Universidade do Minho

Escola de Economia e Gestão

Marcelo Jesus Gonçalves Ribeiro

A IMPLEMENTAÇÃO DA POLÍTICA DE ABERTURA DE DADOS NO BRASIL: CONTRIBUTOS PARA SUA COMPREENSÃO

Abril de 2022

A Implementação da Política de Abertura de Dados no Brasil
Contributos para sua Compreensão

Marcelo Jesus Gonçalves Ribeiro

UMinho | 2022

Universidade do Minho
Escola de Economia e Gestão

Marcelo Jesus Gonçalves Ribeiro

A IMPLEMENTAÇÃO DA POLÍTICA DE ABERTURA DE DADOS NO BRASIL: CONTRIBUTOS PARA SUA COMPREENSÃO

Dissertação de Mestrado
Mestrado em Administração Pública
Especialização em Gestão Pública e Políticas Públicas

Trabalho realizado sob a orientação de
Professor Doutor Joaquim Filipe Ferraz Esteves de Araújo

Abril de 2022

DIREITOS DE AUTOR E CONDIÇÕES DE UTILIZAÇÃO DO TRABALHO POR TERCEIROS

Este é um trabalho académico que pode ser utilizado por terceiros desde que respeitadas as regras e boas práticas internacionalmente aceites, no que concerne aos direitos de autor e direitos conexos. Assim, o presente trabalho pode ser utilizado nos termos previstos na licença abaixo indicada. Caso o utilizador necessite de permissão para poder fazer um uso do trabalho em condições não previstas no licenciamento indicado, deverá contactar o autor, através do RepositóriUM da Universidade do Minho.

Licença concedida aos utilizadores deste trabalho:

Atribuição-Não Comercial-Sem Derivações CC BY

https://creativecommons.org/licenses/by/4.0/

AGRADECIMENTOS

Agradeço a Deus por ter-me ajudado a perseverar diante das dificuldades e a ter-me concedido saúde e disposição para que eu pudesse me dedicar de corpo e alma ao mestrado.

Ao meu orientador, Professor Doutor Joaquim Filipe Ferraz Esteves de Araújo, pela disponibilidade, paciência nas revisões e principalmente por trazer ideias e saber enquadrar o estudo a todo o momento, dando o direcionamento e realizando, com excelência, sua função de orientar e guiar.

À minha família que é a razão de tudo que faço e é meu suporte ao longo de toda a vida: minha esposa Fernanda e meus filhos Lucas, Victor e Maria Eduarda.

À ANP – Agência Nacional do Petróleo, Gás Natural e Biocombustíveis, e em especial aos meus chefes e mentores Sérgio Trigo e José Gutman, que acreditaram na importância desta formação, apoiaram e tornaram viável realizar o mestrado.

DECLARAÇÃO DE INTEGRIDADE

Declaro ter atuado com integridade na elaboração do presente trabalho académico e confirmo que não recorri à prática de plágio nem a qualquer forma de utilização indevida ou falsificação de informações ou resultados em nenhuma das etapas conducente à sua elaboração.

Mais declaro que conheço e que respeitei o Código de Conduta Ética da Universidade do Minho.

RESUMO

Dados abertos são um recurso para a implementação das políticas de transparência que contribuem para a promoção da participação social e democracia. Como parte destas políticas, os governos desenvolvem portais de dados, para também viabilizar maior produtividade e eficiência nas relações e prestação de serviços da Administração Pública para com os cidadãos.

O Brasil possui uma política de transparência de dados operacionalizada através de seu portal de dados abertos. Este portal que compreende 200 órgãos e 11 000 conjuntos de dados, sujeitos a regras sobre formato e atualização, pretende ser uma referência para a sociedade que procura informações sobre a atividade governamental, promovendo o uso dos dados para fomentar a democracia, o controle social e a atividade econômica.

O estudo, realizado sobre o portal de dados brasileiro, procurou compreender como as práticas de governança e de divulgação adotadas influenciam nos resultados da política de transparência de dados, materializados pela qualidade dos dados e o acesso do público à informação.

Foi utilizada uma abordagem metodológica mista: quantitativa, composta de inquérito aplicado à equipes de dados abertos e de dados estatísticos do portal de dados; e qualitativa - entrevistas realizadas em uma amostra dos inquiridos que responderam ao inquérito. Foi utilizada regressão linear e análise de correlações entre as variáveis identificadas a partir da revisão da literatura. No que concerne às entrevistas, as intervenções foram classificadas em categorias e perceções de maneira a se poder interpretar.

O resultado do estudo indica que a aplicação de práticas de governança e de medidas para se divulgar dados abertos influenciam positivamente a qualidade dos dados publicados e o acesso a informações. Indica, também, que o papel de liderança dos dirigentes não é determinante. Por fim, mostra uma preocupação com a qualidade do programa de dados abertos em função de restrições crescentes de recursos, nomeadamente o número de pessoas e orçamento.

Palavras-chave: Transparência, Dados Abertos Governamentais, Governo Eletrônico, Governança.

ABSTRACT

Open data are a resource to implement transparency policies and promote civil participation and democracy. As part of those policies, governments develop data portals to achieve greater productivity and efficiency in the relations and provision of public administration services to citizens.

Brazil has a data transparency policy, made operational through its open data portal. This portal comprising 200 organs and 11 000 datasets, which are bound to rules on formats and updating, aims to be a reference for the society on the search for information on government activity, promoting the use of data to foster democracy, social control and economic activity.

The study, conducted on the brazilian data portal, sought to understand how governance and dissemination practices influence the results of the data transparency policy, in terms of data quality and access to information.

A mixed method approach was applied: quantitative, consisting of a survey applied to open data teams and statistical data from the portal; and qualitative - interviews conducted in a sample of those which attended the survey. Linear regression and correlation analysis between the variables identified from the literature review were done. Regarding the interviews, the quotes were grouped into categories and perceptions to be interpreted.

The result of the study indicates that the application of governance practices and measures to disseminate open data positively influence the quality of published data and access to information. It also indicates that the role performed by agency leaders is not relevant. Eventually, it shows a major concern to the quality of the open data program due to increasing resource constraints in terms of personnel and budget.

Keywords: Transparency, Open Government Data, E-Government, Gouvernance.

ÍNDICE GERAL

DIREITOS DE AUTOR E CONDIÇÕES DE UTILIZAÇÃO DO TRABALHO POR TERCEIROS ii
AGRADECIMENTOS iii
DECLARAÇÃO DE INTEGRIDADE iv
RESUMO v
ABSTRACT vi
Lista de Abreviaturas e Siglas 9
Índice de Figuras 9
Índice de Gráficos 9
Índice de Tabelas 10
Capítulo 1. Introdução 11
1.1 Justificativa do Tema e Formulação do Problema 11
1.2 Objetivo Geral e Específico 12
1.3 Relevância do Estudo 13
1.4 Desenvolvimento e Estrutura do Trabalho 13
Capítulo 2. Revisão da Literatura 14
2.1 Conceitos Relacionados a Dados Abertos 14
2.2 Transparência Governamental e Dados Abertos 15
2.3 Princípios de Governança sobre Dados Abertos 17
2.4 Desafios Encontrados na Disponibilização de Dados Abertos 21
2.5 Agrupamento dos Princípios de Dados Abertos em Dimensões 24
Capítulo 3. Dados Abertos no Brasil 27
3.1 Contextualização dos Mecanismos de Transparência de Dados no Brasil 27
3.2 Open Government Partnership e o Compromisso Brasileiro 28
3.3 O Portal de Dados Abertos 29
Capítulo 4. Design da Investigação 31
4.1 Hipóteses de Investigação e Modelo Conceptual 33
4.2 Técnicas de Análise Utilizadas 37
4.3 Unidades de Análise 38
4.4 Amostra e Procedimentos de Amostragem 39

4.5 Recolha de Dados 41
4.6 Operacionalização dos Conceitos 42
4.7 Validade e Fiabilidade do Instrumento 48
Capítulo 5. Apresentação e Análise dos Dados. 49
5.1 Perfil dos Respondentes ao Questionário 50
5.2 Análises Descritivas 52
5.3 Análise Bivariante e Multivariante 63
5.4 Visão Geral sobre as Entrevistas 68
5.5 Análise das Variáveis do Estudo a partir das Entrevistas 71
Capítulo 6. Considerações Gerais 75
6.1 Discussão dos Resultados e Conclusões 75
6.2 Limitações do Estudo 76
6.3 Pistas para Investigações Futuras 77
6.4 Contributo do Estudo 77
Capítulo 7. Referências Bibliográficas 78

Lista de Abreviaturas e Siglas

CGU.....	Controladoria Geral da União
CKAN.....	*Comprehensive Knowledge Archive Network*
COFOG.....	Classification of the Functions of Government
CV.....	Coeficiente de Variação
IBGE.....	Instituto Brasileiro de Geografia e Estatística
LAI.....	Lei de Acesso â Informação
LGPD.....	Lei Geral de Proteção de Dados
OCDE.....	Organização para a Cooperação e Desenvolvimento Econômico
OGP.....	Open Government Partnership
OKF.....	Open Knowledge Foundation
PDA.....	Plano de Dados Abertos
PNAD.....	Pesquisa Nacional por Amostra de Domicílios
TI.....	Tecnologia da Informação

Índice de Figuras

Figura 1 - Princípios Pioneiros em Dados Abertos 18
Figura 2 - Evolução de Princípios de Dados Abertos 19
Figura 3 – Recomendações dos Autores em Dados Abertos 20
Figura 4 - Dimensões dos Dados Abertos 25
Figura 5 - Painel de Monitoramento de Dados CGU 30
Figura 6 - Números do Portal Dados.gov.br 31
Figura 7 - Design de Investigação 32
Figura 8 - Localização dos Respondentes 40
Figura 9 - Variáveis a partir das Dimensões 44

Índice de Gráficos

Gráfico 1 - Total de Respondentes por Constituição do Órgão 50
Gráfico 2 - Respondentes por Função de Governo 51
Gráfico 3 - Respondentes por Género 51

Gráfico 4 - Respondentes por Faixa Etária 52
Gráfico 5 - Respondentes por Tempo de Serviço no Órgão 52
Gráfico 6 - Histograma da Variável Qualidade dos Dados - Questionário 53
Gráfico 7 - Qualidade dos Dados - Frequência de Ocorrência - Questionário 54
Gráfico 8 – Histograma de Disseminação da Informação - Questionário 55
Gráfico 9 - Frequência das Ações - Disseminação da Informação - Questionário 56
Gráfico 10 - Histograma de Institucionalização - Questionário 56
Gráfico 11 - Institucionalização Macro - Frequência de Ocorrência - Questionário 57
Gráfico 12 - Institucionalização Micro - Frequência de Ocorrência - Questionário 58
Gráfico 13 - Histograma de Recursos de Apoio - Questionário 58
Gráfico 14 - Recursos de Apoio - Frequência de Ocorrência - Questionário 59
Gráfico 15 - Histograma de Liderança - Questionário 60
Gráfico 16 - Liderança - Frequência de Ocorrência - Questionário 60
Gráfico 17 - Histograma Variável Acesso à Informação – Visualizações - Portal 62
Gráfico 18 - Histograma Variável Institucionalização - Respeito ao Plano - Portal 62
Gráfico 19 - Histograma Variável – Disseminação da Informação - Menções - Portal 63
Gráfico 20 - Homocedasticidade e Normalidade dos Resíduos na Regressão - Questionário 65
Gráfico 21 - Homocedasticidade e Normalidade dos Resíduos na Regressão Log - Portal 67
Gráfico 22 - Respondentes por Função de Governo - Entrevistas 68
Gráfico 23 - Respondentes por Constituição do Órgão - Entrevistas 69
Gráfico 24 - Perceção dos Entrevistados - Polaridade 69
Gráfico 25 - Perceção dos Entrevistados - por Categoria 70
Gráfico 26 - Nuvem de Palavras das Entrevistas - Categorias 71

Índice de Tabelas

Tabela 1 - Domicílios com Acesso à Internet - 2016 a 2019 29
Tabela 2 – Hipóteses e Autores para Sustentação 36
Tabela 3 - Amostragem do Questionário 39
Tabela 4 - Variáveis a partir das Dimensões - Questionário 44
Tabela 5 - Variáveis a partir das Dimensões - Dados Secundários 48
Tabela 6 - Valores do Alfa de Cronbach para as Dimensões do Questionário 49
Tabela 7 - Estatística Descritiva - Variáveis - Questionário 53

Tabela 8 - Estatística Descritiva - Variáveis - Dados Secundários 61
Tabela 9 - Resumo Correlação de Spearman para variável Qualidade dos Dados 64
Tabela 10 - Matriz de Correlação das Variáveis Independentes no Questionário 64
Tabela 11 - Resultado do Modelo de Regressão - Questionário..... 65
Tabela 12 - Resumo Correlação de Spearman para variável Acesso à Informação - Visualizações..... 66
Tabela 13 - Resultado do Modelo de Regressão com Log e QuasiPoisson - Portal..... 68
Tabela 14 - Agrupamento Comentários nas Variáveis do Estudo - Entrevistas..... 71

Capítulo 1. Introdução

Dados abertos são um recurso que permite implementar políticas de transparência e promove participação social e democracia. Como parte destas políticas, os governos desenvolvem portais de dados abertos para publicar informações sobre suas atividades, o que permite o controle social e também maior produtividade e eficiência nas relações e prestação de serviço da Administração Pública para com os cidadãos.

O Brasil possui uma política de transparência de dados (Presidência da República, 2016), a qual está baseada em princípios norteadores para publicação e divulgação de dados. A presente investigação tem como objetivo analisar os resultados obtidos na gestão de publicações no portal brasileiro de dados abertos, que é o instrumento de materialização de sua política.

Foi possível identificar no estudo que os princípios de dados abertos presentes na literatura são universais e encontram-se presentes nas iniciativas relatadas, e coincidem com o que foi proposto na política brasileira de transparência de dados.

Percebe-se que há fatores de influência dos princípios na forma de gestão dos programas de dados abertos, conforme relatados na literatura e no caso brasileiro. Este estudo procura identificar estes fatores, relações e resultados, e compor um cenário e análise crítica da experiência brasileira em dados abertos.

1.1 Justificativa do Tema e Formulação do Problema

Dados Abertos foram conceituados em 2004 pela Open Knowledge Foundation (2004), e suas definições e princípios continuaram a evoluir com Tim Berners-Lee (2006). Na sequência, Dados Abertos Governamentais são formulados por um grupo de pesquisadores e representantes de empresas de tecnologia (2007).

A administração Obama (2009) toma a liderança no incentivo à transparência governamental para fortalecimento da democracia, incentivando a colaboração e a participação social como orientação de base para um governo aberto. Em consequência a este movimento, os Estados Unidos aderem aos dados abertos como um dos instrumentos de execução da política, ao criar seu portal de dados. Em seguida, vários países iniciaram um processo progressivo de transparência governamental para compartilhamento e abertura de seus dados. Em um período de tempo menor que 20 anos, a publicação de dados abertos governamentais consolidou-se nos países democráticos. O número de países a possuir portais de dados abertos governamentais aumentou de 46 em 2014 para 153 em 2020 (T. M. Yang & Wu, 2021) (UN, 2020).

Em um contexto nacional, o governo brasileiro implementou uma série de leis e medidas na direção da promoção da transparência de dados. O movimento foi iniciado pela LAI – Lei de Acesso à Informação (Presidência da República, 2011), a qual estabeleceu o processo de gestão transparente da informação e de sua classificação. No contexto específico de dados abertos governamentais, o governo brasileiro promulgou o decreto-lei que instituiu a Política Nacional de Dados Abertos (Presidência da República, 2016), materializada pelo Portal Brasileiro de Dados Abertos. Decorridos 6 anos da criação desta política, importa analisar se o portal responde aos desafios apresentados pela política de transparência de dados. Para o efeito foram formuladas as seguintes perguntas de investigação: "Em que medida o portal de dados abertos governamentais brasileiro atende aos requisitos da política de transparência de dados? No Portal de Dados Abertos são aplicados os princípios encontrados na literatura?"

De maneira a viabilizar a pesquisa e procurar responder à pergunta de investigação, o estudo buscou identificar os princípios de dados abertos na literatura. Em complemento, procurou conhecer, segundo a perspetiva das pessoas que trabalham com dados abertos, os arranjos organizacionais e as práticas adotadas. Esses elementos, agregados, foram utilizados pelo estudo para se testar as hipóteses formuladas e se construir as conclusões.

1.2 Objetivo Geral e Específico

O objetivo geral desta dissertação é entender como a forma de governança adotada e as iniciativas para se disseminar informação influenciam nos resultados da política de transparência de dados, materializados pela qualidade dos dados e o acesso do público à informação.

Em termos de objetivos específicos, pretende-se:

a. Analisar os princípios de dados abertos encontrados na literatura sobre a política de transparência;
b. Compreender, no escopo do portal de dados abertos brasileiro, os processos de publicação, os fatores de influência e os resultados a partir da perspetiva das pessoas que disponibilizam os dados;
c. Avaliar a perceção da utilidade, da efetividade e dos desafios para se implementar as políticas de transparência de dados.

1.3 Relevância do Estudo

Os governos materializam suas políticas de transparências de dados nomeadamente através de portais de dados abertos, realizando investimentos em recursos humanos e de infraestrutura. O estudo do tema é importante para que se possa identificar os mecanismos de gestão utilizados nestes programas e suas consequências. Ainda, poderá permitir conhecer a aderência destas políticas às melhores práticas encontradas na literatura. Os resultados da pesquisa poderão indicar caminhos para se melhorar os programas em termos de efetividade e alcance das políticas de transparência.

1.4 Desenvolvimento e Estrutura do Trabalho

O primeiro capítulo introduz o tema de dados abertos, a importância de sua aplicação nas políticas públicas de transparência.

O segundo capítulo faz a revisão da literatura, dividida em conceitos relacionados a dados abertos, transparência governamental, princípios, desafios e dimensões de dados abertos.

O terceiro capítulo apresenta, no contexto dos mecanismos que viabilizaram as políticas de transparência, as ações do governo brasileiro para governo aberto e o portal de dados, instrumento de realização da política nacional de dados abertos.

O quarto capítulo apresenta o desenho de investigação, detalhando as hipóteses e as ferramentas utilizadas para coletar os dados – inquérito, dados secundários e entrevistas.

O quinto capítulo realiza a análise dos dados do inquérito e dos dados secundários e os resultados encontrados. São apresentados os resultados da análise das entrevistas.

O sexto e último capítulo apresenta as conclusões a partir dos resultados analisados, as limitações do estudo, o contributo esperado e ainda pistas para futuras investigações.

Capítulo 2. Revisão da Literatura

Este capítulo trata do estudo sobre as origens e conceitos relacionados a dados abertos e a relação entre as políticas de transparência governamental e o uso de dados abertos. Em seu desenvolvimento, identifica princípios aceites e desafios apontados pelos autores que os governos devem vencer, assim como as dificuldades enfrentadas para a disponibilização e sucesso dos programas de dados abertos governamentais. Conclui ao apresentar as dimensões que agrupam os princípios de dados abertos e que foram selecionadas para análise nesta dissertação.

2.1 Conceitos Relacionados a Dados Abertos

Para conceituar e definir dados abertos tais como encontramos na literatura, é importante compreender inicialmente o que são dados. Para Davenport & Pruzak (1998), dados são um conjunto discreto de fatos objetivos sobre eventos, e em um contexto organizacional é usualmente descrito como registros estruturados de transações. Dados isolados possuem pouca relevância e propósito. No âmbito da ciência da computação, Viana (2001) define dado como uma sequência de símbolos quantificados ou quantificáveis. Para Miranda (1999), dado é o conjunto de registros qualitativos ou quantitativos. Um texto é um dado, assim como valores, fotos, figuras, sons gravados.

Dados são abertos quando podem ser livremente acessados, modificados e distribuídos por qualquer pessoa para qualquer propósito, preservando suas características (Open Knowledge Foundation, 2004). Os dados abertos são representados como um agrupamento de dois tipos de dados: o conteúdo propriamente dito e registos para melhor compreensão deste conteúdo, chamados "metadados". Os metadados indicam onde o conteúdo está localizado, seu significado, como relacionar o conteúdo com outras fontes de dados, a frequência de atualização e condições de uso. Os metadados são de grande importância para preservação do conteúdo porque fornecem contexto e significado para as coleções de dados (Ubaldi, 2013).

As bases conceituais para o acesso, distribuição, reuso e licenciamento de dados abertos foram criadas pela Open Knowledge Foundation – OKF (Open Knowledge Foundation, 2004). Em 2007, um grupo de pesquisadores e representantes de empresas de tecnologia formulou os princípios orientadores para dados abertos governamentais, baseados no livre acesso, na conectividade, em formatos não proprietários, na necessidade de os dados serem atuais e na possibilidade de sua reutilização e distribuição (Malamud et al., 2007).

Os governos sempre coletaram dados e os armazenaram para desempenhar suas atividades. A necessidade do uso de dados governamentais pela sociedade como forma de controle social e para melhorar os serviços públicos acompanhou a progressiva oferta de dados. Por décadas, a maneira pela qual os dados governamentais são coletados e agregados têm sido de grande interesse para o governo, academia e sociedade (UN, 2020). A maneira pela qual dados governamentais são divulgados e disseminados mudou drasticamente na última década, impulsionada pela revolução tecnológica e a proliferação de aplicações para tratar, transformar e distribuir os dados, entre elas incluídos os portais de dados abertos governamentais. Uma das formas para distribuir os dados governamentais é através de dados abertos.

Dados governamentais são quaisquer dados produzidos ou processados e agrupados por organizações públicas. Dados abertos governamentais são dados produzidos pelos governos que podem ser livremente usados, reutilizados e distribuídos por qualquer pessoa (Ubaldi, 2013). Estes são produzidos e divulgados com o objetivo de entregar informação.

A Informação pode ser descrita como uma "mensagem", normalmente sob a forma de um documento, áudio ou comunicação visível. Espera-se que a informação mude a maneira pela qual seu recetor perceba os fatos. Enfim, a informação deve ser pensada como o dado que faz a diferença (Davenport & Pruzak, 1998). A informação é uma abstração, representada pela combinação e interpretação de dados. É composta por dados organizados de modo significativo, sendo subsídio útil à tomada de decisão (Miranda, 1999). Para Dante (1998), a informação é conceituada como dados ou matéria informacional relacionada ou estruturada de maneira potencialmente significativa.

Para que um dado seja convertido em informação, ele precisa ser interpretado, passando a ter significado. Os dados abertos governamentais apresentados isoladamente são apenas dados, e não informação. Para entregar valor à sociedade e para que possam ser utilizados, devem estar inseridos em um contexto, combinados com outros dados e revestidos de informações de apoio que facilitem sua compreensão.

2.2 Transparência Governamental e Dados Abertos

Vemos em Kaufmann & Bellver (2005, p. 2) que os cidadãos devem poder obter acesso à informação e à transparência. Em linhas gerais, pode-se dizer que, entre outras coisas, uma gestão transparente tem como principais características o acesso a informações compreensíveis para todo cidadão e a abertura para sua participação no governo. Segundo Cruz (2012, p. 156), uma gestão com transparência é marcada pelo acesso a informações compreensíveis para todo cidadão e pela abertura

da possibilidade de se participar na administração, figurando como princípio de governança. A transparência é considerada um elemento chave de um bom sistema de governança corporativa (Tejedo-Romero & Araujo, 2018).

De acordo com Brandsma (2008), a transparência incorpora tanto o direito de cada cidadão para aceder a informações e o dever do governo para garantir que a informação esteja disponível nos websites. Em Araújo & Tejedo-Romero (2016), vemos que alguns países comportam-se de forma pró ativa, com as organizações públicas a exercer papel ativo em prover informação aos atores da sociedade. Em contrapartida, outros países liberam informação apenas de forma passiva, segundo requisição direta dos cidadãos.

Na relação entre governos e sociedade, podemos entender a transparência como a capacidade de os governos divulgarem suas ações e compromissos, tornando públicas suas atividades. Entre as várias maneiras de transparência governamental, está a disponibilização de dados abertos governamentais.

De maneira a obter uma sociedade democrática e que funcione, os cidadãos e demais partes da sociedade precisam estar aptos a monitorar as iniciativas de governo e sua legitimidade. Promover a transparência quer também dizer que que seja possível, além de aceder aos dados, que estes possam ser úteis e reutilizáveis e com isto reduzir o problema da agência.

Segundo a Teoria da Agência, uma relação de agência é um contrato no qual uma ou mais pessoas recebem de outra pessoa um serviço. Esta relação envolve a delegação de autoridade para o agente tomar decisões (Jensen & Meckling, 1976). A seu turno, o governo tem interesse em divulgar seus dados aos cidadãos pois pode assim tornar-se mais eficiente. A publicitação de dados abertos governamentais em larga escala é capaz de reduzir a assimetria de informação entre o agente – governo - e o principal – o cidadão.

Em complemento, segundo Araújo (2004), as melhorias advindas das pressões que viabilizam reformas, em que mudanças acontecem a nível local muitas vezes apresentam ganhos na produção e serviço, podendo melhorar sua qualidade, e conferir transparência e equidade.

As organizações públicas produzem e coletam diferentes tipos de dados de maneira a desempenhar suas atividades. A extraordinária quantidade e centralidade dos dados coletados pelos governos os tornam particularmente significativos como um recurso para aumentar a transparência pública. Dados abertos governamentais podem ajudar os cidadãos a compreender o que os governos fazem, se desempenham corretamente suas funções e assim os responsabilizarem por resultados não atingidos ou ações não realizadas. A disponibilização de dados abertos permite, ainda, obter pistas de como melhorar o desempenho dos governos. O aumento da transparência de dados proporciona as

bases para a participação popular e colaboração para criação e inovação de serviços governamentais (Ubaldi, 2013).

A Internet, de forma geral, e sites oficiais de governo, em particular, mudaram a forma de interação entre cidadãos e governos. Uma quantidade massiva de dados pode ser coletada, distribuída e transformada por empresas privadas, jornalistas, organizações civis e pelo público em geral. Entre as iniciativas de governo aberto, estão os portais de dados que desempenham um papel crucial na promoção de transparência governamental, participação e colaboração (Cuillier & Piotrowski, 2009; Jaeger & Bertot, 2010). A adoção de políticas públicas de dados centradas na perspetiva do cidadão tem sido solicitada como um meio de engajar o público, aumentar a confiança no governo e o valor público das políticas de transparência (Harrison et al., 2012; Marijn Janssen et al., 2012).

