

Dados Abertos Governamentais na América Latina – Um Estudo Multivocal e uma Análise Comparativa de Portais

Open Government Data in Latin America – A Multivocal Study and Comparative Analysis of Portals

Michel Gularte Recondo¹, Nádia Puchalski Kozievitch¹

¹Universidade Tecnológica Federal do Paraná - UTFPR
Curitiba, Paraná – Brasil

michelrecondo@alunos.utfpr.edu.br, nadia@utfpr.edu.br

Abstract. *Despite the potential for transparency and accountability, Open Government Data initiatives face technical and governance challenges. This paper presents a mixed-methods approach to investigate the Latin American landscape. Initially, a Multivocal Literature Review was conducted, identifying gaps in standardization and quality. Secondly, a comparative analysis of three representative portals (Argentina, Chile, and Curitiba-Brazil) using files from 2025. The results showed that, although national portals (Argentina) present higher maturity in GIS formats and metadata, local initiatives (Curitiba) can outperform standardized platforms such as CKAN.*

Keywords. *Open data; Government open data; Analysis.*

Resumo. *Apesar do potencial para transparência e prestação de contas, iniciativas de Dados Abertos Governamentais enfrentam desafios técnicos e de governança. Este artigo apresenta uma abordagem mista para investigar o cenário latino-americano. Inicialmente, conduziu-se uma Revisão Multivocal de Literatura, identificando lacunas em padronização e qualidade. Em segundo lugar, realizou-se uma análise comparativa de três portais representativos (Argentina, Chile e Curitiba-Brasil) usando arquivos de 2025. Os resultados evidenciam que, embora portais nacionais (Argentina) apresentem maior maturidade em formatos GIS e metadados, iniciativas locais (Curitiba) podem superar plataformas como CKAN.*

Palavras-Chave. *Dados abertos; Dados abertos governamentais; Análise.*

1. Introdução

Na área da sistemas de informação, os dados abertos representam um insumo estratégico para o desenvolvimento de aplicações, algoritmos de inteligência artificial e pesquisas

em ciência de dados [European Data Portal 2024, Gobierno de España 2024]. A literatura aponta que, para que esses benefícios se concretizem, é necessário superar barreiras técnicas e metodológicas, especialmente no que diz respeito à qualidade e à interoperabilidade dos dados [Belete et al. 2017].

A publicação de Dados Abertos tem se consolidado como uma prática essencial para promover a transparência, a inovação e a participação cidadã [OECD 2023]. Organizações internacionais, como a *Organisation for Economic Co-operation and Development* (OCDE), a *Open Data Watch* [Open Data Watch 2024] e a *Open Government Partnership* [Open Government Partnership 2024], têm se dedicado à avaliação da implementação dessas políticas, utilizando metodologias distintas que resultam em diferentes escopos e indicadores. Enquanto a iniciativa *OURdata* da OCDE avalia cerca de 40 países [OECD 2023], o *Open Data Inventory* cobre aproximadamente 180 nações¹, evidenciando a diversidade de abordagens e a complexidade do cenário global.

A iniciativa ODIN, em seu relatório bienal de 2024², avalia a abertura dos dados com base em critérios como legibilidade por máquinas, formatos não proprietários, presença de metadados, opções de download e licenças abertas. Os dados revelam uma tendência de melhoria na qualidade da abertura de dados em todas as regiões, com destaque para a Europa. Na América Latina, países como México e Brasil lideram os índices de abertura, enquanto outros, como Haiti e Venezuela, apresentam desempenho inferior.

No Brasil, a promulgação da Lei de Acesso à Informação em 2011 [Brasil 2011], seguida pelo Decretos 11.529 [Brasil 2023] e 8.777 [Brasil 2016], estabeleceu diretrizes para a publicação de dados públicos. A Controladoria-Geral da União atua como órgão central de fiscalização e orientação. Apesar do arcabouço legal, persistem desafios relacionados à consistência, completude e atualização dos dados disponibilizados em portais como o Portal da Transparência³ e o Portal Brasileiro de Dados Abertos⁴. Esses problemas refletem uma realidade comum em diversos países latino-americanos, onde a qualidade e a padronização dos dados ainda são obstáculos significativos [Barni et al. 2022, García-Estrella et al. 2022]. Sob a ótica de Sistemas de Informação, os portais de Dados Abertos Governamentais (DAG) não são apenas repositórios tecnológicos, mas artefatos sociotécnicos que exigem Governança de Dados robusta para garantir a “adequação ao uso” (*fitness for use*), conceito central na teoria de Qualidade da Informação [Wang and Strong 1996]. A literatura de SI aponta que a simples disponibilização de dados (transparência ativa) é insuficiente sem a garantia de interoperabilidade e qualidade técnica.

Nesta direção, o objetivo deste artigo consiste em analisar sistematicamente o estado da arte das iniciativas de DAG na América Latina através de uma análise multivocal de literatura [Garousi et al. 2019], identificando padrões, desafios e oportunidades que possam orientar o aperfeiçoamento das políticas de transparência digital na região. Além disso, é apresentada uma avaliação comparativa de três portais da América La-

¹<https://odin.opendatawatch.com/Methodology>

²<https://odin.opendatawatch.com/Report/biennialReport2024>

³<https://portaldatransparencia.gov.br/>

⁴<https://dados.gov.br>

tina (Curitiba-Brasil, Argentina e Chile), seguindo metodologia previamente estabelecida [Oliveira et al. 2016]. Esta análise empírica permite identificar padrões de qualidade, consistência e acessibilidade dos dados disponibilizados, fornecendo evidências concretas dos desafios documentados na literatura. Este trabalho se diferencia ao contrastar o “estado da arte” (o que a literatura diz sobre DAG na região) com o “estado da prática” (a qualidade real dos dados). O objetivo principal é duplo: (1) mapear sistematicamente os desafios de implementação e governança reportados na literatura; e (2) diagnosticar tecnicamente se esses desafios se manifestam nos artefatos de dados disponibilizados pelos governos, através de uma análise comparativa seguindo a metodologia de [Oliveira et al. 2016].

Por fim, o restante deste trabalho encontra-se organizado da seguinte forma: a Seção 2 apresenta a Fundamentação Teórica, a Seção 3 os Trabalhos Relacionados; e a Seção 4 detalha a metodologia utilizada. Por fim, os resultados são apresentados na Seção 5, e a Seção 6 apresenta as conclusões, as limitações do estudo e as sugestões para trabalhos futuros.

2. Fundamentação Teórica

Esta seção apresenta os conceitos relacionados a Governança de Dados, Avaliação de Dados Abertos Governamentais, Panorama na América Latina e Padrões e Visualização.

2.1. Governança de Dados

Governança de dados pode ser definida como uma abordagem multidisciplinar para a criação e implementação de padrões que gerenciam dados em escala. Ela ajuda organizações a avaliar a origem dos dados, sua qualidade e segurança em todas as etapas de um fluxo de dados [Maffeo and Hogan 2023], e pode impactar as métricas de qualidade.

No contexto dos dados abertos governamentais, a governança de dados assume dimensões particulares que envolvem não apenas aspectos técnicos, mas também questões de transparência pública, responsabilidade democrática e participação cidadã. A implementação de estruturas robustas de governança torna-se fundamental para assegurar que os dados públicos sejam não apenas disponibilizados, mas também mantidos com qualidade adequada ao longo do tempo [Macedo and Lemos 2024].

As estruturas de governança para dados abertos governamentais devem contemplar a definição de responsabilidades claras entre diferentes órgãos públicos, estabelecendo fluxos de trabalho que garantam a atualização regular dos dados e a manutenção de padrões de qualidade. Isso inclui a criação de comitês específicos para supervisionar a publicação de dados, bem como a capacitação contínua de servidores públicos envolvidos no processo [Maffeo and Hogan 2023]. A governança efetiva também deve abordar questões de interoperabilidade entre sistemas governamentais, estabelecendo protocolos padronizados para a coleta, tratamento e publicação de dados [Nogueras-Iso et al. 2021].

Outro aspecto crucial da governança de dados governamentais refere-se ao equilíbrio entre transparência e proteção da privacidade. Para dados que contêm

informações pessoais ou sensíveis, técnicas de anonimização como *k-anonymity*, *l-diversity* e *privacidade diferencial* podem ser aplicadas para remover identificadores diretos e reduzir riscos de reidentificação [Samarati and Sweeney 1998, Dwork et al. 2006].

2.2. Avaliação de Dados Abertos Governamentais

Um dos alicerces estratégicos da governança pública na tomada de decisões são os dados governamentais. Dados abertos governamentais são definidos como informações públicas que podem ser livremente acessadas, utilizadas, modificadas e compartilhadas por qualquer pessoa, para qualquer finalidade [Open Knowledge Foundation 2012]. A qualidade dos dados abertos governamentais é um fator crítico para sua efetiva utilização, e ela representa a adequação ao uso (*fitness for use*) [Wang and Strong 1996]. A qualidade da informação foi estruturada na literatura [Wang and Strong 1996] em quatro categorias: intrínseca (exatidão, objetividade, credibilidade), contextual (completude, atualidade, relevância), representacional (interpretabilidade, formato, consistência) e acessível (acessibilidade, segurança).

Nesta direção, em uma análise de 2023 [Alogaiel and Alrwais 2023], foi identificado que 67% dos portais analisados apresentavam falhas de completude. Outra dimensão crucial é a consistência, que garante a ausência de contradições internas ou entre diferentes conjuntos de dados [Macedo and Lemos 2024] em níveis sintático, semântico e temporal. Os portais ainda podem não disponibilizar dados atualizados [Zainuddin and Akhir 2024], podendo ter informações coletadas há mais de seis meses.

A literatura de Sistemas de Informação aponta que a simples disponibilização de dados (transparência ativa) é insuficiente sem a garantia de interoperabilidade e qualidade técnica [Barni et al. 2022, García-Estrella et al. 2022].

A avaliação de DAGs ainda carece de metodologias específicas para sua análise [Ubaldi 2013], apesar disso, há vários trabalhos avançando nessa direção, como os Dez Princípios para Dados Abertos *Sunlight Foundation*⁵ (analisando completude dos dados, foco prioritário, tempestividade, acesso simples tanto presencial quanto online, compatibilidade com leitura automatizada, igualdade de tratamento, adoção de padrões abertos, definição de licença, manutenção a longo prazo e isenção de custos de uso).

Este estudo utiliza a definição de [Oliveira et al. 2016], com quatro eixos centrais para análise: (i) Acesso aos Dados, verificando a disponibilidade e barreiras técnicas; (ii) Características dos Conjuntos, analisando formatos, volumes e periodicidade; (iii) Metadados, avaliando a padronização e completude das descrições; e (iv) Esquema de Dados, verificando a conformidade estrutural. Além disso, este artigo estenderá a análise fornecendo uma perspectiva de dados espaciais disponibilizados nos DAGs. Este tipo de dado já era considerado um dos cinco principais domínios prioritários para dados abertos na Europa, juntamente com estatísticas, transportes e infraestruturas, registros de empresas e observação da Terra [European Data Portal 2016].

⁵<https://sunlightfoundation.com/policy/documents/ten-open-data-principles/>

2.3. Padrões e Visualização

No âmbito técnico, o conceito de *Linked Open Data* representa a evolução dos dados abertos, adicionando semântica e interconexão [Isotani and Bittencourt 2022]. Para avaliar o nível de abertura, Berners-Lee propôs o modelo '5 Estrelas de Dados Abertos', escala progressiva que parte de dados disponíveis com licença aberta (1 estrela) até dados interligados com outras fontes via *Linked Data* (5 estrelas).

Nesta direção, o portal garante o acesso, a tecnologia provê o tratamento técnico (APIs, CKAN, *Linked Data*, entre outros) e por fim, o ciclo da transparência digital é fechado com a visualização. A visualização eficaz de dados governamentais é essencial para democratizar o acesso à informação e ampliar o impacto das iniciativas de dados abertos. A literatura [Barni et al. 2022] já identificou padrões emergentes em portais latino-americanos, que vão desde visualizações estáticas até interativas e narrativas visuais.

