenticação O conjunto de dados está disponível para acesso e download direto, sem barreiras técnicas ou administrativas?	Avaliação:	Excelente
Fórmula	Índice de qualidade (%) = $\frac{\text{Pontuação obtida}}{\text{Pontuação máxima}} \times 100$	11 12	92
Recursos avaliados	Descrição dos recursos	Avaliação recursos (3 - excelente/ 2 - aceitável/ 1- Crítico)	
Acesso	Acesso anônimo	3	
Solicitações	Sem formulários ou solicitações	3	
Verificação humana	Barreiras de verificação humana ou scripts de bloqueio	3	
Link	Link estável e direto	2	
		11	

Comentários	O acesso ao conjunto de dados não exige <i>login</i> e nem formulários de solicitações. Também não existem barreiras de verificação humana ou scripts de bloqueio.

Métrica:	Interface amigável O Portal oferece uma interface amigável, acessível e eficiente, onde usuários de diferentes tipos de familiaridade com sites conseguem navegar e baixar as informações?	Avaliação:	Excelente
Recursos avaliados	Descrição dos recursos	Avaliação recursos (1 - Sim/ 2 - Não)	

Menu	Presença de menu simples, com categorias bem definidas na página inicial	1
Downloads	Os botões para baixar os conjunto de dados são grandes, claramente identificados e agrupados por tema	1
Cores	Utilização de cores que garantam bom contraste e fácil leitura	1
Títulos	Títulos apresentados em negrito	1
Navegação	Anúncios, pop-ups ou outros elementos que interrompam a navegação	1
		5

Comentários	O Portal oferece uma interface amigável, acessível e eficiente, onde usuários de diferentes tipos de familiaridade com sites conseguem navegar e baixar as informações.

Nota	Os critérios detalhados para a aplicação de cada nota/pontuação estão disponíveis nas orientações da ferramenta, acessíveis através da célula de resultado. Os campos digitáveis da planilha encontram-se destacados em cinza.

Dimensão:	NÃO DISCRICIONÁRIO		
Métrica:	Dados disponíveis diretamente ao usuário Os dados estão disponíveis de forma imediata, sem depender de solicitação por meio de canais como o Serviço de Informação ao Cidadão (SIC) ou requerimentos adicionais?	Avaliação:	Excelente
		Total/Parcial/Não	Total
Comentários	O acesso ao conjunto de dados é imediata sem a necessidade de requerimentos ao Serviço de Informação ao Cidadão ou qualquer outro serviço adicional.		
Métrica:	Custo de acesso O acesso aos dados é gratuito, sem cobrança de taxas, pagamentos ou assinatura para obtenção do conteúdo?	Avaliação:	Excelente
		Total/Parcial/Não	Total
Comentários	Todos os conjuntos de dados são disponibilizados pelo Portal de Dados Abertos de forma gratuita, sem cobrança de taxas, assinatura ou necessidade de pagamento para acesso ou download.		
Nota	Os critérios detalhados para a aplicação de cada nota/pontuação estão disponíveis nas orientações da ferramenta, acessíveis através da célula de resultado. Os campos digitáveis da planilha encontram-se destacados em cinza.		

Dimensão:	PRIMÁRIO		
Métrica:	Fonte de dados O conjunto de dados declara de forma explícita a fonte de origem dos dados?	Avaliação:	Excelente
		Total/Parcial/Não	Total
Comentários	O conjunto de dados disponibilizado pelo INEP indica explicitamente a fonte de origem dos dados. A origem dos dados são de fonte primária pois foram coletados diretamente pelo INEP.		
Métrica:	Processo de dados documentais Há documentação acessível sobre o processo de coleta, tratamento, validação e publicação dos dados?	Avaliação:	Aceitável
		Total/Parcial/Não	Parcial
Comentários	Não foram localizadas informações sobre a coleta, processamento e disponibilização dos dados diretamente no Portal de Dados Abertos. Embora o portal do INEP disponibilize os resumos técnico e estatístico do Censo Escolar da Educação Básica, o acesso é dificultado pela necessidade de navegar por diversas páginas e pela obrigatoriedade de realizar o download dos arquivos.		
Métrica:	Rastreabilidade dos dados É possível acompanhar o histórico dos dados desde sua origem até sua publicação?	Avaliação:	Aceitável
Recursos avaliados	Descrição dos recursos	Avaliação recursos (1 - Sim/ 2 -Não)	