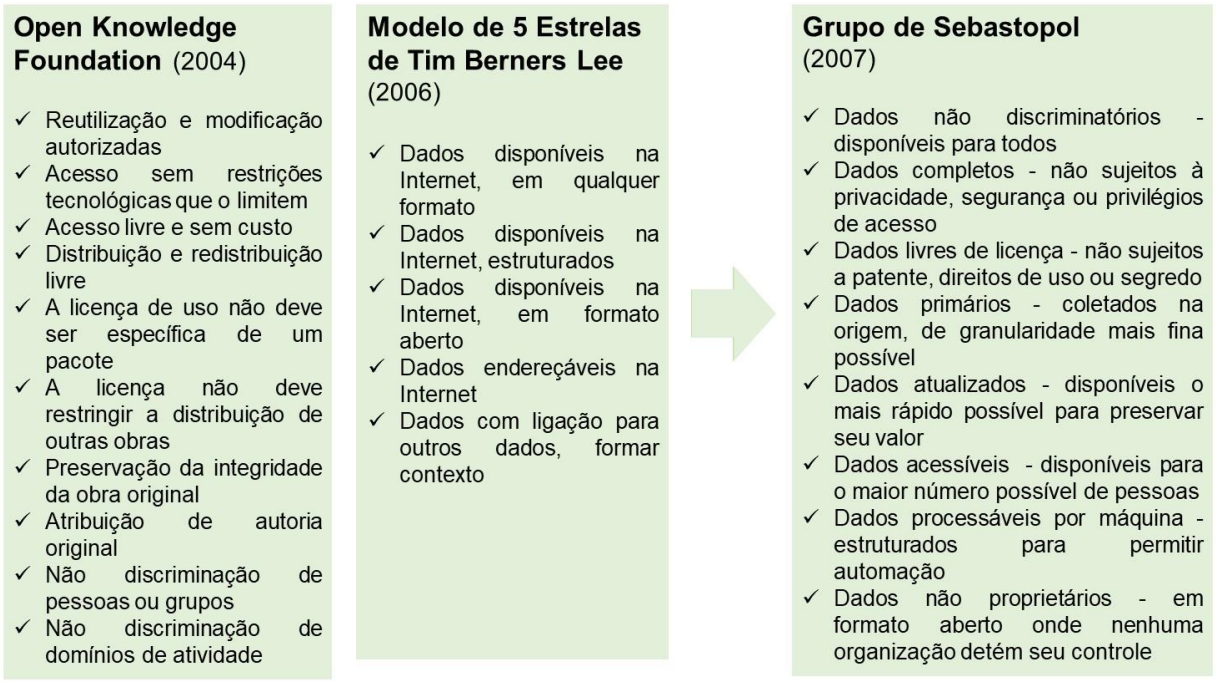
2.3 Princípios de Governança sobre Dados Abertos

A literatura sobre dados abertos identifica uma série de princípios para publicação dos dados. Estes visam garantir que os dados entregues forneçam informação para o público, proporcionem valor e possibilidade de uso, reuso e distribuição. Os dados abertos publicados conforme os princípios identificados são um mecanismo de entrega e execução de políticas de transparência dos governos.

Os princípios enunciados para dados abertos pelos autores complementam-se entre si, tendo evoluído de uma abordagem centrada na natureza dos dados para a prática e processos de governança de dados. Em um primeiro momento, foram estabelecidos princípios técnicos que procuraram estabelecer alguma orientação que pudesse garantir conexão dos dados e entrega de informação (Berners-Lee, 2006; Malamud et al., 2007; Open Knowledge Foundation, 2004). A primeira vez em que dados abertos foram conceituados como princípios técnicos foi em 2004, através de publicação da OKF (Open Knowledge Foundation, 2004). Tim Berners-Lee (2006) propôs em seguida o modelo de 5 estrelas para dados abertos. Em continuidade ao desenvolvimento do tema, um grupo de 30 executivos e técnicos da indústria de Tecnologia da Informação e Comunicações, entusiastas de Open Government, reuniram-se em Sebastopol (Malamud et al., 2007; Ubaldi, 2013), no qual foram formulados mais princípios.

Percebe-se na definição de princípios para dados abertos uma evolução de modelos e orientações mais voltadas aos dados e o seu licenciamento (Berners-Lee, 2006; Open Knowledge Foundation, 2004) e maior detalhamento e objetividade em seu uso, disponibilidade e publicação em (Malamud et al., 2007). A Figura 1 apresenta os princípios propostos pelas três principais abordagens de dados abertos.

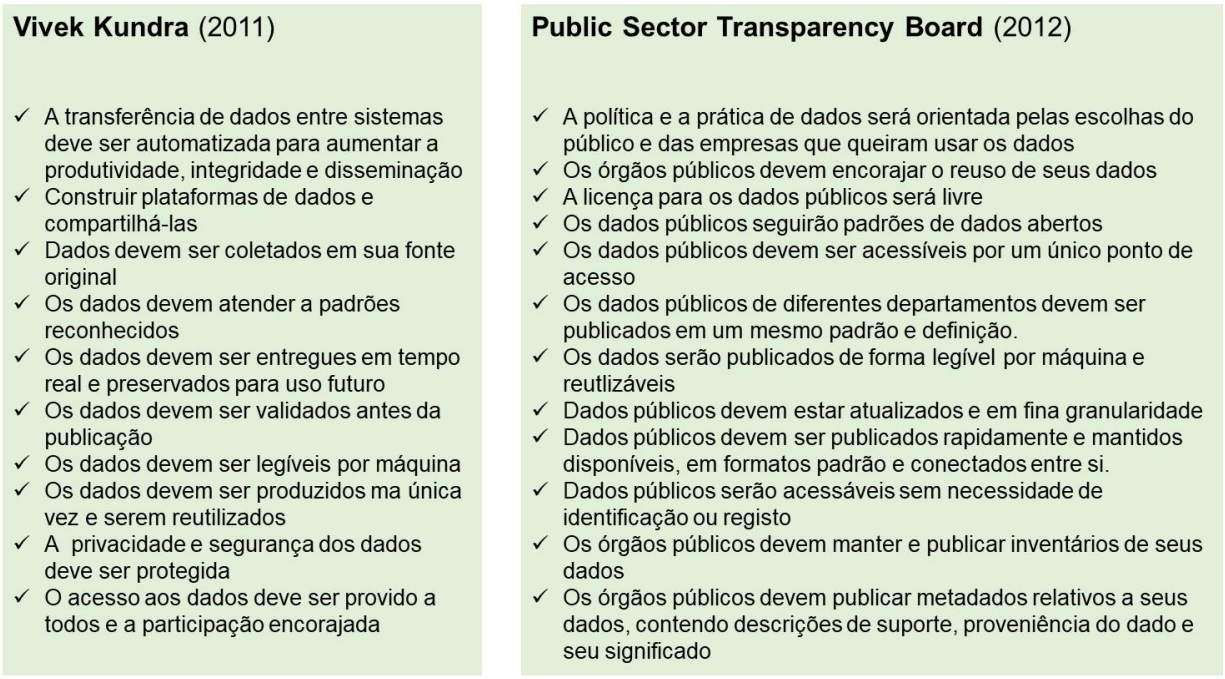
Figura 1 - Princípios Pioneiros em Dados Abertos



Fonte: Elaboração própria a partir da reunião de elementos em (Berners-Lee, 2006; Malamud et al., 2007; Open Knowledge Foundation, 2004)

No campo das diretrizes para implementação de dados abertos governamentais, Vivek Kundra, o Diretor de Tecnologia de Informação do Governo dos Estados Unidos, e o United Kingdom Public Sector Transparency Board redigiram mais princípios e orientações sobre como publicar os dados e promover transparência (Ubaldi, 2013). Estes princípios são de aplicação direta e prática e procuram enquadrar os processos de produção e divulgação de dados abertos, introduzindo orientações voltadas à qualidade, à governança e a formas de disseminação da informação, conforme é mostrado na Figura 2.

Figura 2 - Evolução de Princípios de Dados Abertos



Fonte: Elaboração própria a partir da reunião de elementos em (Ubaldi, 2013)

Os princípios técnicos podem ser considerados como base para o desenvolvimento de programas de dados abertos governamentais. Adicionalmente, identifica-se na literatura recomendações complementares a estes princípios. São estabelecidas diretrizes de promoção de institucionalização, disponibilização de recursos de apoio e de importância da liderança em programas de dados abertos. Estas recomendações introduzem processos de governança em dados abertos e formas de disseminação de informação (Attard et al., 2015; Dawes et al., 2016; Gascó-Hernández et al., 2018; Hartog et al., 2014; Marijn Janssen et al., 2012; Kim & Eom, 2019; Kučera et al., 2013; Lourenço, 2015; Luna-reyes et al., 2016; Purwanto et al., 2019; Ruijer et al., 2020; Ruijer & Meijer, 2020; Safarov, 2019; Ubaldi, 2013; H. J. Wang & Lo, 2016, 2016; V. Wang & Shepherd, 2020; T. M. Yang & Wu, 2016, 2021; Z. Yang et al., 2015; Zuiderwijk et al., 2019; Zuiderwijk & Janssen, 2014b). A Figura 3 faz um resumo das recomendações destes autores.

Figura 3 – Recomendações dos Autores em Dados Abertos

Ubaldi (2013)

- ✓ Identificar conjuntos de dados relevantes através de consultas públicas
- ✓ Buscar voluntários na sociedade civil para aconselhamento sobre os dados a disponibilizar
- ✓ Promover eventos de co-criação de dados junto à sociedade
- ✓ Estruturar, rastrear e personalizar as respostas para as perguntas dos cidadãos sobre dados abertos
- ✓ Estabelecer fluxos de trabalho adequados para coletar, integrar, validar, aprovar e disponibilizar os dados
- ✓ Investir em práticas de administração de dados
- ✓ Criar um responsável por dados abertos que irá representar o órgão e incentivar o uso
- ✓ Garantir a atualidade de dados nos processos de publicação

Dawes (2016)

- ✓ Avaliar a qualidade dos dados a partir dos formatos dos dados e metadados
- ✓ Modernizar os sistemas para oferecer dados de melhor qualidade ao público
- ✓ Adotar padrões técnicos para os formatos de dados que serão publicados
- ✓ Organizar consultas públicas, reuniões e workshops para fomento da comunicação com os cidadãos

Kucera (2013)

- ✓ Obter qualidade pela modificação dos dados
- ✓ Obter qualidade ao redesenhar o processo de criação dos dados

Attard (2015)

- ✓ Encontrar uso inovativo para os dados através de concursos e desafios competitivos
- ✓ Aplicar padrões em larga escala de treinamento para superar estas questões e prever dotação orçamentária
- ✓ Estabelecer um plano de ação para estimular o consumo de dados abertos

Ruijer e Meijer (2020)

- ✓ Obter forte suporte da gestão de topo e da média gerência para o sucesso dos programas de dados abertos
- ✓ Realizar desenvolvimentos em tecnologia da informação para potencializar a publicação de dados

T. M. Yang & Wu (2016)

- ✓ Observar as leis de proteção de dados na decisão de disponibilização
- ✓ Utilizar a liderança dos dirigentes como vetor para impulsionar os funcionários públicos a abrir seus dados

Lourenço (2015)

- ✓ Garantir atualidade dos dados como chave para permitir seu uso e qualidade
- ✓ Garantir visibilidade para os conjuntos de dados em função de sua diversidade e quantidade disponíveis

Marijn Janssen (2012)

- ✓ Investir em ferramentas para melhorar a visualização de dados e auxiliar os cidadãos

Luna-reyes (2016)

- ✓ Obter liderança política e incluir especialistas em dados e da sociedade civil em programas de dados abertos

Gascó-Hernández (2018)

- ✓ Realizar sessões de treinamento como impulsionador do uso de dados abertos

Safarov (2019)

- ✓ Investir em capacitar os cidadãos em habilidades e conhecimentos técnicos para uso dos dados.
- ✓ Melhorar a qualidade dos dados com o envolvimento da sociedade através de melhores mecanismos de feedback

Kim & Eom (2019)

- ✓ Utilizar a liderança para mobilizar recursos tangíveis e intangíveis e aplicá-los nos programas de dados abertos
- ✓ Estabelecer orçamento dedicado aos programas de dados abertos

Z. Yang (2015)

- ✓ Investir em infraestruturas de tecnologia da informação e no apoio da gestão de topo como fatores organizacionais em programas de dados abertos

H. J. Wang & Lo (2016)

- ✓ Viabilizar o acesso a recursos adequados, reengenharia de processos e coordenação dos atores através do apoio da gestão de topo

V. Wang & Shepherd (2020)

- ✓ Diferenciar dado de informação para que o cidadão seja capaz de mais facilmente encontrar o resultado em suas buscas

Zuiderwijk (2014)

- ✓ Dedicar esforços para identificar quais conjuntos de dados são mais relevantes e devem ser abertos com prioridade
- ✓ Estabelecer processos de negócio adequados e divisão de responsabilidades de maneira a desbloquear a publicação dos dados
- ✓ Melhorar os mecanismos de busca de maneira a se encontrar e navegar melhor pelos dados

Hartog (2014)

- ✓ Automatizar as extrações de dados sem ações manuais dos funcionários públicos de forma a obter melhor resultado em programas de dados abertos

Fonte: Elaboração própria a partir da reunião de elementos em (Attard et al., 2015; Dawes et al., 2016; Gascó-Hernández et al., 2018; Hartog et al., 2014; Marijn Janssen et al., 2012; Kim & Eom, 2019; Kučera et al., 2013; Lourenço, 2015; Luna-reyes et al., 2016; Purwanto et al., 2019; Ruijer et al., 2020; Ruijer & Meijer, 2020; Safarov, 2019; Ubaldi, 2013; H. J. Wang & Lo, 2016, 2016; V. Wang & Shepherd, 2020; T. M. Yang & Wu, 2016, 2021; Z. Yang et al., 2015; Zuiderwijk et al., 2019; Zuiderwijk & Janssen, 2014b).

2.4 Desafios Encontrados na Disponibilização de Dados Abertos

Para Ubaldi (2013), os dados abertos governamentais permanecem sendo um território desconhecido. Attard (2015) salienta que por ser este um conceito relativamente novo, é possível não haver qualquer alocação de recursos em iniciativas de dados abertos. A falta de recursos pode dificultar a execução dos programas.

Alguns autores identificam desafios que devem ser vencidos em programas de dados abertos. Os desafios encontrados são de natureza institucional (Attard et al., 2015; Barry & Bannister, 2014; Dawes et al., 2016; Ruijer & Meijer, 2020), de cultura organizacional de transparência de dados (Barry & Bannister, 2014; Dawes et al., 2016; Marijn Janssen et al., 2012; Meijer, 2015; Ruijer & Meijer, 2020; Ubaldi, 2013), de qualidade na produção dos dados (Attard et al., 2015; Dawes et al., 2016; Lourenço, 2015; Máchová & Lněnička, 2017; Ubaldi, 2013) e de disseminação da informação (Attard et al., 2015; Gascó-Hernández et al., 2018; Ho & McCall, 2016; Ruijer & Meijer, 2020).

Do ponto de vista institucional, observa-se que em programas de Dados Abertos as tarefas usualmente são alocadas a funcionários cujas atribuições já estão definidas (Attard et al., 2015), o que dificulta a alocação de tempo para desempenhar atividades ligadas a dados abertos. Daí, tem-se a necessidade de se aplicar padrões em larga escala de treinamento para superar estas questões e prever dotação orçamentária, de forma a se obter capacidade adicional de trabalho. Ao considerar que os processos de publicação de dados abertos é uma tarefa extra que requer esforço, um orçamento específico é necessário, de forma a garantir prioridade, suporte técnico e institucionalização. Para Dawes (2016), a ausência de recursos para lançar e manter um programa de dados abertos e a complexidade em desenvolver processos com as atividades necessárias para identificar, compreender e publicar os dados representam um desafio adicional. Ainda, a falta de regulamentação adequada e de políticas uniformes representam barreiras estruturais.

Para Barry & Bannister (2014), a falta de vontade política e liderança com um plano claro de abertura de dados representam barreiras estruturais. Completam este quadro a legislação de privacidade como barreira para publicação de dados (Ruijer & Meijer, 2020). Ainda para Barry & Bannister (2014, p. 156), barreiras ligadas à privacidade, copyright e sensibilidade comercial podem ser superadas caso os benefícios de dados abertos sejam deixados claros para os funcionários públicos responsáveis pelos dados.

Ao analisar as barreiras culturais para a transparência de dados, observa-se que programas de dados abertos governamentais em uma cultura fechada podem levar ao medo da interpretação, abuso ou segurança dos dados (Ruijer & Meijer, 2020). Barry & Bannister (2014) acrescentam que pode haver medo de responsabilização ou litigância quando dados pessoais são publicados por erro, ou que dados liberados podem ser mal interpretados por terceiros ao tomar decisões de negócios. Barreiras culturais podem levar a critérios para a escolha de dados a publicar, orientando-os para casos em que a segurança e a privacidade não estejam em questão, e não haja risco de erros ao interpretar os dados (Ruijer & Meijer, 2020). De acordo com Janssen (2012), uma das barreiras vem de culturas adversas ao risco. Em culturas deste tipo, o medo da responsabilização domina o capacidade de se ter iniciativa.

Uma cultura burocrática – baseada na formalidade, uniformidade e hierarquia, preserva os modos tradicionais de interação com os cidadãos (Meijer, 2015). Adicionalmente, os funcionários podem temer que as novas tecnologias possam comprometer a robustez e a confiabilidade do governo. Harmonizar a segurança com a abertura é importante porém por vezes representam eixos conflitantes de análise de decisão (Dawes et al., 2016). O compartilhamento de dados pode ser melhor endereçado e motivado através do engajamento dos funcionários públicos que trabalham com dados tanto na linha de frente do

serviço como nas administrações centrais (Ubaldi, 2013). Barry & Bannister (2014, p. 144) sustentam, a partir do resultado de entrevistas, que o nível de abertura de uma organização é influenciado por sua cultura, e que a cultura de uma organização pública é ditada em uma abordagem "*top down*", a partir dos gestores seniores.

No que concerne à qualidade da produção dos dados, Attard (2015) ressalta que alguns órgãos se preocupam em publicar dados sem ter um processo de garantia de boa qualidade, além de poderem publicar dados sem qualquer valor. A qualidade dos dados representa papel importante em dados abertos governamentais. Pode-se identificar dificuldades relativas a problemas com a origem dos dados e sua propriedade, gestão e qualidade de dados e metadados e interoperabilidade semântica (Dawes et al., 2016). O risco de ter metadados de baixa qualidade afeta a descoberta e consumo de dados, e a descrição incorreta dos conjuntos de dados impõe vários desafios para seu processamento e integração com outros dados (Máchová & Lněnička, 2017). Um fator de impedimento para melhorar a exploração completa de dados em iniciativas governamentais é a natureza heterogênea de formatos de dados utilizadas pela administração pública, o que representa uma barreira técnica tanto no lado da oferta como no da demanda (Attard et al., 2015).

A falta ou deficiência de processos de publicação representam barreiras que prejudicam a garantia da atualidade de dados. Custos em recursos humanos são associados à organização e preparo de dados a serem publicados tempestivamente (Ubaldi, 2013). Para Lourenço (2015) a atualidade dos dados deve ser considerada uma característica chave para permitir seu uso e, em termos gerais, está fortemente associada à qualidade dos dados.

Ao analisar as barreiras relativas à disseminação de informação, vê-se que a falta de processos estabelecidos de diálogo com os utentes e a baixa compreensão entre os diferentes níveis de governo tornam difícil a disponibilização de dados. Adicionalmente, os dados podem estar indisponíveis e precisar que sejam produzidos ou encontrados, o que constitui dificuldade adicional (Ruijer & Meijer, 2020). Para Gascó-Hernandez (2018), a falta de diálogo entre os provedores de conteúdo e os utentes e a falta de funcionalidades interativas e centradas nos utentes são identificadas como um fator de limitação ao acesso e uso da informação. De acordo com Ho & McCall (2016), a maioria dos cidadãos comuns não possui as habilidades técnicas necessárias para explorar dados e podem não saber como usá-los de maneira a tornarem-se mais informados e engajados, configurando uma barreira adicional à disseminação da informação.

O movimento de dados abertos foi desde o começo impulsionado pelos governos através da oferta de conjuntos de dados (Gascó-Hernández et al., 2018). Entretanto, a investigação até o momento sugere

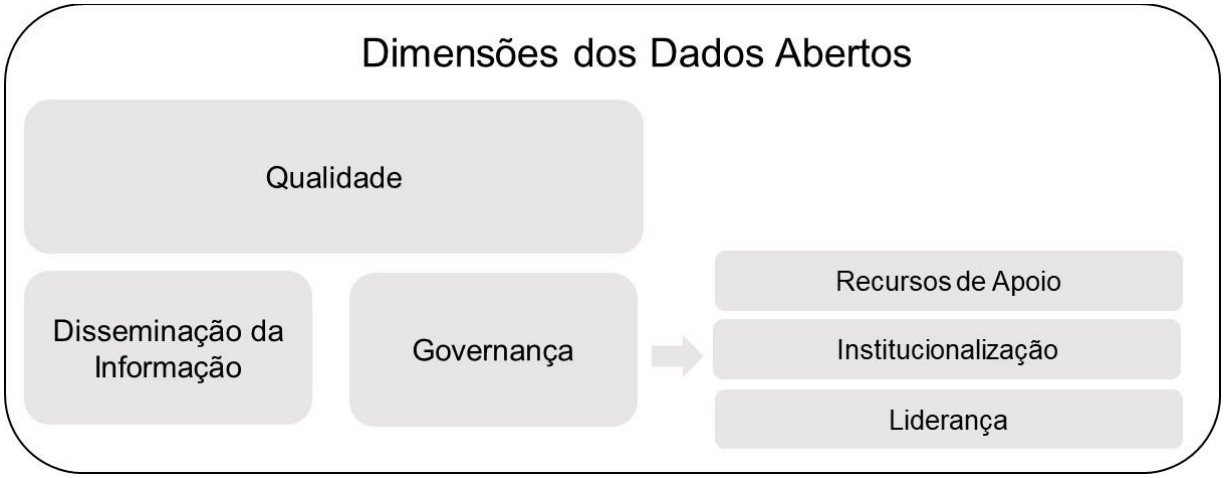
baixo uso dos dados, envolvimento limitado da sociedade em geral e quase não existência de participação dos cidadãos. Mesmo no casos de concursos ou *hackathons*, nos quais o governo estimula a participação dos cidadãos, não se tem um controle direto sobre o resultado da competição, e não pode garantir a qualidade do que for produzido (Attard et al., 2015). A falta de controle e conhecimento sobre quais dados são consumidos ou reutilizados configura também uma barreira para o sucesso de programas de dados abertos. A premissa usual assumida em políticas de dados abertos é que existe uma correspondência entre o fornecimento e a demanda. Em contrapartida, a adequação da oferta de dados com seu consumo é um processo complexo de aprendizado, interação e estabelecimento de redes entre organizações de governo e a comunidade em um sentido mais amplo (Ruijer & Meijer, 2020).

Conhecidos os benefícios esperados e barreiras, as iniciativas em dados abertos terão boa performance apenas se forem planeadas com a apreciação de toda a sua complexidade. As iniciativas devem endereçar não somente as necessidades e capacidades dos governos e utentes, assim como as características dos dados, a natureza da comunidade de utentes interessada e suas relações. Essas considerações demandam uma abordagem mais sistêmica para o desenho e projeto dos programas de dados abertos (Dawes et al., 2016).

2.5 Agrupamento dos Princípios de Dados Abertos em Dimensões

Os princípios e recomendações dos autores de dados abertos possibilitam identificar dimensões nas quais podem ser agrupados. Ao agrupá-los por tema e semelhança, as dimensões identificadas são a qualidade (Attard et al., 2015; Dawes et al., 2016; Lourenço, 2015; Malamud et al., 2007; Matheus & Janssen, 2020; Ubaldi, 2013), a disseminação da informação (Attard et al., 2015; Dawes et al., 2016; Gascó-Hernández et al., 2018; Marijn Janssen et al., 2012; Safarov, 2019; Ubaldi, 2013; V. Wang & Shepherd, 2020; T. M. Yang & Wu, 2021) e a governança. A dimensão governança é abrangente e compreende as sub-dimensões de institucionalização (Kim & Eom, 2019; Kučera et al., 2013; Luna-reyes et al., 2016; Máchová & Lněnička, 2017; Safarov, 2019; Ubaldi, 2013; T. M. Yang & Wu, 2016; Z. Yang et al., 2015; Zuiderwijk et al., 2014; Zuiderwijk & Janssen, 2014a), recursos de apoio (Alzamil & Vasarhelyi, 2019; Dawes et al., 2016; Marijn Janssen et al., 2012; Kim & Eom, 2019; Ruijer et al., 2020; Ubaldi, 2013; Zuiderwijk & Janssen, 2014a) e liderança (Kim & Eom, 2019; Ruijer & Meijer, 2020; H. J. Wang & Lo, 2016; T. M. Yang & Wu, 2016).

Figura 4 - Dimensões dos Dados Abertos



Fonte: Elaboração própria a partir da reunião de elementos da literatura

Entre as definições mais conhecidas de qualidade estão o "*fitness for use*" (adequado ao uso), conforme conceituado por Juran (1974), ou ainda a "conformidade com as especificações", de acordo com Crosby (1987). Em dados abertos, a partir dos princípios encontrados na literatura e listados nos capítulos anteriores, o conceito de qualidade pode ser entendido como os dados estarem de acordo com padrões comuns e não proprietários de dados. Essa abordagem de qualidade para os dados abertos enquadra-se tanto nos conceitos de qualidade aceitos de adequação ao uso e de conformidade às especificações.

Para Ubaldi (2013) e Dawes (2016), deve-se usar padrões técnicos comuns e aceitos para formatos de dados e coletá-los diretamente da fonte. Os governos devem publicar os metadados relevantes de seus conjuntos de dados além de descrições dos formatos, proveniência e significado dos dados. Se os dados publicados têm boa qualidade de metadados, os cidadãos irão mais facilmente descobrir os conteúdos publicados e poder decidir se eles se adequam ao uso pretendido (Attard et al., 2015). Para Matheus & Janssen (2020, p. 515), quanto melhores são descritos os conjuntos de dados e quanto mais atuais forem os dados maior será a promoção da transparência e alcance dos dados abertos. A atualidade dos dados deve ser considerada uma característica chave para permitir seu uso e, em termos gerais, está fortemente associada à qualidade dos dados. (Lourenço, 2015). O acesso a dados com granularidade fina promove também mais transparência, uma vez que apresenta o dado puro, sem agregações ou interpretações (Matheus & Janssen, 2020).

A disseminação de informação em dados abertos diz respeito às ações promovidas pelos governos para divulgar seus dados, com o reconhecimento de que é importante estabelecer diálogo entre o governo e os cidadãos (Ubaldi, 2013). Simpósios de dados abertos, workshops, concursos de exploração de

dados (conhecidos como "*hackatons*") e fóruns são relatados por T.M. Yang & Wu (2021) e Dawes (2016) como oportunidade para interagir e trocar experiências. A natureza destas comunicações pode afetar fortemente como os utentes respondem às oportunidades de usar os dados. Para Attard, Gascó-Hernandez e Safarov (2015; 2018; 2019), a aplicação de treinamentos em larga escala, tanto para os funcionários como para os cidadãos, é um fator impulsionador do uso de dados abertos. De acordo com Janssen (2012) guias para orientar o uso podem ajudar utentes inexperientes. Com relação à forma de divulgação de dados, V.Wang & Shepperd (2020) sugerem uma abordagem alternativa para acesso a dados abertos: claramente diferenciar dado de informação para que o cidadão seja capaz de mais facilmente encontrar ou solicitar conjuntos de dados não agregados e de grande utilidade.