As visualizações estáticas incluem gráficos de barras, mapas, infográficos e tabelas sumarizadas. Já as interativas permitem exploração dinâmica dos dados, como filtros, zoom, *drill-down* e animações temporais. Uma técnica cada vez mais popular é o *storytelling* com dados, que utiliza formatos como o *scrollytelling* [Knight Lab 2024] para guiar o usuário por narrativas visuais.

Nesta direção, dentre as ferramentas, podemos citar: (1) Tableau: ferramenta comercial amplamente utilizada para análise visual e criação de painéis interativos⁶; (2) Power BI: solução da Microsoft⁷, sem custo, para a criação de visualização de dados e integração com fontes governamentais via APIs; (3) Apache Superset: ferramenta *open-source* para exploração visual de dados, com suporte a múltiplos formatos e fontes⁸; e (4) Metabase: plataforma *open-source* voltada para visualização de dados e geração de relatórios acessíveis⁹.

E dentre as Ferramentas de Visualização Espacial, podemos citar ainda: (1) QGIS (Quantum GIS): sistema de informação geográfica *open-source* para visualização, análise e edição de dados geoespaciais¹⁰; (2) GRASS GIS: sistema de informação geográfica *open-source* com capacidades avançadas de análise espacial e modelagem geográfica¹¹; e (3) ARCGIS: sistema de informação geográfica comercial da Esri, amplamente utilizado para análise espacial e criação de mapas interativos¹².

A adoção de boas práticas de design visual, como contraste adequado, responsividade e acessibilidade (WCAG 2.1) [W3C Brasil 2018], também é essencial para garantir que os dados sejam compreensíveis por diferentes perfis de usuários [Sousa et al. 2024].

⁶<https://www.tableau.com/>

⁷<https://powerbi.microsoft.com/>

⁸<https://superset.apache.org/>

⁹<https://www.metabase.com/>

¹⁰<https://qgis.org/>

¹¹<https://grass.osgeo.org/>

¹²<https://www.arcgis.com/>

3. Trabalhos Relacionados

Esta seção apresenta os trabalhos relacionados referentes ao panorama da América Latina, avaliação de Dados Abertos Governamentais e os desafios enfrentados.

3.1. Panorama na América Latina

A institucionalização dos dados abertos na América Latina ocorreu em fases distintas que refletem a evolução das políticas de transparência na região. A primeira geração (2011-2015), liderada pela Lei de Acesso à Informação do Brasil [Brasil 2011], focou na transparência passiva, estabelecendo o direito fundamental de acesso à informação pública. A segunda geração (2016-2020) moveu-se em direção à transparência ativa, com decretos que estabeleceram a obrigatoriedade da publicação proativa de dados governamentais. A terceira geração (2021-presente) vai além da mera disponibilização de dados e os integra à transformação digital do Estado, utilizando-os como insumo estratégico para serviços governamentais inteligentes, interoperabilidade sistêmica e políticas públicas baseadas em evidências.

No Brasil, a promulgação da Lei de Acesso à Informação em 2011 [Brasil 2011], seguida pelos decretos 11.529 [Brasil 2023] e 8.777 [Brasil 2016], estabeleceu diretrizes para a publicação de dados públicos, tendo a Controladoria-Geral da União como órgão central de fiscalização e orientação. Decreto nº 8.777 define formato aberto como “formato de arquivo não proprietário, cuja especificação esteja documentada publicamente e seja de livre conhecimento e implementação, livre de patentes ou qualquer outra restrição legal quanto à sua utilização”, orientando a publicação em formatos como CSV e ODS. Na Argentina, a Lei nº 26.899/2013 estabeleceu repositórios de acesso aberto, enquanto a Colômbia adota a Lei nº 1712/2014 que regulamenta o tema. A Tabela 1 sintetiza os principais marcos legais da região.

Tabela 1. Leis de Dados Abertos por País

País	Nome da Lei/Política	Ano
Chile	Lei nº 20.285	2008
EUA	Open Government Directive	2009
Brasil	Lei nº 12.527 (LAI)	2011
Colômbia	Lei nº 1712	2014
Paraguai	Lei nº 5282	2014
Uruguai	Lei nº 19.355	2015
Brasil	Decreto nº 8.777	2016
Argentina	Política Nacional de Acesso Aberto	2021
Peru	Política de Dados Abertos	2022

Um mapeamento prévio dos 15 portais oficiais da América Latina (Tabela 2) revelou que 84,6% utilizam *Comprehensive Knowledge Archive Network* (CKAN) ou sua variante, DKAN¹³, que integra o Drupal¹⁴ às funcionalidades do CKAN [Barni et al. 2022].

¹³<https://getdkan.org/>

¹⁴<https://drupal.org>

A predominância do CKAN reflete sua consolidação como plataforma *open-source* líder em iniciativas de DAG, oferecendo funcionalidades como catálogo de dados e APIs para acesso [Flausino 2024]. Dois portais, Bolívia e Honduras, apresentaram problemas de acesso – erro HTTP 403 e falha de certificado SSL, respectivamente –, o que limita análises comparativas de cobertura total regional.

Tabela 2. Portais Oficiais de Dados Abertos na América Latina

País	URL	Tecnologia
Argentina	https://datos.gob.ar	CKAN
Bolívia	https://datos.gob.bo	sem acesso
Brasil	https://dados.gov.br	CKAN
Chile	https://datos.gob.cl	CKAN
Colômbia	https://www.datos.gov.co	Socrata
Costa Rica	https://datosabiertos.gob.cr	Wordpress
Equador	https://datosabiertos.gob.ec	CKAN
Guatemala	https://datos.gob.gt	CKAN
Honduras	https://datosabiertos.gob.hn	sem acesso
México	https://www.datos.gob.mx	CKAN
Panamá	https://datosabiertos.gob.pa	CKAN
Paraguai	https://datos.gov.py	DKAN
Peru	https://datosabiertos.gob.pe	Drupal
R. Dominicana	https://datos.gob.do	CKAN
Uruguai	https://catalogodatos.gub.uy	CKAN

Como alternativa comercial, a plataforma proprietária Socrata¹⁵, utilizada na Colômbia, oferece recursos avançados de visualização, embora com custos de licenciamento. Além dessas, existem soluções customizadas, como a do Peru, que desenvolveu uma plataforma própria baseada em Drupal para atender a requisitos legislativos específicos.

A visibilidade dos dados nos portais latino-americanos já foi investigada[Barni et al. 2022], identificando inconsistências na estruturação e padrões emergentes de visualização que vão desde representações estáticas até narrativas interativas. Nesta direção, os fatores institucionais são tão críticos quanto os técnicos no desenvolvimento de ecossistemas de plataformas de dados abertos em metrópoles como Buenos Aires, Cidade do México e Montevidéu[Bonina and Eaton 2020] .

Nos portais brasileiros de Curitiba, Natal e Porto Alegre, foram identificados desafios recorrentes de ausência de metadados, formatos não padronizados, dados desatualizados e problemas de acessibilidade[Flausino 2024].

A Figura 1 apresenta capturas de tela das páginas iniciais dos portais identificados na Tabela 2. A visualização comparativa evidencia diferenças notáveis na interface e na organização das informações entre os portais, reflexo direto das distintas plataformas e

¹⁵<https://www.tylertech.com/products/data-insights/open-data-platform>

políticas de publicação adotadas por cada país. Vale destacar que dois portais não puderam ser capturados: Bolívia, que retornou erro HTTP 403 (*forbidden*), e Honduras, que apresentou problema de certificado SSL.

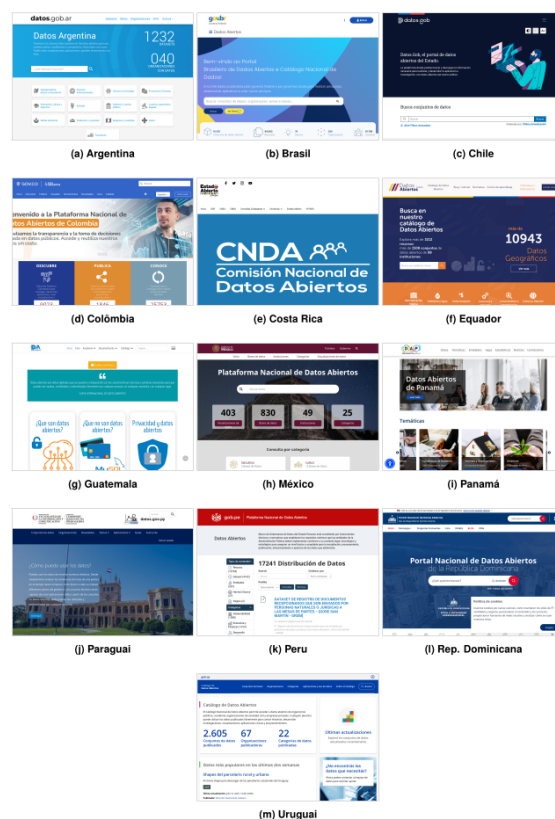


Figura 1. Páginas iniciais dos portais de dados abertos.

A análise visual dos portais reforça a predominância do CKAN como plataforma de referência regional, cujo padrão de interface é reconhecível em Argentina, Brasil, Chile, Equador, Guatemala, México, Panamá, República Dominicana e Uruguai. Destacam-se como exceções: a Colômbia, com a plataforma proprietária Socrata, que apresenta interface distinta e recursos avançados de visualização; a Costa Rica, que utiliza WordPress; e o Peru, com solução customizada baseada em Drupal. Essas variações de plataforma têm implicações diretas na padronização dos metadados e na disponibilidade de API, por exemplo.

3.2. Avaliação de Dados Abertos Governamentais

A garantia da qualidade dos dados publicados é um dos desafios mais citados na literatura [Alogaiel and Alrwais 2023, Andrade et al. 2023, Zainuddin and Akhir 2024, Fagundes and Ribeiro Junior 2020].

A ISO 19157 já foi utilizada como base para revisão de qualidade de metadados [Nogueras-Iso et al. 2021], revelando deficiências em completude e conformidade em portais, com impacto direto na capacidade de descoberta e reuso dos dados.

Macedo e Lemos [Macedo and Lemos 2024] propuseram modelo de maturidade identificando deficiências sistemáticas; Silva [Silva 2024] atualizou a métrica DGABr para avaliar o reuso efetivo de dados no Brasil. No Peru, García-Estrella *et al.* [García-Estrella et al. 2022] documentaram implementação precária em governos locais a despeito da Política Nacional; na Argentina, Zanotti *et al.* [Zanotti et al. 2021] identificaram disparidades entre arcabouço legal e execução prática.

Regionalmente, Moncayo Vives *et al.* [Moncayo Vives et al. 2022] identificaram variações substanciais de qualidade entre Chile, Colômbia, Equador e Peru por meio do Barômetro de Dados Abertos. No Brasil, estudos convergiram para a identificação de metadados ausentes ou não padronizados como principal barreira [Flausino 2024, Camperos-Reyes and Gonçalves Sant’Ana 2021]. Silva *et al.* [Silva et al. 2024] propuseram um catálogo de dados baseado em APIs governamentais para facilitar o acesso e a recuperação de dados no Brasil.

Molodtsov e Nikiforova [Molodtsov and Nikiforova 2024] desenvolveram um framework integrado de usabilidade para avaliar a eficácia dos portais de DAG, identificando problemas de navegação e compreensibilidade, item também indicado por Flausino [Flausino 2024] no contexto brasileiro. Herrera-Melo e González-Sanabria [Herrera-Melo and González-Sanabria M.Sc. 2020] propuseram critérios estruturados para avaliação de portais. Sousa *et al.* [Silva 2024] demonstraram que melhorias de usabilidade e acessibilidade têm impacto mensurável no engajamento dos usuários.

Os desafios de uso de DAG no contexto brasileiro também foram evidenciados [de Oliveira and Silveira 2018], enquanto outros estudos [Attard et al. 2015] mapearam iniciativas globais. Abordagens como o uso de ontologias para integração de dados, indicou a semântica pode revelar relações entre desempenho acadêmico e acesso ao transporte público [Belizario and Berardi 2022].