Modificações	Alterações realizadas nos dados ao longo do tempo (correções, atualizações)	2
Integrações	Combinações com outras fontes de dados ou fusões de bases distintas.	1
Processamento	Transformações aplicadas, como limpeza, agregações, cálculos entre outros.	2
		5

Comentários	Não é possível rastrear os dados, pois o conjunto de dados do INEP não possui um histórico completo documentado de todas as etapas pelas quais os dados passaram, desde a origem até a versão atual publicada.

Nota	Os critérios detalhados para a aplicação de cada nota/pontuação estão disponíveis nas orientações da ferramenta, acessíveis através da célula de resultado. Os campos digitáveis da planilha encontram-se destacados em cinza.

Dimensão:	INTEROPERABILIDADE		
Métrica:	Capacidade de reutilização As informações sobre reuso e a quantidade de downloads realizados do conjunto de dados estão disponíveis para o usuário?	Avaliação:	Excelente
		Total/Parcial/Não	Total

Comentários	O conjunto de dados principal do INEP apresenta informações sobre quantidade de reuso e downloads realizados. A quantidade de downloads realizados pode ser encontrada no valor total dos conjuntos de dados do INEP ou individualizado. É importante esclarecer que o número de reuso que o Portal de Dados Abertos registra são aqueles em que o usuário realiza o cadastro em campo específico de reuso, localiza o conjunto de dados que será reutilizado no portal, confirma os dados e publica o reuso. Muitos dados são reutilizados, mas este número não entra na estatística de reuso, pois o usuário não cadastrou seu reuso. Assim, se fosse adotado formas mais simples de cadastro de reuso esse número poderia vir a ser maior.		

Métrica:	Integração entre sistemas O conjunto de dados apresenta mecanismos técnicos que permitem a integração dos dados a outras plataformas?	Avaliação:	Aceitável
Recursos avaliados	Descrição dos recursos	Avaliação recursos (Sim/Não)	
		Total/Parcial/Não	Parcial

Comentários	O Portal Dados Abertos permite acesso às informações sobre o conjuntos de dados públicos por meio de sua API RESTful oficial: API oficial: Catálogo da API. Autenticação: Para utilizar a API, é necessário gerar um <i>token</i> de acesso, conforme instruções fornecidas pelo próprio portal Como acessar a API com perfil de consumidor. Além disso, o acesso aos dados pode ser facilitado por meio da biblioteca Dados Abertos Brasil, um pacote		

	<i>open-source</i> em Python que permite integrar e manipular dados do portal diretamente em análises com Pandas. Contudo, o INEP não fornece API pública para os indicadores educacionais. A integração com outras plataformas depende que o arquivo seja baixado manualmente.

Nota	Os critérios detalhados para a aplicação de cada nota/pontuação estão disponíveis nas orientações da ferramenta, acessíveis através da célula de resultado. Os campos digitáveis da planilha encontram-se destacados em cinza.

