

# Como a Governança de Dados pode auxiliar na mitigação de barreiras de uso de Portais de Dados Governamentais Abertos? Uma Análise da Literatura

Isela Mendoza<sup>1</sup>, Renan Corrêa<sup>1</sup>, Flavia Bernardini<sup>1</sup>

<sup>1</sup>Instituto de Computação – Universidade Federal Fluminense – Niterói, RJ – Brasil

{imendoza,renancorrea}@id.uff.br, fcbernardini@ic.uff.br

**Abstract.** *Open Government Data (Open Government Data – OGD) portals have shown rapid growth in recent years and increased several problems. To mitigate these problems, as an alternative, Data Governance (Data Governance – DG) is presented in the studies of this research. The main objective of this work is to carry out a literature review to discuss how DG can support the mitigation of the main problems and barriers to use presented in OGD portals. We could initially analyze that there are still problems in OGD portals. Approaches to mitigating these are discussed, as well as good practices to adopt and lessons learned.*

**Resumo.** *O uso de portais de Dados Governamentais Abertos (Open Government Data – OGD) tem apresentado um crescimento acelerado nos últimos anos e com ele também aumentaram vários problemas. Para mitigar esses problemas, como alternativa, a Governança de Dados (Data Governance – DG) é apresentada nos estudos desta pesquisa. O principal objetivo deste trabalho consiste em realizar uma análise da literatura para discutir como a DG pode apoiar na mitigação dos principais problemas e barreiras de uso apresentados em portais de OGD. Pudemos analisar inicialmente que ainda há a presença de problemas em portais de OGD. As abordagens para mitigá-los são discutidas, bem como as boas práticas a serem adotadas e lições aprendidas.*

## 1. Introdução

Os portais de Dados Abertos Governamentais (OGD, do inglês *Open Government Data*) são plataformas para apoiar a iniciativa de dados abertos dos governos. Tais portais facilitam as organizações governamentais a publicar seus conjuntos de dados em formatos abertos para uso público gratuito. Nos últimos anos, o número de portais de OGD tem apresentado um crescimento vertiginoso, o catálogo do portal de dados [DataPortals.org 2023] mostra uma lista abrangente de portais de dados abertos de todo o mundo somando ao todo um total de 598 portais até atualidade, nos últimos 5 anos este número aumentou em 74 portais desde o estudo feito em [Reis et al. 2018] onde foram registrados 524 portais. Com este aumento, cresceram também os problemas que afetam o uso dos dados abertos, como acessibilidade, qualidade e compreensão dos mesmos [Tygel et al. 2016]. Em [Reis et al. 2018] os autores apresentam uma justificativa para a Governança de Dados (DG, do inglês *Data Governance*) como uma abordagem para enfrentar problemas recorrentes em portais de OGD. Para apresentar tal abordagem, os

autores apresentam uma análise da literatura sobre diversos trabalhos que abordam os problemas típicos em portais de OGD e fazem a análise de como boas práticas de DG podem mitigá-los.

DG refere-se à totalidade dos direitos de decisão e responsabilidades em relação ao gerenciamento de ativos de dados. São processos, normas, políticas de tecnologia necessárias para gerenciar, manter e explorar a informação como recurso de negócios [Newman and Logan 2006] [Oracle 2011]. Quando uma empresa decide implementar boas práticas de DG, ela deve decidir qual *framework* se encaixa melhor à sua realidade. Um *framework* de DG é uma estrutura lógica para classificar, organizar e comunicar atividades complexas envolvidas na tomada de decisões, e na execução de ações, sobre dados corporativos [Reis et al. 2018]. No entanto, não é uma tarefa trivial decidir qual *framework* escolher, bem como implementá-lo.

O objetivo do presente estudo consiste em realizar uma análise da literatura sobre os principais problemas que enfrentam atualmente os portais de OGD, como eles podem ser mitigados por meio de uma abordagem de DG, assim como a discussão de boas práticas a serem adotadas e lições aprendidas. Para atender ao nosso objetivo, apresentamos alguns trabalhos citados em uma publicação prévia do nosso grupo de pesquisa eles são [Braunschweig et al. 2012] [Chattapadhyay 2014] [Fioretti 2011] [Gurstein 2011] [Vetrò et al. 2016] [Zuiderwijk et al. 2014], tomando como base o nosso trabalho previamente publicado, estendemos tal estudo por meio da técnica de *forward snowballing*, que consiste em uma busca para frente, para procurar pelas citações de um artigo trabalhos relacionados a um determinado tema. Apresentamos então, como resultado deste processo, os problemas e barreiras recorrentes encontradas na literatura. A partir desses achados, trazemos um alinhamento dos problemas com domínios de decisão de DG. Tal alinhamento é importante para refletir quais áreas da GD devem ter mais enfoque dependendo dos problemas serão tratados dada alguma ordem de prioridade. Por fim, realizamos uma discussão ampliada de lições aprendidas e atividades que podem ser realizadas por um escritório de GD, com base em outros trabalhos publicados pelo nosso grupo de pesquisa. Entendemos que tal discussão apresentada neste trabalho é de fundamental importância para os órgãos governamentais, incluindo prefeituras municipais, pois o ecossistema de dados abertos em tais órgãos cresce cada vez mais, com distintos portais de dados e serviços para a população e muitas leis existentes que devem ser cumpridas.