3.3. Desafios

Ao tratar-se de DAG, os desafios técnicos incluem a heterogeneidade de formatos e, principalmente, a baixa qualidade dos dados, com problemas recorrentes de inconsistência, incompletude e desatualização [Alogaiel and Alrwais 2023] e [Zainuddin and Akhir 2024]. A ausência de padrões uniformes limita a interoperabilidade [Nogueras-Iso et al. 2021], e a documentação, quando existe, é frequentemente incompleta, dificultando a compreensão por parte dos usuários [Oliveira et al. 2016]. No âmbito organizacional, os obstáculos envolvem uma cultura resistente à transparência, a falta de capacitação de servidores e a descontinuidade de políticas. Por fim, os desafios sociais abrangem a baixa literacia digital da população e o engajamento cidadão limitado, sendo necessário equilibrar transparência com segurança e confiança dos usuários [Barati 2023].

A literatura apresenta múltiplas abordagens para superar os desafios identificados. A necessidade de padronização de metadados, dicionários de dados, visualizações interativas e práticas de acessibilidade, além de mecanismos de monitoramento contínuo e transparência pelo design é um dos itens listados [Flausino 2024].

Albuquerque e Berardi [Albuquerque and Berardi 2020] reforçam a relevância de estruturas colaborativas que promovam interoperabilidade, proteção da privacidade e sus-

tentabilidade das iniciativas. Barati [Barati 2023] identificou três correntes na literatura: preocupações com privacidade, proposição de valor dos DAG e frameworks de risco, destacando a necessidade de equilibrar transparência com segurança e confiança.

Tabela 3. Comparativo de Trabalhos Relacionados

Autor(es)/Ano	Foco	Método	Escopo	Lacuna Identificada
Attard <i>et al.</i> (2015)	Iniciativas globais de DAG	Revisão Sistemática	Global	Análise qualitativa
Oliveira <i>et al.</i> (2016)	Portais brasileiros de DAG	Análise técnica comparativa	Brasil	Limitada a portais nacionais
de Oliveira e Silveira (2018)	Desafios de uso de DAG no Brasil	Revisão Sistemática	Brasil	Sem análise técnica direta
Barni <i>et al.</i> (2022)	Visibilidade de dados em portais latino-americanos	Análise de portais	América Latina	Foco em visualização
Macedo e Lemos (2024)	Maturidade de DAG no Brasil	Modelo de diagnóstico	Brasil	Não inclui análise comparativa com outros países
Flausino (2024)	Desafios e oportunidades de DAG no Brasil	Métodos mistos (RSL + análise)	Brasil	Restringe-se a portais nacionais
Zainuddin e Akhir (2024)	Qualidade de dados em portais governamentais	Revisão Sistemática	Global	Sem análise técnica direta
Este trabalho (2025)	Qualidade técnica de portais da América Latina	RML + análise comparativa	América Latina (3 portais)	Confronta estado da arte com estado da prática via análise de 1.938 arquivos

Este trabalho se diferencia das pesquisas anteriores em duas dimensões principais (Tabela 3). Primeiro, adota uma abordagem de métodos mistos que confronta explicitamente o ‘estado da arte’ — o que a literatura reporta sobre DAG — com o ‘estado da prática’ — a qualidade real dos arquivos, mensurada programaticamente sobre 1.938 arquivos coletados em 2025. A denominação ‘América Latina’ no título e no escopo refere-se ao contexto regional que motiva e enquadra a pesquisa, incluindo a RML de 23 estudos da região. Após busca exploratória (Figura 1), apenas três portais foram selecionados, utilizando como critérios a diversidade de temas e disponibilidade e relevância. Em particular, o Portal de Dados Abertos de Curitiba foi escolhido pois destacou-se em categorias como Transparência Financeira e Orçamentária, Legalidade Comunicação/Engajamento ¹⁶.

4. Metodologia

Este trabalho adota uma abordagem de métodos mistos, combinando uma Revisão Multivocal da Literatura [Garousi et al. 2019] (RML) com uma análise técnica comparativa de portais.

4.1. Revisão Multivocal de Literatura

Uma Revisão Multivocal da Literatura (RML ou MLR em inglês) é um método de pesquisa que une a literatura acadêmica formal (literatura branca) com informações práticas

¹⁶<https://www.curitiba.pr.gov.br/noticias/prefeitura-de-curitiba-e-classificada-como-otima-em-ranking-de-transparencia-e-lidera-entre-municipios-do-sul/79793>

e informais de especialistas da indústria (literatura cinza) [Garousi et al. 2019]. A RML seguiu o protocolo PRISMA (*Preferred Reporting Items for Systematic reviews and Meta-Analyses*) [Page et al. 2021], que consiste em uma diretriz para relatar revisões sistemáticas e meta-análises. O protocolo PRISMA 2020 fornece um *checklist* de 27 itens e um fluxograma padronizado que orienta pesquisadores na condução e documentação de revisões sistemáticas, assegurando rigor metodológico e reprodutibilidade dos resultados [Page et al. 2021].

A revisão teve como objetivo três questões: (1) Quais são as similaridades e diferenças entre as iniciativas de dados abertos governamentais dos países latino-americanos em termos de disponibilidade, qualidade e marcos regulatórios? (2) Qual o estado atual dos dados governamentais nos países da América Latina? e (3) Quais metodologias são empregadas para avaliar portais de dados abertos governamentais.

As buscas foram realizadas entre abril e maio de 2025 nas bases de dados (IEEE Xplore, SciELO, Science Direct, ArXiv), abrangendo o período de 2014 a 2024. As bases foram selecionadas de acordo com sua visibilidade e importância nas áreas de informática e educação. A string de busca utilizada foi: "(open data) OR ((open government) OR (digital government)))".

Os seguintes critérios de inclusão e exclusão foram utilizados:

- Inclusão: (1) Foco em DAG (qualidade, portais, avaliação); (2) Estudos Primários em Português, Inglês ou Espanhol; (3) Estudos Primários completos; (4) Relacionados a América Latina ou relevância regional; (5) Estudos Primários que satisfizessem a string de busca;
- Exclusão: (1) Temas marginais ou fora do contexto; (2) Outros idiomas além de PT, EN ou ES; (3) Editoriais, teses ou resumos, ou fora do período; (4) Estudos Primários não relacionados a DAG.

A busca resultou em 277 estudos primários. Para operacionalizar a etapa de triagem inicial do protocolo PRISMA e lidar com o volume de registros recuperados, foi utilizada a ferramenta de Inteligência Artificial Gemini 3 Pro, na fase de *screening* com o seguinte *prompt*: "Análise o título e o resumo e classifique como 'Incluir' apenas se o artigo tratar de Dados Abertos Governamentais no contexto de qualidade, implementação ou portais. Idiomas permitidos: PT, EN, ES."

Como resultado desta filtragem automatizada, seguida de validação manual pelos autores para mitigação de alucinações ou falsos negativos, o *corpus* foi reduzido para 73 registros, os quais avançaram para a etapa de leitura completa e extração de dados.

Foram selecionados 23 estudos primários para análise que satisfazem a todos os critérios de inclusão e não atenderam a nenhum dos critérios de exclusão, tabulados e disponíveis no repositório¹⁷, como ilustra a Figura 2.

4.2. Análise Técnica Comparativa de Portais

Nos últimos anos, surgiram abordagens mais quantitativas para a avaliação de portais de Dados Abertos Governamentais. Dentre as principais, destacam-se duas metodologias

¹⁷<https://doi.org/10.5281/zenodo.18913096>

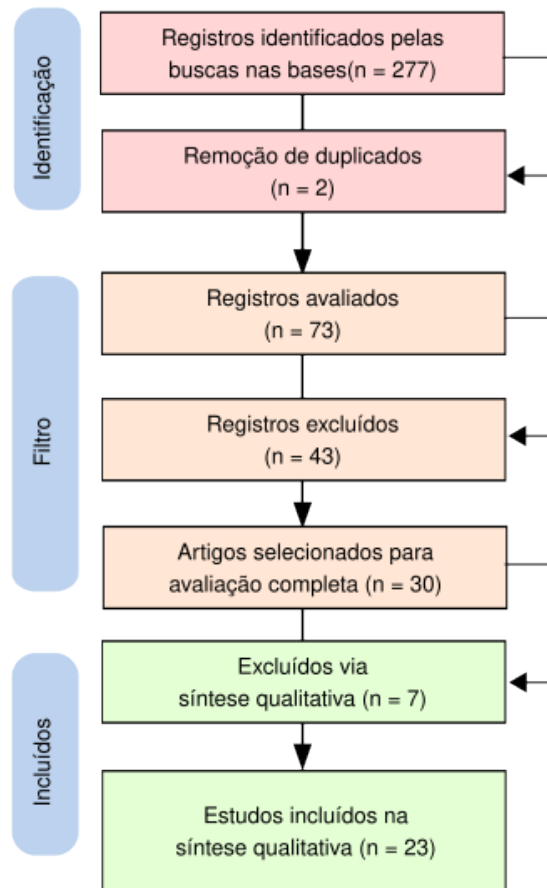


Figura 2. Fluxograma PRISMA.

com métricas mais precisas. A *Open Data Maturity (ODM) Assessment* da Comissão Europeia [European Commission 2025], que utiliza dezenas de indicadores quantificáveis e scores normalizados (0–1) nas dimensões *Policy*, *Portal*, *Quality* e *Impact*, e o *Open Data Quality Measurement Framework* [Vetrò et al. 2016], que oferece métricas granulares de qualidade de dados e metadados em nível de conjunto de dados.

Entretanto, a metodologia de [Oliveira et al. 2016] foi adotada como base principal neste estudo, devido à sua clareza e objetividade, além de já ter sido utilizada em comparações prévias no cenário brasileiro [Flausino 2024][Kozievitch et al. 2022]. Embora reconhecidamente útil, essa abordagem apresenta caráter predominantemente descritivo e menor granularidade métrica.

No presente trabalho, manteve-se a estrutura de quatro eixos de [Oliveira et al. 2016], estendida com um quinto eixo dedicado aos Dados Espaciais (GIS): (i) Acesso aos Dados; (ii) Características dos Conjuntos de Dados; (iii) Metadados; (iv) Esquema de Dados; e (v) Dados Espaciais (GIS). A Tabela 4 apresenta uma síntese dos eixos, suas descrições e os critérios de avaliação utilizados adaptados de [Oliveira et al. 2016].

Tabela 4. Critérios de avaliação utilizados adaptados de [Oliveira et al. 2016].

Eixo	Descrição	Critérios Analisados
1. Acesso aos Dados	Verifica a facilidade de descoberta e obtenção dos dados	• Disponibilidade de <i>download</i> direto • Existência e funcionalidade de API • Barreiras técnicas (login, CAPTCHA, etc.) • Quantidade de temas/categorias • Facilidade de navegação e busca
2. Características dos Conjuntos de Dados	Analisa o volume, variedade e aspectos quantitativos	• Número de datasets e recursos • Volume total de dados (GB) • Formatos utilizados e sua distribuição (%) • Diversidade de formatos • Cobertura temporal e periodicidade de atualização
3. Metadados	Avalia a qualidade descritiva dos dados	• Completude dos metadados (título, descrição, <i>tags</i>, licença) • Padronização (modelo CKAN) • Qualidade das descrições • Licenças declaradas • Proveniência (órgão, data de publicação/atualização)
4. Esquema de Dados	Análise estrutural dos arquivos (foco principal em CSV)	• Conformidade com RFC 4180 • Presença e qualidade de cabeçalhos • Convenção de nomenclatura de campos • Tamanho do esquema (número médio de colunas) • Esparsidade de dados (% de células nulas) • Taxa de duplicação de registros • Erros de <i>parsing</i>/sintaxe
5. GIS	Características adicionais de GIS	• Formatos disponíveis (como <i>shapefiles</i>) • Presença de <i>Dashboards</i>

Foram avaliados três portais de dados abertos latino-americanos: o portal municipal de Curitiba (Brasil) [Prefeitura Municipal de Curitiba 2025], e os portais nacionais da Argentina [Jefatura de Gabinete de Ministros 2025] e do Chile [Gobierno de Chile 2025]. Os dados foram extraídos entre outubro e novembro de 2025. A análise processou um total de 1.938 arquivos distribuídos entre os três repositórios.