PAINEL CONSOLIDADO DE AVALIAÇÃO DA QUALIDADE DOS DADOS GOVERNAMENTAIS ABERTOS

QUADRO 1

Dimensão	Excelente	Aceitável	Crítico	Total de métricas	Excelente(%)	Aceitável(%)	Crítico (%)
Acessibilidade	0	0	1	1	0,00	0,00	100
Compleitude	4	2	0	6	66,67	33,33	0,00
Atualidade	2	0	3	5	40,00	0,00	60,00
Formato de arquivo	1	3	0	4	25,00	75,00	0,00
Disponibilidade	1	2	1	4	25,00	50,00	25,00
Precisão	4	0	0	4	100,00	0,00	0,00
Recuperabilidade	2	2	2	6	33,33	33,33	33,33
Inconsistência	4	0	0	4	100,00	0,00	0,00
Compreensibilidade	1	2	2	5	20,00	40,00	40,00
Conformidade	1	1	1	3	33,33	33,33	33,33
Licença aberta	4	2	0	6	66,67	33,33	0,00
Usabilidade	2	0	0	2	100,00	0,00	0,00
Não discricionário	2	0	0	2	100,00	0,00	0,00
Primário	1	2	0	3	33,33	66,67	0,00
Interoperabilidade	1	1	0	2	50,00	50,00	0,00

QUADRO 2

	Total de métricas	Porcentagem
Excelente	30	52,63
Aceitável	17	29,82
Crítico	10	17,54

QUADRO 3	
Dimensão com maior % de métricas excelentes	Precisão
Maior % Excelente	100
Dimensão com maior % de métricas críticas	Acessibilidade
Maior % Crítica	100

QUADRO 4		
Dimensão	Métricas	Avaliação
Acessibilidade	Conformidade com as normas de acessibilidade	Crítico
Compleitude	Valores ausentes	Aceitável
	Cobertura de dados coletados	Aceitável
	Campos obrigatórios	Excelente
	Comparabilidade	Excelente
	Metadados	Excelente
	Distribuição temporal	Excelente
Atualidade	Idade dos dados	Excelente
	Frequência de atualização	Crítico
	Tempo entre a ocorrência do evento e do registro no sistema	Crítico
	Presença de data e hora da coleta de dados	Crítico

QUADRO 5				
AVALIAÇÃO FINAL DE CADA DIMENSÃO				
Dimensão	Peso Total	Soma da Avaliação real	Cálculo do Índice	Avaliação Final
Acessibilidade	10	0	0	Crítico
Compleitude	60	50	83	Aceitável
Atualidade	50	20	40	Crítico
Formato de arquivo	40	25	63	Aceitável
Disponibilidade	40	20	50	Aceitável
Precisão	40	40	100	Excelente
Recuperabilidade	60	30	50	Aceitável
Inconsistência	40	40	100	Excelente
Compreensibilidade	50	20	40	Crítico
Conformidade	30	15	50	Aceitável
Licença aberta	60	50	83	Aceitável
Usabilidade	20	20	100	Excelente
Não discricionário	20	20	100	Excelente

	Atualização de metadados	Excelente
Formato do Arquivo	Formato dos dados acessíveis	Excelente
	Multiplicidade de formatos	Aceitável
	Dados disponíveis em formato compacto	Aceitável
	Documentação do formato	Aceitável
Disponibilidade	Frequência dos períodos indisponíveis	Excelente
	Hiperlink de downloads funcional	Aceitável
	Disponibilidade de canal de comunicação com o usuário	Aceitável
	Múltiplos idiomas	Crítico
Precisão	Taxa de dados corrompidos	Excelente
	Dados duplicados	Excelente
	Grau de correspondência dos dados com a realidade	Excelente
	Ausência de ambiguidades	Excelente
Recuperabilidade	Função de recuperação de dados	Aceitável
	Eficiência de busca	Excelente
	Navegação	Crítico

Primário	30	20	67	Aceitável
Interoperabilidade	20	15	75	Aceitável

QUADRO 6	
AVALIAÇÃO FINAL DO CONJUNTO DE DADOS	
Peso Total do conjunto de dados	570
Soma da avaliação real	385
Cálculo do índice	68
Avaliação final do conjunto de dados	Aceitável