Os autores identificam uma série de aspetos ligados à governança em dados abertos: institucionalização, recursos de apoio e liderança. A institucionalização de dados abertos é descrita em seu nível macro como a arrumação organizacional para comportar as iniciativas e no nível micro, operacional, como os processos de trabalho adotados para suportar os programas. O uso de recursos de apoio, especialmente a tecnologia, são também mencionados como ferramentas para auxiliar nos projetos de dados abertos. A nível dos incentivos nos programas, a importância da liderança é destacada como fator crítico de sucesso.

A criação de estruturas organizacionais é um dos mecanismos de institucionalização de dados abertos para Ubaldi (2013), que recomenda a criação de um responsável por dados abertos para representar o órgão e incentivar o uso. A institucionalização da Tecnologia da Informação e o reforço de suas infraestruturas têm significativa influência no sucesso de programas de dados abertos (Kim & Eom, 2019; Z. Yang et al., 2015), além da inclusão de especialistas em dados e da sociedade civil nos programas (Luna-reyes et al., 2016).

No nível operacional, a identificação dos conjuntos de dados a disponibilizar e a priorização de sua publicação, observando as leis de proteção de dados nos processos de decisão são ressaltados por Ubaldi (2013), T.M.Yang & Wu (2016) e Zuiderwijk & Janssen (2014a) como necessários em programas de dados abertos. Para Safarov (2019) e Ubaldi (2013), o *feedback* dos utentes também é um fator importante para a priorização dos dados. Os processos internos relativos a dados abertos dos governos podem ser melhorados através do provimento de padrões de qualidade e do redesenho do processo de criação e modificação de dados (Kučera et al., 2013; Máchová & Lněnička, 2017). Processos de negócio adequados e divisão de responsabilidades completam as ações de maneira a desbloquear a publicação dos dados (Zuiderwijk et al., 2014).

A implementação de programas de dados abertos, para ser efetiva, necessita de um conjunto de recursos de apoio, nomeadamente a tecnologia, o orçamento e os recursos humanos.

A disponibilização de infraestruturas e sistemas modernos é citada por Dawes (2016), Ruijer (2020) e Ubaldi (2013) como potencializadores da publicação de dados. Alzamil & Vasarhelyi (2019), Janssen (2012) e Zuiderwijk & Janssen (2014a) ressaltam a importância no investimento em ferramentas de visualização e análise de dados e na melhoria dos mecanismos de busca em dados abertos.

A alocação de um orçamento dedicado a dados abertos garantem prioridade e suporte técnico e têm significativa influência no sucesso dos programas de dados abertos. (Dawes et al., 2016; Kim & Eom, 2019). A organização e o preparo de dados a serem publicados representam custos em recursos humanos (Ubaldi, 2013). Se uma agência possui capacidades como equipa competente, recursos alocados e experiência gerencial pode aumentar sua intenção de publicação de dados e será mais confiante ao lidar com os processos de mudança e dificuldades técnicas que surgem em programas de dados abertos (T. M. Yang & Wu, 2016).

A liderança é relatada pelos autores como um dos principais motores nas iniciativas de dados abertos. Obter forte suporte da gestão de topo e da média gerência (Ruijer & Meijer, 2020) é importante para impulsionar os funcionários públicos a abrirem seus dados (T. M. Yang & Wu, 2016), em particular se os líderes acreditarem nos esforços de implementação de tais políticas. A liderança dos dirigentes é utilizada para mobilizar tangíveis e intangíveis e aplica-los em programas de dados abertos (Kim & Eom, 2019), além de viabilizar o acesso a recursos adequados, reengenharia de processos e coordenação dos atores (H. J. Wang & Lo, 2016).

Capítulo 3. Dados Abertos no Brasil

3.1 Contextualização dos Mecanismos de Transparência de Dados no Brasil

Com a Constituição Federal de 1988, o direito à informação passou a ser um direito fundamental dos cidadãos brasileiros (Presidência da República, 1988). Em seu texto, estabelece que todos têm direito a receber dos órgãos públicos informações de seu interesse particular, ou de interesse coletivo, ressalvadas aquelas cujo sigilo seja imprescindível à segurança da sociedade e do Estado.

O portal da transparência, criado em 2004 pela CGU – Controladoria Geral da União, é um site de acesso livre, no qual o cidadão pode encontrar informações sobre como o dinheiro público é utilizado (Controladoria Geral da União, 2004). Os dados encontrados neste portal dizem respeito às despesas e

receitas do governo federal, nomeadamente aos programas de governo, ao seu orçamento, compras e vencimentos dos funcionários públicos.

Desde a sua criação, a ferramenta aumentou a oferta de dados ano após ano e consolidou-se como importante instrumento de controle social. O Portal da Transparência permite ao cidadão, de forma cada vez mais eficiente, fiscalizar e assegurar a boa e correta aplicação dos recursos públicos federais.

De acordo com Possamai & de Souza (2018), a LAI foi pioneira no mundo ao incorporar o dados abertos no escopo da regulamentação do direito de acesso à informação, e determina a publicidade como regra e o sigilo como exceção (Presidência da República, 2011). Preferencialmente, as informações devem ser divulgadas proactivamente nos sites oficiais na internet e conter ferramentas de pesquisa de conteúdo que facilite o acesso à informação.

A lei em questão definiu informação como sendo dados, processados ou não, que podem ser utilizados para produção e transmissão de conhecimento. A materialização do acesso à informação previsto na Constituição Federal e na Lei de Acesso à Informação passou a acontecer, então, pela publicação de dados governamentais. Os portais governamentais, como o Portal da Transparência, com a especificidade de receitas e despesas federais, e o Portal de Dados Abertos, com a generalidade das atividades e operações dos órgãos públicos federais, foram um instrumento importante de consecução destas leis e políticas de transparência.

3.2 Open Government Partnership e o Compromisso Brasileiro

No contexto internacional, no âmbito dos esforços governamentais para promoção de transparência, o Brasil é membro fundador da *Open Government Partnership* – OGP (OGP, 2011), a qual foi criada como um parceria atualmente composta por 78 governos federais e 76 governos locais. A OGP baseia-se na ideia de que um governo aberto é mais acessível, responsável e capaz de melhor prestar contas a seus cidadãos.

A OGP abrange um conjunto de áreas temáticas em que os países membros aplicam políticas, entre as quais está incluída a governança digital e, mais especificamente, dados abertos. Cada país deve construir planos de ação contendo compromissos e atividades para implementá-los, além de precisarem as estratégias para executá-los (Controladoria Geral da União, 2021a). O atual plano que o Brasil está a cumprir é o quinto o qual é composto de 12 compromissos. O plano inclui dados abertos em 4 objetivos - Meio Ambiente e Floresta, Direitos Humanos, Cadeias Agropecuárias e Dados Eleitorais. Este plano aborda compromissos específicos em segmentos de atividade do governo os quais serão objeto de disponibilização de dados abertos no portal de dados abertos.

Os compromissos assumidos pelo governo brasileiro junto à OGP, nomeadamente no âmbito de dados abertos, têm impacto direto no que é construído e disponibilizado no portal de dados abertos, daí a importância de se estudar no contexto do portal os processos de governança e disponibilização da informação.

3.3 O Portal de Dados Abertos

O acesso a dados em portais governamentais depende da disponibilidade de os cidadãos acederem à internet, o qual se confirmou durante as duas últimas décadas. Este avanço é mostrado pela PNAD, realizada pelo IBGE – Instituto Brasileiro de Geografia e Estatística - em 2019 (Ministério das Comunicações, 2021). De acordo com o levantamento, 84 % dos domicílios nacionais possuem acesso à internet (IBGE, 2021). Como comparação e evidência do crescimento, em 2005 o acesso à internet nos domicílios era de 13,7%, chegando apenas até 27,4% em 2009 (IBGE, 2009).

Tabela 1 - Domicílios com Acesso à Internet - 2016 a 2019

Tabela 7308 - Domicílios e Moradores, por existência de utilização da internet no domicílio							
Variável - Distribuição percentual dos domicílios (%)							
Brasil							
Ano x Existência de utilização da Internet no domicílio							
2016		2017		2018		2019	
Total	Havia utilização de internet	Total	Havia utilização de internet	Total	Havia utilização de internet	Total	Havia utilização de internet
100,0	70,9	100,0	76,4	100,0	80,6	100,0	84,0
Fonte: IBGE - Pesquisa Nacional por Amostra de Domicílios Contínua Anual - 4º trimestre							

Fonte: IBGE

O aumento do acesso à internet pelos cidadãos possibilitou a criação de políticas de transparência voltadas para a divulgação de dados em portais governamentais. O governo brasileiro publicou, sob a forma de decreto-lei, a Política Nacional de Dados Abertos (Presidência da República, 2016). Este diploma estabeleceu para os órgãos públicos a obrigação da publicação em transparência ativa de seus dados, em formato aberto e em consonância com os princípios encontrados na literatura e revistos nesta dissertação.

O Poder Executivo Federal dispõe de política que regula e orienta a publicação de dados abertos governamentais pelos órgãos e entidades do Governo Federal (Controladoria Geral da União, 2016a). Os principais instrumentos da execução da política são os PDAs – Planos de Dados Abertos - e o Portal de Dados Abertos – www.dados.gov.br.

Os órgãos públicos sujeitos à política de dados devem construir o PDA, que é o instrumento que organiza o planejamento das ações de implementação e promoção da abertura de dados. A CGU analisa e aprova ou não os PDAs submetidos (Controladoria Geral da União, 2020).

Enquanto o PDA é um documento interno da Administração Pública, o Portal de Dados Abertos é o resultado da execução da política de divulgação de dados para a sociedade. O portal de dados abertos é administrado pela CGU, que monitora as publicações e compromissos dos órgãos públicos através do Painel Monitoramento de Dados CGU (Controladoria Geral da União, 2021b)

Figura 5 - Painel de Monitoramento de Dados CGU

Fonte: Site CGU

O Portal Brasileiro de Dados Abertos é a ferramenta disponibilizada pelo governo brasileiro para que todos possam encontrar e utilizar os dados e as informações públicas, e tem o objetivo de promover a interlocução entre atores da sociedade e governo para pensar a melhor utilização dos dados, para assim gerar impactos positivos sob os pontos de vista social e econômico (Controladoria Geral da União, 2016a). De acordo com o decreto que estabeleceu a política de dados abertos (2016), os dados publicados devem cumprir requisitos que incluem a publicidade, a participação social e engajamento dos cidadãos, o acesso irrestrito aos dados, a atualização periódica e a completude e interoperabilidade das bases de dados.

O portal de dados abertos governamentais brasileiro conta hoje com 190 órgãos ativos, 10.987 conjuntos de dados e 53.652 ficheiros de dados. Um conjunto de dados é composto de um ou mais ficheiros que podem ser descarregados pelos utentes, sem custo ou necessidade de identificação prévia.

As publicações no portal são realizadas diretamente pelos órgãos participantes através da ferramenta de software livre CKAN - *Comprehensive Knowledge Archive Network* - , disponibilizada pela CGU, a qual administra o site.

Figura 6 - Números do Portal Dados.gov.br

Fonte: Site dados.gov.br

O portal abriga dados dos mais variados segmentos da Administração Pública. Estão presentes setores de atividade como entidades financeiras, agências reguladoras, defesa, segurança e ordem pública, ministérios, proteção ambiental e social, educação, pesquisa, saúde, entre outros. O portal não contém dados sigilosos. Os dados que estiverem sujeitos a restrições de sigilo ou de privacidade conforme definidos na Lei de Acesso à Informação precisam ser previamente classificados pelos seus responsáveis e não são publicados (Controladoria Geral da União, 2016b).

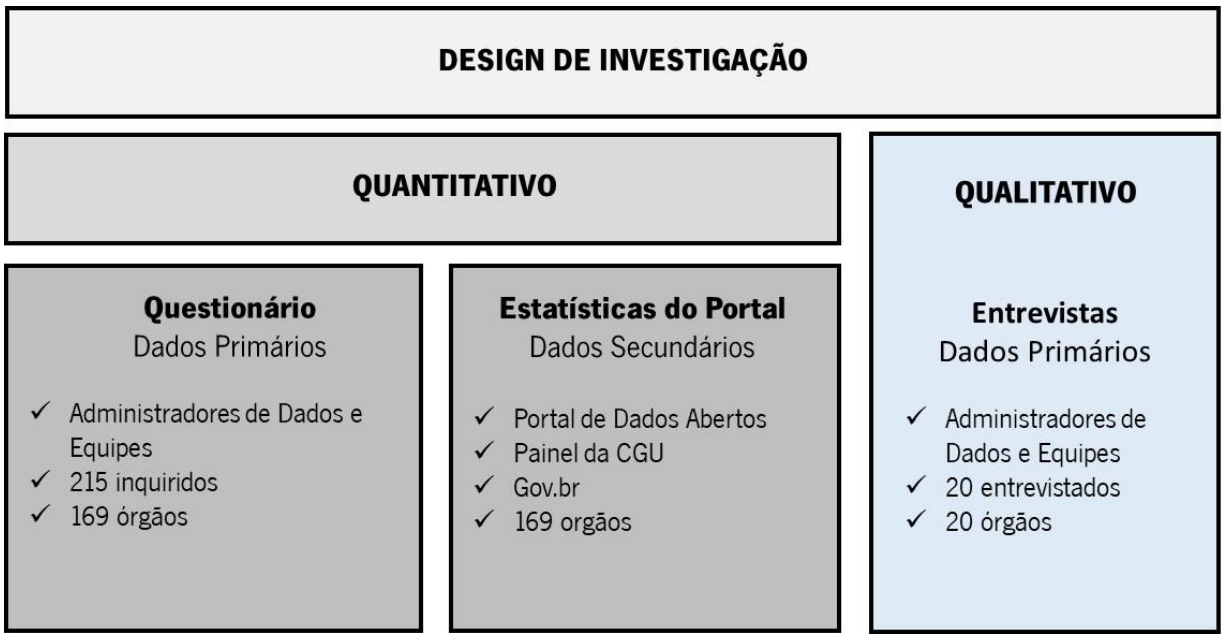
Capítulo 4. Design da Investigação

Segundo Bryman (2012), o design de investigação compreende uma estrutura para a coleta e análise de dados. Sua escolha reflete as decisões sobre as prioridades concedidas a uma série de dimensões do processo de investigação. Isto inclui a importância dada à conexão causal entre variáveis,

generalização para grupos maiores que o estudado e a compreensão do comportamento observado e seu significado em um contexto específico.

O design de investigação proposto neste estudo é não experimental, no qual se irá analisar uma realidade existente, sem a manipular. Tem característica *cross-sectional*, a ser realizado em um único momento com população definida. Constitui abordagem mista: quantitativa, com análise a se realizar a partir de dados quantitativos dos questionários e de estatísticas do portal de dados abertos; e qualitativa, através de entrevistas realizadas com administradores de dados e equipas de produção e publicação de dados abertos.

Figura 7 - Design de Investigação



Fonte: Elaborado pelo autor

Este capítulo discorre sobre a formulação da hipótese central e das hipóteses específicas. Prossegue com o método e procedimentos adotados, o tipo de design escolhido para melhor suportar a investigação e a escolha da população a ser pesquisada e os instrumentos utilizados para coletar os dados. Apresenta em seguida a operacionalização dos conceitos, construídos a partir das dimensões encontradas na revisão literária, e apresenta as variáveis derivadas das dimensões. Em concordância com o método de investigação, são estabelecidos os procedimentos para verificação da validade e fiabilidade do processo.

4.1 Hipóteses de Investigação e Modelo Conceptual

Os autores revistos indicam que as ações realizadas pelos governos e a maneira como se organizam em programas de dados abertos contribuem para aumentar o acesso à informação para os cidadãos e melhorar a qualidade dos dados.

Para implementar estes programas, os governos precisam criar as condições para que tais iniciativas se perenizem e sejam bem-sucedidas. Os autores relatam uma série de atitudes e ações dos governos para gerar valor em programas de dados abertos. Para Ubaldi (2013), deve-se disponibilizar uma infraestrutura simples e pública para expor seus dados, alocar um orçamento dedicado (Attard et al., 2015) e implementar mecanismos de feedback com os cidadãos (Dawes et al., 2016). Para Tai (2021), em sua pesquisa sobre programas de dados abertos executados em uma década, os governos investem mais recursos em dados abertos para atingir um alto nível de transparência. Já de acordo com T.M. Yang & Wu (2021), a liderança de topo nestes programas é um dos maiores vetores para impulsionar os funcionários a abrirem seus dados, estimulando o acesso e uso. A natureza de todas essas comunicações pode afetar fortemente a resposta dos cidadãos no uso dos dados (Dawes et al., 2016). Em seu estudo sobre como planear e desenhar programas de dados abertos, Dawes (2016) relata que as escolhas dos governos sobre como o dado é escolhido, preparado, publicado e comunicado tem grande efeito na quantidade, tipo e qualidade dos dados entregues aos cidadãos.

Conforme foi discutido na revisão da literatura, as ações de governança e de disseminação de informação nos programas de dados abertos influenciam a qualidade dos dados e o acesso à informação. Em consonância com a revisão feita, o objetivo geral do estudo é entender de que maneira, em programas de dados abertos, o modelo de governança adotado e as iniciativas para disseminar informação influenciam a qualidade dos dados e o acesso do público à informação.

O objetivo é formulado como ponto de partida, para em seguida compor hipóteses complementares. Estas dizem respeito aos processos específicos adotados pelos órgãos públicos e sua possível influência nos resultados dos programas de dados abertos, nomeadamente quanto à qualidade e ao acesso à informação. Através da verificação dos dados obtidos no estudo, e com o intuito de responder à pergunta de investigação, pretende-se testar as hipóteses de forma a obter conclusão.

Os governos valem-se de diversas estratégias para divulgar seus programas de dados abertos e aumentar o acesso à informação pelo público: simpósios, workshops, concursos de exploração de dados – hackathons – e treinamentos (Ubaldi, 2013) (T. M. Yang & Wu, 2021) (Attard et al., 2015). Estas ações potencializam a descoberta e uso dos dados (Dawes et al., 2016), chamam a atenção do público

e introduzem os cidadãos na participação social (Attard et al., 2015) tornando-os mais bem informados (Tai, 2021).

Também, o efeito do treinamento dos utentes como impulsionador do uso é relatado por Gascó-Hernandez (2018) no estudo realizado na Itália, Espanha e Estados Unidos. Iniciativas de dados abertos governamentais têm maior chance de sucesso se os utentes possuírem habilidades e conhecimentos técnicos (Safarov, 2019).

Chegamos assim à formulação da seguinte hipótese:

H1: A DISSEMINAÇÃO DA INFORMAÇÃO influencia o ACESSO À INFORMAÇÃO.

O uso de recursos de apoio como suporte à publicação de dados abertos é visto pelos autores como determinante da qualidade dos dados que são divulgados. Alexopoulos (2015) relata, em seu estudo sobre o portal de dados abertos na Grécia, que ao empregar mais tecnologia nos processos de publicação, é possível disponibilizar mais ferramentas para se descobrir, visualizar dados e permitir feedback dos utentes, com o benefício do aumento na qualidade dos dados.

De acordo com Ubaldi (2013), em sua revisão de iniciativas de programas governamentais em dados abertos, a publicação de dados atuais e de alta qualidade depende da alocação de recursos humanos para organizar e preparar os dados. Attard (2015) defende que uma vez que os processos de publicação de dados abertos é uma tarefa extra que requer esforço, um orçamento específico é necessário, de maneira a garantir a publicação de dados através de um processo que garanta sua qualidade.

Assim formulamos a seguinte hipótese:

H2: O USO DE RECURSOS DE APOIO influencia a QUALIDADE DOS DADOS.

Os autores revistos também identificam os recursos de apoio como fator de influência na promoção do acesso à informação. Para Marijn Janssen (2012), Styrin (2017) e Lourenço (2015), as ferramentas de visualização de dados podem ajudar utentes inexperientes no uso de dados. Safarov (2019) complementa que as ferramentas de visualização podem ajudar a identificar tendências e padrões de usos o que permite usar os dados para atingir interesses específicos dos cidadãos. Gascó-Hernandez (2018) relata que funcionalidades interativas e centradas nos utentes são fatores de promoção de acesso e uso da informação.

Kim & Eom (2019) estudaram o sucesso de programas de dados abertos como a medida na qual os dados são acessados pelo público, e identificou a importância de se ter um orçamento dedicado a

estes programas. De acordo com T.M. Yang & Wu (2016), espera-se que uma equipa competente em dados abertos e com recursos alocados atuará de forma positiva em direção a publicar dados abertos, aumentando o acesso à informação.

É possível então identificar a seguinte hipótese:

H3: O USO DE RECURSOS DE APOIO influencia o ACESSO À INFORMAÇÃO.

A institucionalização, uma das dimensões identificadas na revisão da literatura, é representada neste estudo pela estrutura organizacional em dados abertos a nível macro e pelos processos de trabalho a nível micro. Dawes (2016) relata a importância que os processos de seleção, escolha e publicação têm na qualidade da divulgação dos dados. Para Lourenço (2015), em programas de dados abertos, as estruturas de governança e processos formais garantem atualidade e qualidade dos dados. Em complemento, Ubaldi (2013) observa que boas práticas de administração de dados valorizam o esforço em disponibilizar dados legíveis por máquina, que é um dos princípios básicos ligados à qualidade de dados abertos. Hartog (2014) identifica que um melhor resultado nos programas envolve a capacidade de se automatizar as extrações de dados, sem ações manuais dos funcionários públicos.

Sobre os processos de interação com os cidadãos, Safarov (2019) observa em estudo conduzido no Reino Unido que promover o envolvimento da sociedade através do feedback é um das metas prioritárias para melhorar a qualidade dos dados. Os estudos dos autores sobre institucionalização e qualidade nos levam a formular a seguinte hipótese:

H4: A INSTITUCIONALIZAÇÃO influencia a QUALIDADE DOS DADOS.

Os processos de marcação de metadados (Ubaldi, 2013), de garantia de sua qualidade (Attard et al., 2015) e de escolha de dados mais relevantes para publicação (Zuiderwijk et al., 2014) são importantes para que se encontre mais facilmente os dados nas buscas, aumentando o acesso à informação. Dawes (2016) relata que processos de adoção de padrões técnicos proporciona dados mais úteis e possíveis de uso. No que concerne à organização de estruturas que acompanham os programas de dados abertos, percebe-se que a institucionalização da área de Tecnologia da Informação tem significativa influência no nível de acesso ao dados em programas de dados abertos (Kim & Eom, 2019). Os estudos sobre a institucionalização e acesso à informação nos levam a formular a seguinte hipótese:

H5: A INSTITUCIONALIZAÇÃO influencia o ACESSO À INFORMAÇÃO.

O papel exercido pelos dirigentes dos órgãos públicos na implementação de programas de dados abertos e sua influência no acesso do público aos dados está presente nos estudos revistos no âmbito deste estudo. Segundo H.J. Wang (2016), o apoio da gestão de topo é percebido como fator de maior adoção a estes programas. Para escalar o uso dos dados, Ruijer (2020) relata a necessidade de forte comprometimento gerencial. Kim & Eom (2019) reporta que a capacidade gerencial para mobilizar recursos tangíveis e intangíveis e aplicá-los nos lugares e processos corretos sob o comando de líderes capazes é um fator importante para atingir sucesso em projetos de dados abertos. Estes fatores ligados à liderança e ao acesso a dados abertos nos programas nos permitem formular a última hipótese:

H6: A LIDERANÇA influencia o ACESSO À INFORMAÇÃO.

A tabela a seguir reúne os autores citados que, com seus estudos, permitiram que nesta dissertação fossem identificadas as possíveis relações e fatores de influência que deram origem às hipóteses complementares.

Tabela 2 – Hipóteses e Autores para Sustentação

Hipótese	Formulação	# Autores	Citações
H1	A DISSEMINAÇÃO DA INFORMAÇÃO influencia o ACESSO À INFORMAÇÃO	10	(Ubaldi, 2013) (Attard et al., 2015) (Dawes et al., 2016) (Tai, 2021) (Gascó-Hernández et al., 2018) (Ho & McCall, 2016) (Safarov, 2019) (Purwanto et al., 2019) (V. Wang & Shepherd, 2020) (T. M. Yang & Wu, 2021)
H2	O USO DE RECURSOS DE APOIO influencia a QUALIDADE DOS DADOS	3	(Alexopoulos et al., 2015) (Attard et al., 2015) (Ubaldi, 2013)
H3	O USO DE RECURSOS DE APOIO influencia o ACESSO À INFORMAÇÃO	9	(Janssen et al., 2012) (Styrin et al., 2017) (Lourenço, 2015) (Alzamil & Vasarhelyi, 2019) (Gascó-Hernández et al., 2018) (Safarov, 2019) (Kim & Eom, 2019) (T. M. Yang & Wu, 2016) (Zuiderwijk & Janssen, 2014a)
H4	A INSTITUCIONALIZAÇÃO influencia a QUALIDADE DOS DADOS	10	(Lourenço, 2015) (Safarov, 2019) (Ubaldi, 2013) (Dawes et al., 2016) (V. Wang & Shepherd, 2020) (Zuiderwijk & Janssen, 2014b) (Hartog et al., 2014) (Zuiderwijk & Janssen, 2014a) (Possamai & De Souza, 2020) (Vetrò et al., 2016)

H5	A INSTITUCIONALIZAÇÃO influencia o ACESSO À INFORMAÇÃO	11	(Ubaldi, 2013) (Kim & Eom, 2019) (Ruijer et al., 2020) (Safarov, 2019) (Máchová & Lněnička, 2017) (Zuiderwijk et al., 2014) (Ruijer & Meijer, 2020) (Zuiderwijk & Janssen, 2014a) (Attard et al., 2015) (Dawes et al., 2016) (Kaasenbrood et al., 2015)
H6	A LIDERANÇA influencia o ACESSO À INFORMAÇÃO	6	(Luna-reyes et al., 2016) (T. M. Yang & Wu, 2021) (Styrin et al., 2017) (Kim & Eom, 2019) (H. J. Wang & Lo, 2016) (Ruijer & Meijer, 2020)

Fonte: Elaboração própria a partir da reunião de elementos em (Alexopoulos et al., 2015; Alzamil & Vasarhelyi, 2019; Attard et al., 2015; Dawes et al., 2016; Gascó-Hernández et al., 2018; Ho & McCall, 2016; Marijn Janssen et al., 2012; Kaasenbrood et al., 2015; Kim & Eom, 2019; Lourenço, 2015; Luna-reyes et al., 2016; Possamai & De Souza, 2020; Purwanto et al., 2019; Ruijer et al., 2020; Safarov, 2019; Styrin et al., 2017; Tai, 2021; Ubaldi, 2013; Vetrò et al., 2016; V. Wang & Shepherd, 2020; T. M. Yang & Wu, 2016, 2021; Zuiderwijk & Janssen, 2014a)

4.2 Técnicas de Análise Utilizadas

A base de dados do questionário foi utilizada para testar as hipóteses relativas à variável dependente Qualidade dos Dados. Os dados foram carregados no *software RStudio* e foram realizados testes de normalidade de Shapiro Wilk nas variáveis que indicaram a realização de testes não paramétricos, como análises de correlação de Spearman. Foi realizada regressão linear múltipla e previamente foram verificados os pressupostos do modelo ao analisar a normalidade de resíduos, a homocedasticidade e a inexistência de multicolinearidade e homocedasticidade.