A análise comparativa de portais foi executada através das seguintes etapas:

1. **Coleta (Ingestão):** Scripts em *C++* realizam requisições HTTP automatizadas às APIs dos catálogos CKAN (Argentina e Chile) e varreram a estrutura de diretórios da plataforma proprietária de Curitiba, mapeando as URLs diretas de *download* e armazenando os metadados brutos.
2. **Processamento (Parsing):** Arquivos tabulares são submetidos a um *parser* Python 3.13 que valida a estrutura conforme a RFC 4180. A biblioteca *pandas* (v2.3.0) é utilizada para a detecção de anomalias estruturais e cálculo de estatísticas descritivas (esparsidade, duplicatas, contagem de colunas).
3. **Análise dos Eixos:** As análises foram formalizadas por eixo, como por exemplo:
 - *Esquema:* Razão entre colunas válidas e total, e taxa de duplicatas ($\frac{n_{duplicadas}}{n_{total}}$).
 - *Dados Espaciais (GIS):* Contagem absoluta de recursos geoespaciais por formato (Shapefile, GeoJSON, KML, etc.).

Todos os artefatos computacionais produzidos neste trabalho foram disponibilizados em repositório público de acesso aberto sob o DOI: <https://doi.org/10.5281/zenodo.18904936>.

5. Resultados

A abordagem de métodos mistos combina resultados da RML com a análise técnica comparativa de Portais da América Latina.

5.1. Revisão Multivocal de Literatura

Na Questão 1 (similaridades e diferenças entre as iniciativas de dados abertos governamentais dos países latino-americanos), a RML não identificou estudos que realizassem comparação direta entre Brasil e demais países latino-americanos. Contudo, a análise transversal revelou padrões consistentes de desafios compartilhados regionalmente. A falta de padronização emergiu como obstáculo mais recorrente, documentada em 15 dos 23 estudos primários (65%). Foram identificadas inconsistências na estruturação de dados nos portais latino-americanos [Barni et al. 2022], e a ausência de vocabulários controlados comprometendo a interoperabilidade [Nogueras-Iso et al. 2021].

Os problemas na qualidade dos metadados foram evidenciados [Nogueras-Iso et al. 2021] mediante avaliação baseada na ISO 19157, revelando deficiências em completude e conformidade. Oliveira et al. [Oliveira et al. 2016] demonstraram o impacto direto na capacidade de descoberta e reuso dos dados. As dificuldades de acesso aos dados também foram listadas [Molodtsov and Nikiforova 2024], junto com problemas de navegação e compreensibilidade, item também verificado no contexto brasileiro [Flausino 2024].

As diferenças manifestam-se principalmente no grau de maturidade institucional. [García-Estrella et al. 2022] documentaram desafios mais acentuados em governos locais peruanos, enquanto [Moncayo Vives et al. 2022] identificaram variações substanciais entre Chile, Colômbia, Equador e Peru. Quanto aos marcos regulatórios, observa-se convergência temporal nas legislações mas divergências na efetividade [Ashahril et al. 2022].

Na Questão 2 (estado atual dos dados governamentais nos países da América Latina) a análise indica que o estado atual caracteriza-se por avanço normativo significativo mas execução operacional deficiente. Consideram-se avanços relevantes: (i) mudanças institucionais estruturantes como criação de marcos legais; (ii) melhorias mensuráveis na disponibilização quantificadas por métricas de *datasets* e formatos; e (iii) reconhecimento em avaliações internacionais (ODIN, OGP). A efetividade foi mensurada através de análises de qualidade [Macedo and Lemos 2024, Zainuddin and Akhir 2024], métricas de disponibilização [Barni et al. 2022], avaliações de usabilidade [Molodtsov and Nikiforova 2024] e índices internacionais.

No Brasil, oito estudos documentaram marco legal consolidado (LAI 2011, Decreto 8777/2016) mas identificaram desafios operacionais. Flausino [Flausino 2024] revelou ausência de metadados padronizados, dados desatualizados e problemas de acessibilidade em portais municipais. Macedo e Lemos [Macedo and Lemos 2024] propuseram modelo de maturidade identificando deficiências sistemáticas. Silva [Silva 2024] atualizou métrica DGABr para avaliar reuso efetivo.

No Peru, a literatura documentou a implementação como precária em governos locais apesar da Política Nacional [García-Estrella et al. 2022]. Na Argentina, foram identificadas disparidades entre arcabouço legal e execução prática [Zanotti et al. 2021].

Regionalmente, a maioria dos portais (84,6%) utilizam CKAN/DKAN [Barni et al. 2022]. Mas a desatualização (mais que 6 meses), é de aproximadamente 45% dos portais [Zainuddin and Akhir 2024], com falhas de completude em 67% dos portais analisados [Alogaiel and Alrwais 2023]. A Tabela 5 apresenta a síntese do estado atual por país, baseada nos estudos analisados.

Já na Questão 3 (metodologias empregadas para avaliar portais de dados abertos governamentais) a análise revelou seis categorias de metodologias empregadas nos 23 estudos. As abordagens mais recorrentes são sistemas de avaliação por indicadores e revisões sistemáticas (ambos com 23,1% dos estudos). A avaliação via Sistemas de indicadores (23,1%) incluem o *Balanced Scorecard* aplicado a plataformas governamentais [Gao et al. 2022] e o Barômetro de Dados Abertos implementado em Chile, Colômbia, Equador e Peru [Moncayo Vives et al. 2022]. A avaliação via Revisões sistemáticas (23,1%) sintetiza conhecimento acumulado, como revisão sobre qualidade de dados [Zainuddin and Akhir 2024] e desafios de implementação [Ashahril et al. 2022]. A avaliação via Frameworks multidimensionais (19,2%) oferecem avaliações holísticas. Exemplos incluem o *framework* MAAR (*Monitor, Analyze, Act, Review*) [Alogaiel and Alrwais 2023], *framework* de usabilidade considerando diversidade de usuários [Molodtsov and Nikiforova 2024], e o *framework* SuDaMa [Sánchez-Nielsen et al. 2021], focado em sustentabilidade de longo prazo. A Avaliação de metadados (15,4%) utiliza extensão da ISO 19157 [Nogueras-Iso et al. 2021] e análise

Tabela 5. Estado Atual dos DAG por País/Região

País/Região	Est.	Principais Avanços	Principais Limitações
Brasil	8	Marco legal consolidado (LAI 2011, Decreto 8777/2016); Portal nacional estruturado; Iniciativas municipais ativas; Métrica DGABr para avaliação de reúso	Baixa qualidade de metadados; Dados desatualizados (45% com < 6 meses); Problemas de acessibilidade; Descontinuidade de políticas
Peru	3	Política Nacional de Dados Abertos (2022); Plataforma customizada baseada em Drupal; Participação ativa na OGP	Implementação irregular em governos locais; Falta de capacitação técnica; Limitações de recursos em nível subnacional
Argentina	2	Política Nacional de Acesso Aberto (2021); Portal nacional ativo (datos.gob.ar); Marco legal para repositórios	Desafios na implementação institucional; Recursos limitados; Disparidades entre arcabouço legal e execução prática
Chile, Colômbia, Equador	3	Participação ativa na OGP; Marcos legais estabelecidos (Chile: Lei 20.285/2008; Colômbia: Lei 1712/2014); Plataformas nacionais operacionais	Descontinuidade de políticas; Qualidade dos dados variável; Implementação heterogênea entre níveis de governo
América Latina (regional)	4	Uso predominante de CKAN/DKAN (84,6% dos portais); Aumento na disponibilização de dados; Convergência nos marcos legais	Falta de padronização regional; Interoperabilidade limitada; Ausência de vocabulários controlados comuns

de conformidade com DCAT [Camperos-Reyes and Gonçalves Sant’Ana 2021]. Já os Modelos de maturidade (11,5%) estruturam níveis evolutivos, como o modelo de diagnóstico de Macedo e Lemos [Macedo and Lemos 2024]. E a última categoria Métodos automatizados (7,7%) representam fronteira emergente, com identificação automatizada de *High-Value Datasets* [Quarati and Nikiforova 2024].

A análise dos 23 estudos primários selecionados permitiu sua organização em seis categorias temáticas principais. A Figura 3 apresenta a distribuição quantitativa dos estudos dentro de cada uma dessas categorias. A categorização temática foi realizada através da leitura integral e análise crítica dos textos selecionados. Para fins de classificação estatística e devido às limitações de escopo desta revisão, cada artigo foi alocado em uma única categoria predominante.

A Categoria Qualidade de Dados em DAG aborda um dos desafios mais citados na literatura: a garantia da qualidade dos dados publicados [Alogaiel and Alrwais 2023, Andrade et al. 2023, Zainuddin and Akhir 2024, Fagundes and Ribeiro Junior 2020]. As análises dos estudos primários selecionados apresentam desde revisões sobre diretrizes

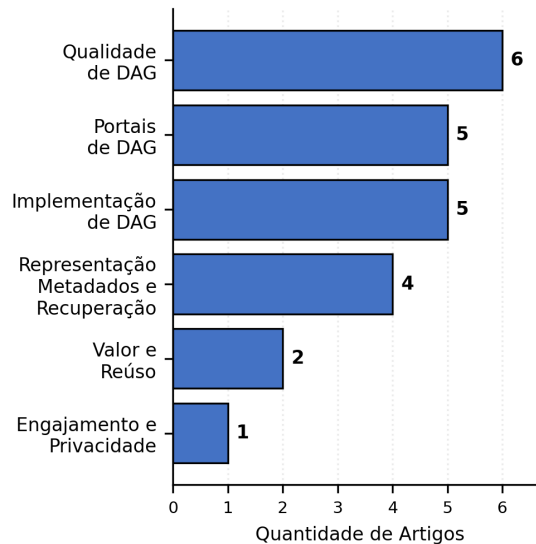


Figura 3. Distribuição dos estudos primários por categoria temática.

de publicação até propostas de modelos de maturidade e *frameworks* para avaliação e melhoria da qualidade de dados em contextos municipais e nacionais.

A categoria Implementação de DAG mostra a análise, sob diferentes perspectivas, dos desafios encontrados por quem implementa essas iniciativas, o nível de implementação e as tendências tecnológicas que moldam a área [Enríquez-Reyes et al. 2021, García-Estrella et al. 2022, Ashahril et al. 2022].

A categoria Representação, Metadados e Recuperação da Informação discute a descoberta e o acesso a informação dependem fortemente da qualidade dos metadados e da estrutura dos portais. Estudos discutem abordagens para metadados e sua avaliação com base em normas como a ISO 19157¹⁸, técnicas de representação e recuperação [Oliveira et al. 2016] e catálogos de dados baseado em APIs governamentais [Silva et al. 2024].

Já a categoria Valor e Reúso de Dados Abertos ilustra o potencial de reúso dos dados [Quarati and Nikiforova 2024], com métodos automatizados para identificar Conjuntos de Dados de Alto Valor (HVDs) e atualização da métrica DAGBr¹⁹ [Silva 2024].

A categoria Engajamento de Usuários e Privacidade explora a relação entre o uso dos dados e a proteção da privacidade [Barati 2023], identificando três correntes principais na literatura: preocupações com privacidade, proposição de valor dos DAG e *frameworks* de risco.

A categoria Portais de Dados Abertos analisa a usabilidade e eficácia dos portais como temas centrais [Barni et al. 2022], a visibilidade dos dados nos portais latino-americanos, e critérios de avaliação [Herrera-Melo and González-Sanabria M.Sc. 2020].

¹⁸<https://www.iso.org/standard/32575.html>

¹⁹<https://www.dgabr.com.br/>

Do ponto de vista de iniciativas, há outros desafios listados: 1) a transição dos portais do foco quantitativo para a qualidade dos dados abertos[European Commission 2022], contrapondo-se aos “portais fakes” e reforçando a necessidade de transparência pelo design[Albuquerque and Berardi 2020], 2) evolução legislativa (ex.: Plano Diretor de São Paulo com LGPD[Brasil 2018]) e 3) arranjos colaborativos como a *Open Data Alliance*[Controladoria-Geral da União 2020] e o Comitê de Dados Abertos de Curitiba[Prefeitura Municipal de Curitiba 2023].

De maneira resumida, os resultados apontam para uma carência generalizada de padrões técnicos consistentes, baixa qualidade dos metadados e ausência de mecanismos de monitoramento contínuo. Tais limitações são recorrentes entre os países da região, sugerindo a necessidade de abordagens colaborativas e regionalmente integradas. As metodologias de avaliação identificadas, como modelos de maturidade e *frameworks* multidimensionais, demonstram potencial para aprimorar a gestão dos dados abertos. No entanto, sua aplicação prática ainda é incipiente, o que reforça a lacuna entre os instrumentos normativos e sua operacionalização.