O presente trabalho está organizado como descrito a seguir: Na Seção 2 são listados os principais problemas identificados como barreiras no uso de portais de OGD em trabalhos publicados anteriormente no nosso grupo de pesquisa, onde sugere o uso do DG para mitigar tais problemas. Nessa mesma seção, apresentamos também o planejamento e execução de nossa metodologia de pesquisa, o *forward snowballing*, com o intuito de de revelar mais problemas, tanto novos como recorrentes, que persistem nos portais de OGD até atualidade. Na Seção 3 é apresentada uma discussão sobre a relevância do uso do DG para mitigar problemas em portais de OGD, analisando o *framework* de Khatri e Brown [Khatri and Brown 2010]. Na Seção 4, são discutidas boas práticas e abordagens para mitigar problemas em portais de OGD, encontradas tanto na literatura como as desenvolvidas em estudos conduzidos pelo nosso grupo de pesquisa. Por fim, na Seção 5, são apresentadas as considerações finais de nosso análise, assim

como os trabalhos futuros e os problemas que ficaram em aberto.

## 2. Barreiras encontradas no uso de portais de ODG

Os principais problemas identificados em trabalhos publicados anteriormente no nosso grupo de pesquisa [Reis et al. 2018] foram resumidos e categorizados da seguinte forma:

- **Metadados pobres:** Ausência ou má qualidade dos metadados (informações sobre conjuntos de dados) [Braunschweig et al. 2012] [Vetrò et al. 2016] [Zuiderwijk et al. 2014].
- **Falta de padrão de dados:** Ausência de padronização das colunas ou atributos, unidades de medida, dados, etc [Braunschweig et al. 2012] [Gurstein 2011] [Zuiderwijk et al. 2014].
- **Problemas de formatos de dados:** Falta de padronização nos formatos de conjuntos de dados. Opção de poucos formatos não proprietários [Braunschweig et al. 2012] [Chattapadhyay 2014] [Vetrò et al. 2016] [Zuiderwijk et al. 2014].
- **Falta de gestão do ciclo de vida dos dados:** Ausência ou falta de gestão da qualidade da vida útil dos dados. Falta de atualização da política, arquivamento e exclusão de conjuntos de dados [Chattapadhyay 2014] [Fioretti 2011] [Vetrò et al. 2016].
- **Falta de completude de dados:** Conjuntos de dados incompletos, dados ausentes, atributos relevantes faltantes, dados preenchidos de forma incorreta [Chattapadhyay 2014] [Vetrò et al. 2016] [Zuiderwijk et al. 2014].
- **Problemas de licença de dados:** Falta de licença de dados abertos, especificando se o uso pode ser comercial, se pode ser compartilhado, divulgado [Chattapadhyay 2014] [Fioretti 2011] [Vetrò et al. 2016] [Zuiderwijk et al. 2014].
- **Dificuldade de acesso e disponibilidade de dados:** Dificuldade de acesso dos usuários ao portal, seja ao tentar identificar os conjuntos de dados que desejam, ou como encontrar o portal por meio de mecanismos de pesquisa, links quebrados, disponibilidade do portal online, etc [Chattapadhyay 2014] [Vetrò et al. 2016] [Zuiderwijk et al. 2014].
- **Reconhecimento da finalidade dos dados:** Conjuntos de dados sem finalidade de uso, dados irrelevantes, sem utilidade para os usuários [Chattapadhyay 2014] [Fioretti 2011] [Gurstein 2011].
- **Coleta de dados pobre:** Poucos conjuntos de dados, pouca variedade de dados em categorias, áreas, etc [Vetrò et al. 2016].
- **Falta de rastreabilidade da fonte de dados:** Falta de informações sobre a fonte dos dados, para verificar a veracidade dos dados [Vetrò et al. 2016].
- **Dados incongruentes:** Falta de consistência nos dados, dados que não refletem a realidade [Chattapadhyay 2014].
- **Má compreensão:** Conjuntos de dados que apresentam algum conteúdo que é difícil de ser interpretado por usuários ou desenvolvedores, ou seja, conteúdo que requer conhecimento avançado, não acessível para todos os usuários [Gurstein 2011] [Vetrò et al. 2016] [Zuiderwijk et al. 2014].

