



## ARTIGO DE PESQUISA

Modelo baseado em *Frictionless Data* aplicado  
aos dados abertos governamentais

Melissa Figueira Fagundes<sup>1</sup>  <https://orcid.org/0000-0003-1082-9245>

Divino Ignácio Ribeiro Junior<sup>2</sup>  <https://orcid.org/0000-0002-9705-6507>

Universidade do Estado de Santa Catarina, Florianópolis, SC, Brasil / e-mail: [melffagundes@gmail.com](mailto:melffagundes@gmail.com)

Universidade do Estado de Santa Catarina, Florianópolis, SC, Brasil / e-mail: [divino@udesc.br](mailto:divino@udesc.br)

**RESUMO**

O presente trabalho discute como um modelo baseado em *Frictionless Data* (FD) pode auxiliar na publicação de dados abertos governamentais (DAGs). FD é uma iniciativa da *Open Knowledge Foundation*, que pretende remover o "atrito" no trabalho com os dados, ou seja, quando se perde muito tempo e recursos para entender e trabalhar com o dado. No âmbito das instituições públicas, a ausência de padrões para publicação e processamento dos DAGs é problema comum quando se trata de abertura dos DAGs. O estudo procurou contribuir ao aplicar o modelo ao conjunto de dados com informações sobre as boas práticas executadas pelo Judiciário relacionadas aos Objetivos de Desenvolvimento Sustentável (ODS). Por meio de levantamento bibliográfico se identificou na literatura as principais características presentes nos modelos para abertura de DAGs e barreiras inerentes ao processo de publicação. Utilizou-se a técnica de pesquisa em laboratório para descrever e analisar a aplicação do modelo proposto em um ambiente controlado para fins específicos do presente trabalho. O modelo se baseou no *Data Publication Workflow*, um fluxo para publicação de dados disponível no site da iniciativa *Frictionless Data* e que aborda etapas como empacotamento, tratamento e publicação dos dados. A aplicação do modelo também utilizou ferramentas de código aberto baseados em *Frictionless Data*. Os resultados mostraram a viabilidade do modelo para a abertura de um conjunto de dados abertos governamentais, assim como demonstrou que as ferramentas disponíveis na iniciativa *Frictionless Data* contribuíram para a verificação das etapas do modelo. Concluiu-se que o modelo precisa de validação em outros contextos como, por exemplo, integração e abertura de diferentes bases públicas relacionadas aos ODS.

**PALAVRAS-CHAVE:** Dados abertos governamentais. Governo. Frictionless data. Open data.

**Frictionless Data-based model applied to open government data****ABSTRACT**

This article discusses how a model based on *Frictionless Data* (FD) can assist in the publication of open government data (DAGs). FD is an initiative of the *Open Knowledge Foundation*, which aims to remove the "friction" in working with the data, that is, when a lot of time and resources is lost to understand and work with the data. Within public institutions, the absence of standards for publishing and processing DAGs is a common problem when it comes to opening DAGs. The study sought to contribute by applying the model to the data set with information on good practices carried out by the Judiciary related to the *Sustainable Development Goals* (SDGs). Through a bibliographic survey, the main characteristics present in the models for opening DAGs and barriers inherent to the publication process were identified in the literature. The laboratory research technique was

used to describe and analyze the application of the proposed model in a controlled environment for specific purposes of this research. The model was based on the Data Publication Workflow, a flow for publishing data available on the website of the *Frictionless Data* initiative and that addresses steps such as packaging, processing and publishing the data. The application of the model also used open source tools based on *Frictionless Data*. The results showed the feasibility of the model for opening a set of open government data, as well as demonstrating that the tools available in the *Frictionless Data* initiative contributed to the verification of stages of the model. It was concluded that the model needs validation in other contexts, for example, integration and opening of different public bases related to the SDGs.

**KEYWORDS:** Open data government. Government. Frictionless data. Open data.



JITA: IM. Open data

## 1 INTRODUÇÃO

As organizações públicas, em especial os governos, estão sujeitas a novas exigências da sociedade, tais como: aumento da transparência e da participação na gestão dos recursos públicos, maior controle sobre a qualidade dos serviços prestados, maior responsabilização dos gestores públicos, entre outras. A partir destas novas exigências políticas, econômicas e sociais, novas formas de gestão e de aproximação dos governos com a sociedade se fazem necessárias. Na busca dessas novas formas de gestão, e especialmente visando maior aproximação com as novas exigências da sociedade, uma das possibilidades é a utilização de ferramentas tecnológicas. Com o advento das tecnologias atuais, especialmente vinculadas à tecnologia da informação e comunicação (TIC) - com destaque para a internet, as organizações públicas estão viabilizando a oferta de uma variada gama de novos produtos e serviços à sociedade (ALBANO; ARAUJO; REINHARD, 2017). Na última década, emergiu e ganhou importância um movimento conhecido como Dados Abertos Governamentais, que estabelece um conjunto de requisitos a serem seguidos pelas instituições públicas para publicação de dados públicos, realizando principalmente essa tarefa através do uso de TIC (CORREA; SOUZA; SILVA, 2019).

O acesso à informação, nos moldes que conhecemos hoje, está diretamente vinculado à adoção da resolução 59 da ONU de 1946, que apresenta esse acesso como sendo um direito humano fundamental e pedra fundamental das liberdades nas quais a ONU é fundamentada. Além disso, o acesso à informação implica no direito de coletar, transmitir e publicar informações (CARDOSO, 2019). Em setembro de 2011, o Brasil tornou-se um membro da *Open Government Partnership (OGP)*, uma iniciativa multinacional para promover mundialmente ampla adoção de *Open Government Data (OGD)*. O compromisso compreende marcos políticos e técnicos e o lançamento do Portal Brasileiro de Dados Abertos. O objetivo inicial era reunir grandes quantidades de dados agregados do governo para publicação digital e criar um catálogo central de informações sobre informações da atividade pública, com intenção de melhorar a governança e monitorar a atividade do governo (BREITMAN *et al.*, 2012).

Em 2017, as Nações Unidas disponibilizaram o Guia de Planejamento de Dados Abertos Governamentais para o Desenvolvimento Sustentável, inspirado na experiência do Brasil, estabelecida pelo Decreto 8.777/2016, que trata da Política de Dados Abertos, onde está definido que cada órgão e entidade da administração pública federal deve fazer o seu próprio Plano de Dados Abertos. O documento contém recomendações e um passo-a-passo para que organizações planejem ações relacionadas à abertura de dados com vistas a promover os Objetivos de Desenvolvimento Sustentável (ODS) da Agenda 2030.

A Agenda Global 2030 é um compromisso assumido por líderes de 193 países, inclusive o Brasil, e coordenada pelas Nações Unidas, por meio do Programa das Nações Unidas para o Desenvolvimento (PNUD), nos termos da Resolução A/RES/72/279.OP32, de 2018, da Assembleia Geral da ONU. São 17 Objetivos de Desenvolvimento Sustentável e 169 metas a serem atingidas de 2016 a 2030, relacionadas com a efetivação dos direitos humanos e promoção do desenvolvimento, que incorporam e dão continuidade aos 8 Objetivos de Desenvolvimento do Milênio, a partir de subsídios construídos na conferência Rio +20 (CONSELHO NACIONAL DE JUSTIÇA, 2018). A Rio +20 foi uma conferência das Nações Unidas sobre o Desenvolvimento Sustentável que aconteceu no Rio de Janeiro em 2012, 20 anos após a conferência das Nações Unidas sobre o Meio Ambiente e Desenvolvimento, conhecida como Rio-92 ou Eco-92 (GALLOPÍN; VESSURI, 2019).

Os ODS são como uma lista de tarefas a serem cumpridas pelos governos, a sociedade civil, o setor privado e todos cidadãos na jornada coletiva para um 2030 sustentável. Nos

próximos anos de implementação da Agenda 2030, os ODS e suas metas irão estimular e apoiar ações em áreas de importância crucial para a humanidade: Pessoas, Planeta, Prosperidade, Paz e Parcerias (CONSELHO NACIONAL DE JUSTIÇA, 2018). Logo, o sucesso da implementação dos ODS depende de parcerias entre sociedade, organizações públicas e privadas. Para ajudar a cumprir as metas, os estados membros da ONU podem usar dados abertos governamentais como um meio para alcançar os ODS, especificamente o ODS 16, que tem como descrição: “Promover sociedades pacíficas e inclusivas para o desenvolvimento sustentável, proporcionar o acesso à justiça para todos e construir instituições eficazes, responsáveis e inclusivas em todos os níveis” (INSTITUTO DE PESQUISA ECONÔMICA APLICADA, 2018). Um exemplo desse uso são os dados públicos como subsídios para aumentar transparência, prestação de contas públicas e participação do cidadão, expondo e prevenindo a corrupção e má administração dos recursos públicos (WORLD BANK, 2015).