A base de dados secundários foi obtida do portal de dados, do site do governo, gov.br e do painel de dados abertos da CGU e foi utilizada para testar as hipóteses relativas à variável dependente Acesso à Informação. Os dados foram carregados no *software RStudio* e foram realizados testes de normalidade de Shapiro Wilk que indicaram a realização de testes não paramétricos, como análises de correlação de Spearman. Foi realizada regressão linear múltipla cujo teste de normalidade dos resíduos também acusou não normalidade. Como as variáveis são do tipo contagem, foi realizada regressão log na variável dependente de forma a normalizar o modelo. Mais uma vez, foi testada a normalidade dos resíduos do modelo com regressão log, o que permitiu validar o modelo. Para tornar mais robusta a análise, foi também realizada a regressão de Poisson, indicada para análise de modelos com variáveis de contagem, e que apresentou resultados semelhantes.

As entrevistas foram realizadas tendo seu conteúdo reunido elementos a partir da revisão geral da literatura, especialmente do framework de avaliação em dados abertos proposto por Ubaldi (2013, pp.

45–52). Também foram utilizadas como base as perguntas sobre qualidade, os critérios de avaliação e os questionamentos sobre objetivos, práticas, desafios e soluções propostos por Dawes (2016, pp. 20, 26). Hartog (2014, p. 493) conduziu entrevistas para apuração de maturidade em programas de dados abertos em agências governamentais. Nesse evento, foram colocadas questões sobre os planos, situação atual das iniciativas, desafios, riscos, benefícios, métricas e melhores práticas, conteúdo que também inspirou a criação de guião para as entrevistas. O conteúdo da entrevista foi validado quanto à sua forma a partir da matriz avaliativa conforme proposto por M. Silvestre (2010, pp. 323–324). Segundo Duarte (2004, p. 222), é necessário ordenar o material coletado nas entrevistas. Deve-se organizá-lo em torno de eixos que permitam o cruzamento com as referências teóricas utilizadas, no caso o que foi revisto na literatura. O investigador deve ser capaz de dar sentido a toda esta junção de informações categorizadas. Nesse sentido, as entrevistas realizadas tiveram o agrupamento das citações em categorias para análise e conclusões. Adicionalmente, foi coletada a impressão neutra, positiva ou negativa, segundo o tema e pergunta. Os comentários foram em seguida classificados segundo esse critério, segundo a primeira abordagem do sistema de codificação proposto por Weston (2001, p. 387).

4.3 Unidades de Análise

De maneira a concretizar a abordagem quantitativa escolhida para esta dissertação, os instrumento para recolha de dados escolhido foi o de questionário e entrevistas – dados primários – complementado com dados secundários do portal de dados abertos.

A unidade de análise do questionário é o indivíduo, representado pelos administradores de dados e equipes de produção e publicação de dados abertos. Esta unidade de análise foi operacionalizada no questionário para as dimensões de variáveis independentes – disseminação da informação, institucionalização, recursos de apoio e liderança – e para a variável dependente qualidade dos dados.

A unidade de análise dos dados secundários é o órgão publicador. Esta unidade de análise foi operacionalizada para as dimensões de variáveis independentes – disseminação da informação e institucionalização– e para a variável dependente acesso à informação. Foi necessário utilizar dados secundários pois a variável acesso à informação somente está disponível através das estatísticas de visualizações de página do portal por órgão.

A unidade de análise das entrevistas é o indivíduo, representado pelos administradores de dados e equipas de produção e publicação de dados abertos para os órgãos no portal.

4.4 Amostra e Procedimentos de Amostragem

A investigação é composta por um universo dos 190 órgãos ativos que publicam no portal. O governo brasileiro dispõe de uma ferramenta unificada, o Fala.BR - Plataforma Integrada de Ouvidoria e Acesso à Informação (cgu.gov.br) – a qual permite entrar em contato com cada órgão e realizar uma solicitação. O prazo para resposta é de, no máximo, 30 dias úteis. Em meados de Dezembro de 2021, foi solicitada a cada um dos 190 órgãos, através da ferramenta, autorização de resposta ao questionário e o fornecimento dos nomes, e-mails e telefones dos responsáveis de dados abertos para que se enviasse o questionário, caso o órgão em questão concordasse em participar do estudo. Foram aguardadas respostas até o final de Janeiro de 2022.

Dos 190 órgãos para os quais houve a solicitação, 164 autorizaram (86%) e forneceram ou os e-mails dos responsáveis e equipes ou um e-mail genérico do departamento responsável por dados abertos. 5 órgãos no portal de dados são administrados por outro órgão que autorizou e também receberam o questionário (3%). 5 órgãos não autorizaram ou optaram por não participar do estudo, e 16 órgãos não responderam ao pedido de participação a tempo. Procurou-se amostrar a totalidade dos órgãos do portal de dados, atingindo-se 89% deste universo com envio de questionários. Os 164 órgãos que autorizaram o envio do questionário forneceram 393 e-mails de contato, para os quais foi enviado na totalidade dos contatos. Houve 215 respostas completas, o que configura uma taxa de resposta de 55%.

O questionário foi enviado de forma completamente anônima através de link online para resposta, usando o software Qualtrics, licenciado para a Universidade do Minho. O questionário esteve disponível para preenchimento de 15 de Dezembro de 2021 até 3 de Fevereiro de 2022. Durante este período, foram enviados 2 e-mails de reforço para solicitar respostas, com intervalo de 2 semanas entre cada envio. Foi observado que os e-mails de reforço tiveram efeito direto no aumento da taxa de respostas. Os dias que registaram mais respostas foram aqueles imediatamente após o envio do e-mail de solicitação.

Tabela 3 - Amostragem do Questionário

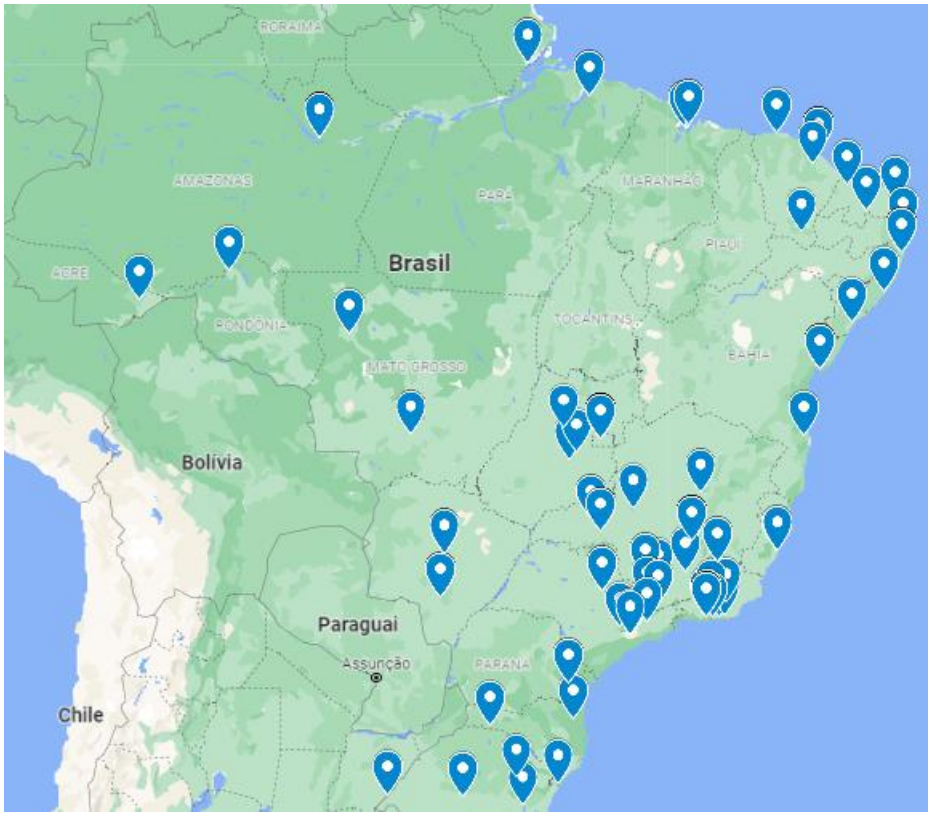
Universo da Pesquisa (órgãos públicos, que receberam pedido de autorização)	190	
Órgãos que autorizaram e receberam o questionário	164	86%
Órgãos duplicados no portal (administrado por outro órgão que recebeu o questionário)	5	3%
Não autorizaram o envio do questionário	5	3%
Não responderam a tempo	16	8%

Interlocutores nomeados individualmente pelos órgãos que autorizaram	297	
Interlocutores genéricos - e-mails de departamento - fornecidos pelos órgãos que autorizaram	96	
Total de interlocutores que receberam o questionário	393	
Total de respostas completas recebidas	215	55%

Fonte: extraído dos dados de envio e resposta, software Qualtrics

Os questionários foram enviados para órgãos públicos presentes no portal de dados abertos que se encontram situados em todas as regiões do território brasileiro. A ferramenta *Qualtrics*, na qual se fez o questionário anônimo, captura a localização dos respondentes. A figura a seguir apresenta o local de onde foram preenchidos os questionários, o que dá uma ideia da abrangência geográfica do estudo. Em virtude do tamanho do mapa, não é possível ver todos os pontos onde há uma quantidade maior de respondentes, como por exemplo na região Sudeste do Brasil.

Figura 8 - Localização dos Respondentes



Fonte: dados de localização das respostas plotados no Google Maps

Antes do envio do questionário, foi realizado um pré-teste com 5 servidores tendo sido realizadas alterações de acordo com os comentários recebidos. As alterações solicitadas e efetivadas contribuíram para dar clareza do questionário.

Para os dados secundários, foram coletadas estatísticas para os mesmos 169 órgãos que receberam o questionário, de maneira a manter a coerência com o universo do estudo.

As entrevistas foram realizadas entre 16 de fevereiro e 11 de março de 20220 através de uma amostragem por conveniência, composta por 20 pessoas que responderam ao questionário e trabalham diretamente com dados abertos em seus órgãos de lotação.

4.5 Recolha de Dados

Os dados foram recolhidos a partir de questionários, entrevistas e dados secundários. O questionário foi destinado aos administradores de dados abertos dos órgãos e suas equipes como ferramenta de recolha de dados. A extração de dados secundários foi realizada diretamente do portal de dados, do painel de dados abertos e do site gov.br. As entrevistas foram realizadas com funcionários dos órgãos, estes responsáveis de dados abertos ou pertencentes às equipas de produção e publicação de dados.

O questionário é composto por uma primeira parte com 3 questões para identificação no geral do órgão público respondente, em específico o nível de governo, a classificação quanto à constituição e a função de governo. Estas questões iniciais possibilitam a identificação da natureza do órgão, o que é importante para a apresentação das análises de dados. Ao mesmo tempo, não permitem individualizar o entrevistado e o órgão público no qual é lotado. Esta abordagem garante a liberdade do respondente para manifestar-se com franqueza, uma vez que não poderá ser identificado pelas respostas. Algumas das questões tratam do apoio de dirigentes aos programas de dados abertos, e podem ser consideradas sensíveis, sendo importante manter o anonimato.

A segunda parte do questionário é representada por três perguntas de caráter demográfico – gênero, tempo de colaboração no órgão de lotação e faixa etária dos respondentes. Estas perguntas foram administradas como forma de se obter dados censitários a serem considerados na análise dos dados das dimensões.

A terceira parte é composta de 18 questões em que cada seção representa uma dimensão e suas variáveis. A estrutura do questionário foi desenhada de forma a representar as dimensões identificadas na literatura, e cada pergunta foi formulada segundo a revisão bibliográfica de maneira a formar o construto. Para as dimensões de disseminação de informação, institucionalização a nível macro e

qualidade dos dados, foram usadas perguntas dicotômicas de maneira a formar índices de conformidade. Já para as dimensões institucionalização a nível micro, recursos de apoio e liderança, foram utilizadas perguntas com respostas baseadas em uma escala de Likert de 5 pontos, onde 1 representa "Discordo Totalmente" e 5 representa "Concordo Totalmente", tendo o 3 como valor neutro "Não Concordo nem Discordo". Para a dimensão acesso à informação foram extraídos os registos de número de visualizações e tempo de permanência nas páginas dos órgãos constantes do portal de dados abertos.

A extração de dados secundários foi realizada em três repositórios diversos: a área de estatísticas do portal dados.gov.br, o painel de dados abertos da CGU e pesquisa google no domínio gov.br. A área de dados estatísticos do portal respondeu pelas variáveis de total de visualizações e de total de conjuntos de dados publicados. As estatísticas foram recolhidas e somadas para o período de julho de 2021 a janeiro de 2022, de maneira a obter o maior período possível para análise. O painel de dados abertos respondeu pela variável taxa de respeito ao plano de dados, e foi calculada a partir dos dados dos planos dos órgãos publicadores no site da CGU, ao dividir o total de bases publicadas sobre o total de bases previstas até janeiro de 2022. A pesquisa google no domínio gov.br foi efetuada para a contagem do total de menções ao órgão e ao portal dados.gov.br, tendo sido executada em fevereiro de 2022.

As entrevistas foram realizadas para complementar, de forma qualitativa, o conhecimento adquirido previamente pelos questionários e dados secundários. O objetivo das entrevistas foi compreender de maneira mais aprofundada a opinião das pessoas sobre o tema, suas dificuldades e seu papel na promoção da transparência governamental.

As entrevistas foram realizadas de forma semiestruturada com o uso da ferramenta online Zoom, uma vez que os entrevistados exercem seus cargos públicos e trabalham no Brasil. O critério para não se identificar no estudo individualmente nem o inquirido nem o órgão público objeto da entrevista foi mantido da mesma maneira com o que foi feito para o questionário. As entrevistas não foram gravadas porém foram tomadas notas detalhadas das respostas ao longo das entrevistas. A duração das entrevistas oscilou entre 40 minutos e 1 hora e 15 minutos.

4.6 Operacionalização dos Conceitos

Este estudo apresenta a pergunta de investigação "Em que medida o portal de dados abertos governamentais brasileiro atende aos requisitos da política de transparência e aos princípios de divulgação de dados encontrados na literatura?", a qual se pretende responder com as hipóteses formuladas.

A política de transparência e a observância dos princípios de dados abertos são relatados na literatura como determinantes para a consecução dos programas de dados abertos. Cruz (2012) e Tejedo-Romero & Araújo (2018) relacionam a transparência à governança, e para Araújo (2004), as melhorias na administração em função da pressão pelas reformas resultam em ganhos de produtividade e transparência. Já Ubaldi (2013) associa a transparência ao aumento do acesso à informação e Cuillier & Piotrowski (2009) e Jaeger & Bertot (2010) e à implantação de portais de dados como instrumento.

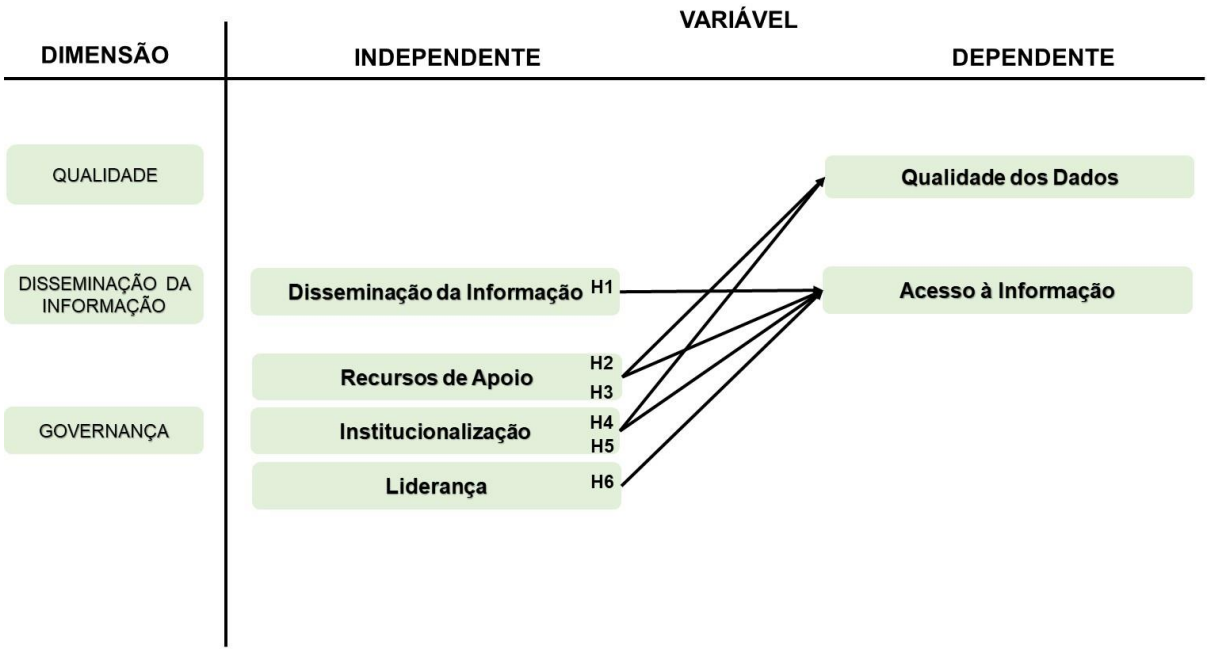
Ubaldi (2013) traz em seu estudo uma coletânea de princípios de dados abertos e advoga sua aplicação como forma de promoção de transparência. Os princípios enumerados em Berners-Lee (2006), Malamud (2007) e pela Open Knowledge Foundation, mais os estudos dos autores que se seguiram relacionaram a implementação de portais de dados abertos às práticas de disseminação de informação e de governança, conforme já relatados na revisão da literatura.

As medidas adotadas pelos governos para aumentar a governança em dados abertos, mais as ações de disseminação da informação sobre estes programas contribuem para promover qualidade dos dados (Attard et al., 2015; Dawes et al., 2016; Matheus & Janssen, 2020; Ubaldi, 2013) e o acesso à informação, materializado pelas visualizações e interações nos portais de dados (Attard et al., 2015; Tai, 2021; Ubaldi, 2013; T. M. Yang & Wu, 2021).

A pergunta de investigação pretende compreender se o portal de dados abertos brasileiro atende às políticas de transparência e aos princípios de dados abertos. Vemos, pela revisão da literatura, que a transparência aplicada a divulgação de dados e a boa aplicação de princípios de publicação são atingidos com a governança adotada e com as ações para divulgar os dados abertos. Estas ações resultam, também pelo que foi revisto no estudo, em mais acesso e qualidade dos dados.

No que diz respeito aos conceitos das variáveis, o estudo identificou dimensões relacionadas à disseminação de informação e à governança, e dimensões relacionadas à qualidade dos dados. As variáveis independentes são a disseminação da informação, os recursos de apoio, a institucionalização e a liderança e as variáveis dependentes são a qualidade dos dados e o acesso à informação.

Figura 9 - Variáveis a partir das Dimensões



Fonte: Elaboração própria a partir da reunião de elementos da literatura

Para operacionalizar a captura de dados relativos às dimensões identificadas foram identificadas as variáveis que estão associados a cada uma das dimensões. As variáveis das dimensões disseminação da informação, institucionalização, recursos de apoio, liderança e qualidade de dados foram recolhidas da literatura e basearam a criação do questionário.

Tabela 4 - Variáveis a partir das Dimensões - Questionário

Dimensão Institucionalização	
Dimensão	Variáveis
Institucionalização nível macro	Existe uma unidade responsável por dados abertos
	Um Chefe de Dados foi nomeado e representa o órgão em assuntos de dados abertos
	A unidade de dados abertos possui orçamento para realizar suas atividades
	Uma equipe está alocada para a gestão de dados abertos
	Existem instrumentos de governança para gerir a política de dados abertos do órgão
	Voluntários da sociedade civil contribuem para o programa de dados abertos

Institucionalização nível micro	A publicação de dados abertos está automatizada - não depende de intervenções dos servidores públicos para transformar os dados
	As requisições dos cidadãos sobre dados abertos são respondidas dentro do prazo legal previsto
	Os dados abertos passam por um controle de qualidade antes de serem publicados
	Os servidores públicos responsáveis pela publicação de dados abertos foram adequadamente treinados
	Existe um processo definido para selecionar os dados que serão publicados

Dimensão Disseminação da Informação	
Dimensão	Variáveis
Disseminação da Informação	Concursos de exploração de dados ("hackathons")
	Consultas públicas
	Audiências públicas
	Workshops
	Publicações em redes sociais
	Palestras
	Treinamentos - online ou presenciais
	Publicação de material de divulgação no site do órgão
	Elaboração de guias ou tutoriais para uso de dados abertos
	Outras ações de divulgação aqui não listadas

Dimensão Recursos de Apoio	
Dimensão	Variáveis
Recursos de Apoio	As ferramentas de tecnologia da informação disponíveis para publicação são adequadas
	A documentação sobre dados abertos é conhecida pelos servidores públicos envolvidos na publicação

	O número de pessoas envolvidas na publicação de dados abertos é adequado
	As pessoas envolvidas na publicação de dados abertos possuem a competência necessária
	O orçamento alocado a dados abertos é adequado para realizar as atividades

Dimensão Liderança	
Dimensão	Variáveis
Liderança	Foi dado conhecimento ao(s) dirigente(s) das atividades do programa de dados abertos
	O(s) dirigente(s) participa do planejamento do programa de dados abertos
	O(s) dirigente(s) participa das decisões tomadas no programa de dados abertos
	O(s) dirigente(s) exerce liderança direta no programa de dados abertos
	A ação do(s) dirigente(s) no programa de dados abertos motiva a participação dos servidores públicos

Dimensão Qualidade dos Dados	
Dimensão	Variáveis
Qualidade dos Dados	Os dados estão publicados em formato aberto
	São publicados metadados (arquivos que descrevem os dados)
	Os metadados estão atualizados em relação aos dados publicados
	A periodicidade de publicação dos dados está declarada no portal
	Os dados são publicados de acordo com a publicidade declarada no portal

	O responsável pelos dados abertos para suporte ao cidadão está indicado no portal
	É possível realizar o download dos dados abertos
	Os dados abertos contém "tags" - palavras-chave para facilitar a pesquisa de dados por assunto
	Os dados podem ser visualizados de outra forma além do download de arquivos
	Os dados encontram-se em granularidade fina, ou seja, são apresentados sem agrupamento em totais ou categorias, preservando sua atomicidade

Fonte: Elaboração própria a partir de elementos reunidos da literatura

As questões apresentadas no questionário foram tratadas como variáveis intervalares. Quando as questões de formato Likert estão agrupadas a formar um mesmo construto, podemos dizer que se trata de uma Escala Likert. Carifio & Perla (2008) relatam que nesse caso é apropriada tratar as variáveis de um questionário como intervalares, e podem ser analisadas de forma paramétrica. Boone & Boone (2012) diferenciam "itens Likert" – que isolados só podem ser representados como variáveis ordinais – de "escala Likert". Em uma escala Likert, é possível usar ferramentas de estatística inferencial, médias e desvios padrão, enquanto que em itens Likert isolados o mais recomendado é usar modas, medianas e frequências como ferramentas de análise.

Quando se faz uma investigação empírica, a pergunta de partida deve apontar os dados necessários para se trabalhar. Entretanto, nem sempre se consegue recolhê-los em uma única fonte. Deve-se levar em conta a informação que está disponível (H. C. Silvestre & Araújo, 2012). Em adição aos dados obtidos do questionário, foi avaliada a possibilidade de cruzá-los com os dados secundários levantados no portal de dados abertos, onde se encontram dados que representam a dimensão acesso à informação. Entretanto, a unidade de análise do questionário é o individuo e a dos dados secundários e o órgão publicador. Logo, seria necessária uma agregação pela unidade comum aos dois instrumentos, a função de governo, o que resultou em apenas 9 categorias, correspondentes ao número de funções, insuficiente para se montar um modelo estatístico significativo.

A solução encontrada foi usar, além da variável visualização número de acessos ao portal, representando a dimensão acesso à informação, outras variáveis obtidas de dados secundários com a mesma unidade de análise, o órgão público. Em consonância com a revisão da literatura, foram encontradas variáveis que representam a dimensão de institucionalização e disseminação da informação.

Tabela 5 - Variáveis a partir das Dimensões - Dados Secundários

Dimensão Acesso à Informação	
Dimensão	Variável
Acesso à Informação	Número de acessos ao portal de dados para cada órgão pertencente ao universo pesquisado

Dimensão Disseminação da Informação	
Dimensão	Variável
Disseminação da Informação	Menções ao portal no domínio gov.br - ações de divulgação

Dimensão Institucionalização	
Dimensão	Variável
Institucionalização	Taxa de respeito ao plano de publicação - capacidade de gerir a política de dados

Fonte: Elaboração própria a partir de elementos reunidos da literatura

4.7 Validade e Fiabilidade do Instrumento

A fiabilidade versa sobre o quanto um instrumento é estável, consistente e preciso para realizar uma medição. Para avaliação do trabalho quanto à fiabilidade, Souza (2017) relata que a maioria dos pesquisadores avalia a consistência interna de instrumentos por meio do coeficiente alfa de Cronbach.

Outro método utilizado é o do SplitHalf o qual consiste em dividir n vezes em duas metades a amostra e realizar comparações, de forma a estabelecer um coeficiente de fiabilidade de um instrumento. Segundo Callender & Osburn (1977), estudos teóricos e práticos sugerem que o coeficiente splithalf máximo pode ser mais apropriado que o Alfa de Cronbach. Guttman (1945) mostrou que a verdadeira fiabilidade de um instrumento é a maior entre os coeficientes splithalf calculados entre todas as combinações possíveis. O Alfa de Cronbach utiliza o mesmo método de cálculo que o splithalf, porém neste caso o valor apurado como Alfa é a média dos coeficientes splithalf calculados.