Para mitigar tais problemas, propusemos uma abordagem que consiste na adoção de um *framework* de DG para auxiliar na gestão eficaz de dados públicos. Na maioria

dos trabalhos apresentados é possível identificar um padrão de estrutura pela DG. Assim, apresentamos também uma discussão que alinhava os principais problemas encontrados nos portais de OGD com os cinco domínios de decisão na DG e aplica o framework da Notre Dame [Chapple 2013] no escopo do portal de dados abertos. Com a utilização da metodologia de DG, é possível gerenciar todas as etapas importantes do gerenciamento de dados, desde a concepção até a entrega ao cliente, neste caso a população.

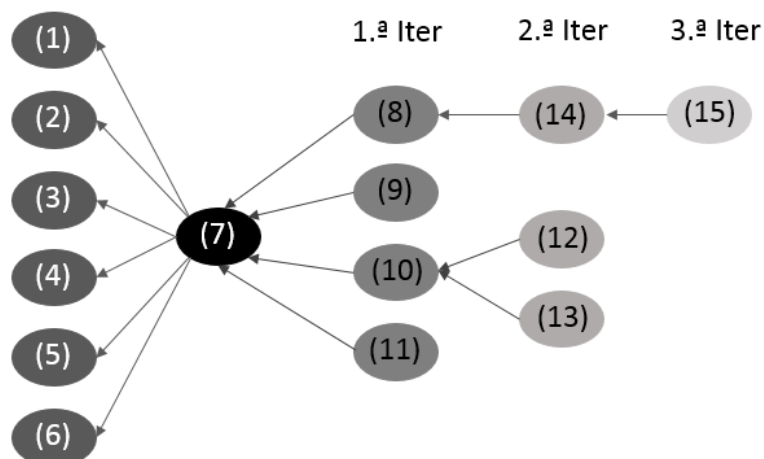
A partir de tal trabalho, realizamos um processo do *forward snowballing*. Para selecionar os trabalhos, utilizamos os seguintes Critérios de Inclusão (CI): foram usados para selecionar os artigos que vão formar parte da pesquisa seguido da descrição dos filtros aplicados; (CI1) O idioma dos estudos deve ser o português, inglês e espanhol; (CI2) Estudos devem estar publicados em conferências, workshop e jornais; (CI3) Estudos devem estar relacionados ao tema do artigo principal (dados governamentais abertos); (CI4) Estudos que identifiquem problemas em portais de OGD e usem DG para mitigá-los. Também, utilizamos os seguintes Critérios de Exclusão (CE): (CE1) Teses, relatórios técnicos, powerpoint, resumos de procedimentos e fontes irrelevantes (conhecidos como “white paper”); (CE2) Estudos duplicados ou indisponíveis; (CE3) Estudos totalmente fora da área e tema de pesquisa do trabalho principal. Por fim, aplicamos também dois filtros (F): (F1) Seleção dos artigos pela leitura dos títulos e resumos; e (F2) Seleção dos artigos finais e definitivos que formam parte da pesquisa pela sua leitura completa. Em ambos os filtros são aplicados os critérios de inclusão e exclusão antes mencionados. no qual identificamos 4 novos problemas em portais de OGD, além da reiteração e presença de outros já existentes.

Seguindo a metodologia apresentada, foram coletados inicialmente 16 artigos na primeira iteração do *forward snowballing*. Aplicando os critérios de inclusão e exclusão e os filtros de seleção, obtivemos 4 novos artigos na primeira iteração a serem analisados na pesquisa. Na segunda iteração, coletamos 13 trabalhos sobre os artigos resultantes da primeira iteração e, após a aplicação dos critérios de exclusão e inclusão e filtros, restaram 3 novos artigos. Por último, aplicamos mais uma vez o processo e, de dois trabalhos coletados, ficamos com mais um artigo, totalizando no final 8 novos artigos para nossa análise. Na Tabela 1 é resumido o processo para cada iteração o número de trabalhos coletados (# Art. colet.) e selecionados (# Art. selec.).

**Tabela 1. Resultados das iterações forward snowballing**

<b>Iterações</b>	<b>1.<sup>a</sup> Iter</b>	<b>2.<sup>a</sup> Iter</b>	<b>3.<sup>a</sup> Iter</b>	<b>Total</b>
<b># Art. colet./# Art. selec.</b>	16/4	13/3	2/1	31/8