Por meio da Portaria CNJ n. 133/2018 foi instituído Comitê Interinstitucional, destinado a avaliar a integração das metas do Poder Judiciário às metas e aos indicadores dos ODS, Agenda 2030 e elaborar relatório de trabalho com apoio de todos os tribunais do país, cuja composição consta da Portaria CNJ n. 148/2018. Segundo a Portaria CNJ n.133/2018, o alinhamento da atuação do Poder Judiciário à Agenda 2030, da ONU, pode representar avanço no campo na concretização dos direitos fundamentais dos cidadãos, pois o Judiciário poderá fornecer informações relevantes e necessárias – cuja base de dados é produzida e mantida pelo próprio Poder Judiciário – para o cumprimento das metas dos ODS, que vincula todas as pessoas, de todos os países, por se traduzir em um Pacto Global para erradicar a pobreza e promover vida digna para todos, dentro dos limites do planeta (CONSELHO NACIONAL DE JUSTIÇA, 2018)

Para mostrar que é possível essa correlação do Poder Judiciário aos ODS, o Comitê mapeou todas as Resoluções do CNJ, bem como as recomendações, os provimentos e as orientações da Corregedoria Nacional de Justiça, além dos atos normativos e boas práticas dos Tribunais Superiores, Tribunais Regionais Federais, Seções Judiciárias, Tribunais de Justiça dos Estados e do Distrito Federal, Tribunais Regionais do Trabalho, Tribunais Regionais Eleitorais, Tribunais Militares, sendo este último conjunto de dados utilizado no presente artigo para validação do modelo baseado em *Frictionless Data*.

Os órgãos públicos em geral criam, armazenam e divulgam uma grande variedade de informações, desde dados demográficos e econômicos, até renda familiar. Esses dados podem assumir formatos diferentes, divulgados como dados primários ou processados. Podem também estar relacionados a serviços públicos ou processos internos (REN; GLISSMANN, 2012). Por exemplo, a informação judicial disponibilizada pelos sites dos Tribunais de Justiça, acessada diariamente pelo cidadão comum e por todos os operadores do Direito, devido ao seu grande volume e importância, necessita ter boa qualidade (ALBUQUERQUE; BASTOS; LINO, 2009). Diversas iniciativas surgiram para resolver o problema da qualidade da informação referente às barreiras técnicas que envolvem a publicação de DAG. Em 2016 surge uma iniciativa chamada *Frictionless Data*, criada pela *Open Knowledge Foundation*, que se apresenta como uma alternativa para remover o “atrito” no trabalho com os dados, com foco na simplicidade e uma estrutura básica que permita qualquer pessoa publicar e usar conjuntos de dados com mais facilidade, sem passar por um intermediário. Diante do exposto, este artigo se propôs a responder ao seguinte questionamento: “como um modelo baseado na especificação *Frictionless Data* pode contribuir para a abertura dos dados abertos governamentais?”.

As seções seguintes deste artigo estão divididas da seguinte forma: contexto da pesquisa, trabalhos relacionados e os fundamentos teóricos com principais conceitos sobre a problemática do trabalho com dados abertos governamentais sob a ótica da Qualidade da informação; um breve levantamento sobre Dados Abertos e Dados Abertos Governamentais; uma introdução

aos princípios FAIR (*Findable, Accessible, Interoperable and Reusable*); introdução aos conceitos fundamentais da presente pesquisa sobre *Frictionless Data* e *Data Package*; apresentação da metodologia, estrutura do modelo e sua aplicação no contexto do judiciário brasileiro; considerações finais e trabalhos futuros.

### 1.1 O problema da Qualidade da Informação no âmbito dos Dados Abertos Governamentais

No contexto das instituições públicas, foco do presente artigo, tem-se que os órgãos públicos criam, armazenam e divulgam uma grande variedade de informações, desde dados demográficos e econômicos, até renda familiar. Esses dados assumem formatos diferentes, podendo ser divulgados como dados primários ou processados. Podem também estar relacionados a serviços públicos ou processos internos (REN; GLISSMANN, 2012). Por exemplo, a informação judicial disponibilizada pelos sites dos Tribunais de Justiça, acessada diariamente pelo cidadão comum e por todos os operadores do Direito, devido ao seu grande volume e importância, necessita ter boa qualidade (ALBUQUERQUE; BASTOS; LINO, 2009).

De acordo com Quarati e De Martino (2019), a disseminação dos dados abertos governamentais é considerada a força motriz do crescimento econômico e social, além de ser um fator essencial para publicidade das ações governamentais. Porém, a alta disponibilidade de dados não garante a qualidade das informações presentes nessas bases de dados públicos. A divulgação de dados pode mostrar que a qualidade dos dados sobre os quais são tomadas decisões importantes é ruim. O sucesso na abertura de dados governamentais requer mais do que a simples provisão de acesso aos dados. Também são necessários o aprimoramento da qualidade das informações governamentais, a criação e a institucionalização de uma cultura de governo aberto e o fornecimento de ferramentas e instrumentos com os quais os dados serão utilizados. Essa perspectiva mais ampla precisa ser adotada pelos governos que agora estão apenas abrindo um portal para tornar os dados acessíveis. (JANSSEN; CHARALABIDIS; ZUIDERWIJK, 2012).

No entanto, a tarefa de definir objetivamente o que é qualidade da informação não é algo trivial. Na área da Ciência da Informação, por exemplo, observa-se que a definição de qualidade da informação ainda não é um consenso. Nehmy e Paim (1998) apresentam uma análise sobre o conceito de qualidade da informação, de forma a “focalizar a discussão nas formas de abordagem da qualidade da informação na literatura, no esforço de se desvelarem limitações e desafios para a construção teórica do conceito”. De acordo com o estudo, não existe uma definição geralmente aceita sobre qualidade da informação, sendo um conceito vago e subjetivo (NEHMY; PAIM, 1998 *apud* SCHWUCHOW 1990, P. 55).

Para Paim (1996), o tema qualidade da informação ainda esbarra na inerente dificuldade de conceituar o termo qualidade, não havendo um consenso na literatura sobre definições teóricas e operacionais da qualidade da informação. Os autores fazem uma análise sobre as dificuldades existentes em torno da qualidade da informação e a tendência na literatura a estudá-la sob três vertentes: qualidade transcendente (ou filosófica, ou metafísica) da informação; uma vertente que se baseia em aspectos intrínsecos, a qual alguns autores entendem que é sinônimo da transcendente, e outra nos atributos contingenciais.

Já trabalhos como de Trindade e Oliveira (2007) procuraram definir atributos para avaliação da qualidade da informação em sistemas que disponibilizam grandes quantidades de informação que podem ser compartilhadas e transformadas em conhecimento. No entendimento desses autores, alguns atributos que possam ajudar a avaliar a qualidade da informação são: acessibilidade, clareza, completeza, concisão, consistência, credibilidade, disponibilidade,

empacotamento, facilidade de uso, histórico, interface, interpretabilidade, objetividade, precisão, prontidão, quantidade, relevância, reputação, segurança, temporalidade e valor.

Para Bobrowski, Marré e Yankelevich (1999), os problemas relacionados à qualidade da informação derivam de práticas equivocadas na engenharia de software. Foram levantados sete atributos para mensurar a qualidade da informação: completude, relevância, confiabilidade, consistência, temporalidade, precisão e concisão. Outros autores mostram a evolução sobre o tema qualidade da informação e sua aplicação no contexto organizacional. Segundo Calazans (2008), a falta de qualidade da informação em uma organização pode ter impactos sociais e no negócio. Fatores como informações com múltiplas origens, utilização de julgamentos subjetivos, sistemáticos erros na produção da informação, além da grande quantidade de informação em diferentes fontes influenciam a qualidade da informação (CALAZANS, 2008, *apud* STRONG; LEE; WANG, 1997).