Para esta dissertação, o método de Cronbach e o método de SplitHalf foram aplicados às variáveis do questionário através dos pacotes ltm e psych do software RStudio e foram obtidos os seguintes resultados:

Tabela 6 - Valores do Alfa de Cronbach para as Dimensões do Questionário

Dimensão	Alfa Cronbach	SplitHalf Maximal
Todo o questionário	0.89	0.93
Institucionalização	0.67	0.78
Disseminação da Informação	0.55	0.71
Recursos de Apoio	0.76	0.81
Liderança	0.89	0.91
Qualidade dos Dados	0.77	0.85

Fonte: apuração de resultados do modelo no software RStudio

De uma forma geral, as dimensões para as quais foram calculados os Alfa de Cronbach apresentaram valores satisfatórios – 89% para o instrumento como um todo, e valores de 67%, 76%, 77% e 89% para as dimensões Institucionalização, Recursos de Apoio, Qualidade e Liderança. A exceção ficou por conta da dimensão "Disseminação da Informação", com Alfa de 55%, calculada a partir de variáveis dicotômicas a compor um índice, no qual se percebeu maior variação nas respostas. Ainda assim, o splifhalf máximo calculado para esta dimensão foi de 71%.

A recolha dos dados foi feita a partir de questionário com as questões colocadas de forma precisa tal como foram revistas de forma extensiva na literatura, de maneira que podemos afirmar que as questões estão validadas pelo estudo feito.

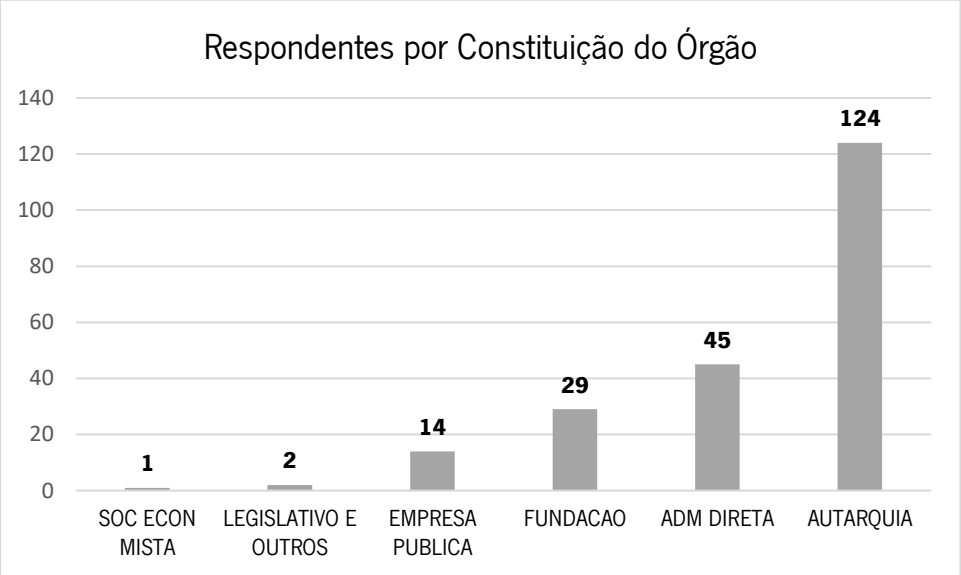
Capítulo 5. Apresentação e Análise dos Dados.

Este capítulo contém a apresentação e análise dos resultados conforme apurados nos 3 instrumentos escolhidos para o *design*: questionário, análise de dados secundários e entrevistas. O questionário forneceu elementos para construção de modelo contendo a variável dependente Qualidade dos Dados e teste das hipóteses H2 e H4. Os dados secundários permitiram a construção de modelo contendo a variável Acesso à Informação e teste das hipóteses H1 e H5. As hipóteses H3 e H6, relativas também à variável Acesso à Informação, não puderam ser testadas pela ausência de dados secundários de Recursos de Apoio e Liderança. As entrevistas foram realizadas para complementar o conhecimento sobre as dimensões do estudo, reforçar as conclusões e obter novas ideias e perceções sobre o tema estudado.

5.1 Perfil dos Respondentes ao Questionário

Esta seção tem como objetivo apresentar o perfil dos funcionários públicos que responderam ao questionário. A quase totalidade das respostas veio de funcionários federais (214, contra 1 estadual), além de configurar como maioria os funcionários deste grupo de órgãos afetados pelo decreto. Os funcionários públicos de autarquias, da administração direta e de fundações respondem juntos, por 92% de todos os respondentes.

Gráfico 1 - Total de Respondentes por Constituição do Órgão

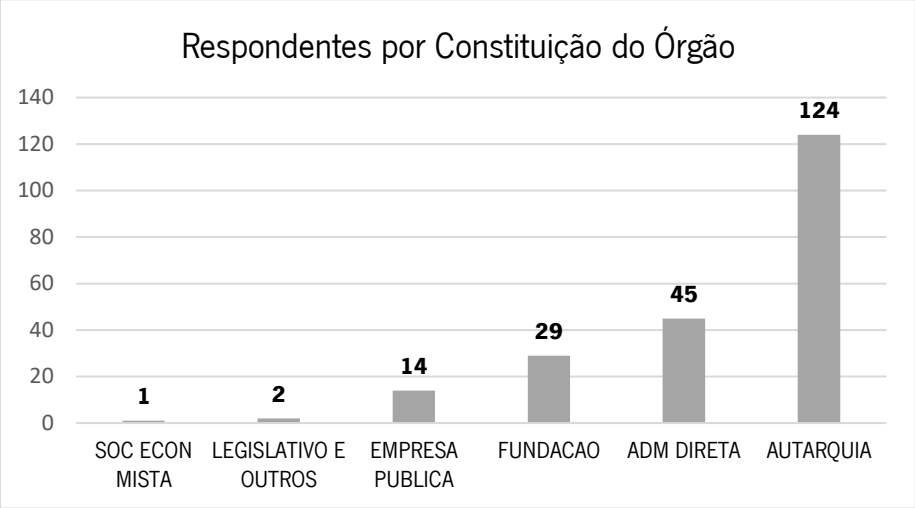


Fonte: apuração de resultados do questionário aplicado

O questionário compreendeu a pergunta sobre a função de governo desempenhada pelo órgão ao qual pertence o respondente. Observamos que o segmento de Educação é o responsável pelo maior número de respondentes, a representar 42% do total, seguido por Serviços Públicos Gerais, a 27% e por Assuntos Econômicos, a 15%. Juntas, estas funções de governo totalizam 84% de todos os respondentes. Se levarmos em conta os órgãos que publicam no portal, suas funções respondem juntas por 78% do total do portal, o que denota uma coerência entre o perfil respondente e sua representatividade no portal de dados abertos.

A Educação é representada no portal em grande maioria pelas universidades, porém também compreende uma única escola secundária. Já os Serviços Públicos Gerais compreendem os órgãos da Administração Direta como os ministérios e as agências reguladoras. Já os Assuntos Econômicos respondem pela atividade governamental neste segmento, como indústria e serviços, energia e transporte, entre outros.

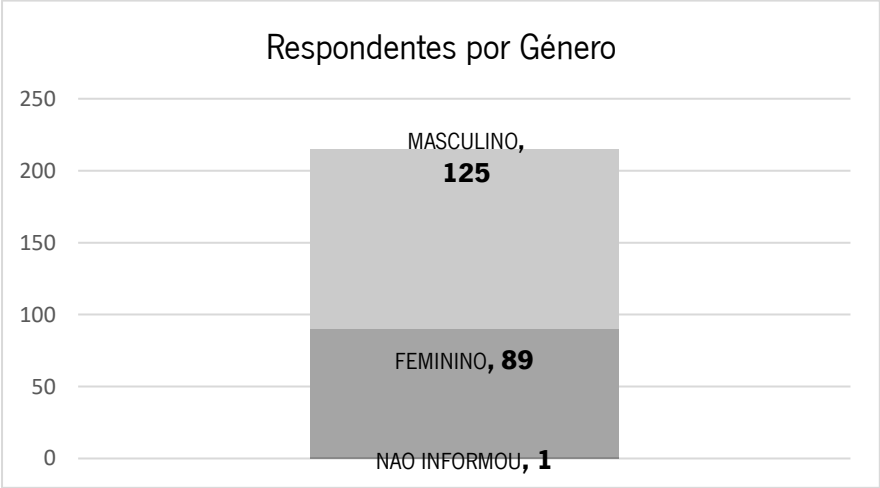
Gráfico 2 - Respondentes por Função de Governo



Fonte: apuração de resultados do questionário aplicado

No que diz respeito ao género, apenas 1 respondente optou por não informar. 125 são do género masculino, o que representa 58% da amostra, e 89 são do gênero feminino, correspondente a 42%. A título de comparação, em seu relatório sobre género e funcionalismo público federal(Secretaria de Gestão e Desempenho de Pessoal, 2021), o Ministério da Economia computa em 2020 os homens a 54%, e as mulheres a 46% para toda a administração pública federal.

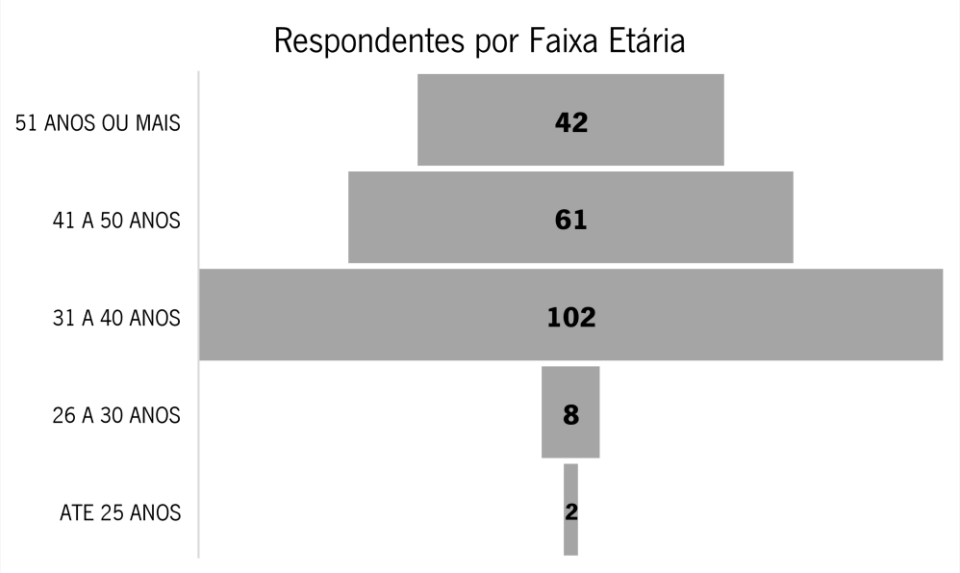
Gráfico 3 - Respondentes por Género



Fonte: apuração de resultados do questionário aplicado

Os respondentes ao questionário estão, em sua maioria (47%), na faixa dos 31 a 40 anos, seguidos pela faixa 41-50 (28%) e 51-60 (20%). Os mais jovens, até 30 anos, são minoria nos funcionários públicos que responderam ao questionário (5%). Em efeito, nos últimos 8 anos poucos concursos públicos foram realizados no Brasil, o que vem impedindo a entrada dos mais jovens no serviço público.

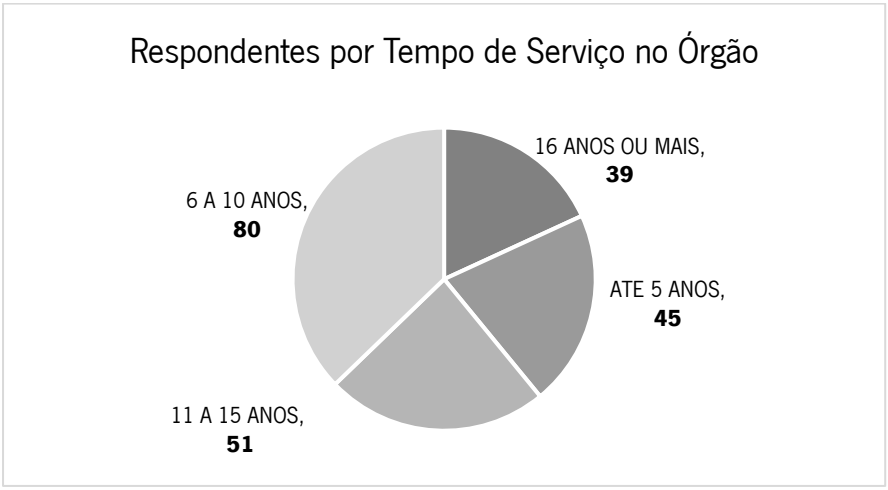
Gráfico 4 - Respondentes por Faixa Etária



Fonte: apuração de resultados do questionário aplicado

Ao se questionar o tempo de serviço em seu órgão atual, percebemos que a maioria (37%) está lotada entre 6 e 10 anos. O percentual de funcionários novos nos órgãos (até 5 anos, 21%) pode ser explicado pela crescente movimentação entre os órgãos, estimulada por programas governamentais de mobilidade, que têm por objetivo repor os funcionários que vão à reforma e oxigenar as repartições, em função da baixa atividade do governo em promover concursos públicos.

Gráfico 5 - Respondentes por Tempo de Serviço no Órgão



Fonte: apuração de resultados do questionário aplicado

5.2 Análises Descritivas

Esta seção tem como objetivo apresentar a estatística descritiva das variáveis apuradas no questionário e nos dados secundários. Apresenta mínimos, máximos, medianas, médias e desvios-padrão.

Análises Descritivas - Questionário

O instrumento operacionalizou a coleta da variável dependente qualidade de dados e das variáveis independentes disseminação da informação, institucionalização, recursos de apoio e liderança. A Tabela 7 faz um resumo da estatística descritiva para todas as variáveis do questionário. Os testes de Shapiro-Wilk indicaram não-normalidade das variáveis, com exceção da Institucionalização

Tabela 7 - Estatística Descritiva - Variáveis - Questionário

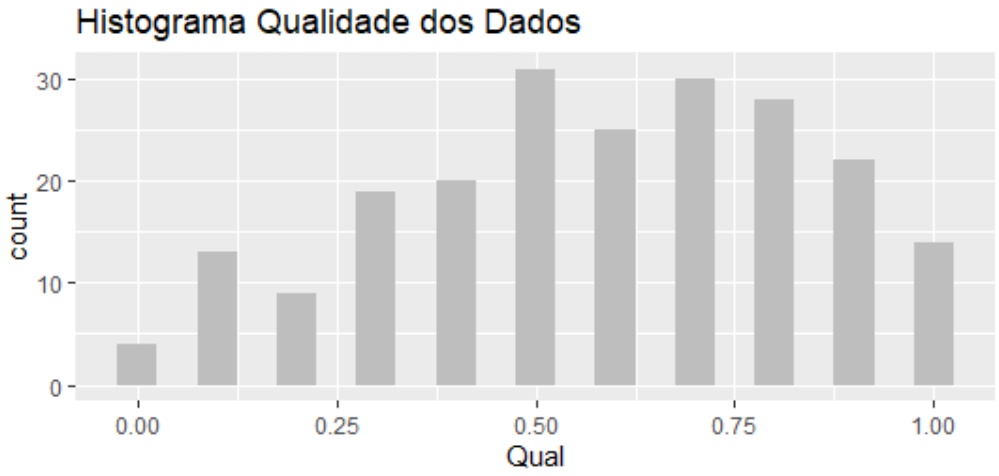
	Variável	Obs.	Mínimo	Máximo	Mediana	Média	Desvio-Padrão	P-value teste Shapiro-Wik
Dependente	Qualidade dos Dados	215	0.00	1.00	0.60	0.58	0.26	8.91E-06
Independente	Disseminação da Informação	215	0.00	0.80	0.20	0.24	0.17	2.53E-08
	Institucionalização		0.03	0.89	0.48	0.48	0.17	0.82
	Recursos de Apoio		0.05	1.00	0.60	0.60	0.18	1.03E-02
	Liderança		0.00	1.00	0.60	0.59	0.23	7.17E-03

Fonte: execução de rotinas no banco de dados do software RStudio

A Qualidade divulgada dos Dados

O histograma indica que as avaliações de Qualidade indicam perceção de qualidade de dados tanto negativas e positivas, sendo que as avaliações se compensaram na média, porém as encontramos em todos os escores, de 0 a 1.

Gráfico 6 - Histograma da Variável Qualidade dos Dados - Questionário

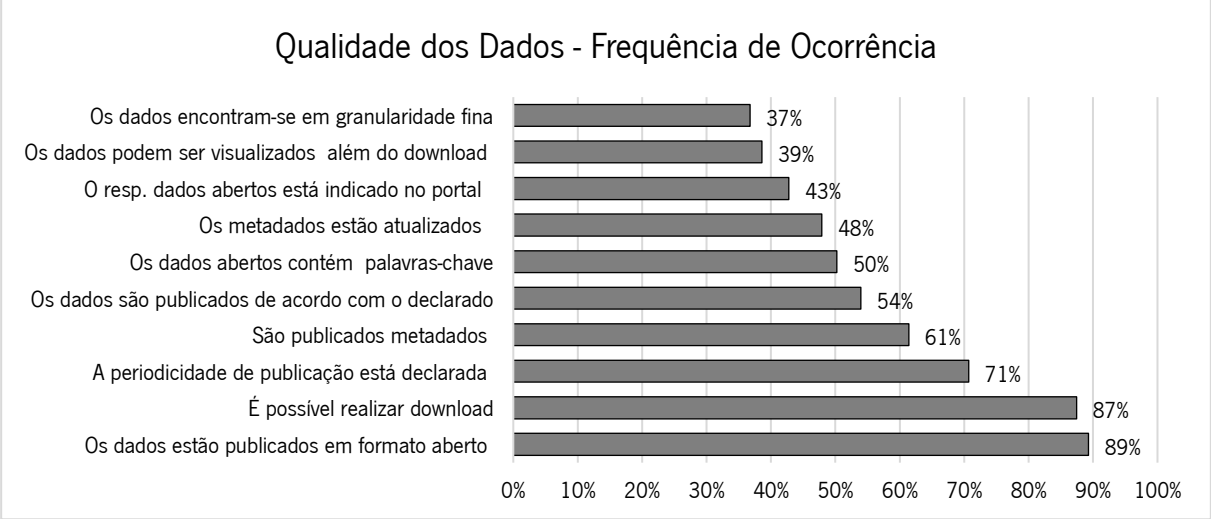


Fonte: execução de rotinas no banco de dados do software RStudio

A variável Qualidade dos Dados teve seu valor aferido para cada respondente a partir das respostas do questionário. Agora iremos analisar o comportamento das respostas nesta dimensão. A maioria dos 215 respondentes (89% e 87%, respetivamente) indicou que os dados são publicados em formato aberto e que o download dos dados é possível, duas características básicas de dados abertos, o que denota o reconhecimento do respeito às regras de base do portal. Adicionalmente, mas com maior moderação (71% e 61%), aponta-se que a periodicidade está declarada e os metadados são publicados, também condições importantes para se consumir os dados.

De uma forma geral, os quesitos de qualidade apresentaram desempenho de médio (em torno de 50% a 60%) a bom. A exceção está na constatação de os dados estarem em granularidade fina a 37% e a presença de outras formas de visualização a 39%. Estas características demandam ferramentas de TI mais sofisticadas para que estejam presentes, o que não é uma realidade como se pode ver na pergunta feita para a variável de institucionalização micro, sobre a automação nas extrações de dados.

Gráfico 7 - Qualidade dos Dados - Frequência de Ocorrência - Questionário

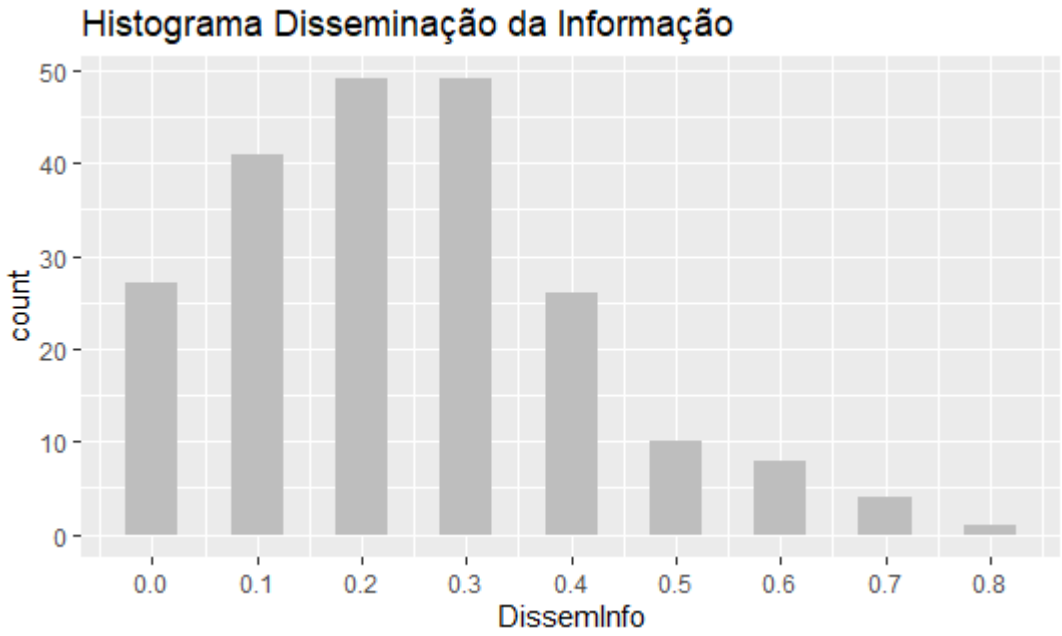


Fonte: execução de rotinas no banco de dados do software RStudio e tabuladas no software Excel

A Disseminação da Informação sobre o Portal de Dados

O histograma mostra que as avaliações muito positivas sobre a variável Disseminação da Informação são pouco frequentes, e apresenta concentração em valores mais baixos (até 0.4), e com cauda à direita.

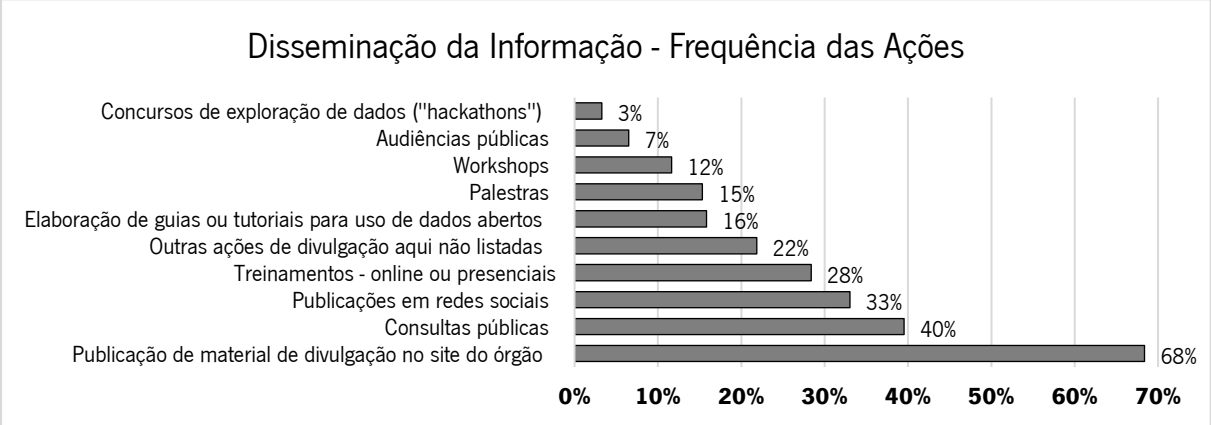
Gráfico 8 – Histograma de Disseminação da Informação - Questionário



Fonte: execução de rotinas no banco de dados do software RStudio

A variável Disseminação da Informação teve seu valor aferido para cada respondente a partir das respostas do questionário. Agora iremos analisar o comportamento das respostas nesta dimensão. A maioria dos 215 respondentes (68%) assinalou a publicação de material sobre dados abertos no site institucional como ação de divulgação de informação, seguida pelas consultas públicas e publicações nas redes sociais. Estas são recomendações da CGU para preparação e divulgação de dados abertos nos PDAs dos órgãos públicos (Controladoria Geral da União, 2020). Por outro lado, a ação menos mencionada na pesquisa foi os concursos de exploração de dados ("hackatons", apenas 7 respostas e 3% do total de 215), a qual é objeto de várias citações na literatura como ação efetiva de divulgação para dados abertos. A publicação de material em site e as consultas públicas demandam menos esforço do que a organização de concursos, que dependem de orçamento, tempo para planejar e seleção prévia de dados.

Gráfico 9 - Frequência das Ações - Disseminação da Informação - Questionário

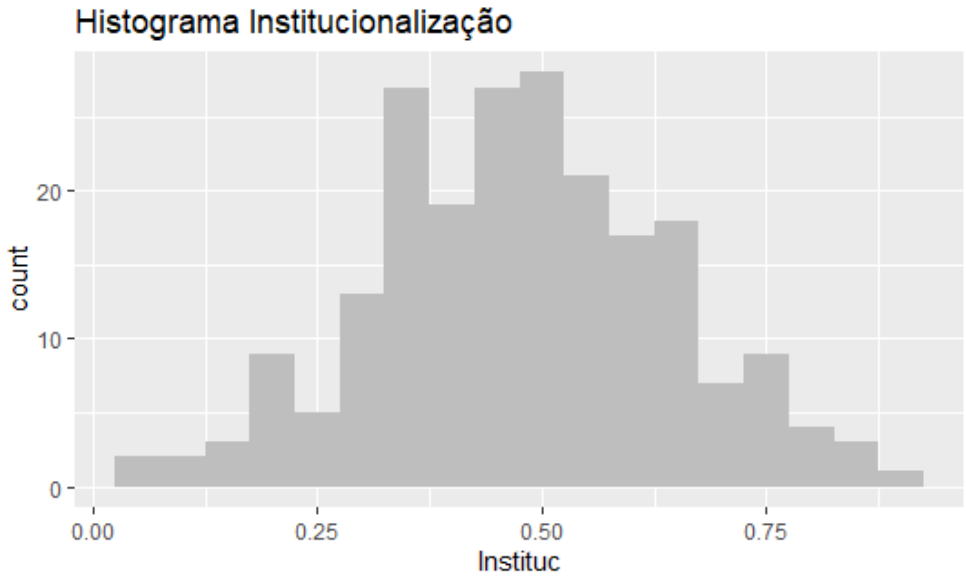


Fonte: execução de rotinas no banco de dados do software RStudio e tabuladas no software Excel

A Institucionalização da Política de dados abertos

A institucionalização é descrita no estudo com uma face macro, que diz respeito ao arranjo institucional, e uma face micro, relativa aos processos de trabalho. A variável Institucionalização para cada respondente é então computada através de seis perguntas dicotômicas ao nível macro, e cinco perguntas do tipo Likert ao nível micro. O histograma mostra que as avaliações para a variável Institucionalização ficaram em torno de 0.5, e os valores extremos - baixa e alta perceção de institucionalização - não foram frequentes.