Nesta direção, os desafios listados aqui dialogam com diversas outras iniciativas e marcos conceituais que buscam aprimorar a disponibilização de dados abertos. Entre eles, destacam-se os princípios FAIR (Localizável, Acessível, Interoperável e Reutilizável), que estabelecem diretrizes para a gestão da informação [Wilkinson et al. 2016], e normas como a ISO 37122, que apresenta critérios específicos para Cidades e Comunidades Sustentáveis [International Organization for Standardization 2019].

5.2. Análise de Três Portais da América Latina

A análise técnica comparativa foi estruturada em cinco eixos, adaptando a metodologia de Oliveira *et al.* [Oliveira et al. 2016]: (i)Acesso aos Dados; (ii)Características dos Conjuntos de Dados; (iii)Metadados; (iv)Esquema de Dados; e (v)Dados Espaciais (GIS). A análise contemplou 109 arquivos de Curitiba, 1.567 da Argentina e 262 do Chile, totalizando 1.938 *datasets* avaliados.

5.2.1. Acesso aos Dados

De acordo com os Princípios de Dados Abertos ²⁰, os conjuntos de dados devem ser acessíveis e fáceis de obter por qualquer meio. É crucial que a existência e a localização desses dados sejam de conhecimento público e facilmente descobertas.

As características dos portais (como plataformas utilizadas, disponibilidade para *download* e de APIs, quantidade de temas e termos padronizados) estão listadas na Tabela 6. O CKAN permanece sendo a plataforma mais comumente utilizada em dois dos três portais analisados (Argentina e Chile), enquanto Curitiba implementa uma solução proprietária. A maioria permite acesso tanto através de *downloads* de arquivos quanto por APIs, com exceção de Curitiba, que disponibiliza apenas downloads diretos de arquivos, tanto de maneira manual quanto programaticamente.

²⁰https://public.resource.org/8_principles.html

Tabela 6. Características de Acesso dos Portais

Portal	Plataforma	Download	API	Temas	Termos Padronizados
Curitiba	Proprietária	SIM	NÃO	37	Não
Argentina	CKAN	SIM	SIM	13	SIM
Chile	CKAN	SIM	SIM	23	SIM

A maioria dos dados permite download, porém sua divisão entre temas de dados é apresentada com categorias diferentes (menor quantidade de temas no portal da Argentina, e maior quantidade no portal de Curitiba). Entretanto, a maioria dos termos utilizados nos portais estão padronizados (a plataforma CKAN provê esta característica).

5.2.2. Características dos Conjuntos de Dados

Esta etapa avalia as características dos conjuntos de dados, incluindo quantidade, tamanho, formatos, domínios e cobertura temporal, baseada nos 1.938 arquivos analisados.

Para evitar ambiguidades na apresentação dos resultados quantitativos, este trabalho adota as definições hierárquicas comuns ao ecossistema CKAN e a portais de dados abertos. Define-se ‘Dataset’ (ou conjunto de dados) como a unidade lógica que agrupa metadados e conteúdos sobre um tema específico; ‘Recurso’ refere-se a cada objeto individual ou link listado dentro de um dataset; e ‘Arquivo’ denota o arquivo individual, sem ligação com estruturas definidas e efetivamente baixado e processado.

Quantidade de Arquivos e Tamanho

A análise de quantidade de arquivos e tamanho revela diferenças fundamentais entre os três portais (Tabela 7). Curitiba opera um modelo baseado em repositório de arquivos brutos, disponibilizando 147.390 arquivos individuais com 3,35 terabytes de dados – volume 197 vezes superior à Argentina e 165 vezes ao Chile. A predominância absoluta de CSV (98,7%) reflete padronização em dados tabulares. O portal mantém 8 formatos distintos. A escala temporal (10,8 anos) abrange 50 diretórios temáticos, com os maiores sendo:

- Legisladoc: 1,76 TB (7.297 arquivos legislativos);
- SIGMU (Sistema de Mobilidade Urbana): 1,57 TB (127.576 arquivos de transporte público); e
- CliqueEconomia: 76 GB (2.145 arquivos econômicos).

Tabela 7. Volume e Escala do Universo de Dados Disponíveis nos Portais

Portal	Arquivos/ <i>Datasets</i>	Tamanho (GB)	CSV (%)	Formatos Únicos	Período (anos)
Curitiba	147.390	3.354	98,7	8	10,8
Argentina	1.199	17	56,2	40	6
Chile	2.793	20	39,7	28	10

Já a Argentina adota arquitetura CKAN com 1.199 *datasets* organizados por 41 entidades governamentais. Cada *dataset* agrupa em média 5,58 recursos, totalizando 6.693

arquivos. A diversidade de formatos (40 tipos únicos) reflete abordagem multicamada: CSV para dados tabulares (56,2%), ZIP para arquivos consolidados (19,8%), PDF para documentação (9,5%), e formatos geoespaciais nativos (SHP, GeoJSON, KML = 3,6%). Dentre as principais organizações contribuidoras podemos citar o Sistema de Segurança Pública (sspm: 271 *datasets*), Agricultura (145), Energia (123) e Produção (81).

Já o Chile opera o maior catálogo CKAN da análise, com 2.793 *datasets* distribuídos por 227 organizações. A estrutura descentralizada resulta em média menor de recursos por *dataset* (3,83). A diversidade é moderada de formatos (28 tipos), com predominância de CSV (39,7%), XLSX (21,6%) e RAR (15,5%). Entre a liderança de dados, podemos citar a de Orçamentos (436 *datasets*), Aeronáutica Civil (111) e municípios (Ponte Alto: 91, Maule: 71).

Domínio dos Conjuntos de Dados

Os portais analisados apresentam diferentes perfis de cobertura temática (ilustrado como Temas na Tabela 6). Curitiba demonstra o escopo mais abrangente com 37 categorias voltadas principalmente para a gestão municipal, incluindo saúde, serviços ao cidadão, licitações, servidores públicos, cultura e educação. A Argentina organiza seus dados em 13 categorias com foco setorial, utilizando abreviações para itens como energia, transporte, saúde e governo. O Chile, por sua vez, oferece 23 categorias, abrangendo desde ciência, educação e saúde até transporte e turismo, refletindo uma abordagem mais detalhada na categorização temática.

Formatos de Dados

Os Princípios de Dados Abertos exigem que os dados disponibilizados sejam legíveis por máquina, convenientes e livres de barreiras tecnológicas. Considerando o recorte amostral desta pesquisa, a Tabela 8 apresenta a distribuição dos formatos de dados nos três portais analisados.

Curitiba demonstra uma predominância acentuada do formato CSV, com 91 arquivos (83,5% do total), seguido por XLSX com 15 arquivos (13,8%). A Argentina apresenta o maior volume absoluto de arquivos, totalizando 1.567 *datasets*, dos quais 1.338 (85,4%) estão em formato CSV. O portal argentino também oferece 168 arquivos XLSX (10,7%) e demonstra maior diversidade ao incluir outros formatos como ODS, representando uma abordagem mais flexível na disponibilização de dados. O Chile, por sua vez, apresenta um padrão distinto com maior prevalência de arquivos XLSX (144 arquivos, 55,0%), enquanto CSV representa apenas 18,3% dos *datasets* (48 arquivos). O portal chileno também inclui formatos alternativos como Parquet, indicando uma abertura para tecnologias mais recentes de armazenamento de dados.

A predominância do formato CSV em Curitiba e Argentina está consistente com as boas práticas de dados abertos, uma vez que CSV é um formato não proprietário, de fácil leitura por máquinas e amplamente suportado por diferentes ferramentas e linguagens de programação.

Criação do *Dataset* e Atualização Regular

Este critério trata da evolução e o histórico de atualização dos conjuntos de dados

Tabela 8. Distribuição dos Formatos de Dados na Amostra Analisada.

Formato	Curitiba	Argentina	Chile
CSV	91 (83,5%)	1.338 (85,4%)	48 (18,3%)
XLSX	15 (13,8%)	168 (10,7%)	144 (55,0%)
ODS	0 (0%)	1 (0,1%)	0 (0%)
Parquet	0 (0%)	0 (0%)	1 (0,4%)
Outros	3 (2,7%)	60 (3,8%)	69 (26,3%)
Total	109	1.567	262

ao longo do tempo. Observou-se que a maioria dos portais publica uma grande quantidade de dados rapidamente na fase inicial. Contudo, após esta fase de lançamento, a atualização e a publicação de novos conjuntos de dados tendem a tornar-se muito esparsas.

Curitiba demonstra uma cobertura temporal significativa, com dados que remontam a maio de 2015 e se estendem até dezembro de 2024, totalizando 81 datas únicas de atualização ao longo de aproximadamente 9,5 anos. A maioria dos dados possuem atualizações periódicas (geralmente a cada três meses). A Argentina apresenta dados que abrangem o período de dezembro de 2011 a 2025, porém a data de atualização depende do tema: educação, por exemplo, tem dados desde 2011 a 2024. Já transporte só possui 4 anos de dados. Não há uma frequência fixa de atualização para todos os temas. No caso do Chile, os dados variam de 2015 a 2025, porém eles historicamente não se encontram agrupados, e não há uma frequência fixa de atualização.

5.2.3. Metadados

A disponibilização de metadados é essencial sempre que se publica informação na internet. Os metadados fornecem mais informações e recursos para ajudar os editores e consumidores de dados a publicar, descobrir e consumir dados [Lóscio et al. 2016].

Padronização de Termos e Estrutura dos Cabeçalhos

Metadados são informações estruturadas cruciais para descrever e facilitar o uso e a gestão de qualquer recurso de dados. O principal desafio técnico reside na padronização da definição e nomenclatura dos seus campos. Essa padronização deve garantir que o consumidor tenha informação suficiente e não ambígua para compreender e processar eficientemente os dados descritos.

Tabela 9. Características de Metadados dos Portais (Amostra Analisada)

Portal	Amostra CSV	Média Colunas	Delimitador Principal	Idioma Principal
Curitiba	91	17,1	; (91,2%)	PT (83,5%)
Argentina	1.338	14,9	, (68,7%)	ES (71,7%)
Chile	48	19,3	; (85,4%)	ES (42,4%)

Como ilustra a Tabela 9, Curitiba mantém a média de 17,1 colunas (97,8% dos arquivos), com *outliers* apenas em *datasets* emergenciais de COVID-19 (995 e 16.384 colunas). Argentina demonstra maior consistência (14,9 colunas), enquanto Chile apresenta estruturas ligeiramente complexas (19,3). Diferenças regionais de delimitadores (vírgula na Argentina vs. ponto-e-vírgula em Curitiba/Chile) refletem diferenças na padronização de dados sul-americanos.

Informações de Licença e Proveniência

Os conjuntos de dados abertos não devem estar não podem ser restringidos por direitos autorais, patentes, marcas ou segredos comerciais. Por essa razão, com o objetivo de assegurar uma licença aberta, é importante disponibilizar todas as informações que possibilitem o uso, o reuso e a redistribuição dos dados.

A especificação de licença e informações de proveniência varia entre os portais. Os portais baseados em CKAN (Argentina e Chile) declaram explicitamente licenças para cada conjunto de dados, utilizando principalmente *Open Data Commons Open Database License* (ODbL) e *Creative Commons Attribution*. Curitiba, apesar de utilizar uma plataforma proprietária, documenta sua licença de dados abertos no nível do portal através do Decreto Municipal 1023/2014 [Prefeitura Municipal de Curitiba 2014].

Quanto à proveniência, Argentina e Chile fornecem informações detalhadas através dos metadados CKAN, incluindo organizações responsáveis e datas de criação/atualização. Curitiba apresenta não só o setor, mas também o responsável e o contato via telefone.

5.2.4. Esquema de Dados

O esquema de dados é crucial para estruturar e organizar a informação, facilitando a sua gestão e compreensão. No entanto, mesmo em formatos estruturados, persistem problemas comuns como a falta de metadados de esquema, erros de sintaxe ou a presença de registros inválidos (duplicados ou vazios). Tais problemas comprometem a qualidade e tornam o processamento dos dados altamente difícil. O foco desta seção foi na análise de arquivos CSV, pois é um dos tipos de dados mais utilizados para automatização.