Em estudo que realizou um levantamento sobre *frameworks* para avaliação da qualidade da informação, foram identificadas vinte estruturas de qualidade da informação que definem e categorizam critérios de qualidade das informações em variados contextos (EPPLER; WITTIG, 2000). Vale ressaltar que o estudo adota a definição de *framework* assumida por (PORTER, 1991), o qual considera-se uma forma de ajudar a identificar variáveis relevantes e as perguntas que o usuário deve responder para desenvolver conclusões personalizadas para um determinado setor e empresa (PORTER, 1991, *apud* EPPLER; WITTIG, 2000).

Nas próximas seções será retomada a discussão sobre qualidade da informação, dessa vez sob a perspectiva dos dados abertos governamentais, mas antes serão apresentados conceitos sobre o que faz um dado ser aberto. Segundo Davenport (2001), o compartilhamento de informações é definido como o ato voluntário de colocá-las à disposição dos outros e qualquer fornecedor de informação pode agregar valor à informação ao torná-la acessível. Com o paradigma das Tecnologias de Informação e Comunicação (TIC), tornou-se um desafio para as organizações gerenciar grande produção informacional, ou seja, a informação precisa ser gerenciada e compartilhada, visando à construção do conhecimento organizacional e, conseqüentemente, promover a sobrevivência em um mercado altamente competitivo (VALENTIM, 2010).

A *Open Knowledge Foundation* (OKF), no *Open Data Handbook*, define dados abertos como dados que podem ser livremente usados, reutilizados e redistribuídos por qualquer pessoa – sujeitos, tendo como única exigência a atribuição da fonte e compartimento pelas mesmas regras (OPEN KNOWLEDGE FOUNDATION, 2011). Alguns pontos são importantes relacionados a definição completa de aberto são:

- a) **Disponibilidade e Acesso:** os dados devem estar disponíveis como um todo e sob custo não maior que um custo razoável de reprodução, preferencialmente possíveis de serem baixados pela internet; devem também estar disponíveis de uma forma conveniente e modificável;
- b) **Reutilização e Redistribuição:** os dados devem ser fornecidos sob termos que permitam a reutilização e a redistribuição, inclusive a combinação com outros conjuntos de dados;
- c) **Participação Universal:** todos devem ser capazes de usar, reutilizar e redistribuir - não deve haver discriminação contra áreas de atuação ou contra pessoas ou grupos; por exemplo, restrições de uso ‘não-comercial’ que impediriam o uso ‘comercial’, ou restrições de uso para certos fins (ex.: somente educativos) excluem determinados dados do conceito de ‘abertos’.

Duas qualidades importantes tornam os dados "abertos": os dados devem estar "legalmente" abertos, ou seja, significa que é irrestrito por termos restritivos de direitos autorais e pode ser compartilhado legalmente e usado por indivíduos, empresas, academia, organizações sem fins lucrativos e outros; os dados também devem ser abertos no sentido "técnico" da palavra, que se refere à entrega dos dados em um formato padrão e bem definido, geralmente via Internet (AYRE; CRANER, 2017). Outro conceito relevante relacionado aos dados abertos é a interoperabilidade, o qual significa "a capacidade de diversos sistemas e organizações trabalharem juntos (interoperar). Ou seja, a capacidade de interoperar - ou combinar - diferentes conjuntos de dados. Conforme o Manual dos dados abertos (2011), um dado é definido como aberto quando permite que as informações sejam publicadas e disseminadas de maneira que terceiros consigam reutilizá-lo. Considera-se, ainda, que são três as leis dos dados abertos, conforme lista abaixo:

- 1) Os dados precisam ser encontrados e indexados na *Web*
- 2) Devem estar em formato aberto e compreensíveis por máquina
- 3) Sem restrições quanto à utilização

Dados abertos se referem aos dados coletados e compartilhados com outras pessoas para usarem como desejam, sem restrições de direitos autorais ou uso. Exemplos tradicionais de dados abertos incluem dados coletados pelo governo (por exemplo, boletins meteorológicos, relatórios de incidentes criminais, códigos postais), bem como algumas fontes acadêmicas (periódicos de acesso aberto, pesquisas e dados de pesquisa, resultados de experimentos científicos). Outros exemplos de aplicação dos dados abertos são dados disponibilizados por governos estaduais, nacionais, agências de fomento e organizações não governamentais, que são usados para reduzir custos e melhorar os serviços que oferecem aos cidadãos (AYRE; CRANER, 2017).

Sob a perspectiva da qualidade da informação, a W3C criou o *Data on the Web Best Practices*, um documento que fornece as melhores práticas relacionadas à publicação e uso de dados na *Web*, criado para ajudar e apoiar um ecossistema autossustentável para dados abertos. De acordo com o guia, composto por 35 melhores práticas, os dados precisam ser descobertos e compreensíveis por seres humanos e máquinas, sendo o objetivo dessas práticas recomendadas facilitar a interação entre o consumidor e o publicador de dados abertos. Ainda conforme o *Data on the Web Best Practices*, nem todos os dados e metadados devem ser compartilhados abertamente, pois questões como segurança, sensibilidade comercial e, acima de tudo, privacidade dos indivíduos precisa ser considerada e cabe aos produtores de dados determinar quais políticas são adequadas e quais dados devem ser compartilhados e sob quais circunstâncias (MANUAL..., 2011).

Diante o exposto, faz-se necessário diferenciar dados abertos dos demais conceitos existentes na literatura. De acordo com Albano e Reinhard (2014), governo aberto, dados abertos e dados abertos governamentais têm significados diferentes:

- a) **Governo aberto** é a disponibilização de informações em qualquer formato por parte dos governos e outras ações que visem promover maior transparência;
- b) **Dados abertos** são a disponibilização de informações, em alguns formatos (pré-estabelecidos), por governos, organizações privadas, com ou sem fins lucrativos, ou por outros atores de uma sociedade;
- c) **Dados governamentais abertos** são a disponibilização de dados por parte dos governos, nos mesmos formatos estabelecidos para dados abertos

Dados abertos governamentais também são caracterizados sob três perspectivas (KUMAR, 2016 apud GONZALEZ-ZAPATA; HEEKS, 2015):

- a) **Dados do governo:** os governos coletam e geram uma grande quantidade de dados e conjuntos de dados; esses conjuntos de dados contêm dados relacionados, por exemplo, a pagamentos de assistência social, impacto da estrutura de assistência social, dados de identificação exclusiva de cidadãos e dados relacionados ao serviço público;
- b) **Dados abertos:** preocupações em superar barreiras à acessibilidade, trazendo uniformidade na maneira como os dados são capturados, armazenados, processados e gerados; dados abertos são quaisquer dados que podem ser usados livremente, reutilizados e redistribuídos por qualquer pessoa;
- c) **Governo Aberto:** um governo pode ser denominado Governo Aberto se mostrar uma tendência de governança, priorizando a abertura de seus dados; a abertura também envolve a criação de atos de direito à informação, mais liberdades civis e mais transparência.

A abertura de dados públicos está intrinsecamente relacionada à criação da *Freedom of Information Act (FOIA)*, ou Lei de Liberdade de informação, que pode ser definida como o direito de acessar informações mantidas por órgãos públicos, sendo parte integrante do direito fundamental à liberdade de expressão, conforme reconhecido em Resolução 59 da Assembleia Geral da ONU, adotada em 1946 e pelo Artigo 19 da Declaração Universal dos Direitos Humanos (1948). As leis associadas ao *FOIA* refletem a premissa fundamental de que todas as informações mantidas por governos e instituições governamentais são, em princípios, públicas e só podem ser retidas por razões como privacidade e segurança. Quanto à adesão ao *FOIA*, em 1990, apenas 13 países adotaram leis nacionais de *FOIA*, enquanto atualmente existem mais de 90 leis adotadas em todo o mundo.

As leis *FOIA* determinam quem pode acessar os dados públicos governamentais, quais informações estão disponíveis, quanto tempo os governos devem responder e quais penalidades resultam quando o governo descumpre a legislação. Além disso, a implementação das leis *FOIA* frequentemente acompanha movimentos de democratização e anticorrupção em todo o mundo e a seu início se deu com o que se chamou “Década da Abertura”, que durou desde o final da Guerra Fria até aos ataques terroristas nos EUA (FINK, 2018). Dados públicos abertos são uma ferramenta poderosa e também um derivado da *FOIA*. Sem o *FOIA*, os dados abertos não podem cumprir o que prometem: a verdadeira liberdade de informação (IZDEBSKI, 2015).