A Figura 1 mostra um mapa de citações partindo do artigo principal (7) [Reis et al. 2018] no círculo centro da imagem. Os artigos representados por círculos à esquerda de (7) são os artigos relacionados que foram citados no artigo principal, que são: (1) [Braunschweig et al. 2012], (2) [Chattapadhyay 2014], (3) [Fioretti 2011], (4) [Gurstein 2011], (5) [Vetrò et al. 2016] e (6) [Zuiderwijk et al. 2014]. Os artigos representados por círculos nos níveis 1, 2 e 3 à direita de (7) são os artigos que foram recuperados na execução do processo de forward snowballing e suas respectivas iterações, eles são: na 1.<sup>a</sup> Iter (8) [Lněnička et al. 2021], (9) [Cadena-Vela et al. 2020], (10) [Cadena-Vel et al. 2019] e (11) [Rangel et al. 2020]. Na 2.<sup>a</sup> Iter foram obtidos os artigos: (14) [Nikiforova and Lnenicka 2021], (12)



**Figura 1. Mapa de citações**

[Martínez et al. 2021] e (13) [Rangel-Carrillo et al. 2020]. Por último na 3.ª Iter temos o artigo (15) [Masoumi et al. 2022].

Os autores [Lněnička et al. 2021] apresentam os problemas de falta de Transparência e de rastreabilidade da fonte de dados, e propõe para mitigá-lo um fornecimento de mecanismos de transparência relacionados a recursos e capacidades relevantes para as partes interessadas. O artigo do [Cadena-Vela et al. 2020] sugere que para os problemas de má compreensão, metadados pobres ou ausentes, um framework para considerar metadados de fontes de dados dentro de uma abordagem de gerenciamento de dados mestre, já [Cadena-Vel et al. 2019] sugere para dados incongruentes, reconhecimento da finalidade dos dados, falta de completude de dados e má compreensão, um framework para publicação de dados abertos com um nível adequado de qualidade. Por outro lado, [Rangel et al. 2020] propõe a realização de uma categorização automática de conjuntos de dados em portais de OGD utilizando aprendizado de máquina supervisionado dado o problema de para a falta de categorização de dados. No estudo feito pelo [Martínez et al. 2021], para a falta de completude de dados, dados incongruentes e dados redundantes é proposta uma ferramenta de validação: Open Data Quality Validation Tool (HEVDA), que implementa métricas de qualidade que medem o conjunto de dados selecionados. No estudo do [Rangel-Carrillo et al. 2020] para dados incongruentes, dados irrelevantes e reconhecimento da finalidade propõe um modelo que permite desenvolver uma avaliação da qualidade de dados governamentais abertos para viabilizar, por meio de sua publicação e reutilização, os benefícios de converter a matéria-prima dos dados em produtos que se materializam em informações relevantes e coerentes. O autor [Nikiforova and Lnenicka 2021] para a dificuldade na reutilização de dados, apresentam uma abordagem multiperspectiva, isso é, uma análise a partir de: perspectiva do cidadão, do usuário, do especialista e do estado da arte, por meio do reaproveitamento de dados e conhecimentos que sustentem a ideia do reaproveitamento. Por último [Masoumi et al. 2022] sugere para a falta de transparência, dados pobres, incongruentes, irrelevantes, reconhecimento da finalidade e má compreensão uma abordagem sistemática, de cima para baixo, baseada em ontologia combinada com uma nova arquitetura e seus processos correspondentes. Os novos

problemas em portais de OGD encontrados nestes artigos são:

- **Falta de transparência:** Ausência total ou parcial de informação relevante entre duas partes interessadas [Cadena-Vela et al. 2020].
- **Falta de categorização de dados:** Quando os conjuntos de dados não estão distribuídos corretamente pelas diversas categorias que representam os tópicos abordados de um portal [Rangel et al. 2020].
- **Dados redundantes:** A redundância de dados é uma situação que ocorre dentro dos sistemas de banco de dados e envolve a criação não intencional de dados duplicados que não são necessários. Pode ter um efeito adverso no funcionamento do sistema, retornando dados que são menos úteis, afetando a reutilização dos dados e a interoperabilidade [Martínez et al. 2021].
- **Dificuldade na reutilização de dados:** Quando a fonte dos dados não está disponível nem utilizável, dificultando a reutilização dos dados para o maior número possível de partes interessadas [Nikiforova and Lnenicka 2021].

### 3. DG para mitigar problemas de portais de OGD

Há diferentes definições de DG. Para Newman e Logan, a DG envolve um conjunto de decisões, processos, padrões, políticas e tecnologias necessárias para gerenciar, manter e explorar informações como um recurso corporativo [Newman and Logan 2006]. Já para a Oracle [Oracle 2011], a DG envolve a especificação de direitos de decisão e uma estrutura de responsabilidade para incentivar o comportamento desejável na avaliação, criação, armazenamento, uso, arquivamento e exclusão de dados e informações. Inclui os processos, funções, padrões e métricas que garantem o uso eficaz e eficiente de dados e informações para permitir que uma organização atinja seus objetivos. As boas práticas de DG define qual processo uma organização deve seguir para garantir que os dados existentes de uma organização tenham alta qualidade durante todo o seu ciclo de vida e deve evoluir dentro da empresa para alterar a forma de trabalho garantindo que as informações possam ser utilizadas por toda a organização. Por isso, é função da GD coordena o gerenciamento de dados.