Aderência às Regras Sintáticas e Estruturais

Apesar do formato CSV ser bastante utilizado, não há um padrão fixo. Assim como em [Oliveira et al. 2016], o arquivo CSV foi analisado baseado no formato RFC 4180²¹, o padrão mais usado atualmente, como ilustra a Tabela 10.

Arquivos que geraram exceções de *parsing* (*ParserError*) durante esta etapa automatizada foram contabilizados como não conformes ou estruturalmente corrompidos. A maioria dos portais baseados em CKAN tem um *header* apropriado (geralmente uma linha única).

A análise estrutural revelou que todos os arquivos CSV processados com sucesso possuem cabeçalhos, com média de 8-12 colunas por arquivo. Nenhum arquivo sem

²¹<https://www.rfc-editor.org/rfc/rfc4180>

header foi identificado (0,00%), superando problemas de estruturação frequentemente relatados na literatura sobre dados abertos governamentais latino-americanos.

Todos os três portais indicaram conformidade com RFC4180. Os Níveis de duplicação baixos (< 0,06%) demonstram limpeza rigorosa pré-publicação. A presença de colunas com somente com vocabulário em língua nativa, em todos os portais evidencia validação de esquemas antes da disponibilização.

Tabela 10. Aderência a Padrões Sintáticos (RFC4180)

Portal	Duplicação Média	Tipos Mistos
Curitiba	0,058%	0,0
Argentina	0,019%	0,0
Chile	0,022%	0,0

Convenção de Nomenclatura de Campos

Adotar convenções de nomenclatura claras e autoexplicativas para os atributos do esquema é uma boa prática essencial na gestão de bases de dados. Nomes bem definidos são cruciais, pois facilitam a descoberta dos dados pelos consumidores.

Curitiba utiliza convenção distinta com MAIÚSCULAS (COD, NOME, DTHR), como ilustra a Tabela 11, mantida consistentemente através de 50 diretórios. Argentina e Chile seguem padrão internacional com minúsculas (*indice_tiempo*, *cod_detector*). Todos usam línguas nativas.

Tabela 11. Análise de Nomenclatura de Campos

Portal	Média	Mín.	Máx.	Convenção
Curitiba	17,1	1	54	MAIÚSCULAS_underscore
Argentina	14,9	1	841	minúsculas_underscore
Chile	19,3	2	317	minúsculas_underscore

Dimensão do Esquema e Esparsidade de Dados

A análise do tamanho do esquema de dados (número de campos) é importante, pois fornece uma indicação inicial da complexidade e dimensão de um conjunto de dados. Esta métrica ajuda a avaliar a escala da informação definida estruturalmente. A esparsidade de dados é relacionada com a quantidade de células com valor nulo. A Tabela 12 apresenta as estatísticas descritivas sobre a quantidade de cabeçalhos identificados em cada portal.

Tabela 12. Estatísticas dos Headers por Portal

Portal	Média	Mín.	Máx.	Observações
Curitiba	183,87	1	16.384	Outliers em arquivos de COVID-19
Argentina	14,18	1	841	Distribuição mais concentrada
Chile	19,54	1	421	Variabilidade moderada

Uma alta esparsidade pode indicar problemas de qualidade ou características estruturais de *datasets* opcionais. A Tabela 13 contrasta a esparsidade global com uma visão ajustada, onde arquivos *outliers* foram removidos utilizando o método do Intervalo Interquartil (IQR).

Tabela 13. Comparativo de esparsidade com e sem *Outliers*

Fonte	Arquivos	esparsidade Global (%)	Outliers Removidos	esparsidade Ajustada
Curitiba	103	59,72	8	17,15
Chile	185	31,26	16	1,86
Argentina	1.473	9,80	235	1,57

A Argentina demonstra melhor completude global e a menor esparsidade (9,8%). Curitiba apresenta completude inferior (83,8%) e alta esparsidade aparente (59,7%), contudo, a análise de distribuição mostra que este valor é distorcido por outliers. A esparsidade ajustada de Curitiba é de 17,2%, sugerindo que a lacuna de dados não deriva de campos opcionais por categoria temática, mas sim de poucos arquivos com falhas estruturais.

5.2.5. GIS e Dados Espaciais

Além da análise de parâmetros utilizados em [Oliveira et al. 2016], o artigo também avaliou algumas características adicionais de GIS, como formatos disponíveis e disponibilidade de *Dashboards* (Tabela 14).

Tabela 14. Capacidades GIS dos Portais

Portal	Recursos GIS	Formatos Nativos	Sistema Referência	Suporte Nativo
Curitiba	0	Não	WGS84	Limitado
Argentina	254	Sim	SIRGAS2000	Completo
Chile	40	Limitado	WGS84	Básico

A Argentina lidera com 254 tipos de arquivos GIS: 169 Shapefiles (formato padrão para SIG profissional), 46 GeoJSON (compatível com web mapping), 27 KML + 7 KMZ (compatível com Google Earth), e 2 GeoPackage (formato OGC moderno). Chile oferece 40 tipos de arquivos GIS (16 Shapefile, 14 KMZ, 8 KML, 2 WMS), suporte básico mas funcional. Já Curitiba não disponibiliza formatos GIS nativos, mas incorpora coordenadas geográficas em estruturas JSON/CSV. Note que a cidade de Curitiba possui dados espaciais disponibilizados por outros órgãos, como IPPUC²², porém não são informados nem direcionados pelo portal oficial da cidade.

5.3. Resumo da Avaliação

Curitiba destaca-se pela escala – 147.390 arquivos e 3,35 TB, volume 197 vezes superior à Argentina – e boa consistência estrutural (99,32%), mantida ao longo de 10,8 anos em

²²<https://www.ippuc.org.br/>

plataforma proprietária. A padronização do formato CSV (98,7%) facilita processamento automatizado, mas a ausência de formatos GIS nativos exige conversão adicional.

Embora o portal de Curitiba demonstre robustez em volume e consistência estrutural, há desafios de padronização decorrentes de sua plataforma proprietária. Uma análise comparativa com o estudo de [Kozievitch et al. 2022] revela uma evolução significativa na higienização dos dados: enquanto em 2022 foram identificados problemas recorrentes de cabeçalhos duplos e ausência de metadados técnicos, a presente análise encontrou 100% de conformidade nos *headers*. No entanto, a fraca padronização permanece um obstáculo para a interoperabilidade automática; diferentemente dos portais baseados em CKAN (como Argentina e Chile), Curitiba não adota vocabulários controlados universais nem disponibiliza nativamente formatos geoespaciais (Shapefile/GeoJSON).

Argentina possui 254 recursos GIS nativos (Shapefile, GeoJSON, KML) que proporcionam compatibilidade com ferramentas profissionais. A diversidade de 40 formatos equilibra acessibilidade técnica, com CSV (56,2%) para dados tabulares, PDF (9,5%) para documentação e ZIP (19,8%) para consolidação. Ainda há problemas entre a diferença temporal entre domínios temáticos: enquanto áreas como educação possuem séries históricas longas (2011-2024), setores críticos como transporte apresentam janelas de dados reduzidas (apenas 4 anos), sem uma frequência de atualização transversalmente estabelecida. Adicionalmente, o uso significativo de arquivos compactados (ZIP, 19,8%) e documentos estáticos (PDF, 9,5%) impõe barreiras no acesso direto, exigindo etapas manuais de extração que reduzem a pontuação de acessibilidade imediata.

Chile opera o maior catálogo CKAN (2.793 *datasets*) com maior descentralização organizacional (227 entidades), incluindo municípios pequenos ao lado de ministérios nacionais. A heterogeneidade de formatos (28 tipos, Shannon 1,711 [Shannon 1948]) reflete modelo federado que democratiza publicação, embora reduza homogeneidade técnica. A pulverização da publicação entre 227 organizações distintas gera uma heterogeneidade difícil de integrar, resultando em dados historicamente não agrupados. Além disso, o portal apresenta um desvio das melhores práticas de dados abertos ao privilegiar formatos proprietários de planilha (XLSX, 55,0%) em detrimento de padrões abertos como CSV (18,3%), limitando a interoperabilidade direta com sistemas automatizados e contribuindo para a redução de sua nota global de qualidade.

A complementaridade dos três perfis sugere estratégia de triangulação para pesquisas regionais: Curitiba para estudos longitudinais profundos (10,8 anos, granularidade diária), Argentina para análises geoespaciais avançadas (254 GIS) e Chile para estudos comparativos multiorganizacionais (227 entidades). Criticamente, Curitiba demonstra que governança pode superar vantagens de frameworks padronizados, alcançando boa consistência, mesmo sem plataformas como o sem CKAN.

De maneira resumida, desafios apresentados em 2016 [Oliveira et al. 2016] e em 2022 [Kozievitch et al. 2022], ainda estão presentes em 2026, como uma grande quantidade de dados proprietários ou não estruturados (PDF), documentação faltando sobre diferentes versões dos dados e metadados ao longo do tempo, ou não integração com dados GIS (com serviços como visualização no próprio portal).

A RML identificou a falta de padronização regional como um dos principais desafios dos Dados Abertos Governamentais na América Latina. Os resultados empíricos reforçam essa conclusão ao evidenciar diferenças significativas na padronização de formatos, convenções de nomenclatura, qualidade dos metadados e mecanismos de acesso aos dados. Os resultados sugerem que muitos portais latino-americanos permanecem concentrados nos níveis iniciais do modelo de maturidade de dados abertos de cinco estrelas proposto por [Berners-Lee 2009]. Embora os dados sejam disponibilizados para download, ainda existem limitações relacionadas à interoperabilidade semântica, integração entre sistemas e reutilização automatizada.

Segundo [Janssen et al. 2012], a interoperabilidade constitui um dos principais fatores críticos para o sucesso das iniciativas de *Open Government Data*, pois influencia diretamente a reutilização dos dados por cidadãos, empresas e governos. Em ambientes urbanos inteligentes, a integração entre domínios como mobilidade, segurança pública, energia e meio ambiente depende da existência de modelos semânticos compatíveis e mecanismos padronizados de intercâmbio de informações. Nesse contexto, iniciativas como o DCAT-AP da União Europeia ²³, o *European Data Space Framework* ²⁴, e a arquitetura *FIWARE Smart Data Models* ²⁵ representam referências importantes para a construção de ecossistemas interoperáveis de dados urbanos.

Considerando os dados espaciais, por exemplo, os resultados da análise indicam níveis variados de maturidade na disponibilização de dados espaciais pelos portais estudados. Iniciativas internacionais como a *INSPIRE Directive* da União Europeia ²⁶, a *National Spatial Data Infrastructure* (NSDI) ²⁷ e a plataforma *Singapore Smart Nation Sensor Platform* ²⁸ demonstram que a integração entre dados espaciais e governança digital constitui um elemento fundamental para a transformação digital das cidades.

Além disso, com base nos resultados obtidos e nas boas práticas identificadas na literatura, recomenda-se: 1) Adoção de padrões, como DCAT-AP e FAIR [Wilkinson et al. 2016]; 2) disponibilização de APIs para acesso; 3) Publicação de dados geoespaciais em padrões OGC; 4) Monitoramento automático da qualidade dos dados; e 5) Preparação dos portais para aplicações avançadas baseadas em Inteligência Artificial, análise de dados urbanos (*Urban Analytics*) e Digital Twins, por exemplo. Em síntese, os resultados demonstram que a evolução dos portais latino-americanos deve transcender a simples disponibilização de dados, avançando em direção à construção de ecossistemas digitais interoperáveis capazes de sustentar serviços urbanos inteligentes, transparência governamental e inovação baseada em dados.

5.4. Ameaças a Validade e Limitações

Apesar do rigor metodológico adotado, os resultados deste estudo devem ser interpretados considerando algumas ameaças à validade, conforme recomendações para estudos

²³<https://op.europa.eu/pt/web/eu-vocabularies/dcat-ap>

²⁴<https://digital-strategy.ec.europa.eu/en/policies/data-spaces>

²⁵<https://www.fiware.org/smart-data-models/>

²⁶https://knowledge-base.inspire.ec.europa.eu/legislation/inspire-directive_en

²⁷<https://www.fgdc.gov/nsdi/nsdi.html>

²⁸<https://www.developer.tech.gov.sg/products?topics=sensor-platforms-and-internet-of-things>

empíricos em Sistemas de Informação [Wohlin et al. 2012].