Em 2009, o governo Obama esboçou um conjunto de princípios abertos do governo dos EUA que abrangem três objetivos fundamentais: colaboração, participação e transparência. Com o objetivo de divulgar os dados do governo americano ao contribuinte, de modo que possam ser usadas para gerar valor social e econômico, instituiu-se o *Open Government Directive (OGD)*, um conjunto de diretivas as quais exigem que todas as agências do governo americano publiquem conjuntos de dados em um site disponível ao público, fornecendo contexto e metadados importantes por meio de relatórios, resumo de serviços públicos, links para os principais programas e conjuntos de dados para download (DAWES; HELBIG, 2010).

No Brasil, de acordo com Oliveira e Lóscio (2014), a Lei de Acesso à Informação Pública (BRASIL, 2014), sancionada em 18 de novembro de 2011 e que começou a vigorar em 16 de maio de 2012, permitiu a qualquer pessoa ter acesso a documentos e informações que estejam sob a guarda de órgãos públicos, em todos os poderes e níveis do governo. O lançamento do Portal Brasileiro de Dados Abertos possibilitou a interlocução entre atores da



sociedade e governo, viabilizando a localização dos dados em formato aberto a partir de uma única fonte e lugar, por meio de uma ferramenta de busca.

Diversas iniciativas surgiram com a criação da Lei de Acesso à Informação e muitas são as possibilidades com relação ao uso de dados abertos públicos como, por exemplo, promover a transparência quanto às ações dos governos. Segundo o entendimento de Barbalho (2018), algumas funções da transparência para o setor público são: confiança no governo e nas instituições; participação cidadã; planejamento estratégico; direcionamento de produtos e serviços; e exercício da cidadania. As plataformas de dados abertos podem ajudar na criação de políticas públicas e se trata de uma poderosa abordagem para tornar os dados localizáveis, interoperáveis e legíveis por máquina, melhorando drasticamente a eficiência de análises e insights (KLEIN; KLEIN; LUCIANO, 2018). No entanto, no Brasil, apesar de existirem programas bastante avançados de transparência pública – como, por exemplo, de transparência orçamentária –, ainda são raríssimos os órgãos ou secretarias que disponibilizam dados abertos. No melhor dos casos, há dados disponíveis para visualização, mas existem inúmeras barreiras técnicas, e até políticas, para que sejam reutilizados pela sociedade na criação de novos projetos e serviços (MANUAL..., 2011).

É notório que dados abertos implementados e mantidos por governos e suas agências, organizações da sociedade civil e setor de negócios oferecem diversos benefícios, como a criação de novos produtos e serviços e geração de novas oportunidades econômicas (ALBANO; REINHARD, 2014). Quando se fala em mercado de dados, estima-se que o seu valor se aproxima de 40 a 140 bilhões de euros anualmente (IZDEBSKI, 2015). Diante disso, surgiram alguns exemplos da utilização de dados abertos com a finalidade econômica, como a iniciativa do Banco Mundial de instituir um concurso para indicar a melhor solução (aplicativo desenvolvido) criada a partir de seus dados (dados que o banco publica em formato aberto) (ALBANO; REINHARD, 2014).

A respeito da qualidade dos dados abertos governamentais, alguns trabalhos abordam a avaliação da qualidade das informações no setor público brasileiro como o de Albuquerque, Bastos e Lino (2009), que apresenta um modelo de avaliação para aferição de qualidade das informações publicadas em portais dos tribunais de justiça, os quais têm vital importância para a comunidade que acessa diariamente o conteúdo judiciário e necessita de mecanismos e metodologias para avaliar a qualidade das informações disponíveis em seus sites.

As definições de qualidade no contexto de dados abertos variam consideravelmente. O Portal Europeu de Dados, por exemplo, considera como alta qualidade dos dados abertos se “os seres humanos puderem entendê-los e as máquinas puderem manipulá-los” e estiverem alinhados com o sistema de classificação *5-Star Open Data*. Outros, como o *G8 Open Data Charter* e os *Open Data Institute Certification Badges*, se concentram no fornecimento de metadados, descrições de esquemas de dados, uso de dicionários de dados compartilhados, licença usada, formato de arquivo e suporte ao editor para interagir com os usuários de dados (CORSAR; EDWARDS, 2017).

Melhorar a qualidade de acesso aos dados abertos também pode ser um facilitador essencial para ciência, tecnologia e inovação (OECD, 2018). Um exemplo são iniciativas para compartilhar dados de pesquisa, aumentando o ritmo da descoberta de conhecimento e do progresso científico. A reutilização de dados de pesquisa tem o potencial de evitar a duplicação de conjuntos de dados e trazer novas perspectivas a partir de múltiplas análises do mesmo conjunto de dados. Um exemplo são os estudos que utilizam dados de diversas fontes para analisar variações genômicas associadas ao câncer, com o objetivo de selecionar qual terapia mais apropriada para um paciente específico (FIGUEIREDO, 2017).

Em 2016, a *Scientific Data*, uma publicação focada em dados, publicou o *FAIR Guiding Principles for Scientific Data Management and Stewardship*, cujo objetivo era fornecer

diretrizes para melhorar a localização, acessibilidade, interoperabilidade e reuso dos ativos digitais. Os princípios enfatizam a capacidade dos sistemas computacionais de encontrar, acessar, interoperar e reutilizar os dados com pouca ou nenhuma intervenção humana, uma vez que se exige um maior suporte computacional para lidar com o crescente volume de dados.

Os princípios *FAIR* podem ser definidos como:

Os princípios orientadores do *FAIR* precedem as escolhas de implementação e não sugerem nenhuma tecnologia, padrão ou solução de implementação específica; além disso, os Princípios não são, eles próprios, um padrão ou uma especificação. Eles atuam como um guia para editores de dados e administradores para ajudá-los a avaliar se suas opções de implementação específicas estão tornando seus artefatos de pesquisa digital localizáveis, acessíveis, interoperáveis e reutilizáveis (WILKINSON *et al.*, 2016).

Os princípios *FAIR* defendem o aumento da localização, acessibilidade, interoperabilidade e reutilização de dados de pesquisa e objetos digitais acadêmicos em geral. Sob a égide da sigla *FAIR*, um acrônimo para *Findable, Accessible, Interoperable and Reusable*, foram formulados 15 princípios para orientar as ações dos editores, administradores de dados e outras partes interessadas.

O ponto central do conceito *FAIR* é sua aplicação para atividades guiadas por humanos e orientadas por máquinas. Além disso, *FAIR* não é binário (ou seja, *FAIR*/não *FAIR*), mas sim um espectro ao longo do qual são possíveis variáveis “*FAIRness*”, isto é, o máximo quanto for possível sua aplicação (HIGMAN; BANGERT; JONES, 2019).

Destaca-se que os princípios *FAIR* se aplicam aos dados e seus metadados, ou seja, registros sobre conjunto de dados. Por isso o termo “metadados” é declarado nos princípios. Outro aspecto importante sobre os princípios *FAIR* é que não se referem apenas a dados abertos, sendo possível trabalhar com dados privados, por exemplo. Os princípios *FAIR* não representam um padrão de qualidade para avaliação de ferramentas, dados, políticas, entre outros, como também sua implementação pode ser uma adaptação gradual e sistemática de novas rotinas de trabalho (HANSEN, 2018).

No ano de 2010, com o objetivo de classificar o grau de abertura e conexão dos dados, Tim Berners-Lee propôs cinco princípios conhecidos como “Sistema de 5 estrelas” ou *5-Star OPEN DATA principles*. Os princípios apresentam um esquema, o qual determina que quanto maior o número de estrelas, maior será o grau de abertura e conexão dos dados (CAMPOS; VILELA, 2018). As abordagens *5-Star* e *FAIR* possuem pontos convergentes, no entanto vale ressaltar que os princípios *FAIR* cobrem um escopo mais amplo e incluem dados não abertos. Já o esquema *5-Star* são orientados para abertura dos dados. Outro ponto de distinção está relacionado aos requisitos de acessibilidade como um elemento essencial de reutilização, sendo esse aspecto diretamente atrelado as infraestruturas de TI que atendem os dados. Nesse sentido, os princípios *FAIR* são mais orientados por considerações de implementação do que os princípios *5-Star* (HASNAIN; REBHOLZ-SCHUHMAN, 2018).

Na próxima seção será abordada a iniciativa *Frictionless Data*, a qual tem sido usada para tratar problemas de publicação de dados abertos.