Quando um *framework* de GD é implementado, ele deve ser capaz de entregar à organização políticas de dados, padrões de dados, problemas resolvidos, dados e informações de qualidade, valor de dados reconhecido (problemas de dados mapeados), projetos e serviços de gerenciamento de dados. Políticas de dados são coleções de declarações que controlam a integridade, segurança, qualidade e uso de dados durante seu ciclo de vida e mudança de estado. Padrões de dados incluem padrões de nomenclatura, padrões de modelagem de dados e outros padrões de arquitetura de dados [Oracle 2011].

Um *framework* de DG é em geral apresentado como uma estrutura lógica para classificar, organizar e comunicar atividades complexas envolvidas na tomada de decisões e ações sobre dados corporativos. Quando uma companhia decide implementar um *framework* de DG, ela deve decidir qual *framework* se adapta melhor à sua realidade. A adoção de uma estruturação de um escritório de governança de dados permite a colaboração de vários níveis das organizações para gerenciar dados em toda a organização e fornece a capacidade de alinhar vários programas relacionados a dados com objetivos corporativos [Cheong and Chang 2007]. Tal característica é particularmente importante em órgãos governamentais, como prefeituras, já que a estruturação em secretarias e órgãos detentores de dados é bastante segmentado.

Escolhemos para nossa discussão o *framework* de Khatri e Brown [Khatri and Brown 2010], devido à organização lógica que a DG é apresentada. Para os autores, a DG inclui cinco domínios de decisão inter-relacionados: princípios de dados, qualidade de dados, metadados, acesso a dados e ciclo de vida dos dados. Os princípios de dados explicam o papel dos dados como um ativo do negócio. Tal domínio estabelece até que ponto os dados são bons em toda a empresa e, portanto, quais políticas, normas e diretrizes específicas são apropriadas. Estabelece e promove também oportunidades de compartilhamento e reutilização de dados. Os princípios de dados de uma organização também levam em consideração o ambiente regulatório que pode influenciar os usos de dados nos negócios. O domínio qualidade dos dados estabelece os requisitos do uso pretendido dos dados. Embora a qualidade dos dados tenha várias dimensões, como precisão, pontualidade, integridade e credibilidade, essas dimensões são relativas e precisam ser definidas no contexto do uso final dos dados. Os metadados estabelecem a semântica ou conteúdo dos dados a serem interpretados pelos usuários. Ele explica o que são os dados e fornece o mecanismo que descreve consistentemente a representação dos dados, ajudando assim a interpretar o significado ou a semântica dos dados. Ao fornecer um conjunto de mapeamentos de uma linguagem de representação para concordar com conceitos no mundo real, os metadados conectam um banco de dados ao mundo real. O acesso aos dados tem como premissa a capacidade dos beneficiários dos dados de atribuir um valor a diferentes categorias de dados. Ele fornece padrões em nível físico e lógico. Os padrões de integridade de dados físicos garantem que os dados sejam imunes a danos físicos, como falta de energia. Padrões para integridade lógica de dados garantem que a estrutura de um banco de dados seja preservada. Por fim, o ciclo de vida dos dados é um conceito de compreensão de que todos os dados passam pelos estágios do ciclo de vida e é fundamental para projetar a governança de dados. Ao entender como os dados são usados e por quanto tempo devem ser retidos, as organizações podem desenvolver abordagens para mapear padrões de uso para a mídia de armazenamento ideal, minimizando assim o custo total de armazenamento de dados ao longo de seu ciclo de vida.

Podemos assim relacionar os problemas típicos de portais de dados abertos a tais domínios de DG. Na Tabela 2 são mostrados os problemas encontrados nos artigos analisados anteriormente e dos quais emergiram os 16 problemas frequentes encontrados na literatura, divididos nos domínios de decisão de governança de dados. Vale a pena notar que os princípios e a qualidade dos dados abrangem 11 dos 16 problemas encontrados (o problema Baixa compreensão dos dados pode ser tratado nos dois domínios). Portanto, podemos ver que se a governança de dados puder ser aplicada, ela permite um ganho em diferentes aspectos e ataca as principais dificuldades relatadas pelos autores citados.