QUADRO 7				
CONTRIBUIÇÃO GERAL DE CADA DIMENSÃO PARA A AVALIAÇÃO				
Dimensão	Desempenho da Dimensão	Contribuição da dimensão no índice geral	Peso Atribuído	Peso Determinado
Acessibilidade	0,00	0	1	0
Completude	0,83	0,06	1	0,056
Atualidade	0,40	0,03	1	0,0267
Formato de arquivo	0,63	0,04	1	0,04
Disponibilidade	0,50	0,03	1	0,03

	Taxa de falhas na recuperação de dados	Aceitável
	Metadados recuperáveis	Excelente
	Vocabulários controlados	Crítico
Inconsistência	Coerência lógica	Excelente
	Consistência entre os conjuntos de dados	Excelente
	Coerência em dados agregados	Excelente
	Uso de classificações	Excelente
Compreensibilidade	Presença de dicionários notas explicativas ou legendas	Aceitável
	Consistência na linguagem	Excelente
	Descrições claras e detalhadas dos dados	Crítico
	Linguagem adequada a diferentes públicos	Crítico
	Valores padronizados	Aceitável
Conformidade	Proteção de dados pessoais conforme legislação	Excelente
	Origem dos dados	Aceitável
	Responsável em caso de violação de direitos ou uso inadequado dos dados	Crítico

Precisão	1,00	0,07	1	0,07
Recuperabilidade	0,50	0,03	1	0,033
Inconsistência	1,00	0,07	1	0,0667
Compreensibilidade	0,40	0,03	1	0,0267
Conformidade	0,50	0,03	1	0,03
Licença aberta	0,83	0,06	1	0,056
Usabilidade	1,00	0,07	1	0,067
Não discricionário	1,00	0,07	1	0,067
Primário	0,67	0,04	1	0,044
Interoperabilidade	0,75	0,05	1	0,050
SOMA		15		

QUADRO 8	
CONTRIBUIÇÃO GERAL	
Peso total das dimensões	1500
Soma do Peso real	1001
Cálculo do índice	67
Avaliação Final	Aceitável

Licença aberta	Existência de licença	Aceitável
	Tipo de licença	Excelente
	Permissão para reutilização	Excelente
	Período de validade para uso dos dados	Excelente
	Ausência de restrições	Excelente
	Declaração de domínio público	Aceitável
Usabilidade	Conjunto de dados disponível para download sem necessidade de autenticação	Excelente
	Interface amigável	Excelente
Não discricionário	Dados disponíveis diretamente ao usuário	Excelente
	Custo de acesso	Excelente
Primário	Fonte de dados	Excelente
	Processo dos dados documentais	Aceitável
	Rastreabilidade dos dados	Aceitável
Interoperabilidade	Capacidade de reutilização	Excelente
	Integração com sistemas externos	Aceitável