## 1.2 *Frictionless Data*

*Frictionless Data* é uma iniciativa da *Open Knowledge Foundation* que pretende remover o “atrito” no trabalho com os dados. Entende-se como “atrito” ou “fricção” quando os consumidores de dados perdem tempo e recursos demais apenas para entender e trabalhar com as informações. Por outro lado, os produtores de informação encontram dificuldades para

descrever os tipos de dados permitidos em cada coluna de uma planilha, por exemplo (OPEN KNOWLEDGE FOUNDATION, 2018).

Ainda segundo o site oficial da iniciativa, *Frictionless Data* é um *progressive framework* para construção de infraestrutura de dados – gerenciamento, integração, fluxo, entre outros, ou seja, ao contrário de outros *frameworks*, FD foi projetado para ser adotado de forma incremental e “progressiva”. O objetivo do *framework* é trabalhar, desenvolver e aprimorar os dados e ferramentas existentes usadas no trabalho com dados, ao invés de substituí-los. Logo, não se trata de um grande banco de dados ou um sistema de gerenciamento de dados, mas sim uma abordagem que auxilia na automatização de tarefas repetitivas sobre conjunto de dados, como obtenção, limpeza e utilização dos dados.

Uma boa analogia para entender o atrito no trabalho com os dados é o processo de fazer um bolo, conforme exemplifica Rufus Pollock, criador e fundador do *Open Knowledge Foundation*: ao decidir cozinhar, tem-se os ingredientes prontamente disponíveis no mercado ou na sua própria cozinha. Não é necessário viajar para uma fazenda, coletar ovos, moer o milho, fazer o bacon. Ao invés disso, graças à logística, transporte (*containers*) e pagamento, os ingredientes estão disponíveis em um supermercado, sem necessidade de buscá-los na fonte. O problema é que quando se trata de dados, ainda existe grande custo para obter e trabalhar com os dados. Ao realizar uma análise ou construir um aplicativo, por exemplo, dependendo do conjunto de dados necessário, precisa-se descobrir a fonte de dados, extrai-los, limpá-los e prepará-los antes que possa iniciar o trabalho.

O *Frictionless Data* oferece, ainda, uma gama de ferramentas, padrões e boas práticas para publicação de dados. O *Frictionless Data* se baseia em princípios como:

- a) foco: concentra em tipos específicos de dados;
- b) orientado à *Web*: usa formatos que funcionam bem na *Web*, como o *JSON* (*Javascript Object Notation*), formato comumente usados para troca de dados por *APIs* (*Application Programming Interface*);
- c) distribuição: projeto com visão de um ecossistema sem centralização de dados;
- d) abertura: qualquer pessoa pode usar e reusar livremente e abertamente os dados;
- e) uso de ferramentas existentes: integração com ferramentas tecnológicas existentes;
- f) simplicidade: manutenção dos formatos e metadados de forma simples e leve, fácil de interpretar e usar (OPEN KNOWLEDGE FOUNDATION, 2019).

A ideia central da iniciativa *Frictionless Data* é o conceito de "containerização de dados", ou seja, a forma como o dado é compartilhado segue o conceito de contêiner como se entende no transporte de mercadorias em um navio de carga, por exemplo. O objetivo é desenvolver padrões especificando os *containers* de modo a reduzir os custos e tempo para transporte de dados (OPEN KNOWLEDGE FOUNDATION, 2019).

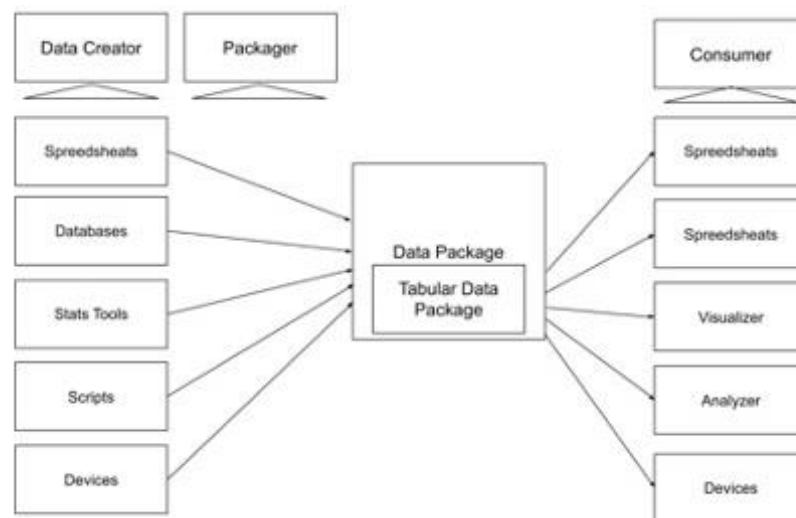
Um conceito chave presente na iniciativa *Frictionless Data* é o *Data Package* ou pacote de dados. Trata-se de uma forma para simplificar o agrupamento de dados e suas descrições em um único local, com a finalidade de tornar os dados mais facilmente compartilhados e usados. O formato do pacote de dados é muito simples, amigável e extensível (OPEN KNOWLEDGE FOUNDATION, 2019).

Existem diversas iniciativas de “empacotamento” de dados que compartilham um modelo de “arquivos em disco” para compactar dados com metadados. Estes incluem *BagIt* (Kunze et al., 2016) e *CSV na Web* do *W3C* (*CSVW*) (W3C, 2016). *BagIt* é um formato de pacote de conjunto de dados popular entre as bibliotecas, incluindo a Biblioteca do Congresso nos EUA, projetado para oferecer suporte ao armazenamento baseado em disco e à transferência de rede de conteúdo digital arbitrário. *BagIt* se concentra no arquivamento de ativos digitais

(principalmente conteúdo), enquanto o *Data Package* se concentra em descrever a estrutura dos dados empacotados. Além disso, um elemento-chave da abordagem de *Data Package* é a integração com as ferramentas existentes e a extensibilidade para muitos tipos de dados, ambos os quais não são prioridades para *BagIt*. CSVW é o resultado de um grupo de trabalho W3C concluído em 2016, é mais diretamente comparável. O CSVW começou como um esforço para padronizar a especificação do então protótipo do *Tabular Data Package* com várias modificações para garantir a compatibilidade com a *Web*. (FOWLER; BARRATT; WALSH, 2018).

Um *Data Package*, conforme mostra a figura 1, consiste em metadados que descrevem a estrutura e o conteúdo do pacote, além de recursos como arquivo de dados que formam o conteúdo do pacote. Os metadados do pacote são armazenados em um "descriptor", que por sua vez transforma uma coleção de dados em um pacote de dados. A estrutura deste descriptor é o conteúdo principal da especificação do pacote de dados. A especificação do *Data Package* não impõe nenhum requisito à sua forma ou estrutura e, portanto, pode ser usada para "empacotar" qualquer tipo de dado, contribuindo para a publicação de dados, por exemplo (OPEN KNOWLEDGE FOUNDATION, 2019).

**Figura 1.** Processo de compartilhamento de dados usando Data Package



Fonte: Adaptado de *Open Knowledge Foundation* (2019)

O processo de compartilhamento de dados usando *Data Package* possui, em apenas uma camada, uma estrutura de transporte de dados básica que reduz consideravelmente o "atrito" na integração e distribuição de dados, apoiando a automação, sem impor mudanças drásticas nos dados para serem empacotados. A estrutura, por sua natureza simples e leve, torna fácil sua adoção pelos publicadores de conteúdo, usuário de dados e criadores de ferramentas de software (OPEN KNOWLEDGE FOUNDATION, 2019).

Na próxima seção será apresentado o modelo baseado em *Frictionless Data* para colaborar com a publicação e compartilhamento de dados abertos governamentais, além de uma contextualização sobre os detalhes do modelo e sua aplicação.

### 1.3 Trabalhos relacionados

Segundo Geraldo e Dalenogare (2019), existe um consenso na literatura sobre a necessidade da definição de um framework ou modelo padrão de disponibilização de dados dos governos, a fim de que seja proporcionada maior unicidade e respeito a determinados critérios, aceitos internacionalmente. De acordo com Janssen, Charalabidis e Zuiderwijk (2012), os gestores públicos simplesmente preferem disponibilizar os dados, sem se preocuparem com outras atividades que envolvam a publicação de dados. No entanto, os dados geralmente não podem ser usados em estado bruto, podendo ser necessário antes uma avaliação da qualidade, modificação e processamento desses dados. Os autores supracitados afirmam, ainda, que a padronização de métodos e o desenvolvimento de metadados robustos podem aumentar o acesso aos dados. Outro problema é a falta de mecanismos de feedback que mostre o que é feito com os dados abertos. Comumente os governos que publicam conjuntos de dados enfrentam críticas como má usabilidade, falta de *feedback* de dados e mecanismos de melhoria e metadados inadequados (DAWES & HELBIG, 2010).