#### **4. Boas práticas a serem adotadas e lições aprendidas**

Na maioria dos estudos que propõem mecanismos e *frameworks* para mitigar os problemas [Reis et al. 2018] [Lněička et al. 2021] [Cadena-Vela et al. 2020] [Cadena-Vel et al. 2019], é afirmado que é muito importante no futuro a implementação desses marcos em projetos reais em desenvolvimento, assim como realizar diversos estudos de caso (e não somente um), com o objetivo de criar indicadores para medir a eficácia das propostas. No caso do estudo que faz uma categorização automática dos dados utilizando aprendizado de máquina [Rangel et al. 2020], é observado uma

Domínios de decisão da DG	Problemas e barreiras em portais de OGD
Princípios de Dados	Dificuldade de Acesso aos dados Reconhecimento da finalidade dos dados Falta de rastreabilidade dos dados Baixa compreensão dos dados
Qualidade de Dados	Falta de padrão de dados Problemas de formatos de dados Falta de completude de dados Baixa compreensão dos dados Dados incongruentes Falta de transparência Dados redundantes Dificuldade na reutilização dos dados
Metadados	Metadados pobres
Acesso aos Dados	Falta de categorização de dados Problemas de licença de dados
Ciclo de Vida dos Dados	Falta de gestão do ciclo de vida dos dados

**Tabela 2. Domínios de decisão de DG de acordo com Kathri e Brown alinhados aos problemas e barreiras encontrados em portais de OGD**

limitação na extração de termos em casos em que o atributo de um arquivo de dados possui um nome que consiste de duas ou mais palavras, mostrando-se ineficiente. Sendo indicado como um possível trabalho futuro, inferir categorias do sub-conjunto abrangente para os conjuntos de dados referidos no trabalho após a geração do classificador proposto para em seguida avaliar se a classificação faz sentido ou não. No que diz respeito a qualidade dos dados, o estudo [Martínez et al. 2021] ressalta a importância de implementar mais métricas de qualidade e o desenvolvimento de ferramentas como HEVDA [Martínez et al. 2021], para a análises e detecção de problemas em conjuntos de dados. Melhorando assim, a qualidade do conteúdo e a interoperabilidade. Já no estudo [Rangel-Carrillo et al. 2020] os autores argumentam que a qualidade dos OGD deve ocorrer como um processo prévio à abertura dos dados, para agregar valor na criação e geração de novos modelos de negócios na esfera empresarial e, por sua vez, facilitar a inovação no setor público, para construir em ideias e melhores práticas do setor privado. Por outro lado, sobre o que refere ao estado da reutilização de dados abertos [Nikiforova and Lnenicka 2021] ainda requer de mais análise, enquanto a identificação de conjuntos de dados de alto valor. Por último, o estudo [Masoumi et al. 2022] indica que projetar o modelo de dados, bem como explorar uma abordagem baseada em ontologia, são pilares essenciais para criar conjuntos de dados ricos, relevantes e bem descritos.

A partir de tais alinhamentos, em um trabalho prévio [Reis et al. 2018], exploramos a adoção de boas práticas de GD para melhoria da qualidade de disponibilização de dados de contratos e licitações em um portal de transparência em um município brasileiro. Uma grande dificuldade encontrada foi a falta de um escritório de GD na prefeitura para apoiar tais atividades. Elaboramos então uma metodologia de atuação, por meio de modelagem As-Is dos processos de disponibilização dos dados, que muitos são realizados manualmente. Apresentamos assim propostas de atualização do



processo de disponibilização dos dados por meio de automação dos processos. Um grande fator dificultador foi que os processos de licitação e contratação são de secretarias que não gerem o portal da transparência. Assim, para além de aspectos tecnológicos, observamos que aspectos organizacionais precisam ser melhorados para a disponibilização de dados seja mais efetiva. Tais lições aprendidas reforçam a necessidade de um escritório de DG nas prefeituras municipais.

Além desses apontamentos encontrados na literatura, citamos a seguir diversas iniciativas acadêmicas e em parceria com a prefeitura municipal de Niterói que estão contextualizados no tratamento das barreiras dos portais de OGD e que descrevemos a seguir, bem como a maneira que podem ser apropriadas por um escritório de GD:

- Em [Barcellos et al. 2022] apresentamos uma discussão da dificuldade de definição do significado de baixa compreensão dos dados. Definições precisas são necessárias para estabelecimento de requisitos não-funcionais e funcionais de software. Apresentamos um *framework* para determinar a interpretabilidade de dados abertos, no qual definimos quais são aspectos importantes para melhoria da interpretabilidade. Os portais de OGD devem atender à lei de acesso à informação, porém também carecem de mecanismos para melhoria do entendimento dos dados por parte da população. Como o cenário de serviços de governo digital envolve perfis de usuários de muitos diversos tipos, diferentes mecanismos devem ser oferecidos, como mecanismos de recuperação dos dados, visualização dos dados, dentre outros. Tais tecnologias e tarefas também podem ser incorporadas por um escritório de GD;
- Em [Reis et al. 2022] apresentamos uma proposta de um mecanismo para verificação de completude de metadados conforme padrões estabelecidos na literatura. Tal mecanismo vai ao encontro do domínio de metadados e pode ser implementado em um escritório de GD;
- Em [Interian et al. 2022] apresentamos uma proposta de padronização de dados para diários oficiais, de maneira que os diários oficiais também sejam vistos como portais de OGD. Definição de padrões de disponibilização de dados e concepção de APIs para tal disponibilização são atividades de um escritório de GD para permitir melhor reuso dos dados;
- Em [Pinto et al. 2022], apresentamos uma proposta de alinhamento de categorias de portais de OGD para facilitar a integração de dados oriundos de diferentes portais de OGD. Tal mecanismo pode auxiliar na integração de diferentes portais que oferecem dados para a população, como portais de transparência, indicadores, dados abertos, dentre outros. Além disso, pode ainda facilitar a integração de portais de prefeituras municipais distintas, por exemplo, para apoiar o processo de tomada de decisão por parte dos gestores. Tal proposta vai ao encontro do problema da dificuldade na reutilização dos dados, o que também pode ser parte das atribuições de um escritório de GD;
- Por fim, em um projeto em parceria com a prefeitura municipal de Niterói, RJ, ao modelar o ecossistema de dados abertos da prefeitura, incorporando os diferentes portais de dados abertos (transparência, indicadores, dados georreferenciados e dados abertos), percebemos a falta de alinhamento entre os portais e a dificuldade das pessoas que buscam informações em saber onde encontrar o que procuram. Portanto, é importante que tal escritório de GD construa mecanismos

de interoperabilidade entre os serviços para apoiar os cidadãos a encontrarem informações.

## 5. Conclusões

Ao analisar os trabalhos citados, fundamentalmente os novos encontrados, pudemos observar que além dos problemas existentes em portais de OGD há mais 4 novos problemas que emergiram e eles são: falta de transparência [Cadena-Vela et al. 2020], falta de categorização de dados [Rangel et al. 2020], dados redundantes [Martínez et al. 2021] e dificuldade na reutilização de dados [Nikiforova and Lnenicka 2021] em portais de OGD que não foram mencionados em estudos anteriores, assim como a reiteração de problemas que já existiam. No entanto, apesar do trabalho de [Reis et al. 2018] apontar o uso de DG para mitigar problemas de portais de OGD, não observamos ainda nenhum estudo que aponte o uso de boas práticas de DG para mitigação dos problemas. Nosso trabalho discute o framework de Khatri e Brown [Khatri and Brown 2010], devido estrutura e a organização lógica que a DG é apresentada para a classificação e comunicação de atividades complexas envolvidas na a toma de decisões, analisando os cinco domínios de decisão inter-relacionados que são: princípios de dados, qualidade de dados, metadados, acesso a dados e ciclo de vida dos dados. Embora existam outros framework de DG e cada companhia é a que decide qual se adapta melhor à sua realidade, o framework de Khatri e Brown [Khatri and Brown 2010] mostra-se como um dos mais completos e o que mais se alinha aos problemas e barreiras encontrados em portais de OGD neste estudo.

Sobre os trabalhos futuros e problemas aberto nos trabalhos analisados, aponta a necessidade de validar suas abordagens, com o propósito de medir a eficácia das propostas, o que é uma tarefa complexa de longo prazo. Isso indica que a comunidade científica ainda tem muito trabalho pela frente para evoluir e aprimorar as abordagens utilizadas para mitigar os problemas existentes hoje em dia nos portais de OGD, assim como a proposta de novas soluções. Para isso, neste estudo foram apresentadas diversas iniciativas acadêmicas de nossa equipe de pesquisa em parceria com uma prefeitura municipal no contexto de tratamento das barreiras dos portais de OGD, ressaltando a importância de implementar escritórios de GD por parte de órgãos governamentais como boa prática, discussão que constitui uma das principais contribuições de nosso análises. Outro caminho possível é investigar o uso de tecnologias recentes de inteligência artificial e aprendizado de máquina para construção de métodos para essa validação.

## Referências

- Barcellos, R., Bernardini, F., and Viterbo, J. (2022). Towards defining data interpretability in open data portals: Challenges and research opportunities. *Information Systems*, 106:101961.
- Braunschweig, K., Eberius, J., Thiele, M., Lehner, W., and Dresden, T. U. (2012). The state of open data limits of current open data platforms.
- Cadena-Vel, S., Fuster-Guilló, A., and Maz, J.-N. (2019). Publishing open data considering quality criteria. (publicando datos abiertos considerando criterios de calidad). *RISTI - Iberian Journal of Information Systems and Technologies*, pages 295–308.