Gráfico 10 - Histograma de Institucionalização - Questionário

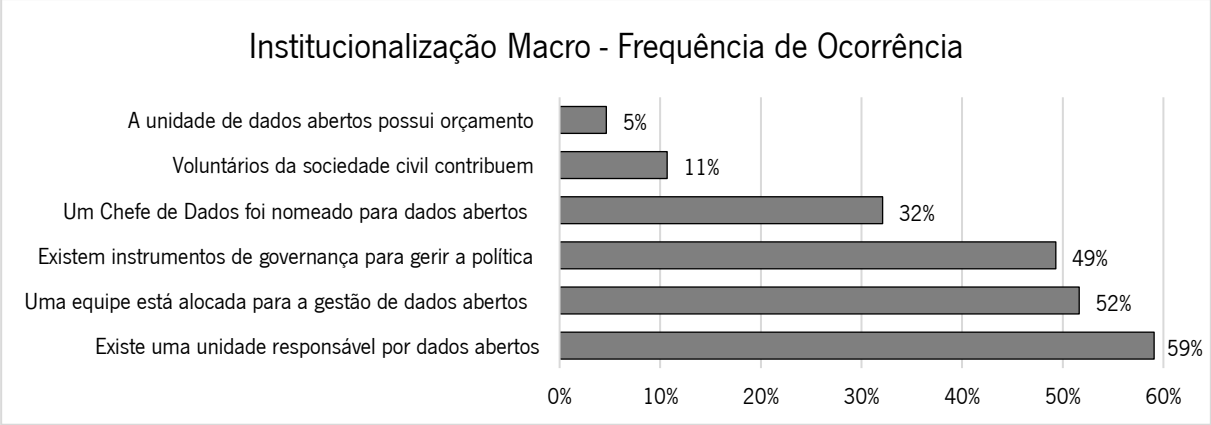


Fonte: execução de rotinas no banco de dados do software RStudio

A variável Institucionalização teve seu valor aferido para cada respondente a partir das respostas do questionário nestas duas faces Agora iremos analisar o comportamento das respostas nesta dimensão. 59% dos respondentes indicaram a existência da unidade de dados abertos, o que é um primeiro passo para a institucionalização de uma área de dados em um órgão público. Ainda 52% reconhecem uma equipe alocada especificamente para esta tarefa, e 49% identificam instrumentos de governança para a gestão da política de dados.

Todavia, a alocação de orçamento específico e a presença de voluntários da sociedade civil a contribuir ainda é tido como de baixa frequência. A participação da sociedade e a disponibilização de recursos são condições também importantes para que se desenvolvam os programas de dados abertos, apontando ações a serem exploradas para melhor sucesso destes programas.

Gráfico 11 - Institucionalização Macro - Frequência de Ocorrência - Questionário

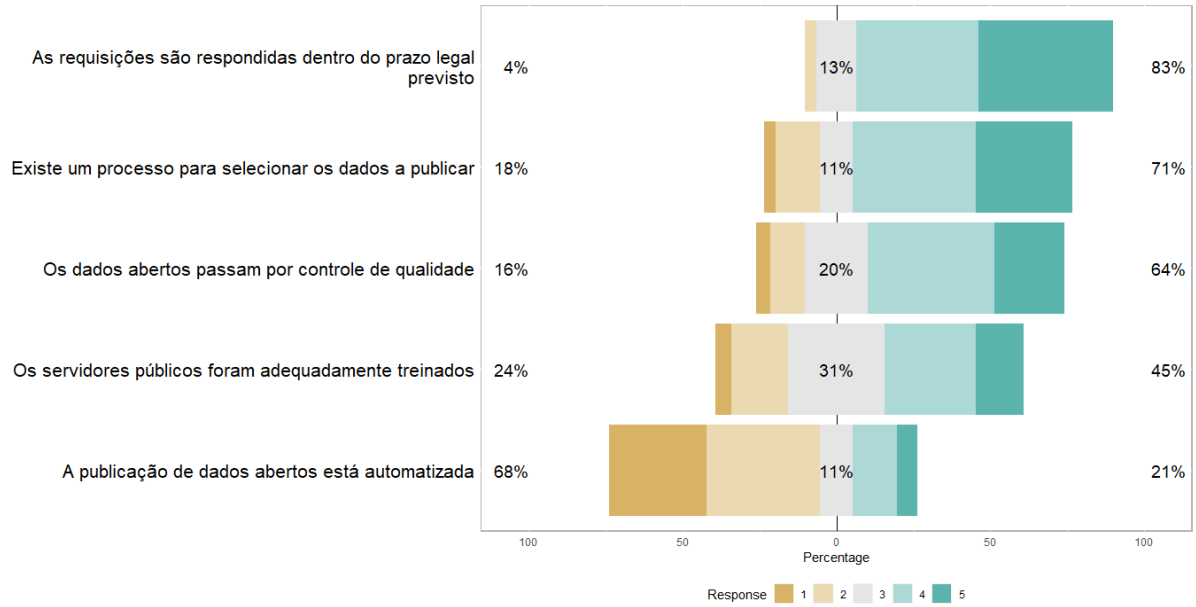


Fonte: execução de rotinas no banco de dados do software RStudio e tabuladas no software Excel

O Gráfico 12 mostra que a perceção dos respondentes é positiva para a existência de processos de tratamento sobre dados abertos – respostas aos cidadãos no prazo recebem 83% de respostas entre Concordo e Concordo Totalmente (itens em azul-claro e azul). Também é positiva a visão sobre os processos de seleção e controle de qualidade dos dados.

O treinamento e a preparação dos funcionários para lidar com dados abertos é vista como neutra. Já a perceção sobre a automação das extrações, reconhecida na literatura como um dos vetores de promoção de qualidade, é baixa – 68% dos respondentes discordam sobre a existência de processos de automação para se extrair os dados.

Gráfico 12 - Institucionalização Micro - Frequência de Ocorrência - Questionário

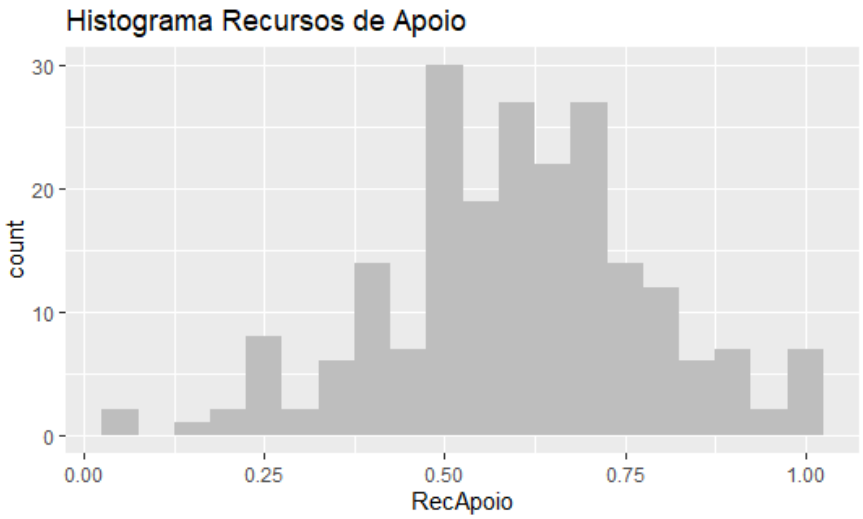


Fonte: execução de rotinas no banco de dados do software RStudio

Os Recursos de Apoio à publicação de dados abertos

O histograma mostra que as avaliações para a os Recursos de Apoio ficaram em torno da média de 60%, com concentração à esquerda e à direita, e apresentando valores extremos de avaliação baixa e alta. É a variável do estudo com maior avaliação média, o que reforça a importância dada pelos órgãos públicos em fornecer ferramentas para publicação de dados abertos, seja como treinamento, documentação ou tecnologia da informação.

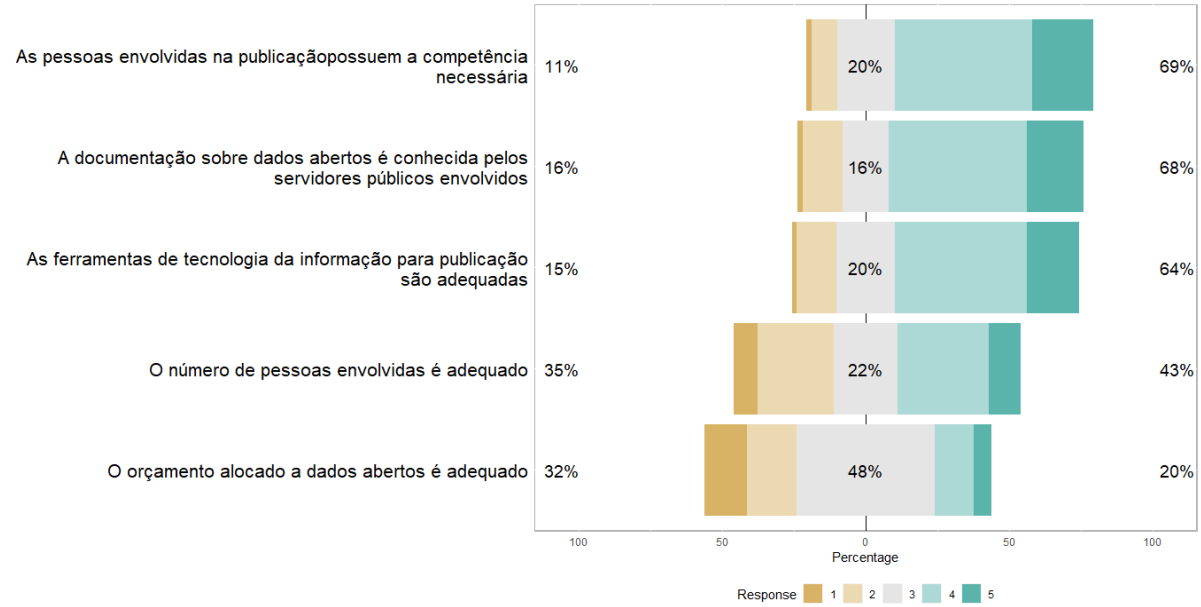
Gráfico 13 - Histograma de Recursos de Apoio - Questionário



Fonte: execução de rotinas no banco de dados do software RStudio

A variável Recursos de Apoio teve seu valor aferido para cada respondente a partir das respostas do questionário nestas duas faces. Agora iremos analisar o comportamento das respostas nesta dimensão. O Gráfico 14 mostra que a perceção dos respondentes é positiva em relação à competência das pessoas que lidam com dados abertos, à documentação e às ferramentas de publicação – todas recebem mais de 60% de respostas entre Concordo e Concordo Totalmente. Já no que concerne ao número de pessoas envolvidas na publicação não existe uma uniformidade de pensamento, a opção neutra é menos frequente, o que pode levar-se a inferir que parte dos órgãos tem equipe suficiente e outra parte tem equipe pequena, porém competente. Sobre o orçamento dedicado, a visão é mais neutra, o que pode se justificar pelo fato de os responsáveis de dados abertos não serem com frequência os controladores de despesas ou orçamento, atendo-se a função mais técnica.

Gráfico 14 - Recursos de Apoio - Frequência de Ocorrência - Questionário

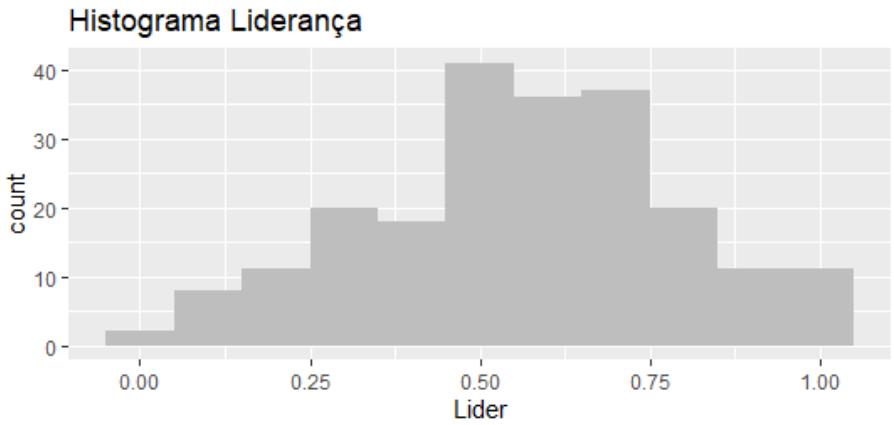


Fonte: execução de rotinas no banco de dados do software RStudio

A Liderança na disponibilização de Dados Abertos

O histograma acima mostra as avaliações de Liderança em torno da média de 59%, com maior peso para avaliações positivas, maiores que 50%. Porém a análise individual sobre as questões da dimensão identifica a informação dos dirigentes como motor de elevação da avaliação, como é visto a seguir na análise das respostas do questionário.

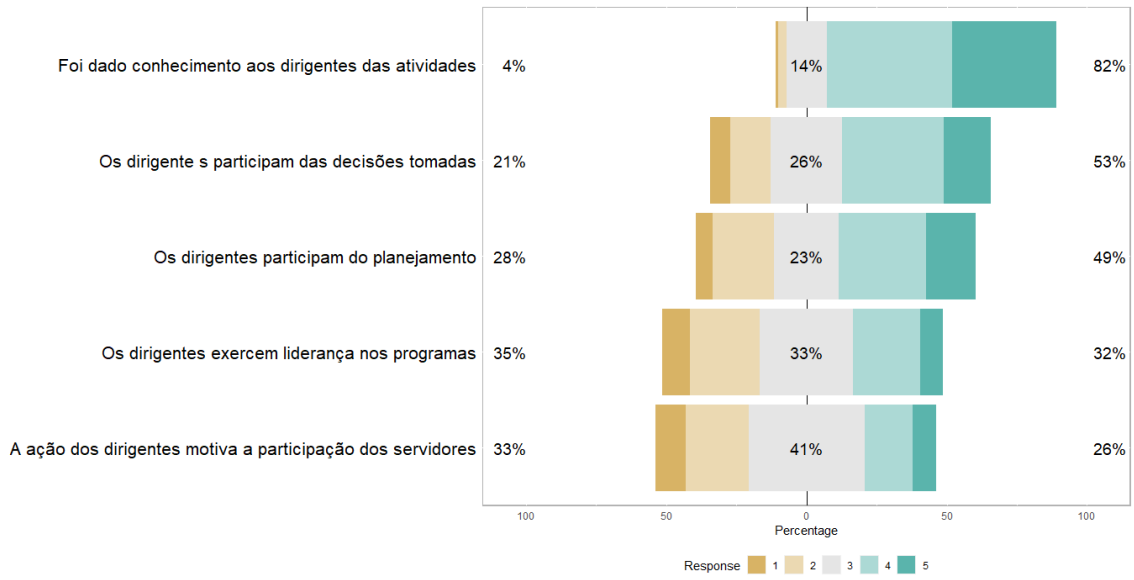
Gráfico 15 - Histograma de Liderança - Questionário



Fonte: execução de rotinas no banco de dados do software RStudio

O Gráfico 16 mostra com maior destaque a perceção de que os dirigentes foram informados sobre as atividades em dados abertos (82% de concordância). Esse resultado é esperado uma vez que os planos de dados abertos precisam ser formalmente aprovados pelos dirigentes de cada órgão para que sejam aceitos pela CGU. Também a participação dos dirigentes na tomada de decisão e no planejamento é vista como positiva (53% e 49%), porém em menor grau que o conhecimento dos dirigentes sobre o tema. A perceção de liderança direta e motivação permanece mais neutra, sem expressão de concordância ou discordância em maior grau. Esse resultado também é esperado uma vez que as estruturas dos órgãos públicos são fortemente baseadas na hierarquia, e as equipes de dados abertos podem ter pouco ou mesmo nenhum contato com os dirigentes, o que pode vir a dificultar a expressão de sua perceção.

Gráfico 16 - Liderança - Frequência de Ocorrência - Questionário



Fonte: execução de rotinas no banco de dados do software RStudio

Análises Descritivas - Dados Secundários

O instrumento operacionalizou a coleta da variável dependente acesso à informação, representada pela contagem de visualizações das páginas de dados abertos de cada órgão e das variáveis independentes disseminação da informação e institucionalização, representadas respetivamente pelo número de menções aos dados abertos de cada órgão segundo o portal governamental gov.br e pela taxa de respeito ao plano de dados abertos. A Tabela 8 faz um resumo da estatística descritiva para todas as variáveis do questionário. Os testes de Shapiro-Wilk indicaram não-normalidade de todas as variáveis do modelo.

Tabela 8 - Estatística Descritiva - Variáveis - Dados Secundários

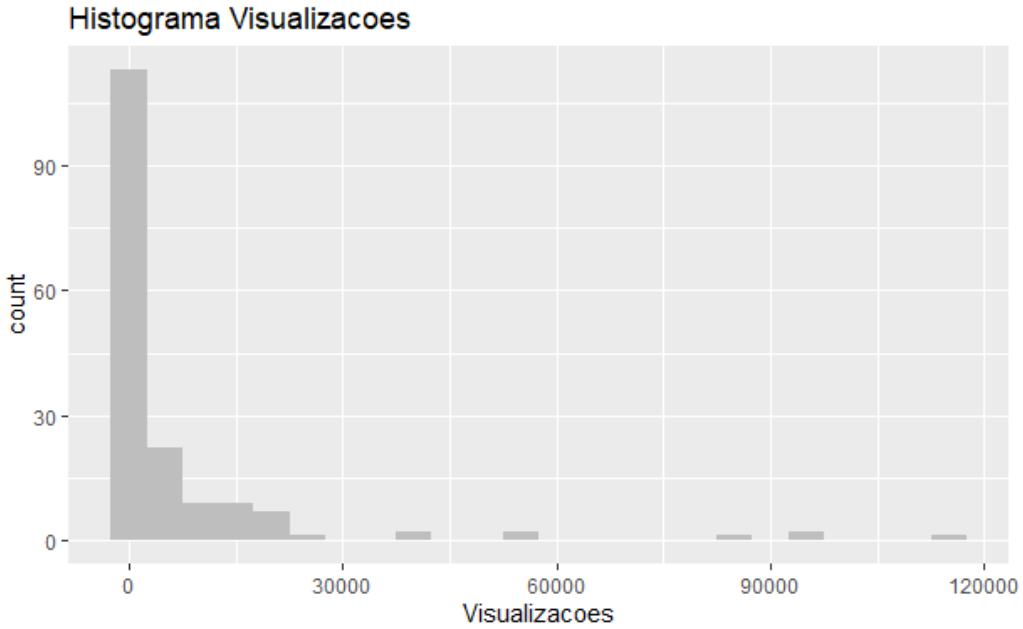
	Variável	Obs.	Mínimo	Máximo	Mediana	Média	Desvio-Padrão	P-value teste Shapiro-Wik
Dependente	Acesso à Informação - Visualizações	169	5	115 026	954	6 783	16 673	2.20E-16
Independente	Disseminação da Informação - Menções	169	61	37 500	1 560	2 700	3 853	2.20E-16
	Institucionalização - Respeito ao Plano		0.00	1.00	0.00	0.41	0.46	2.20E-16

Fonte: execução de rotinas no banco de dados do software RStudio

O Acesso à Informação através das visualizações de páginas

É a contagem do total de visualizações de páginas no site dados.gov.br para cada órgão que lá publica, coletado através das estatísticas mensais disponíveis no portal dados.gov.br. O histograma abaixo mostra uma cauda longa à direita, o que demonstra um pequeno número de ocorrências com alto número de visualizações, e uma concentração à esquerda com baixa contagem de visualizações.

Gráfico 17 - Histograma Variável Acesso à Informação – Visualizações - Portal

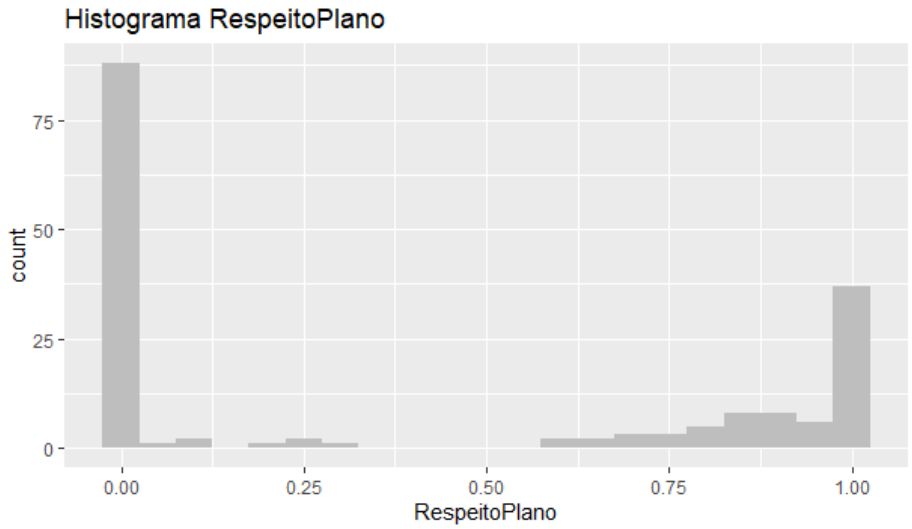


Fonte: execução de rotinas no banco de dados do software RStudio

A Institucionalização como indicador de respeito ao plano de dados abertos

É o resultado do total de bases de dados publicadas dividido pelo total de bases declaradas a serem publicadas no PDA, medido em uma escala de 0 a 1. O histograma mostra um comportamento de um maior número de órgãos com taxa 0, uma vez que não possuem PDA ativo, mas também uma boa parcela de órgãos com excelência, a respeitar os cronogramas de publicação de seus dados.

Gráfico 18 - Histograma Variável Institucionalização - Respeito ao Plano - Portal

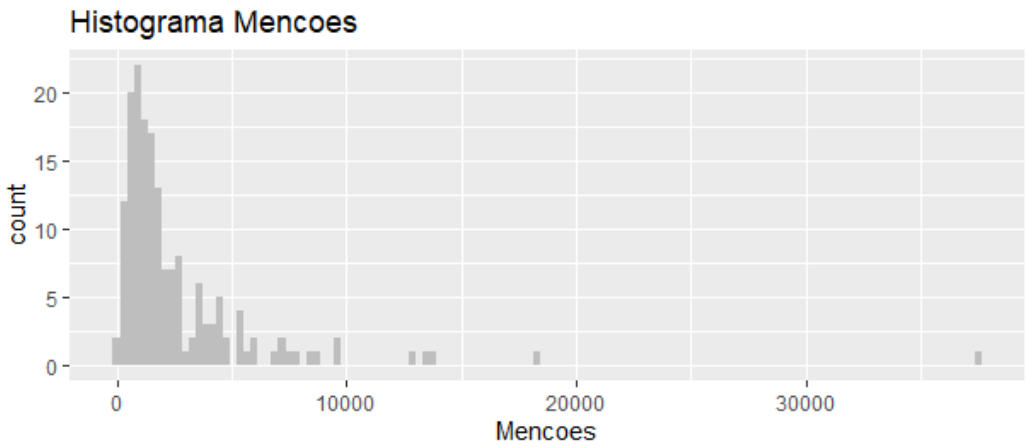


Fonte: execução de rotinas no banco de dados do software RStudio

A Disseminação da Informação pelas menções a dados abertos do órgão

É o resultado do total de páginas encontrados em pesquisa do google para o domínio gov.br, com o nome do órgão e o termo dados.gov.br. (pesquisado como site:gov.br + nome do órgão + dados.gov.br. O histograma mostra uma cauda longa à direita, e uma concentração de órgãos com baixo número de menções de divulgação.

Gráfico 19 - Histograma Variável – Disseminação da Informação - Menções - Portal



Fonte: execução de rotinas no banco de dados do software RStudio

5.3 Análise Bivariante e Multivariante

Esta seção realiza a análise bivariada e multivariada para as variáveis dependentes Qualidade dos Dados e Acesso à Informação, relativas aos modelos de dados do questionário e dos dados secundários.

Qualidade dos Dados

O objetivo aqui proposto é apresentar as análises de dados realizadas para testar as hipóteses H2 – os recursos de apoio influenciam a qualidade dos dados, e H4 – a institucionalização influencia a qualidade dos dados.

Iniciamos por realizar as análises de correlação entre todas as variáveis independentes e a variável dependente, Qualidade dos Dados. Uma vez que as distribuições das variáveis não são normais, os testes a realizar são do tipo não paramétricos. Então, foi utilizado o método de correlação de Spearman no software RStudio, com os seguintes resultados:

Tabela 9 - Resumo Correlação de Spearman para variável Qualidade dos Dados

	p-value	rho de Spearman	Variância Compartilhada rho^2
Disseminação da Informação	1.74E-06	0.32	0.10
Recursos de Apoio	1.90E-12	0.46	0.21
Liderança	3.99E-05	0.28	0.08
Institucionalização	1.53E-13	0.48	0.23

Fonte: elaborado pelo autor a partir da execução de rotinas do software RStudio

Os testes rejeitaram a hipótese nula de ausência de correlação (=0), com p-value < 0.001 para todas as variáveis, e indicam relação direta positiva das variáveis com a qualidade. O índice de variância compartilhada – rho^2 - ficou entre 8% e 23%. Os coeficientes de correlação (rho) oscilaram em torno de 32% a 48%, indicando correlação moderada a forte, segundo Cohen (1988, p. 115).

Em seguida, foi verificada a correlação de Spearman entre as variáveis independentes. O objetivo é estabelecer se existe um risco de multicolinearidade, que é quando as variáveis apresentam um alto grau de correlação entre elas, representando o mesmo conceito, logo em uma investigação não são desejáveis. Segundo Biaggi (2017), um fator maior que 80% indica multicolinearidade. Para tal análise, deve-se observar a matriz de correlações. No caso deste estudo, a maior relação é de 0,62, o que permite a evolução da análise para um modelo de regressão.

Tabela 10 - Matriz de Correlação das Variáveis Independentes no Questionário

	Disseminação da Informação	Recursos de Apoio	Liderança
Recursos de Apoio	0.35		
Liderança	0.34	0.58	
Institucionalização	0.46	0.62	0.50

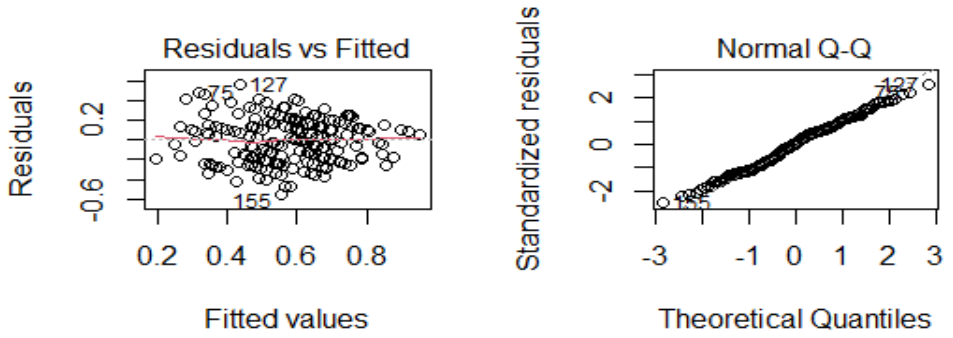
Fonte: elaborado pelo autor a partir da execução de rotinas no banco de dados do software RStudio

Foi feito o modelo de regressão com todas as variáveis. Para que o modelo de regressão seja válido, alguns pressupostos devem ser verificados: os resíduos devem seguir uma distribuição normal, a variância deve ser constante (homocedasticidade) e os resíduos devem ser independentes – distribuídos de forma aleatória.