Por outro lado, vale ressaltar que a definição de padrões pode proporcionar diversos benefícios pois, de acordo com Kassen (2013), os DAGs podem criar um ambiente favorável para o engajamento cívico proativo, fornecendo uma oportunidade real para desenvolvedores independentes criarem aplicativos usando conjuntos de DAGs, juntamente com a participação dos cidadãos, criando um novo ambiente de cooperação entre governo local e os cidadãos, tendo como consequência uma verdadeira transformação na forma de comunicação entre governo e sociedade.

Para realizar o levantamento dos trabalhos relacionados à presente pesquisa, utilizou-se os mecanismos de busca no portal de Periódicos da CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior). Através do uso de palavras-chave como *modelo*, *open government data*, *open data* e *Frictionless Data*, identificou-se estudos com propostas de modelos para publicação de dados abertos governamentais. O levantamento na literatura sobre publicação de dados abertos governamentais mostrou trabalhos, como o de Ávila (2015), que propôs um modelo de processo de publicação de dados considerando o nível de maturidade da instituição publicadora, contribuindo com as estratégias de produção e disponibilização de dados abertos e dados abertos conectados do setor público. No entanto, apesar do trabalho considerar o nível de maturidade das instituições e contemplar várias etapas do processo de publicação de dados abertos, “não houve maior detalhamento técnico (envolvendo ferramentas, por exemplo) de como deve ser desenvolvida a publicação de dados abertos. Consequentemente, não contempla nenhuma recomendação referente as boas práticas para publicação de dados abertos de natureza mais técnica” (ÁVILA, 2015).

Trabalhos, como o de Dzikrullah e Rinjani (2017), desenvolveram um modelo chamado *SmartGov* para integração de sistemas de dados baseados em Big Data com foco em *e-gov*. A proposta é facilitar a troca de dados abertos governamentais entre agências do governo, criando um ambiente de serviços integrados. O foco do modelo foi resolver problemas de interoperabilidade de dados e, apesar do modelo contemplar diversas etapas inerentes à publicação de dados abertos, como uma camada chamada *Open Data Layer*, a qual utiliza o CKAN (*Comprehensive Knowledge Archive Network*), plataforma desenvolvida pela *Open Knowledge Foundation*, não possui uma etapa de *feedback* para reportar problemas de qualidade encontrado nos dados.

Já outro trabalho apresenta um modelo heurístico para compreender a relação entre dados abertos governamentais e as atividades dos cidadãos e governos do ponto de vista dos processos democráticos monitorial, deliberativo e participativo. O modelo adapta a abertura de

dados abertos de acordo com o tipo de processo democrático e ajuda a identificar desafios do uso de dados abertos para a democracia. O trabalho tem como ponto forte considerar a complexidade dos processos democráticos, mas não apresenta etapa para tratamento de qualidade dos dados, apesar de mencionar que a importância do fornecimento de informações e a qualidade dos dados sejam cruciais (RUIJER; GRIMMELIKHUIJSEN; MEIJER, 2017).

Saxena (2016) propôs um modelo para avaliar a intenção comportamental para o uso dos DAGs. Realizou-se uma análise quantitativa usando regressão múltipla para analisar 341 catálogos com dados governamentais. O ponto forte desse trabalho foi verificar a popularidade de um conjunto de dados abertos entre usuários finais. O modelo proposto por Zhu e Freeman (2019) forneceu uma estrutura para avaliar o desempenho dos portais que publicam DAGs, com foco nas possíveis interações e experiência de uso pelos usuários. O ponto forte do estudo foi a avaliação de 34 portais, baseada em critérios abrangentes, que tinham como objetivo avaliar senso de confiança dos usuários nos portais, além da capacidade de compreender o conteúdo e de integrar dados e formas de participação do cidadão. Porém, ambos modelos mencionados com foco em participação cidadã não contemplam outras etapas importantes para publicação de dados abertos, como tratamento da informação publicada.

No artigo de Tambouris (2016), foi delineado um processo para publicação, expansão e consumo de dados governamentais abertos usando um modelo multidimensional, como a abordagem de cubo de dados juntamente com ferramentas de *Linked Data*. Como ponto forte do trabalho, seguindo o processo proposto, os produtores de dados podem transformar dados brutos em cubos de dados em formato RDF. Com isso, os consumidores de dados podem encontrar, visualizar, mesclar e analisar dados com maior facilidade. No entanto, o trabalho não abrange outros processos fundamentais para publicação de dados abertos governamentais, como tratamento dos dados e *feedback* dos usuários.

Outro trabalho apresenta um modelo de dados abertos conectados (*Linked Open Data - LOD*) para um conjunto de dados abertos legislativos da Câmara dos Deputados. O modelo elaborado na pesquisa foi feito com base no conjunto real de dados fornecidos pela Câmara dos Deputados usando RDF para estruturar os dados. Como ponto forte do trabalho, houve a contribuição para áreas da Ciência da Informação como, por exemplo, para a representação da informação e do conhecimento, com o uso de vocabulários e ontologias para representar e trazer significado aos relacionamentos entre as entidades modeladas. No entanto, a pesquisa não contempla questões como alinhamento com melhores práticas para publicação de dados e uso de metadados (BRANDT et al., 2018).

O modelo proposto por Kalampokis, Tambouris e Tarabanis (2011), possui duas dimensões principais: complexidade organizacional, tecnológica e valor agregado para os consumidores de dados. A ideia era complementar os modelos de governo eletrônico existentes para reuso de DAGs, com o objetivo de fornecer um roteiro para reutilização e DAGs, além de permitir a avaliação das iniciativas relevantes. O ponto forte do modelo foi a integração dos dados que, além das duas dimensões mencionadas, possui quatro fases: agregação de dados governamentais, integração de dados governamentais, integração de dados governamentais com dados sociais e formais não governamentais. Entretanto, não houve implementação de um protótipo para cada estágio do modelo, como também não foi contemplado o tratamento dos dados.

Os autores Wiese *et al.* (2019) usaram *Frictionless Data* para resolver o problema da grande quantidade de dados fornecidos pelas empresas pertencentes ao Sistema Europeu de Energia com documentação insuficiente, além de dados em formatos inconsistentes. O trabalho elencou alguns desafios que, com o uso do *Frictionless Data*, obtiveram resultados satisfatórios para resolução de problemas relacionados: à identificação de fontes apropriadas, a diferentes

fontes de dados, à falta de nomenclatura padronizada e classificação de fontes de energia, à falta de metadados, à qualidade dos dados (inconsistências, erros, lacunas nos dados).

O *Frictionless Data* já tem sido usado em iniciativas internacionais no âmbito governamental como a *Workforce Data Initiative*, a qual pretende modernizar a força de trabalho dos Estados Unidos por meio do uso dos dados. O objetivo dessa iniciativa é ajudar os conselhos estaduais e locais de força de trabalho a coletar, agregar e distribuir estatísticas sobre a eficácia dos provedores de treinamento e programas que esses provedores. O Departamento de Trabalho dos EUA exige que os provedores de treinamento elegível (ETP) trabalhem com as *Workforces* dos conselhos estaduais para acompanhar os resultados de seus alunos a fim de receber financiamento federal. Para tal tarefa, estão construindo um conjunto de ferramentas de código aberto baseado na especificação *Frictionless Data* chamada *etp-uploader*, uma ferramenta que permite aos provedores de treinamento carregar os dados em uma única plataforma. São muitas centenas ou milhares de provedores de treinamento sob a alçada de cada conselho e cada um deve carregar com segurança seus dados e, ao mesmo tempo, deve ser fácil para a própria placa processar e validar automaticamente os conjuntos de dados. Uma ferramenta como o *etp-uploader* poderia ser desenvolvida baseada no modelo proposto e usada para enviar os dados das boas práticas relacionadas aos Objetivos de Desenvolvimento Sustentável dos vários tribunais de justiça para o CNJ em formato *Data Package* (FRICTIONLESS DATA, 2017).