- Cadena-Vela, S., Mazón, J.-N., and Fuster-Guilló, A. (2020). Defining a master data management approach for increasing open data understandability. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 11878 LNCS*. Springer-Verlag.
- Chapple, M. (2013). Speaking the same language: Building a data governance program for institutional impact. *Educause Review*, 48(6):14.
- Chattapadhyay, S. (2014). Access and use of government data by research and advocacy organizations in india: A survey of (potential) open data ecosystem. In *In Proceedings of the 8th International Conference on Theory and Practice of Electronic Governance*, page 361–364. ACM.
- Cheong, L. and Chang, V. (2007). The need for data governance: A case study. In *In Proc. 18th Australasian Conference on Information Systems, Toowoomba (Australia)*, pages 999–1008. AIS Electronic Library (AISeL).
- DataPortals.org (2023). A comprehensive list of open data portals from around the world. Online reference at <https://datacatalogs.org/>.
- Fioretti, M. (2011). Open data: Emerging trends, issues and best practices. Laboratory of Economics and Management of Scuola Superiore Sant’Anna, Pisa.
- Gurstein, M. B. (2011). Open data: Empowering the empowered or effective data use for everyone? *First Monday* 16, 2, 16(2).
- Interian, R., Mendoza, I., Bernardini, F., and Viterbo, J. (2022). Unified vocabulary in official gazettes: An exploratory study on procurement data. In *Proceedings of the 15th International Conference on Theory and Practice of Electronic Governance, ICEGOV ’22*, page 195–202, New York, NY, USA. Association for Computing Machinery.
- Khatri, V. and Brown, C. V. (2010). Designing data governance. *ACM*, 53(1):148–152.
- Lněnička, M., Machova, R., Volejníková, J., Linhartová, V., Knezackova, R., and Hub, M. (2021). Enhancing transparency through open government data: the case of data portals and their features and capabilities. *Online Information Review*, 45(6):1021–1038.
- Martínez, R., Pons, C., Rodríguez, R., and Vera, P. (2021). Quality evaluation of government open data sets in argentina using the hevda validation tool. *Journal of Science and Research*, 6(2).
- Masoumi, H., Farahani, B., and Shams Aliee, F. (2022). Systematic and ontology-based approach to interoperable cross-domain open government data services. *Transforming Government: People, Process and Policy*, pages 110–127.
- Newman, D. and Logan, D. (2006). Governance is an essential building block for enterprise information management. *Gartner Research*.
- Nikiforova, A. and Lnenicka, M. (2021). A multi-perspective knowledge-driven approach for analysis of the demand side of the open government data portal. *Government Information Quarterly*, 38(4).
- Oracle (2011). Enterprise information management: Best practices in data governance. *An Oracle white paper on enterprise architecture*.

- Pinto, H., Barcellos, R., Bernardini, F., and Viterbo, J. (2022). A methodology for aligning categories from open government data portals to a comprehensive set of categories. In *Electronic Government: 21st IFIP WG 8.5 International Conference, EGOV 2022, Linköping, Sweden, September 6–8, 2022, Proceedings*, page 258–273, Berlin, Heidelberg. Springer-Verlag.
- Rangel, M., Bernardini, F., Viterbo, J., Monteiro, R., Seixas, E., and Pinto, H. d. S. (2020). Use of machine learning for automatic categorization of datasets from open data portals. *Workshop on Computing Applied to E-Government (WCGE)*, 8.
- Rangel-Carrillo, A. M., Maestre-Góngora, G. P., and y M. A. Osorio-Sanabria (2020). Principles, guidelines, dimensions, and attributes for the quality evaluation of open government data. (principios, lineamientos, dimensiones y atributos para la evaluación de calidad de datos abiertos de gobierno). *Aibi revista de investigación, administración e ingeniería - UDES*, 8(S1):54–65.
- Reis, J. R., Bernadini, F., and Viterbo, J. (2022). A new approach for assessing metadata completeness in open data portals. *International Journal of Electronic Government Research (IJEGR)*, 18(1):1–20.
- Reis, J. R., Viterbo, J., and Bernardini, F. (2018). A rationale for data governance as an approach to tackle recurrent drawbacks in open data portals. In *Proceedings of the 19th Annual International Conference on Digital Government Research: Governance in the Data*, page 1–9. ACM.
- Tygel, A., Auer, S., Debattista, J., Orlandi, F., and Campos, M. L. M. (2016). Towards cleaning-up open data portals: A metadata reconciliation approach. In *2016 IEEE Tenth International Conference on Semantic Computing (ICSC)*, page 71–78. IEEE.
- Vetrò, A., Canova, L., Torchiano, M., Minotas, C. O., Iemma, R., and Morando, F. (2016). Open data quality measurement framework: Definition and application to open government data. *Government Information Quarterly*, 33(2):325–337.
- Zuiderwijk, A., Helbig, N., Gil-García, J. R., and Janssen, M. (2014). Special issue on innovation through open data: Guest editors' introduction. *Journal of theoretical and applied electronic commerce research*, 9(2):i–xiii.