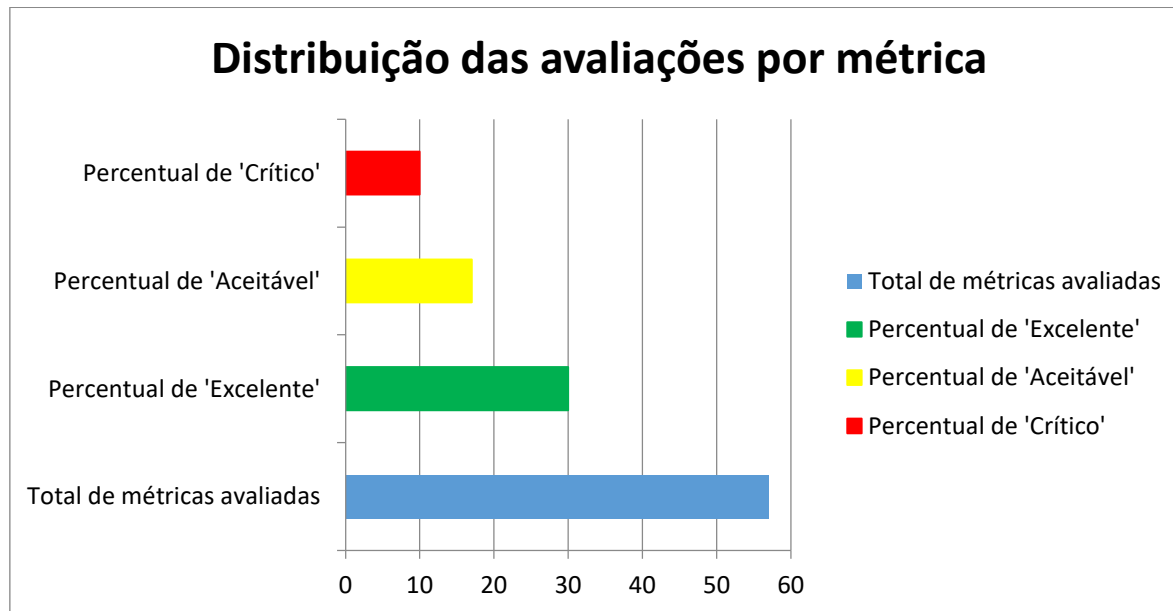
Notas	1 - O Quadro 1 detalha, para cada dimensão, a quantidade de métricas enquadradas nas categorias excelente, aceitável e crítico, juntamente com a porcentagem relativa a cada avaliação.

	2 - O Quadro 2 sintetiza o total de métricas avaliadas como excelentes, aceitáveis e críticas referentes ao conjunto de dados.
	3 - O Quadro 3 apresenta a dimensão com a maior porcentagem de métricas classificadas como excelentes e aquela com a maior porcentagem de métricas avaliadas como críticas. Além disso, evidencia a maior proporção de métricas excelentes e a maior proporção de métricas críticas dentre todas as dimensões analisadas.
	4 - O Quadro 4 apresenta os resultados consolidados da avaliação das métricas por dimensão.
	5 - Para calcular o desempenho de cada dimensão, siga as etapas abaixo utilizando a escala de pesos: Excelente (10), Aceitável (05) e Crítico (00). Identificação do Peso Total: Multiplique a quantidade de métricas da dimensão pelo peso máximo (10). Este valor representa o potencial máximo (100%). Soma da Avaliação Real: Some os pesos atribuídos a cada métrica conforme o desempenho observado (10, 05 ou 00). Cálculo do Índice: Aplique a fórmula: Soma da avaliação dividido pelo peso total multiplicado por 100. Classificação Final: Excelente: De 90% a 100% Aceitável: De 50% a 89% Crítica: Menor que 49%
	6 - O desempenho geral é consolidado através da média ponderada de todas as dimensões avaliadas: Soma do Peso Total (Potencial): Soma-se o peso máximo de todas as dimensões do conjunto (o valor que seria obtido se tudo fosse nota 10). Soma da Avaliação Real: Soma-se o valor real obtido em cada uma das dimensões. Cálculo do Índice: Aplique a fórmula: Soma da avaliação dividido pelo peso total multiplicado por 100. Classificação do Conjunto: Excelente: 90% a 100% Aceitável: 50% a 89% Crítica: Menor que 49%
	7 - No quadro 7 é calculado a contribuição geral de cada dimensão para a avaliação da qualidade do conjunto de dados: Peso: Foi atribuído peso 1 para todas as dimensões multiplicadas por 100. Desempenho de cada dimensão: Por meio do Quadro 5 (Avaliação final de cada dimensão) foi possível descobrir o desempenho de cada dimensão pelo cálculo das métricas (nota 6). Repete-se os dados encontrados na coluna desempenho da dimensão. Peso efetivo de cada dimensão: para encontrar o peso efetivo de cada dimensão foi dividido o valor do desempenho da dimensão pelo peso total. Contribuição de cada dimensão no índice geral: Para o cálculo da contribuição de cada dimensão na avaliação foi dividido o peso efetivo de cada