## 2 METODOLOGIA, APRESENTAÇÃO DO MODELO, APLICAÇÃO E RESULTADOS

Segundo Marconi e Lakatos (2002) um problema, para ser considerado relevante, precisar ser analisado sob a ótica de seu valor. De acordo com os autores, um problema de pesquisa deve ser viável, relevante, novo, exequível e oportuno. O presente trabalho se fundamentou em uma documentação direta sobre as abordagens para publicação de dados abertos governamentais. Buscou-se, por intermédio dos mecanismos de busca presentes no portal de Periódicos da CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior), base de dados utilizada na presente pesquisa, o uso de palavras-chave como *modelo*, *open government data*, *open data* e *Frictionless Data*, com o objetivo de identificar estudos com propostas de modelos para publicação de dados abertos governamentais. Por meio do levantamento bibliográfico, identificou-se na literatura as principais características presentes nos modelos para abertura de dados abertos governamentais e as barreiras inerentes ao processo de publicação.

A técnica de pesquisa de laboratório oferece um resultado mais exato, apesar de ser considerado um procedimento de investigação mais difícil. Além disso, exige instrumental específico, preciso e ambientes adequados (MARCONI E LAKATOS, 2003, pg. 190), o que se alinha aos objetivos delineados na presente pesquisa. Por essas razões, a técnica de pesquisa de laboratório foi selecionada para descrever e analisar a aplicação do modelo proposto em um ambiente controlado para fins específicos do presente trabalho, usando as ferramentas online disponíveis na iniciativa *Frictionless Data* para realizar os experimentos usando a base de dados do CNJ relacionadas aos Objetivos de Desenvolvimento Sustentável e verificar todas as etapas propostas no presente modelo. A iniciativa *Frictionless Data* disponibiliza um conjunto de ferramentas, especificações e práticas recomendadas para descrever, publicar e validar dados.

O modelo tem como base, além dos estudos de trabalhos relacionados, o guia com os princípios *FAIR* mencionados anteriormente no artigo como suporte para atender às

características principais de interoperabilidade; o *Frictionless Data Spec*, um conjunto de especificações para uso e publicação de dados usando *Frictionless Data*; os requisitos para o processo de dados abertos propostos por Zuiderwijk, Janssen e Jeffery; os elementos-chave para a criação de um ecossistema de dados abertos (ZUIDERWIJK; JANSSEN; DAVIS, 2014) e a aderência às melhores práticas para publicação de dados abertos.

As etapas do modelo foram delineadas de forma a responder ao problema delimitado nesse artigo. Ou seja, o modelo pretende auxiliar a publicação de dados abertos governamentais, tendo em vista diminuir o tempo que se gasta trabalhando com os dados abertos governamentais, assim como promover a qualidade da informação publicada e fornecer uma estrutura básica que permita qualquer cidadão publicar e usar conjuntos de dados com mais facilidade e sem passar por um intermediário.

A figura 2 mostra as etapas do modelo baseado em *Frictionless Data* para publicação de dados abertos governamentais:

**Figura 2.** Modelo baseado em *Frictionless Data* para publicação de dados governamentais



Fonte: Elaborado pelos autores

No modelo proposto, a etapa Empacotar se refere à transformação de um conjunto de dados em um pacote de dados, de modo a permitir a interoperabilidade e o reuso dos dados. Essa etapa tem como fundamentação a especificação *Frictionless Data*, que contém o conceito chamado *Data Package*, explicado em seções anteriores com mais detalhes. Um *Data Package* consiste em metadados, que descrevem a estrutura e o conteúdo do pacote de dados, e recursos como arquivos de dados, que formam o conteúdo do pacote. Os metadados do *Data Package* são armazenados em um descritor. Esse descritor é o conteúdo principal da especificação *Frictionless Data*. Além do descritor, um *Data Package* incluirá outros recursos, como arquivo de dados. Um dos pontos fortes da especificação é que o *Data Package* não impõe nenhum requisito em sua forma ou estrutura e, portanto, pode ser usada para empacotar qualquer tipo de dados.



Como os órgãos públicos disponibilizam grande quantidade de informações e essas bases geralmente estão em diferentes formatos ou em formatos que impossibilitam acesso pelas partes interessadas aos dados (ARAUJO *et al.*, 2012), essa etapa auxilia na disponibilização dos dados, independente do formato. Além disso, a etapa atende ao requisito de Zuiderwijk, Janssen e Jeffery (2014), “vinculação e combinação de dados”. Outra vantagem dessa etapa é o alinhamento aos 15 princípios *FAIR*, os quais defendem o aumento da localização, acessibilidade, interoperabilidade e reutilização de dados. Apesar do foco inicial dos princípios *FAIR* serem os dados de pesquisa e objetos digitais acadêmicos em geral, podem ser usados para qualquer tipo de dado. Logo, trata-se de uma contribuição relevante da presente pesquisa a aplicação inédita dos princípios *FAIR* aos dados abertos governamentais através do modelo baseado em *Frictionless Data*.

Os resultados decorrentes da verificação do modelo mostraram que a etapa “Empacotar” permitiram a transformação da base de boas práticas relacionadas aos Objetivos de Desenvolvimento Sustentável em um pacote de dados, conceito chamado *Data Package*, dentro da especificação *Frictionless Data*, através do uso da ferramenta *Data Package Creator*. O *Data Package Creator* é um serviço on-line que facilita a criação e edição de um *data package*. O serviço gera automaticamente um arquivo *datapackage.json*, à medida que é possível adicionar e editar dados que fazem parte da base de dados (FRICTIONLESS DATA, 2017). A etapa Empacotar da presente pesquisa disponibilizou os dados independente do formato, uma vez que o *Frictionless Data* não impõe nenhum requisito em sua forma ou estrutura, podendo ser usado para empacotar qualquer tipo de dado. Logo, considera-se que a etapa “Empacotar” é o coração do modelo proposto e os resultados mostraram que o objetivo da etapa foi atingido de maneira satisfatória.

Outra problemática relevante no contexto dos órgãos públicos é com relação à qualidade das informações presentes nas bases de dados abertos governamentais. A difusão dos DAGs por parte das instituições públicas se manteve em um ritmo muito acelerado, no entanto, a divulgação de dados sem controle de qualidade adequado prejudica a sua reutilização e afeta negativamente a participação cívica (VETRÒ *et al.*, 2016).

Pesquisas mostram que existem diversas interpretações e maneiras diferentes de medir a qualidade de dados, como comentado anteriormente no presente trabalho. Geralmente as abordagens incluem interpretabilidade de dados, precisão, oportunidade de publicação, confiabilidade, acessibilidade, capacidade de descoberta e processamento ou integridade. No entanto, uma vez que pessoas usam dados para finalidades distintas, certas qualidades de dados são mais importantes para um grupo de usuários que para outros. Muitas métricas de qualidade de dados são focadas nos usuários, porém é fundamental que o governo, como produtor de dados, compreenda, monitore e melhore a qualidade inerente dos dados que produz. A *Open Knowledge International* tem como alvo os produtores de dados e problemas de qualidade dos arquivos de dados, principalmente por meio da iniciativa *Frictionless Data*. Na presente pesquisa, ao propor o modelo baseado nas especificações contidas na iniciativa *Frictionless Data*, sugere que a etapa “Tratar” contemple aspectos essenciais de qualidade para arquivos de dados tabulares contidos no *Data Quality Spec*, projeto apoiado pela *Open Knowledge International*. Por fim, para essa etapa se recomenda atender ao requisito qualidade de dados – contemplando dimensões da qualidade da informação tais como concisão, consistência e precisão (TRINDADE; OLIVEIRA, 2007).

Os resultados apresentados nessa etapa sugerem que a etapa “Tratar” contemplaram dimensões essenciais de qualidade da informação como concisão, consistência e precisão, ao realizar a validação do conjunto de dados, verificando problemas de qualidade da informação como cabeçalhos em branco, linhas duplicadas e dados com tipo errado.

A etapa “Publicar” sugere que se contemple aspectos fundamentais para publicação e disponibilidade dos DAGs. Com isso, pretende-se que se atenda aos seguintes requisitos propostos por Zuiderwijk, Janssen e Jeffery quanto aos processos necessários para publicação dos DAGs: acesso, pesquisa, navegação, upload, download e visualização. Para isso, na presente pesquisa, utilizou-se a ferramenta *DataHub*, plataforma de gerenciamento de dados criada pela *Open Knowledge International*, baseada no sistema de gerenciamento de dados CKAN. O *Datahub* fornece acesso gratuito a muitos dos principais recursos do CKAN, permitindo pesquisa de dados, o registro de conjuntos de dados publicados a criação e gerenciamento de grupos de conjuntos de dados e acompanhamento das atualizações dos conjuntos de dados e grupos de interesse. Esses recursos estão disponíveis através de uma interface *Web* ou uma *API*, a *CKAN API*.