Para verificar a normalidade da distribuição, utilizamos o teste de Shapiro-Wilk, que com p-valor = 0.1765 (> 0.05) rejeita a hipótese nula e confirma a normalidade dos resíduos. A seguir, no primeiro gráfico, temos os resíduos em função dos valores estimados. Os resíduos estão distribuídos de maneira aleatória e com amplitude em torno do 0, conferindo a independência e a homocedasticidade. O segundo

gráfico confirma a normalidade dos resíduos, de forma visual, ao não apresentar grande fuga em relação à reta teórica, na diagonal.

Gráfico 20 - Homocedasticidade e Normalidade dos Resíduos na Regressão - Questionário



Fonte: elaborado pelo autor a partir da execução de rotinas do software RStudio

Verificados os pressupostos, o modelo de regressão múltipla, com todas as variáveis dependentes e a variável independente, mostrou pelo R^2 ajustado uma explicação de 28.76% da variância da Qualidade dos Dados em função das demais variáveis, quando agem em conjunto no modelo.

Tabela 11 - Resultado do Modelo de Regressão - Questionário

Variável Dependente

Qualidade dos Dados

Hipótese	Variável Independente	Sinal	Coeficiente	
	(Intercept)	+	0.12646	*
H2	Recursos de Apoio	+	0.42198	***
H4	Institucionalização	+	0.43095	***
	Disseminação da Informação	+	0.23684	*
	Liderança	-	-0.11226	

R2 ajustado = 28.76%, p-value: 1.543e-15

'***' < 0.001 '**' < 0.01 '*' <0.05 '.' < 0.1

Fonte: elaborado pelo autor a partir da execução de rotinas no banco de dados do software RStudio

Assim, foi possível confirmar estatisticamente as hipóteses H2 – os recursos de apoio influenciam a qualidade dos dados, e H4 – a institucionalização influencia a qualidade dos dados. Pela análise do modelo, a disseminação da informação também contribui para a qualidade dos dados, em um grau menor que recursos de apoio e institucionalização. É possível também perceber que a variável liderança não foi considerada significativa e não se pode afirmar sua influência.

Acesso à Informação

O objetivo aqui proposto é apresentar as análises de dados realizadas para testar as hipóteses H1 – a disseminação da informação influencia o acesso à informação e H5 - a institucionalização influencia o acesso à informação.

Iniciamos por realizar as análises de correlação entre cada uma das variáveis independentes – Menções, representando a Disseminação da Informação, e Taxa de Respeito ao Plano, representando a Institucionalização – com a variável dependente Visualizações, que representa o Acesso à Informação. Uma vez que as distribuições das variáveis não são normais, os testes a realizar são do tipo não paramétricos. Então, foi utilizado o método de correlação de Spearman no software RStudio, com os seguintes resultados:

Tabela 12 - Resumo Correlação de Spearman para variável Acesso à Informação - Visualizações

	p-value	rho de Spearman	Variância Compartilhada rho^2
Institucionalização - Respeito ao Plano	3.61E-07	0.38	0.14
Disseminação da Informação - Menções	9.09E-09	0.42	0.18

Fonte: elaborado pelo autor a partir da execução de rotinas no banco de dados do software RStudio

Os testes rejeitaram a hipótese nula de ausência de correlação (=0), com p-value < 0.001 para todas as variáveis, e indicam relação direta positiva das variáveis com o acesso à informação. O índice de variância compartilhada – rho^2 - ficou entre 14% e 18%. Os coeficientes de correlação (rho) oscilaram em torno de 38% a 42%, indicando correlação moderada a forte, segundo Cohen (1988, p. 115).

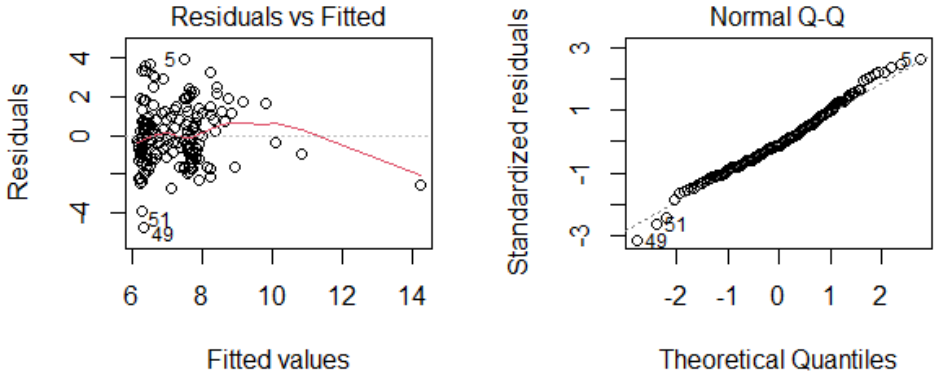
Em seguida, foi verificada a correlação de Spearman entre as variáveis independentes. O objetivo é estabelecer se existe um risco de multicolinearidade, que é quando as variáveis apresentam um alto grau de correlação entre elas, representando o mesmo conceito, logo em uma investigação não são desejáveis. Segundo Biaggi (2017), um fator maior que 80% indica multicolinearidade. Para tal análise, deve-se observar a matriz de correlações. No caso deste estudo, a relação entre as variáveis disseminação da informação e institucionalização é muito baixo, de 0,0085, o que permite a evolução da análise para um modelo de regressão.

Segundo Keene (1995), os dados devem ser analisados sem transformação e seus resíduos devem ser confrontados com a presença de outliers e desvios de normalidade. Caso existam problemas

no modelo, como a ausência de normalidade nos resíduos, transformações devem ser consideradas como a logarítmica de forma a buscar validar o modelo e sua interpretação. A regressão múltipla sem transformação foi realizada e acusou no teste de normalidade dos resíduos de Shapiro Wilk um valor de 2.2e-16, muito inferior ao mínimo de 0.05 que permitiria validar o modelo.

A variável Acesso à Informação – visualizações de páginas – é uma variável de contagem que possui uma longa cauda à direita indicando a possibilidade de fazer a transformação log para análise do modelo. Um novo modelo foi então criado, com resultado de R^2 ajustado de 30.5% e desta vez com o resultado do teste de normalidade de resíduos de Shapiro Wilk de 0.12, o que confere validade ao modelo. Os gráficos a seguir confirmam visualmente a normalidade dos resíduos e a homocedasticidade, com os resíduos distribuídos de maneira aleatória e com amplitude em torno do 0.

Gráfico 21 - Homocedasticidade e Normalidade dos Resíduos na Regressão Log - Portal



Fonte: elaborado pelo autor a partir da execução de rotinas do software RStudio

Verificados os pressupostos, o modelo de regressão múltipla, com todas as variáveis dependentes e a variável independente, mostrou pelo R^2 ajustado uma explicação de 30.49% da variância do Acesso à Informação em função da Disseminação da Informação e da Institucionalização, quando agem em conjunto no modelo.

Adicionalmente, como forma de tornar mais robusta a análise, foi também feita uma regressão de Poisson, indicada para modelos nos quais existem variáveis de contagem como é o caso da variável visualizações (UCLA Statistical Consulting Group, 2021). Uma vez que a variância é maior que a média, a regressão de Poisson utilizada foi a de "QuasiPoisson", que ajusta os erros standard de acordo com a relação entre variância e média, de forma a simular a condição de Poisson – variância = média.

Tabela 13 - Resultado do Modelo de Regressão com Log e QuasiPoisson - Portal

Variável Dependente **Acesso à Informação**							
		Regressão Log Múltipla			Regressão QuasiPoisson		
Hipótese	Variável Independente	Sinal	Coeficiente		Sinal	Coeficiente	
	(Intercept)	+	6.123	***	+	7.692	***
H1	Disseminação da Informação	+	1.85E-04	***	+	8.36E-05	***
H5	Institucionalização	+	1.418	***	+	1.345	***

R^2 ajustado = 30.49%, p-value: 2.885e-14
'***' < 0.001 '**' < 0.01 '*' <0.05 '.' < 0.1

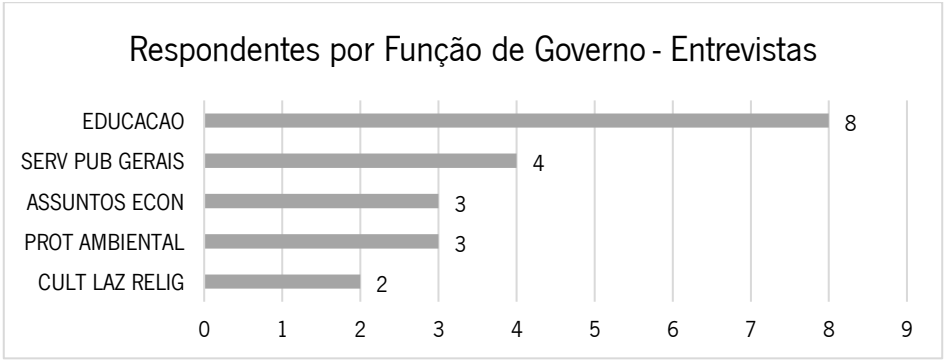
Fonte: elaborado pelo autor a partir da execução de rotinas no banco de dados do software RStudio

Assim, foi possível confirmar estatisticamente as hipóteses H1 – a disseminação da informação influencia o acesso à informação e H5 - institucionalização influencia o acesso à informação. Tanto a variável Menções, representando a disseminação da informação, quanto a variável Taxa de Respeito ao Plano, representando a Institucionalização – contribuem de forma positiva para o aumento percentual na variável Acesso à Informação, representada pelas visualizações no portal. O aumento é na razão percentual para aumento de unidades uma vez que o modelo é transformado em log para a variável dependente.

5.4 Visão Geral sobre as Entrevistas

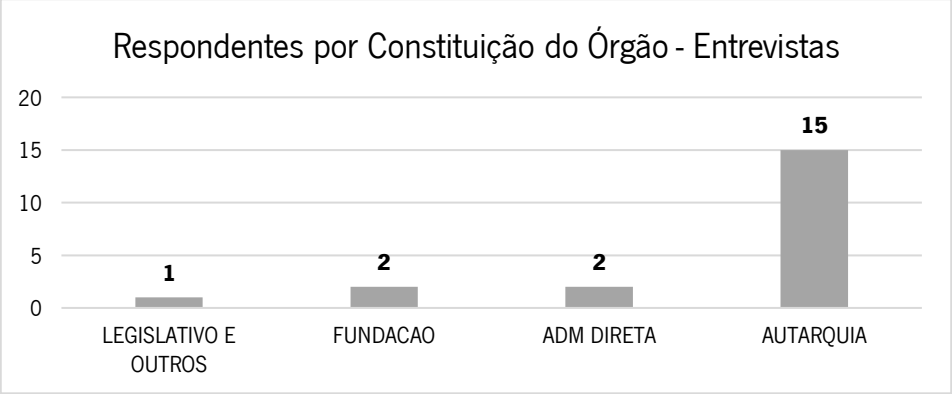
O perfil dos entrevistados seguiu uma distribuição parecida com a dos respondentes ao questionário. A função de governo mais representada é a Educação, com 40% dos entrevistados. A título de comparação, no questionário a Educação responde por 42% das respostas. Para a constituição do órgão, as Autarquias são 75%, contra 58% no questionário, representando também maioria nos dois instrumentos.

Gráfico 22 - Respondentes por Função de Governo - Entrevistas



Fonte: apuração de resultados das entrevistas realizadas

Gráfico 23 - Respondentes por Constituição do Órgão - Entrevistas

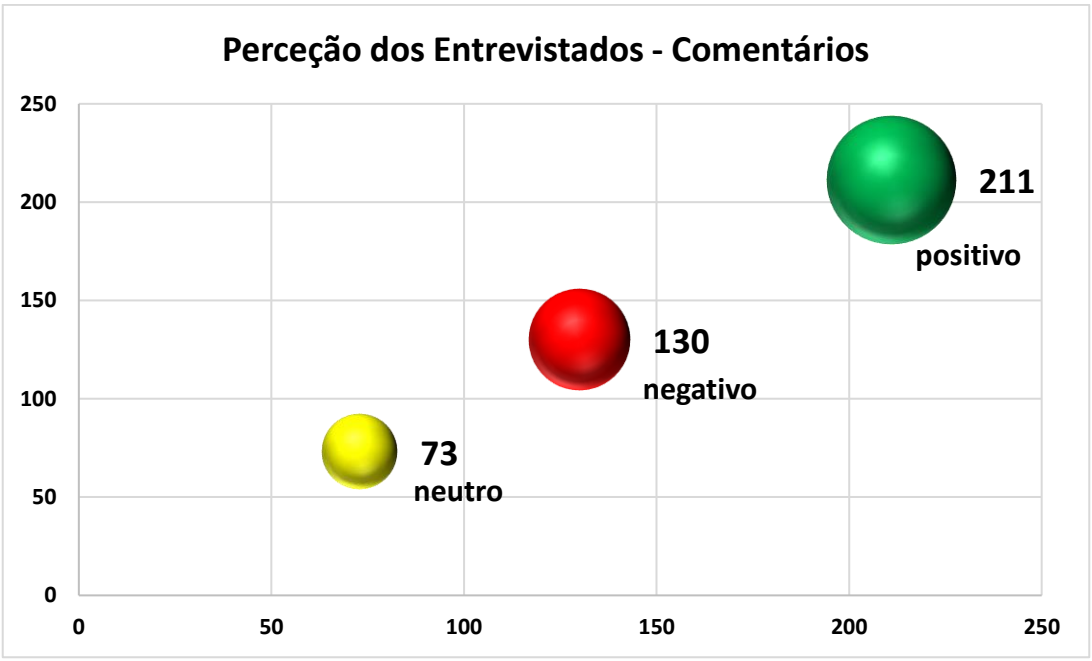


Fonte: apuração de resultados das entrevistas realizadas

Foram realizadas 20 entrevistas, tendo sido anotados 414 comentários. Os comentários foram registados a partir das respostas fornecidas durante as entrevistas. Embora a entrevista tenha seguido um roteiro de perguntas que remontam às dimensões de dados abertos, os funcionários que participaram tiveram a liberdade de colocar suas opiniões, expectativas e de levantar aspectos na generalidade ligados às políticas de dados abertos.

Percebe-se que as visões emitidas são de forma geral positivas – 211 comentários; porém vemos que representam apenas 51% do total das opiniões, representando assim também uma visão critica dos entrevistados sobre os dados abertos em seus órgãos.

Gráfico 24 - Perceção dos Entrevistados - Polaridade



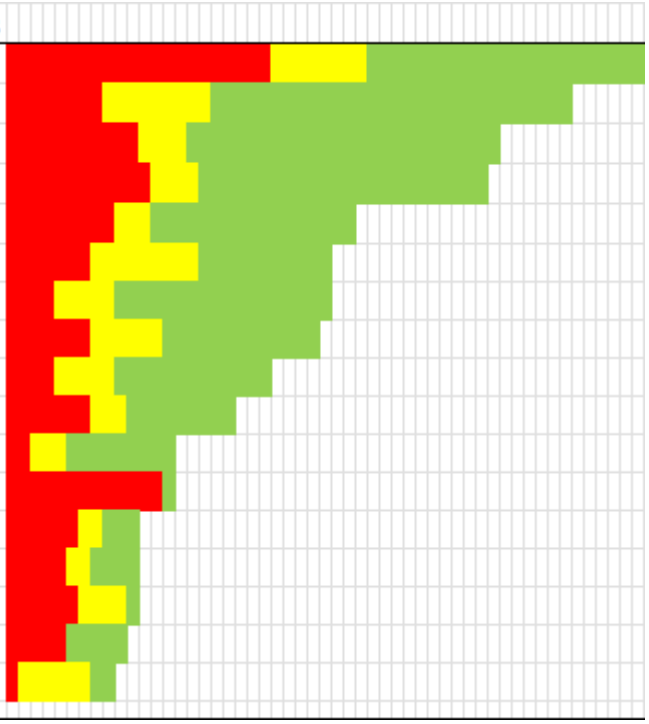
Fonte: apuração de resultados das entrevistas realizadas

Os comentários foram, na sequência, agrupados por tema, e formaram 17 categorias. Das 414 observações, o tema mais frequente foi o de Qualidade dos Dados – 55 citações, seguido por Processos – 47, Estrutura Organizacional – 41, Planejamento – 40, e Extração de Dados - 29. A prevalência de menções estes temas pode ser explicada em função de serem pilares para a publicação de dados abertos, conforme foi visto na revisão da literatura. A menção repetida ao planejamento é esperada uma vez que os órgãos públicos precisam construir e executar PDAs a cada 2 anos, segundo as leis brasileiras.

O gráfico abaixo mostra todas as categorias apuradas em um mapa de cores onde o vermelho significa perceção negativa, o amarelo neutra, e o verde positiva. Ao observar o gráfico, é notável o equilíbrio entre a perceção negativa e positiva para a Qualidade, e a visão maioritariamente positiva sobre Processos, Estrutura Organizacional e Planejamento. Recursos Humanos aparece como destaque de perceções negativas e Orçamento de perceção neutra.

Gráfico 25 - Perceção dos Entrevistados - por Categoria

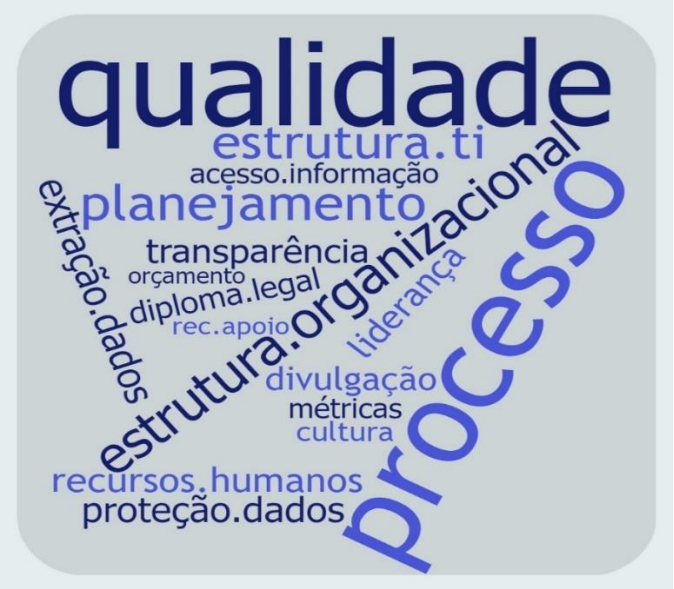
Categoria	Comentários
Qualidade	55
Processo	47
Estrutura Organizacional	41
Planejamento	40
Extração de Dados	29
Liderança	27
Transparência	27
Divulgação	26
Acesso à Informação	22
Estrutura de TI	19
Proteção dos Dados	14
Recursos Humanos	14
Diploma legal	12
Métricas	11
Cultura Organizacional	11
Recursos de Apoio	10
Orçamento	9



Fonte: apuração de resultados das entrevistas realizadas

O gráfico a seguir ilustra, em forma de nuvem de palavras, a frequência dos comentários e sua relevância relativa uns aos outros.

Gráfico 26 - Nuvem de Palavras das Entrevistas - Categorias



Fonte: apuração de resultados das entrevistas tabuladas em wordclouds.com

5.5 Análise das Variáveis do Estudo a partir das Entrevistas

De forma a permitir a análise dos dados das entrevistas vis-à-vis as dimensões do estudo, as 17 categorias de comentários foram agrupadas, por afinidade e coerência com a revisão da literatura, nas variáveis identificadas no estudo, como é mostrado na tabela abaixo.

Tabela 14 - Agrupamento Comentários nas Variáveis do Estudo - Entrevistas

Categoria	Variável do Estudo	Comentários
Acesso à Informação	Acesso à Informação	22
Qualidade	Qualidade	55
Processo	Institucionalização	47
Estrutura Organizacional		41
Planejamento		40
Extração de Dados		29
Proteção dos Dados		14
Métricas		11
Liderança	Liderança	27
Divulgação	Disseminação da Informação	26
Estrutura de TI	Recursos de Apoio	19
Recursos Humanos		14
Recursos de Apoio		10
Orçamento		9
Transparência	Não mapeada	27
Diploma legal		12
Cultura Organizacional		11

Fonte: apuração de resultados das entrevistas realizadas

As categorias de transparência, do diploma legal – atuação dos órgãos de controle, e da cultura organizacional não foram mapeadas no estudo, porém foram mencionadas e assim registadas. O aparecimento destes temas de forma espontânea nas entrevistas mostra a complexidade e abrangência dos programas de Dados Abertos, cuja realidade vai além do que foi revisto na literatura.

Acesso à Informação

De um total de 22 comentários, 6 salientam que o acesso à informação pode ser potencializado por ações de divulgação, o que reforça a hipótese H1 – a disseminação da informação influencia o acesso à informação. Para 3 pessoas, o uso de recursos de apoio tais como a apresentação dos dados em painéis – "*dashboards"*, a disponibilidade de serviços digitais e a construção de portais específicos de metadados podem aumentar o acesso à informação, indicando a possibilidade de a hipótese H3 – o uso de recursos de apoio influencia o acesso à informação – ser válida. Entretanto, são apenas 3 pessoas a fazer esta afirmação em um grupo de 20.

Para outros 2 entrevistados, o desafio no acesso à informação parece continuar a ser como tornar os dados abertos conhecidos. Eles desconhecem se a sociedade tem ou não acesso aos dados, pois recebem muito pouco ou nenhum *feedback*.

Qualidade dos Dados

De 55 comentários, 21 foram sobre o tema da atualidade dos dados, sendo que 11 relataram preocupação com dados desatualizados, e 10 estão satisfeitos. Essa divisão de opiniões reflete bem o que foi respondido no questionário, que apontou 48% de dados atualizados. A atualidade dos dados é um dos principais requisitos para os dados serem consumidos e serem úteis.

Sobre o comportamento da Qualidade dos Dados como variável dependente, 7 pessoas identificaram os recursos de apoio como influenciadores da qualidade, o que reforça a hipótese H2 – o uso de recursos de apoio influencia a qualidade dos dados. 13 entrevistados relacionaram os processos de dados abertos à qualidade, o que reforça a hipótese H4 – A institucionalização influencia a Qualidade dos Dados.

Recursos de Apoio

Em um total de 52 comentários, 2 foram feitos ligando diretamente os recursos de apoio à promoção do acesso a dados abertos – necessidade de ferramentas para evoluir o portal e proporcionar

feedback e promoção de treinamentos para capacitação. Esses comentários reforçam mas não podem confirmar a hipótese H3, já que são poucos em relação ao universo de comentários sobre os recursos de apoio. Em contrapartida, foram 7 os comentários sobre a pouca estrutura de recursos humanos, e 13 sobre a pouca estrutura de TI – Tecnologia da Informação, com possível impacto negativo no acesso à informação. Os entrevistados salientam as dificuldades de se publicar dados abertos com poucos recursos mas não estabelecem relação causal com a dificuldade de se obter acesso à informação. Entretanto, é razoável pensar que seja mais difícil obter acesso à informação de dados abertos publicados por estruturas com recursos escassos. Em adição, o número de reformas em crescimento, poucos ou nenhum concurso público para reposição de funcionários e fuga para a iniciativa privada e para o trabalho no estrangeiro são mencionados como um fator de diminuição da capacidade em prestar serviços, com impacto nos programas de dados abertos.

Liderança

Dos 27 comentários feitos sobre liderança, 10 identificaram o papel dos dirigentes apenas como formal, em que aprovam os planos e estabelecem comitês ou delegam o comando de programas a áreas como as ouvidorias. Esse comportamento neutro dos dirigentes sobre a liderança em dados abertos também já havia sido identificado no questionário. Ainda, 4 entrevistados relataram atraso nos programas de dados abertos e pouco envolvimento dos dirigentes pela falta de ação ou delegação. Já 9 pessoas, responsáveis por 11 comentários, salientaram a participação positiva dos dirigentes na condução direta dos programas de dados abertos, porém apenas 1 comentário foi feito no sentido de que os dirigentes incentivam diretamente o uso dos dados. Assim, não houve elementos para se verificar a hipótese H6 – a liderança influencia o acesso à informação.

Disseminação da Informação

Em um universo de 26 comentários, 7 foram de que falta divulgação dos dados abertos. Entre os motivos, coloca-se realizar consulta pública sem se saber quais dados estão disponíveis, divulgação somente para público interno e em 2 casos ausência de comunicação. Em contrapartida, 3 entrevistados relataram ações que efetivamente promoveram o acesso à informação, como consultas públicas bem sucedidas, construção de painéis de dados abertos concursos de dados que gerou ótima experiência de interação com a comunidade consumidora de dados. Este fato reforça a hipótese H1 – a disseminação da informação influencia o acesso à informação.

Institucionalização

A variável Institucionalização, por agregar os comentários das categorias de processo, estrutura organizacional, planejamento, extração e proteção de dados e métricas, é a que reúne o maior número de comentários, 182. Foram feitas 17 observações que fazem referência direta à qualidade dos dados em razão da institucionalização, as quais versam sobre os programas de dados abertos serem sustentáveis e com processos definidos. Estes processos incluem a extração automática dos dados, a garantia de atualização dos dados e a elaboração de documentação que garanta a perenidade e eficiência dos programas. Especificamente, 1 entrevistado mencionou estar a elaborar processos para seus dados abertos de maneira a estabelecer modelo de maturidade em gestão de dados. Estas afirmações reforçam a hipótese H4 – A institucionalização influencia a Qualidade dos Dados.

No que se refere à influência da institucionalização no acesso à informação, houve 8 comentários que indicam esta relação, como já ter estrutura prévia de divulgação de dados, publicar para áreas finalísticas e ter metas ligadas à redução de pedidos de informação. Ainda, foram observadas a existência de processos de resposta a solicitação de dados estruturados e de processos de seleção de dados. Estas afirmações reforçam a hipótese H5 – a institucionalização influencia o acesso à informação.