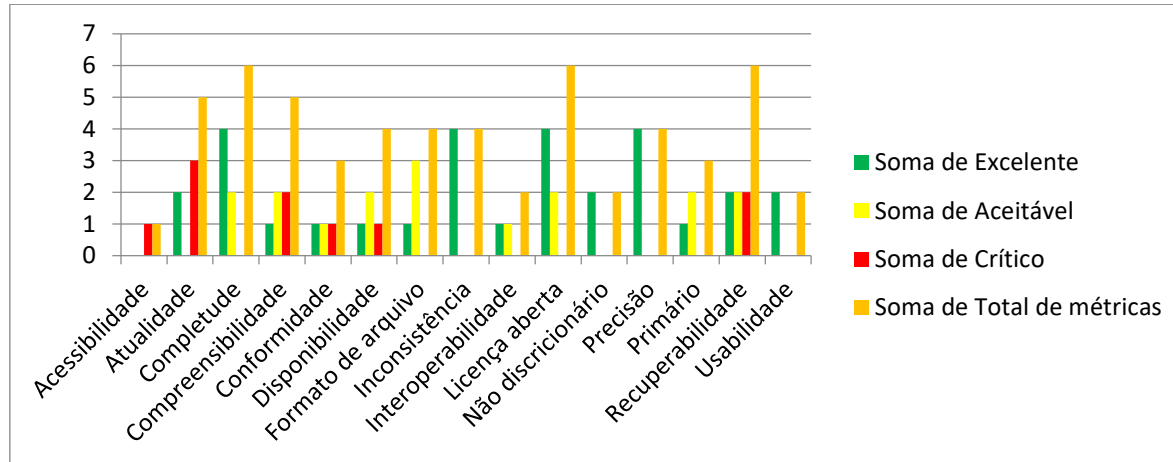
	dimensão pela quantidade total de dimensões (15).
	8 - A contribuição geral de cada dimensão levando em conta o mesmo peso para todas as dimensões é realizada da seguinte forma: Peso total das dimensões: Soma total dos pesos das dimensões. Soma do peso real: Soma do peso total efetivo de cada dimensão multiplicado por 100. Cálculo do índice: Soma da contribuição de cada dimensão no índice geral. Avaliação Final: Excelente: 90% a 100% Aceitável: 50% a 89% Crítica: Menor que 49%

PAINEL CONSOLIDADO

Indicador	Valor
Total de métricas avaliadas	57
Percentual de 'Excelente'	30
Percentual de 'Aceitável'	17
Percentual de 'Crítico'	10



Rótulos de Linha	Soma de Excelente	Soma de Aceitável	Soma de Crítico	Soma de Total de métricas
Acessibilidade	0	0	1	1
Atualidade	2	0	3	5
Compleitude	4	2	0	6
Compreensibilidade	1	2	2	5
Conformidade	1	1	1	3
Disponibilidade	1	2	1	4
Formato de arquivo	1	3	0	4
Inconsistência	4	0	0	4
Interoperabilidade	1	1	0	2
Licença aberta	4	2	0	6
Não discricionário	2	0	0	2
Precisão	4	0	0	4
Primário	1	2	0	3
Recuperabilidade	2	2	2	6
Usabilidade	2	0	0	2
Total Geral	30	17	10	57



AVALIAÇÃO GERAL DAS DIMENSÕES

Soma de Cálculo do Índice				
Rótulos de Linha	Aceitável	Crítico	Excelente	Total Geral
Acessibilidade		0		0
Atualidade		40		40
Completude	83			83
Compreensibilidade		40		40
Conformidade	50			50
Disponibilidade	50			50
Formato de arquivo	63			63
Inconsistência			100	100
Interoperabilidade	75			75
Licença aberta	83			83
Não discricionário			100	100
Precisão			100	100

Primário	67			67
Recuperabilidade	50			50
Usabilidade			100	100