Com relação a validade de construção, na RML, a utilização da ferramenta Gemini 3 Pro na etapa inicial com 277 estudos recuperados pode introduzir vieses de classificação. Embora os resultados tenham sido validados manualmente pelos autores, não se pode descartar a exclusão inadvertida de estudos relevantes. Adicionalmente, a metodologia adotada para avaliação técnica dos portais [Oliveira et al. 2016] enfatiza aspectos estruturais dos DAG, como acesso, metadados e esquemas, não contemplando dimensões como valor público, participação cidadã ou efetivo reuso dos dados.

Com relação a validade interna, a análise automatizada dos arquivos pressupõe conformidade mínima com especificações sintáticas, especialmente a RFC 4180 para arquivos CSV. Embora os *parsers* desenvolvidos em Python permitam identificar inconsistências estruturais, eles podem não capturar nuances semânticas específicas dos diferentes domínios analisados. Além disso, os 1.938 arquivos foram coletados entre outubro e novembro de 2025, representando um retrato temporal dos portais. Atualizações posteriores ou variações sazonais podem alterar os resultados obtidos. Diferenças tecnológicas entre os portais — CKAN na Argentina e no Chile e solução proprietária em Curitiba — também podem influenciar indicadores como padronização de metadados e disponibilidade de APIs.

Com relação a validade externa, a seleção dos portais foi intencional, buscando diversidade quanto à escala administrativa e à infraestrutura tecnológica. Contudo, os resultados não devem ser generalizados para toda a América Latina, especialmente para países com diferentes capacidades institucionais e níveis de maturidade em dados abertos. Adicionalmente, a inclusão do caso de Curitiba, reconhecido por iniciativas consolidadas de transparência digital, pode não refletir a realidade média dos governos da região. Ainda assim, a convergência entre os resultados empíricos e os desafios identificados na RML fortalece a plausibilidade analítica dos achados.

Com relação a confiabilidade e reprodutibilidade, todos os artefatos computacionais foram disponibilizados publicamente, favorecendo transparência e replicação. Entretanto, devido à natureza dinâmica dos portais governamentais, futuras replicações podem acessar conjuntos de dados distintos daqueles analisados neste estudo.

Além disso, o estudo concentrou-se predominantemente na qualidade técnica dos DAG, não avaliando o reuso efetivo dos dados por cidadãos, pesquisadores ou empresas. A RML abrangeu o período de 2014 a 2024, podendo não contemplar avanços recentes ocorridos após esse intervalo. Também não foram consideradas métricas de impacto, como estatísticas de acesso e *downloads*, nem realizadas entrevistas com gestores ou testes de usabilidade com usuários finais. Essas limitações não comprometem os principais achados do estudo, mas indicam oportunidades para pesquisas futuras, incluindo estudos longitudinais, ampliação da amostra geográfica e triangulação com métodos qualitativos.

6. Conclusão

Na área de sistemas de informação, dados abertos são um recurso fundamental para impulsionar a criação de novos aplicativos, treinamento de algoritmos de inteligência artificial e a pesquisa científica de dados. No entanto, para que esses avanços sejam realmente

alcançáveis, é crucial superar desafios relacionados à qualidade e à interoperabilidade dos dados.

Este artigo analisou o estado da arte das iniciativas de DAG na América Latina, através de uma RML de estudos fundamentada no protocolo PRISMA, identificando padrões, desafios e oportunidades. Além disso, é apresentada uma avaliação comparativa de três portais de dados governamentais da América Latina (Curitiba-Brasil, Argentina e Chile), seguindo metodologia previamente estabelecida [Oliveira et al. 2016].

A análise revelou padrões consistentes de desafios compartilhados na literatura. A falta de padronização emergiu como obstáculo mais recorrente (65% dos estudos), comprometendo a interoperabilidade entre sistemas governamentais. Problemas na qualidade dos metadados e dificuldades de acesso foram documentados consistentemente, impactando a capacidade de descoberta e reúso dos dados. Apesar de marcos legais consolidados – como a LAI no Brasil (2011) e legislações similares em Chile, Colômbia e Peru –, indicadores quantitativos evidenciam lacunas operacionais: 67% dos portais apresentam falhas de completude e 45% disponibilizam dados desatualizados [Alogaiel and Alrwais 2023, Zainuddin and Akhir 2024], mesmo com adoção predominante de plataformas consolidadas (84,6% utilizam CKAN/DKAN).

A análise comparativa dos três portais revelou três modelos distintos, mas que enfrentam barreiras específicas para a maturidade plena. Argentina destacando-se pela diversidade de formatos (40 tipos únicos) e superior capacidade GIS (254 recursos nativos em Shapefile, GeoJSON e KML), embora sua excelência técnica seja mitigada por inconsistências temporais nas séries históricas de dados.

Curitiba revelou boa escala – 147.390 arquivos totalizando 3,35 TB, volume 197 vezes superior à Argentina – com consistência estrutural boa (99,32%) mantida ao longo de 10,8 anos através de plataforma proprietária. A padronização rígida em CSV (98,7%) facilita processamento automatizado.

O Chile, com o maior catálogo e descentralização (227 entidades), democratizou a publicação, mas o uso predominante de formatos proprietários (XLSX) e a fragmentação histórica limitaram sua pontuação de consistência.

Apesar dos avanços, o estudo apresenta limitações, como o uso assistido de IA na RML, a amostra representativa e o caráter temporal dos 1.938 arquivos, e a seleção de três DAGs, que embora estratégica, não abrange todos os tipos de portais da América Latina.

Como trabalhos futuros, propõem-se ações alinhadas a cada um dos dois objetivos deste estudo. Em relação ao primeiro objetivo (RML), recomenda-se: (a) replicar em bases adicionais (Scopus, Web of Science) para ampliar a cobertura; (b) conduzir meta-análise quantitativa para estimar a magnitude dos desafios de qualidade por país e região; e (c) atualizar a revisão periodicamente para monitorar a evolução das iniciativas de dados abertos.

Em relação ao segundo objetivo (análise de portais), recomenda-se: (a) expandir a análise comparativa para portais adicionais da América Latina, como México, Colômbia e Uruguai, para ampliar a validade externa dos resultados; (b) conduzir estudo longitudinal

repetindo a análise técnica nos mesmos portais em intervalos regulares, para monitorar a evolução da qualidade ao longo do tempo; (c) investigar o impacto efetivo dos dados na transparência governamental e no engajamento cidadão, por meio de análises de reuso dos *datasets*; e (d) desenvolver e validar um *framework* automatizado DAG.

7. Agradecimentos

Os autores agradecem à Universidade Tecnológica Federal, ao Instituto de Pesquisa e Planejamento Urbano de Curitiba (IPPUC) e à prefeitura de Curitiba pelo compartilhamento de parte dos dados utilizados neste estudo.

Referências

- [Albuquerque and Berardi 2020] Albuquerque, J. M. d. and Berardi, R. d. C. (2020). Governo aberto: caminhos para transparência, dados abertos, participação, colaboração e accountability. *Cadernos Gestão Pública e Cidadania*, 25(81).
- [Alogaiel and Alrwais 2023] Alogaiel, N. F. and Alrwais, O. A. (2023). An Assessment of the Quality of Open Government Data in Saudi Arabia. *IEEE Access*, 11:61560–61599.
- [Andrade et al. 2023] Andrade, M. C. d., Moreno, M. J. B., and Pastor-Sánchez, J.-A. (2023). Qualidade dos dados na web: Revisão integrativa sobre diretrizes para publicação web. *Perspectivas em Ciência da Informação*, 28.
- [Ashahril et al. 2022] Ashahril, S. M., Isa, A. M., and Anwar, N. (2022). Challenges in Open Government Data Implementation: A Systematic Literature Review. In *2022 International Conference on Artificial Intelligence of Things (ICAIoT)*, page 1–6.
- [Attard et al. 2015] Attard, J., Orlandi, F., Scerri, S., and Auer, S. (2015). A systematic review of open government data initiatives. *Government Information Quarterly*, 32(4):399–418.
- [Barati 2023] Barati, M. (2023). Open Government Data Programs and Information Privacy Concerns: A Literature Review. *JeDEM*, 15(1):73–123.
- [Barni et al. 2022] Barni, G. d. A. C., Moro, M. F., Bornia, A. C., and Tezza, R. (2022). Visibilidade dos dados nos portais de dados governamentais abertos da América Latina. *Innovar*, 32:187–203.
- [Belete et al. 2017] Belete, G. F., Voinov, A., and Laniak, G. F. (2017). An overview of the model integration process: From pre-integration assessment to testing. *Environmental Modelling & Software*, 87:49–63.
- [Belizario and Berardi 2022] Belizario, M. and Berardi, R. (2022). Linked open data in smart cities: An application in the domains of mobility and education. In *Anais da XVII Escola Regional de Banco de Dados*, pages 61–70, Porto Alegre, RS, Brasil. SBC.
- [Berners-Lee 2009] Berners-Lee, T. (2009). Linked Data – Design Issues, Disponível em: <https://www.w3.org/DesignIssues/LinkedData.html>. Acessado em: 10 de fevereiro de 2026.

- [Bonina and Eaton 2020] Bonina, C. and Eaton, B. (2020). Cultivating open government data platform ecosystems through governance: Lessons from Buenos Aires, Mexico City and Montevideo. *Government Information Quarterly*, 37(3):101479.
- [Brasil 2011] Brasil (2011). Lei nº 12.527, de 18 de novembro de 2011. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2011-2014/2011/lei/l12527.htm. Regula o acesso a informações previsto na Constituição Federal, conhecida como Lei de Acesso a Informação LAI. Acessado em: 10 de fevereiro de 2026.
- [Brasil 2016] Brasil (2016). Decreto nº 8.777, de 11 de maio de 2016. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2016/decreto/d8777.htm. Institui a Política de Dados Abertos do Poder Executivo Federal. Acessado em: 10 de fevereiro de 2026.
- [Brasil 2018] Brasil (2018). Lei geral de proteção de dados pessoais (lgpd) - lei nº 13.709, de 14 de agosto de 2018. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/L13709.htm. Acessado em: 10 de fevereiro de 2026. Diário Oficial da União.
- [Brasil 2023] Brasil (2023). Decreto nº 11.529, de 16 de maio de 2023. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2023-2026/2023/decreto/D11529.htm. Institui o Sistema de Integridade, Transparência e Acesso à Informação da Administração Pública Federal e a Política de Transparência e Acesso à Informação. Acessado em: 10 de fevereiro de 2026.
- [Camperos-Reyes and Gonçalves Sant’Ana 2021] Camperos-Reyes, J. T. and Gonçalves Sant’Ana, R. C. (2021). Abordajes de metadatos para publicadores de datos gubernamentales en el contexto de Iberoamérica. *Revista Interamericana de Bibliotecología*, 44.
- [Controladoria-Geral da União 2020] Controladoria-Geral da União (2020). Modelo de referência para governo aberto. Disponível em: <https://www.gov.br/cgu/pt-br/governo-aberto/a-ogp/planos-de-acao/4o-plano-de-acao-brasileiro/compromisso-2-docs/modelo-de-referencia-de-abertura-de-dados-versao-final-2.pdf>. Compromisso 2 do 4º Plano de Ação Brasileiro na OGP. Acessado em: 10 de fevereiro de 2026.
- [de Oliveira and Silveira 2018] de Oliveira, E. F. and Silveira, M. S. (2018). Open government data in Brazil: a systematic review of its uses and issues. In *Proceedings of the 19th Annual International Conference on Digital Government Research: Governance in the Data Age*, dg.o ’18, pages 60:1–60:9, New York, NY, USA. Association for Computing Machinery.
- [Dwork et al. 2006] Dwork, C., McSherry, F., Nissim, K., and Smith, A. (2006). Calibrating noise to sensitivity in private data analysis. In *Theory of Cryptography*, volume 3876 of *Lecture Notes in Computer Science*, pages 265–284. Springer.
- [Enríquez-Reyes et al. 2021] Enríquez-Reyes, R., Cadena-Vela, S., Fuster-Guilló, A., Mazón, J.-N., Ibáñez, L. D., and Simperl, E. (2021). Systematic Mapping of Open

Data Studies: Classification and Trends From a Technological Perspective. *IEEE Access*, 9:12968–12988.