Dimensões não mapeadas no estudo porém presentes nas entrevistas

As entrevistas foram conduzidas de forma semiestruturada, o que levou à descoberta de novas dimensões que não foram objeto do estudo e revisão literária. 8 entrevistados ligaram dados abertos a transparência ativa. Foi percebida uma conscientização sobre a importância de se abrir os dados, também por uma razão prática: especificamente 2 entrevistados ressaltaram que o aumento da transparência ativa faz diminuir os pedidos de acesso à informação, reduzindo a carga administrativa e conferindo maior produtividade aos departamentos. Ainda se pode observar uma preocupação dos dirigentes sobre o que será publicado, e um cuidado e análise anteriores às decisões de publicação, conforme relatado em 4 entrevistas.

A cultura organizacional foi associada a dados abertos com frequência, como um meio que pode tanto facilitar quanto dificultar os programas. Apenas 1 entrevistado reportou que a cultura do órgão está baseada em transparência de dados mesmo antes dos dados abertos, e 3 pessoas relataram a hierarquia e a complexidade dos órgãos como desafios para conduzir, implementar e manter os programas de dados abertos.

A CGU aparece como impulsionador e suporte técnico para formar os órgãos na disponibilização de dados e para cobrar a execução dos planos de dados abertos, conforme 4 relatos, o que indica um

avanço pela obrigatoriedade imposta pelo controle externo. Entretanto, 2 entrevistados informaram que o controle exercido pela CGU funcionou como incentivo para mudar de patamar em gestão de dados, citando exemplos de publicação de dados abertos em painéis de inteligência de dados - "*business intelligence",* e de criação de área de governança de dados.

Capítulo 6. Considerações Gerais

6.1 Discussão dos Resultados e Conclusões

O estudo procurou responder à pergunta de investigação : "em que medida o portal de dados abertos governamentais brasileiro atende aos requisitos da política de transparência de dados? São aplicados os princípios de dados tais como os encontramos na literatura?"

A revisão da literatura indicou que, nos programas de dados abertos, quando os princípios são aplicados nas ações de gestão dos governos para implementar uma política de transparência de dados, resultados são produzidos para o sucesso destes programas, nomeadamente a promoção do acesso à informação e a entrega de qualidade nos dados à sociedade. Esta dissertação teve como objetivo geral compreender a existência destas relações, aplicadas ao portal de dados abertos brasileiro, e as hipóteses específicas foram daí derivadas.

O modelo conceptual para análise dos dados foi construído com base nas dimensões identificadas na revisão de literatura, e é composto de 2 variáveis dependentes - qualidade dos dados e acesso à informação - e 4 independentes - disseminação da informação, recursos de apoio, institucionalização e liderança. Desse modelo foram derivadas as hipóteses H1 a H6.

Conforme apresentado no capítulo de Análise de Dados, as análises de regressão múltipla realizadas indicam que a qualidade dos dados é influenciada positivamente pela institucionalização (H4) e pelos recursos de apoio (H2), e o acesso à informação é influenciado positivamente pela disseminação da informação (H1) e pela institucionalização (H5). As observações advindas das entrevistas também indicam a confirmação destas hipóteses, conforme detalhado no capítulo anterior.

As hipóteses H3 – o uso de recursos de apoio influencia o acesso à informação e H6 – a liderança influencia o acesso à informação - não puderam ser testadas devido à indisponibilidade de dados secundários sobre liderança e recursos de apoio. A validade destas hipóteses foi então explorada na análise dos dados das entrevistas. Em um universo de 20 entrevistas, para a hipótese H3 encontrou-se indícios de estar correta através de 3 depoimentos, mas não foi possível confirmar a mesma. Já para a hipótese H6, apenas um entrevistado mencionou sua ocorrência, podendo ser um caso isolado.

O estudo indicou que mais ações de gestão devem influenciar no resultado dos programas de dados abertos. Os coeficientes de determinação dos 2 modelos de regressão – para o questionário e para os dados secundários, ficaram em torno de 30%, o que indica uma influência moderada das variáveis independentes. Os programas de dados abertos, conforme de pode aferir das entrevistas, também são impactados por recursos humanos, restrições orçamentárias, aspetos da cultura organizacional e mudanças de estrutura e comando.

A ação dos dirigentes, da dimensão Liderança, não foi significativa no modelo de regressão para a variável dependente qualidade. Nas respostas ao questionário, a liderança aparece com perceção positiva no aspeto formal, quando da aprovação dos planos, mas neutra quanto à condução dos programas e à motivação dos funcionários. As entrevistas também confirmaram essa neutralidade de perceção a respeito do papel dos dirigentes.

O questionário indicou que a existência de unidade com equipa competente e de processos para tratar dados abertos parece estar consolidada. Porém, aparece como pouco frequente a participação da sociedade civil nestas estruturas, assim como orçamentos dedicados a dados abertos não são comuns. Faltam técnicos de TI e investimento em sistemas que possam melhorar em geral o resultado de programas de dados abertos. A baixa automação das extrações obriga a processos manuais de verificação de dados o que tende a criar bases de dados desatualizadas.

Os princípios de dados abertos que versam sobre a qualidade dos dados são efetivamente seguidos no que diz respeito aos formatos e à sua disponibilidade para uso. Entretanto, ferramentas de visualização para facilitar o uso e dados não agregados são menos frequentes.

As divulgações sobre dados abertos são feitas maioritariamente nos canais tradicionais – site do órgão, redes sociais – e poucas ações interativas, de maior impacto potencial como concursos ou seminários, são realizadas.

Ao final, após as análises estatísticas e qualitativas, entende-se que a pergunta de investigação foi respondida de maneira afirmativa. Os princípios de dados abertos de forma geral são aplicados no portal de dados brasileiro, e o portal implementa a política de transparência de dados ao, através das ações de gestão, promover acesso e qualidade dos dados na entrega. Entretanto, restam desafios a superar em recursos humanos, em automação e em tornar efetiva a liderança como impulsionador dos programas.

6.2 Limitações do Estudo

Ao estudar de forma mais aprofundada o assunto de dados abertos, percebe-se na literatura ser este um tema complexo: observa-se múltiplos fatores de influência e de sucesso nos programas assim

como uma multiplicação de iniciativas dos governos de vários países. Assim, faz-se necessário enquadrar, no âmbito temporal de uma dissertação, um escopo limitado.

Então, este estudo foi limitado aos princípios e práticas de gestão aplicados ao portal brasileiro. Outras variáveis presentes na literatura como a influência da cultura organizacional ou os impactos das recentes leis de proteções de dados não foram incluídas no estudo, embora tenham aparecido nas entrevistas e sido registadas.

6.3 Pistas para Investigações Futuras

A literatura sobre dados abertos é vasta e identifica temas importantes que podem estar a determinar uso de dados abertos e o sucesso destes programas. Estudar os dados abertos pela ótica da procura, de forma a confrontar as perceções dos produtores com as dos consumidores pode ser um caminho de investigação que permita conhecer como os dados são consumidos. O desafio a ser vencido, entretanto, é identificar os consumidores de dados abertos para que sejam estudados.

Alguns aspetos surgidos de forma espontânea nas entrevistas forneceram linhas de investigação que poderiam ser exploradas. Por exemplo, estudar o efeito das reformas, que estão a reduzir o número de funcionários na capacidade de prestação de serviço público em geral e especificamente nos programas de dados abertos.

Em um contexto mais abrangente, no qual dados abertos comportam-se como um ecossistema que envolve a sociedade e os governos, uma sugestão é estudar os impactos económicos gerados pelas aplicações produzidas a partir dos portais de dados abertos e avaliá-los contra os investimentos e despesas realizados pelos governos nestes programas. Ainda, pode-se procurar compreender a relação entre diferentes estágios de democracia nos países e a presença e uso de dados abertos nestas sociedades.

6.4 Contributo do Estudo

Espera-se que a investigação realizada possa ser replicada para portais de dados abertos de outros países ao aproveitar o mesmo método – inquérito – dados secundários – entrevistas. Esta dissertação almejou aumentar o conhecimento sobre as práticas de gestão e os desafios a serem vencidos em programas de dados abertos. Finalmente, pretendeu identificar as ações necessárias para promoção da eficácia destes programas pelo confronto da realidade com as melhores práticas identificadas na literatura.

Capítulo 7. Referências Bibliográficas

Alexopoulos, C., Loukis, E., Mouzakitis, S., Petychakis, M., & Charalabidis, Y. (2015). Analysing the Characteristics of Open Government Data Sources in Greece. In *Journal of the Knowledge Economy* (Vol. 9, Issue 3). https://doi.org/10.1007/s13132-015-0298-8

Alzamil, Z. S., & Vasarhelyi, M. A. (2019). A new model for effective and efficient open government data. *International Journal of Disclosure and Governance*, *0123456789*. https://doi.org/10.1057/s41310-019-00066-w

Araujo, J. F. F. E. de, & Tejedo-Romero, F. (2016). Local government transparency index: determinants of municipalities' rankings. *International Journal of Public Sector Management*, *29*(4), 327–347. https://doi.org/10.1108/IJPSM-11-2015-0199

Araujo, J. F. (2004). A Reforma da Gestão Pública: do mito à realidade. *Actas Do Seminário Internacional Luso-Galaico*, 31–41. https://repositorium.sdum.uminho.pt/bitstream/1822/3312/1/ConeferenicaIGAP2004.pdf

Attard, J., Orlandi, F., Scerri, S., & Auer, S. (2015). A systematic review of open government data initiatives. *Government Information Quarterly*, *32*(4), 399–418. https://doi.org/10.1016/j.giq.2015.07.006

Barry, E., & Bannister, F. (2014). Barriers to open data release: A view from the top. *Information Polity*, *19*(1–2), 129–152. https://doi.org/10.3233/IP-140327

Berners-Lee, T. (2006). *Linked Data 5 star model*. W3C. https://www.w3.org/DesignIssues/LinkedData.html

Biaggi, M. L. S. de, Medvid, M., & Assis, C. M. C. de. (2017). Fator de inflação da variância e regressões auxiliares para diagnóstico do problema de multicolinearidade nos modelos de regressão. *Cobreap Xix*. https://ibape-nacional.com.br/biblioteca/wp-content/uploads/2017/08/025.pdf

Boone, H. N., & Boone, D. A. (2012). Analyzing Likert Data. *Journal of Extension*, *50*(2). https://doi.org/10.1007/s11172-017-1908-3

Brandsma, G. J., Curtin, D., & Meijer, A. (2008). How transparent are EU 'Comitology'committees in practice? *European Law Journal*, *14*(6), 819–838. https://doi.org/10.1111/eulj.2008.14.issue-6

Bryman, A. (2012). *Social Research Methods* (4th ed.). Oxford.

Callender, J. C., & Osburn, H. G. (1977). A METHOD FOR MAXIMIZING SPLIT-HALF RELIABILITY COEFFICIENTS. *Educational and Psychological Measurement*, 819–825.

Carifio, J., & Perla, R. (2008). Resolving the 50-year debate around using and misusing Likert scales. *Medical Education*, *42*(12), 1150–1152. https://doi.org/10.1111/j.1365-2923.2008.03172.x

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed., Vol. 1, Issue 2). LAWRENCE ERLBAUM ASSOCIATES, PUBLISHERS.

Controladoria Geral da União. (2004). *Portal da Transparência*. Portal Da Transparência Na Internet. https://www.portaltransparencia.gov.br/sobre/o-que-e-e-como-funciona

Controladoria Geral da União. (2016a). *Portal Brasileiro de Dados Abertos*. https://dados.gov.br/pagina/dados-abertos. https://dados.gov.br/pagina/dados-abertos

Controladoria Geral da União. (2016b). *Sobre Dados Abertos*. https://dados.gov.br/pagina/sobre

Controladoria Geral da União. (2020). *Manual de Elaboração de Planos de Dados Abertos (PDAs)*. https://www.gov.br/cgu/pt-br/centrais-de-conteudo/publicacoes/transparencia-publica/arquivos/manual-pda.pdf

Controladoria Geral da União. (2021a). *5º Plano de Ação Nacional em Governo Aberto*. 60. https://www.opengovpartnership.org/wp-content/uploads/2022/01/Brazil_Action-Plan_2021-2023_PT.pdf

Controladoria Geral da União. (2021b). *Painel de Dados Abertos CGU*. Site Do Governo Federal Brasileiro - CGU.

Crosby, P. (1987). *Quality Process Improvement Management College*.

Cruz, C. F., Ferreira, A. C. de S., da Silva, L. M., & Macedo, M. Á. da S. (2012). Transparência da gestão pública municipal: Um estudo a partir dos portais eletrônicos dos maiores municiápios brasileiros. *Revista de Administracao Publica*, *46*(1), 153–176. https://doi.org/10.1590/S0034-76122012000100008

Cuillier, D., & Piotrowski, S. J. (2009). Internet information-seeking and its relation to support for access to government records. *Government Information Quarterly*, *26*(3), 441–449. https://doi.org/10.1016/j.giq.2009.03.001

Dante, G. P. (1998). Gestión de información en las organizaciones: principios, conceptos y aplicaciones. *Centro de Información En Capacitación*.

Davenport, T. H., & Pruzak, L. (1998). *Working Knowledge : How Organizations Manage What They Know* (Vol. 2, Issue 5). Harvard Business School Press.

Dawes, S. S., Vidiasova, L., & Parkhimovich, O. (2016). Planning and designing open government data programs: An ecosystem approach. *Government Information Quarterly*, *33*(1), 15–27. https://doi.org/10.1016/j.giq.2016.01.003

Duarte, R. (2004). Entrevistas em pesquisas qualitativas. *Educar Em Revista*, *24*, 213–225. https://doi.org/10.1590/0104-4060.357

Gascó-Hernández, M., Martin, E. G., Reggi, L., Pyo, S., & Luna-Reyes, L. F. (2018). Promoting the use of open government data: Cases of training and engagement. *Government Information Quarterly*, *35*(2), 233–242. https://doi.org/10.1016/j.giq.2018.01.003

Guttman, L. (1945). A basis for analyzing test-retest reliability. *Psychometrika*, *10*(4), 255–282. https://doi.org/10.1007/BF02288892

Harrison, T. M., Burke, G. B., Cook, M., Cresswell, A., Helbig, N., Hrdinova, J., & Pardo, T. (2012). Open Government and E-government : Democratic Challenges from a Public Value Perspective. *Information Polity*, *17*, 83–97.

Hartog, M., Mulder, B., Spée, B., Visser, E., & Gribnau, A. (2014). Open data within governmental organisations. *EJournal of EDemocracy & Open Government*, *6*(1), 49–61. http://ezproxy.csu.edu.au/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=tsh&AN=99675140&site=ehost-live

Ho, A. T.-K., & McCall, B. (2016). *Ten Actions to Implement Big Data Initiatives: A Study of 65 Cities*. IBM Center for the Business of Government. http://businessofgovernment.org/report/ten-actions-

implement-big-data-initiatives-study-65-cities

IBGE. (2009). *Série Histórica Acesso à Internet 2005 a 2009*. Site Do IBGE - Instituto Brasileiro de Geografia e Estatística.

IBGE. (2021). *Pesquisa PNDA IBGE Acesso à Internet 2016 a 2019*. Site Do IBGE - Instituto Brasileiro de Geografia e Estatística. https://sidra.ibge.gov.br/tabela/7308#resultado

Jaeger, P. T., & Bertot, J. C. (2010). Transparency and technological change: Ensuring equal and sustained public access to government information. *Government Information Quarterly*, *27*(4), 371–376. https://doi.org/10.1016/j.giq.2010.05.003

Janssen, M, Charalabidis, Y., & Zuiderwijk, A. (2012). Benefits, Adoption Barriers and Myths of Open Data and Open Government. *Information Systems Management*, *29*(4), 258–268. https://doi.org/10.1080/ 10580530.2012.716740

Janssen, Marijn, Charalabidis, Y., & Zuiderwijk, A. (2012). Benefits, Adoption Barriers and Myths of Open Data and Open Government. *Information Systems Management*, *29*(4), 258–268. https://doi.org/10.1080/10580530.2012.716740

Jensen, C., & Meckling, H. (1976). Theory of the firm: managerial behavior, agency costs and ownership structure. *Journal of Finantial Economics*, *3*, 305–360.

Juran, J. M. (1974). *Jurans Quality Control Handbook* (F. M. Gryna (ed.); 4th ed.). Mc Graw Hill.

Kaasenbrood, M., Zuiderwijk, A., Janssen, M., de Jong, M., & Bharosa, N. (2015). Exploring the Factors Influencing the Adoption of Open Government Data by Private Organisations. *International Journal of Public Administration in the Digital Age*, *2*(2), 75–92. https://doi.org/10.4018/ijpada.2015040105

Kaufmann, D., & Bellver, A. (2005). Transparenting Transparency: Initial Empirics and Policy Applications. *SSRN Electronic Journal*, 1–72. https://doi.org/10.2139/ssrn.808664

Keene, O. N. (1995). The log transformation is special. *Statistics in Medicine*, *14*(8), 811–819. https://doi.org/10.1002/sim.4780140810

Kim, J. H., & Eom, S. J. (2019). The managerial dimension of open data success: Focusing on the open data initiatives in Korean local governments. *Sustainability (Switzerland)*, *11*(23). https://doi.org/10.3390/su11236758

Kučera, J., Chlapek, D., & Martin Nečaský. (2013). Open Government Data Catalogs: Current Approaches and Quality Perspective. *Springer*, *8061*.

Lourenço, R. P. (2015). An analysis of open government portals: A perspective of transparency for accountability. *Government Information Quarterly*, *32*(3), 323–332. https://doi.org/10.1016/j.giq.2015.05.006

Luna-reyes, L. F., Harrison, T. M., & Styrin, E. (2016). Open Data and Open Government : From Abstract Principles to Institutionalized Practices. *17th International Digital Government Research*, 76–85. https://doi.org/https://doi.org/10.1145/2912160.2912161

Máchová, R., & Lněnička, M. (2017). Evaluating the quality of open data portals on the national level. *Journal of Theoretical and Applied Electronic Commerce Research*, *12*(1). https://doi.org/10.4067/S0718-18762017000100003

Malamud, C., O´Reilly, T., Elin, G., Sifry, M., Holovaty, A., O´Neil, D. X., Migruski, M., Allen, S., Tauberer, J., Lessig, L., Newman, D., Geraci, J., Bender, E., Steinberg, T., Moore, D., Shaw, D., Needham, J., Hardi, J., Zuckerman, E., ... Swartz, A. (2007). *Open Government Data Principles*. Open Government Data Principles. https://opengovdata.org/

Matheus, R., & Janssen, M. (2020). A Systematic Literature Study to Unravel Transparency Enabled by Open Government Data: The Window Theory. *Public Performance and Management Review*, *43*(3), 503–534. https://doi.org/10.1080/15309576.2019.1691025

Meijer, A. (2015). E-governance innovation: Barriers and strategies. *Government Information Quarterly*, *32*(2), 198–206. https://doi.org/10.1016/j.giq.2015.01.001

Ministério das Comunicações. (2021). *Pesquisa Acesso à Internet*. Site Do Governo Federal Brasileiro. https://www.gov.br/mcom/pt-br/noticias/2021/abril/pesquisa-mostra-que-82-7-dos-domicilios-brasileiros-tem-acesso-a-internet

Miranda, R. C. da R. (1999). O uso da informação na formulação de ações estratégicas pelas empresas. *Ciência Da Informação*, *28*(3), 286–292. https://doi.org/10.1590/s0100-19651999000300006

Obama, B. H. (2009). Memorandum for Heads of Departments and Agencies. *The White House*, *open Government*, 1–2. https://obamawhitehouse.archives.gov/the-press-office/transparency-and-open-government

Open Knowledge Foundation. (2004). *Open Definition*. Open Knowledge Foundation. http://opendefinition.org/

Possamai, A. J., & de Souza, V. G. (2018). Lei de Acesso à Informação e dados abertos: possibilidades e desafios a partir da experiência federal. *Carta de Conjuntura FEE*, *12*(2), 9–10. http://carta.fee.tche.br/article/lei-de-acesso-a-informacao-e-dados-abertos-possibilidades-e-desafios-a-partir-da-experiencia-federal/

Possamai, A. J., & De Souza, V. G. (2020). Transparência e Dados Abertos Governamentais: Possibilidades e Desafios a Partir da Lei De Acesso À Informação. *Administração Pública e Gestão Social*, 523–528. https://doi.org/10.21118/apgs.v12i2.5872

Presidência da República. (1988). *Constituição Federal do Brasil*. Diario Oficial Da União. http://www.planalto.gov.br/ccivil_03/constituicao/constituicao.htm

Presidência da República. (2011). Lei de Acesso à Informação. *Diario Oficial Da União*, 1–13.

Presidência da República. (2016). *Decreto 8.777 de 2016*. Diario Oficial Da União. https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2016/decreto/D8777.htm

Purwanto, A., Zuiderwijk, A., & Janssen, M. (2019). Citizens' Motivations for Engaging in Open Data Hackathons. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics): Vol. 11686 LNCS*. Springer International Publishing. https://doi.org/10.1007/978-3-030-27397-2_11

Ruijer, E., Détienne, F., Baker, M., Groff, J., & Meijer, A. J. (2020). The Politics of Open Government Data: Understanding Organizational Responses to Pressure for More Transparency. *American Review of Public Administration*, *50*(3), 260–274. https://doi.org/10.1177/0275074019888065

Ruijer, E., & Meijer, A. (2020). Open Government Data as an Innovation Process: Lessons from a Living Lab Experiment. *Public Performance and Management Review*, *43*(3), 613–635.

https://doi.org/10.1080/15309576.2019.1568884

Safarov, I. (2019). Institutional Dimensions of Open Government Data Implementation : Evidence from the Netherlands, Sweden and the UK. *Public Performance & Management Review*, *0*(0), 1–24. https://doi.org/10.1080/15309576.2018.1438296

Secretaria de Gestão e Desempenho de Pessoal. (2021). *Série Relatórios Gerenciais Servidores classificados por gênero Poder Executivo Federal*. https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&ved=2ahUKEwjVrL2yuLv2AhWYQfEDHS7GBr4QFnoECAUQAQ&url=https%3A%2F%2Fwww.gov.br%2Fservidor%2Fpt-br%2Fobservatorio-de-pessoal-govbr%2Fserie-relatorios-gerenciais-genero%2F%40%40edit%2F%2B%2Bwidget

Silvestre, H. C., & Araújo, J. F. (2012). *Metodologia para a investigação social*. Escolar Editora.

Silvestre, M. J., Fialho, I., & Saragoça, J. (2010). Da palavra à construção de conhecimento. *V. 3: Atas - Investigação Qualitativa Em Ciências Sociais*, *3*, 321–330.

Souza, A. C. de, Alexandre, N. M. C., & Guirardello, E. de B. (2017). Propriedades psicométricas na avaliação de instrumentos: avaliação da confiabilidade e da validade. *Epidemiologia e Servicos de Saude : Revista Do Sistema Unico de Saude Do Brasil*, *26*(3), 649–659. https://doi.org/10.5123/S1679-49742017000300022

Styrin, E., Luna-Renyes, L., & Harrisson, T. M. (2017). Transforming government: People, process, and policy. *University of Newcastle*, *1*(4), 5–6. https://doi.org/10.1108/tg.2007.32601daa.001

Tai, K. (2021). Open government research over a decade : A systematic review. *Government Information Quarterly*, *38*(2), 101566. https://doi.org/10.1016/j.giq.2021.101566

Tejedo-Romero, F., & Araujo, J. F. F. E. (2018). Transparency, Social Responsibility and Corporate Governance: Human capital of companies. *Cuadernos de Gestion*, *18*(2), 133–162. https://doi.org/10.5295/cdg.160693ft

Ubaldi, B. (2013). Open Government Data: Towards Empirical Analysis of Open Government Data Initiatives. *OECD Working Papers on Public Governance*, *NO.22*(22), 61. https://doi.org/10.1787/5k46bj4f03s7-en

UCLA Statistical Consulting Group. (2021). *Poisson Regression - R Data Analysis Examples*. https://stats.oarc.ucla.edu/r/dae/poisson-regression/

UN. (2020). *E-Government Survey 2020 - Digital Government in the Decade of Action for Sustainable Development: With addendum on COVID-19 Response*. https://publicadministration.un.org/egovkb/en-us/Reports/UN-E-Government-Survey-2020

Vetrò, A., Canova, L., Torchiano, M., Minotas, C. O., Iemma, R., & Morando, F. (2016). Open data quality measurement framework: Definition and application to Open Government Data. *Government Information Quarterly*, *33*(2), 325–337. https://doi.org/10.1016/j.giq.2016.02.001

Viana, T. (2001). Dado , Informação , Conhecimento e Competência. *Editora Escrituras*, *10*.

Wang, H. J., & Lo, J. (2016). Adoption of open government data among government agencies. *Government Information Quarterly*, *33*(1), 80–88. https://doi.org/10.1016/j.giq.2015.11.004

Wang, V., & Shepherd, D. (2020). Exploring the extent of openness of open government data – A critique of open government datasets in the UK. *Government Information Quarterly*, *37*(1), 101405.

https://doi.org/10.1016/j.giq.2019.101405

Weston, C., Gandell, T., Beauchamp, J., Mcalpine, L., Wiseman, C., & Beauchamp, C. (2001). Analyzing Interview Data : The Development and Evolution of a Coding System. *Qualitative Sociology*, *24*(3).

Yang, T. M., & Wu, Y. J. (2016). Examining the socio-technical determinants influencing government agencies' open data publication: A study in Taiwan. *Government Information Quarterly*, *33*(3), 378–392. https://doi.org/10.1016/j.giq.2016.05.003

Yang, T. M., & Wu, Y. J. (2021). Looking for datasets to open: An exploration of government officials' information behaviors in open data policy implementation. *Government Information Quarterly*, *38*(2), 101574. https://doi.org/10.1016/j.giq.2021.101574

Yang, Z., Sun, J., Zhang, Y., & Wang, Y. (2015). Understanding SaaS adoption from the perspective of organizational users: A tripod readiness model. *Computers in Human Behavior*, *45*, 254–264. https://doi.org/10.1016/j.chb.2014.12.022

Zuiderwijk, A., & Janssen, M. (2014a). Barriers and Development Directions for the Publication and Usage of Open Data: A Socio-Technical View. *M. Gascó-Hernández (Ed.), Open Government, Public Administration and Information Technology - Springer Science+Business Media New York*, 1–216. https://doi.org/10.1007/978-1-4614-9563-5

Zuiderwijk, A., & Janssen, M. (2014b). Open data policies, their implementation and impact: A framework for comparison. *Government Information Quarterly*, *31*(1), 17–29. https://doi.org/10.1016/j.giq.2013.04.003

Zuiderwijk, A., Janssen, M., Choenni, S., & Meijer, R. (2014). Transforming Government: People, Process and Policy Design principles for improving the process of publishing open data. *Iss Process and Policy*, *8*(3), 185–204. https://doi.org/10.1108/TG-07-2013-0024

Zuiderwijk, A., Romer, M., & Kroesen, M. (2019). Open data policy development: How can municipalities take account of residents' perspectives? *Proceedings of the 15th International Symposium on Open Collaboration, OpenSym 2019*. https://doi.org/10.1145/3306446.3340819