- [European Commission 2022] European Commission (2022). Open data maturity report 2022. Available at: <https://data.europa.eu/en/publications/open-data-maturity-report-2022-out>. Acessado em: 10 de fevereiro de 2026.
- [European Commission 2025] European Commission (2025). Open data maturity report 2025, disponível em <https://data.europa.eu/en/open-data-maturity/2025>. acessado em: 10 de fevereiro de 2026. Technical report, European Commission.
- [European Data Portal 2016] European Data Portal (2016). Open Data Maturity in Europe 2016: Insights into the European State of Play. Landscaping insight report no. 2, European Commission. Prepared by Capgemini Consulting for the European Data Portal. Disponível em https://data.europa.eu/sites/default/files/edp_landscaping_insight_report_n2_2016.pdf, Acessado em: 10 de fevereiro de 2026.
- [European Data Portal 2024] European Data Portal (2024). Open Data and AI: A Symbiotic Relationship in Progress. Disponível em: <https://data.europa.eu/en/publications/datastories/open-data-and-ai-symbiotic-relationship-progress>. Acessado em: 10 de fevereiro de 2026.
- [Fagundes and Ribeiro Junior 2020] Fagundes, M. F. and Ribeiro Junior, D. I. (2020). Modelo baseado em frictionless data aplicado aos dados abertos governamentais. *RDBCI: Revista Digital de Biblioteconomia e Ciência da Informação*, 18.
- [Flausino 2024] Flausino, M. G. (2024). *Dados abertos: desafios e oportunidades*. Dissertação de Mestrado, Universidade Tecnológica Federal do Paraná.
- [Gao et al. 2022] Gao, T., Yu, T., and Yan, J. (2022). Research on Performance Evaluation of Government Data Open Platform. In *2022 8th Annual International Conference on Network and Information Systems for Computers (ICNISC)*, page 943–948.
- [García-Estrella et al. 2022] García-Estrella, C., Santa-Maria, J. C., and Celis Hernandez, M. R. (2022). Datos abiertos y gobierno abierto en los gobiernos regionales y locales del Perú. *Enfoque UTE*, 13:68–82.
- [Garousi et al. 2019] Garousi, V., Felderer, M., and Mäntylä, M. V. (2019). Guidelines for including grey literature and conducting multivocal literature reviews in software engineering. *Information and Software Technology*, 106:101–121.
- [Gobierno de Chile 2025] Gobierno de Chile (2025). Datos Chile. Disponível em: <https://datos.gob.cl/>. Acessado em: 10 de fevereiro de 2026.
- [Gobierno de España 2024] Gobierno de España (2024). Combining the benefits of open data and artificial intelligence. Disponível em: <https://datos.gob.es/en/blog/combining-benefits-open-data-and-artificial-intelligence>. Acessado em: 10 de fevereiro de 2026.

- [Herrera-Melo and González-Sanabria M.Sc. 2020] Herrera-Melo, C.-A. and González-Sanabria M.Sc., J.-S. (2020). Proposal for the evaluation of open data portals. *Revista Facultad de Ingeniería*, 29.
- [International Organization for Standardization 2019] International Organization for Standardization (2019). ISO 37122:2019 - sustainable cities and communities – indicators for smart cities. Disponível em: <https://www.iso.org/standard/69050.html>. Acessado em: 10 de fevereiro de 2026.
- [Isotani and Bittencourt 2022] Isotani, S. and Bittencourt, I. I. (2022). *Dados abertos conectados*. Novatec Editora.
- [Janssen et al. 2012] Janssen, M., Charalabidis, Y., and Zuiderwijk, A. (2012). Benefits, adoption barriers and myths of open data and open government. *Information Systems Management*, 29(4):258–268.
- [Jefatura de Gabinete de Ministros 2025] Jefatura de Gabinete de Ministros (2025). Datos Argentina. Disponível em: <https://datos.gob.ar/>. Acessado em: 10 de fevereiro de 2026.
- [Knight Lab 2024] Knight Lab (2024). Scrollytelling: A storytelling format for the web. Disponível em: <https://knightlab.northwestern.edu>. Ferramenta de narrativa digital interativa. Acessado em: 10 de fevereiro de 2026.
- [Kozievitch et al. 2022] Kozievitch, N. P., Fonseca, K. V. O., Rosa, M. d. O., Berardi, R., Gutierrez, M. B., Belizario, M., Luz, M. G., and Liu, E. (2022). Assessment of open data portals: A Brazilian case study. In *2022 IEEE International Smart Cities Conference (ISC2)*, pages 1–7.
- [Lóscio et al. 2016] Lóscio, B. F., Burle, C., and Calegari, N. (2016). Data on the web best practices. W3c working draft, World Wide Web Consortium (W3C).
- [Macedo and Lemos 2024] Macedo, D. F. and Lemos, D. L. d. S. (2024). Open government data: maturity diagnosis model for quality data published on the web. *Em Questão*, 30.
- [Maffeo and Hogan 2023] Maffeo, L. and Hogan, B. P. (2023). *Designing Data Governance from the Ground up: Six Steps to Build a Data-Driven Culture*. Pragmatic Express. the Pragmatic bookshelf.
- [Molodtsov and Nikiforova 2024] Molodtsov, F. and Nikiforova, A. (2024). An Integrated Usability Framework for Evaluating Open Government Data Portals: Comparative Analysis of EU and GCC Countries. In *Proceedings of the 25th Annual International Conference on Digital Government Research*, page 899–908. ACM.
- [Moncayo Vives et al. 2022] Moncayo Vives, G., Ortiz Chaves, L. T., and Rodríguez, I. F. (2022). Implementación del barómetro de datos abiertos 2020: desafíos y oportunidades en Chile, Colombia, Ecuador y Perú. *Estado & comunes, revista de políticas y problemas públicos*, 2:59–76.
- [Nogueras-Iso et al. 2021] Nogueras-Iso, J., Lacasta, J., Ureña-Cámara, M. A., and Ariza-López, F. J. (2021). Quality of Metadata in Open Data Portals. *IEEE Access*, 9:60364–60382.

- [OECD 2023] OECD (2023). *2023 OECD Open, Useful and Re-usable data (OURdata) Index: Results and Key Findings*. Number 43 in OECD Public Governance Policy Papers. OECD Publishing, Paris. <https://doi.org/10.1787/a37f51c3-en>.
- [Oliveira et al. 2016] Oliveira, M. I. S., de Oliveira, H. R., Oliveira, L. A., and Lóscio, B. F. (2016). Open government data portals analysis: The Brazilian case. In *Proceedings of the 17th International Digital Government Research Conference on Digital Government Research*, dg.o '16, page 415–424, New York, NY, USA. Association for Computing Machinery.
- [Open Data Watch 2024] Open Data Watch (2024). Open data inventory. Disponível em: <https://odin.opendatawatch.com/>. Acessado em: 10 de fevereiro de 2026.
- [Open Government Partnership 2024] Open Government Partnership (2024). Open data. Disponível em: <https://www.opengovpartnership.org/open-data/>. Acessado em: 10 de fevereiro de 2026.
- [Open Knowledge Foundation 2012] Open Knowledge Foundation (2012). *Open Data Handbook*. Disponível em: <https://opendatahandbook.org/guide/en/what-is-open-data/>. Defines open data as data that can be freely used, re-used and redistributed by anyone - subject only, at most, to the requirement to attribute and sharealike. Acessado em: 10 de fevereiro de 2026.
- [Page et al. 2021] Page, M. J. et al. (2021). Prisma 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews. *BMJ*, 372:n160.
- [Prefeitura Municipal de Curitiba 2014] Prefeitura Municipal de Curitiba (2014). Aprova o regulamento de política de dados abertos do município de Curitiba. Disponível em: <https://legisladoexterno.curitiba.pr.gov.br/VisualizarHTML.aspx?id=186383>. Acessado em: 10 de fevereiro de 2026.
- [Prefeitura Municipal de Curitiba 2023] Prefeitura Municipal de Curitiba (2023). Comitê de dados abertos de Curitiba. Disponível em: <https://dadosabertos.curitiba.pr.gov.br/>. Informações disponíveis no portal de dados abertos da cidade. Acessado em: 10 de fevereiro de 2026.
- [Prefeitura Municipal de Curitiba 2025] Prefeitura Municipal de Curitiba (2025). Portal de dados abertos de Curitiba. Disponível em: <https://dadosabertos.c3sl.ufpr.br/curitiba/>. Acessado em: 10 de fevereiro de 2026.
- [Quarati and Nikiforova 2024] Quarati, A. and Nikiforova, A. (2024). Automating the identification of high-value datasets in open government data portals. *arXiv preprint arXiv:2406.10541*.
- [Samarati and Sweeney 1998] Samarati, P. and Sweeney, L. (1998). Protecting privacy when disclosing information: k-anonymity and its enforcement through generalization and suppression. Technical Report SRI-CSL-98-04, SRI International, Computer Science Laboratory, SRI International.
- [Shannon 1948] Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27(3):379–423.

- [Silva 2024] Silva, P. N. (2024). Reúso de dados abertos do governo brasileiro: atualização da métrica dgabr. *Em Questão*, 30.
- [Silva et al. 2024] Silva, P. N., Silva, G. V. P. d., and Fernandes Filho, D. R. (2024). Catálogo de dados para descoberta e recuperação de dados abertos: uma solução baseada em apis governamentais do Brasil. *Transinformação*, 36.
- [Sousa et al. 2024] Sousa, F. K. d., Fonseca, K. V. O., Kozievitch, N. P., Liu, E., Naves, T., Patuzzo, G. V., and Beuren, A. T. (2024). Rediseño de portales de datos abiertos: Mejorando la usabilidad y accesibilidad. *Cuadernos del CIMBAGE*, 2(26):26–40. Number: 26.
- [Sánchez-Nielsen et al. 2021] Sánchez-Nielsen, E., Morales, A., Mendo, O., and Chávez-Gutiérrez, F. (2021). SuDaMa: Sustainable Open Government Data Management Framework for Long-Term Publishing and Consumption. *IEEE Access*, 9:151841–151863.
- [Ubaldi 2013] Ubaldi, B. (2013). Open government data: Towards empirical analysis of open government data initiatives. Technical Report No. 22, OECD Publishing.
- [Vetrò et al. 2016] Vetrò, A., Canova, L., Torchiano, M., Orozco, C., and Morando, F. (2016). Open data quality measurement framework: Definition and application to open government data. *Government Information Quarterly*, 33(2):325–337.
- [W3C Brasil 2018] W3C Brasil (2018). Web content accessibility guidelines (wcag) 2.1 - tradução autorizada em português do Brasil. Disponível em: <https://www.w3c.br/traducoes/wcag/wcag21-pt-BR/>. Acessado em: 10 de fevereiro de 2026.
- [Wang and Strong 1996] Wang, R. Y. and Strong, D. M. (1996). Beyond accuracy: What data quality means to data consumers. *Journal of Management Information Systems*, 12(4):5–33.
- [Wilkinson et al. 2016] Wilkinson, M., Dumontier, M., Aalbersberg, I., et al. (2016). The fair guiding principles for scientific data management and stewardship. *Scientific Data*, 3:160018. <https://doi.org/10.1038/sdata.2016.18>.
- [Wohlin et al. 2012] Wohlin, C., Runeson, P., Höst, M., Ohlsson, M. C., Regnell, B., and Wesslén, A. (2012). *Experimentation in Software Engineering*. Springer, Berlin, Heidelberg, 1 edition.
- [Zainuddin and Akhir 2024] Zainuddin, Z. and Akhir, E. A. P. (2024). Systematic Literature Review of Data Quality in Open Government Data: Trend, Methods, and Applications. *IEEE Access*, 12:148466–148487.
- [Zanotti et al. 2021] Zanotti, A., Isoglio, A., and Piccotto, N. (2021). Implementación de la política nacional de acceso abierto en tres instituciones argentinas. *Palabra Clave*, 10:123–123.