Os resultados da etapa “Publicar” do presente modelo buscaram atender o principal desafio: os dados abertos governamentais não têm valor em si mesmos, senão quando permite que os usuários utilizem de forma eficaz. Logo, faz-se necessário disponibilizar os dados em formato aberto e com qualidade. Durante a presente pesquisa, após a verificação das etapas de “Empacotar” e “Tratar”, expostas em seções anteriores, foi possível publicar os dados através da ferramenta *Data.io*, pertencente à iniciativa *Frictionless Data*. A etapa “Publicar” cumpriu o seu objetivo ao contemplar aspectos fundamentais para publicação e disponibilidade, ao permitir o acesso, upload e download da base de boas práticas do CNJ relacionadas aos Objetivos de Desenvolvimento Sustentável. No entanto, vale ressaltar que requisitos como visualização não foram atendidos com o uso da ferramenta *Data.io*, sendo necessário no futuro a criação de tais funcionalidades para que se consiga verificar tais requisitos importantes para publicação de dados abertos governamentais.

A etapa “*Feedback*”, proposta no presente modelo, foi estabelecida com o objetivo de atender a lacuna encontrada na literatura com relação aos modelos existentes para publicação e abertura de dados abertos, ao acrescentar a participação e *feedback* sobre as publicações. A intenção da etapa era uma possível contribuição através da criação de ciclos de *feedback*, com os quais o governo pode aprender com o público e promover melhorias sobre a qualidade dos dados abertos. Durante o decorrer da aplicação do modelo, não foi possível verificar a etapa “*Feedback*” utilizando ferramentas da iniciativa *Frictionless Data*, posto que a especificação não possui uma ferramenta que atenda ao ciclo de *feedback* dos dados. No entanto, a etapa “*Feedback*”, no presente modelo, foi concebida sem a intenção de determinar uma ferramenta específica ou processo fixo, de modo a flexibilizar a escolha da tecnologia ou ferramenta conforme necessidade do órgão público. Logo, a ausência de uma ferramenta na iniciativa *Frictionless Data* que permita acompanhar o ciclo de *feedback* não compromete o uso do modelo, uma vez que a presença da etapa atende a lacuna identificada na literatura, e se torna uma oportunidade para desenvolvimento futuro de uma ferramenta pela comunidade.

### 3 CONSIDERAÇÕES FINAIS E TRABALHOS FUTUROS

Diante do crescimento nos últimos anos da quantidade de dados produzidos pelas organizações públicas e a divergência presente no formato desses dados disponíveis para consulta pelo cidadão, o que dificulta o trabalho com os dados e seu compartilhamento, além da baixa qualidade dos dados abertos governamentais descrita na literatura, tendo como possíveis consequências o prejuízo à participação popular e enfraquecimento da democracia, que este artigo se propôs a responder o questionamento “como um modelo baseado na especificação *Frictionless Data* pode contribuir para a abertura dos dados abertos

governamentais?”. Ao decorrer dessa pesquisa, foi realizado um levantamento nas bases de produção científica nacionais e internacionais, além de uma busca sobre as boas práticas para abertura de dados abertos governamentais. Conceitos como qualidade da informação sob a perspectiva da Ciência da Informação, dados abertos governamentais, princípios *FAIR* e *Frictionless Data* foram abordados no presente trabalho, com o objetivo de embasar a proposição do modelo para auxiliar a publicação e o compartilhamento de dados abertos governamentais.

Os trabalhos relacionados ao contexto dessa pesquisa mostraram um consenso sobre a necessidade de um *framework* ou modelo de disponibilização de dados dos governos, com o objetivo de trazer maior unicidade e respeito a determinados critérios internacionalmente aceitos. Além disso, na etapa Empacotar do presente modelo, com o uso da ferramenta *Data Package Creator*, ao padronizar métodos e desenvolvimento de metadados, por exemplo, pode-se aumentar o acesso ao dado. Outro problema presente no cenário de abertura de dados é a falta de feedback que mostre o que é feito com os dados abertos. Percebeu-se uma lacuna relacionada a possibilidade de *feedback* para reportar problemas na qualidade dos dados, assim como viabilizar que os cidadãos enviem sugestões de melhorias nos conjuntos de dados abertos governamentais.

Os resultados do presente trabalho foram considerados positivos e conduziram a apresentação do modelo e respectiva verificação das etapas. Um os pontos diferenciais com relação aos modelos para publicação de dados abertos governamentais encontrados na literatura estão o uso da especificação *Frictionless Data* como base para um modelo de abertura dos dados abertos governamentais no contexto do judiciário brasileiro, como também a inclusão da etapa de “*Feedback*” para atender a lacuna encontrada nos demais modelos.

Outra possibilidade como trabalhos futuros seria o uso do *Data Package Pipelines*, uma estrutura declarativa baseada em fluxo para a construção de pipelines de processamento de dados tabulares. Pipelines são definidos como uma forma de paralelização de tarefas em um sistema que seja executado de forma decomposta, ou seja, permite que um fluxo de tarefas possa dividir suas etapas de trabalho em unidades ou estágio e as executa simultaneamente. Cada estágio do pipeline é normalmente a entrada de sua fila, que é a fila de saída do estágio anterior, e depois de processado o dado, serve entrada do próximo estágio. No caso do *Data Package Pipelines*, poderia ser usado para todas as tarefas de extração, transformação e carregamento (*ETL*) como estrutura para facilitar a implementação do modelo proposto, pois é particularmente adequado para trabalhar com fontes de dados diversas e heterogêneas de qualidade variável e desconhecida. Existem muitas ferramentas e estruturas para fazer trabalho de *ETL* com dados, no entanto, o *Data Package Pipelines* tem como foco aqui organizar e controlar dados confusos de várias fontes e usar as ferramentas de dados sem fricção para transformar esses dados em um fluxo de dados limpo e consistente (*FRICITIONLESS DATA*, s.d.).

Ainda sobre outras possíveis contribuições, o presente trabalho pode colaborar, por exemplo, de forma semelhante com outra iniciativa que utiliza os ODS para acompanhamento de resultados, Portal PPA Cidadão, que foi concebido para permitir ao cidadão acompanhar a implantação de metas e objetivos do Plano Plurianual do Governo Federal e os Objetivos de Desenvolvimento Sustentável (ODS), sendo uma parceria do Governo Federal com o *Open Government Partnership* (OGP) e coordenado pela Corregedoria Geral da União (BRASIL, 2018). Logo, a aplicabilidade e a generalização desta pesquisa têm grande potencial de contribuição científica e social.

O modelo proposto se mantém aberto a novas aplicações e sugestão de melhorias, principalmente no que tange a sua aplicabilidade em outros contextos como, por exemplo, integração e abertura de bases públicas relacionadas aos Objetivos de Desenvolvimento

Sustentável. O aperfeiçoamento da etapa de “*Feedback*”, com o uso de outras ferramentas e adição de metodologias como *Net Promoter Score*, poderia contribuir para medir a satisfação do cidadão com relação aos dados abertos governamentais publicados, podendo permitir um aprimoramento da qualidade dos dados por parte dos governos. O NPS foi apresentado por Reichheld (2003), uma forma simples de medir a fidelidade do cliente. O NPS se baseia na seguinte pergunta: Qual a probabilidade de você recomendar uma marca ou empresa a um amigo ou colega? Classificados em uma escala de 0 a 10. Os clientes são agrupados em ‘promotores’ (classificação de 9-10), ‘passivamente satisfeitos’ (classificação de 7-8) e ‘detratores’ (classificação de 0-6). Segundo Reichheld (2003), os promotores são considerados clientes fiéis que provavelmente continuaram usando o serviço.

A presente proposta de pesquisa pode ainda, ao contribuir com o reuso e compartilhamento dos dados, permitir a construção de modelos para reconhecimento de padrões de DAG e, no caso do uso das bases de boas práticas do judiciário, possibilitar a classificação automática das boas práticas e as temáticas de desenvolvimento sustentável, sendo a sua aplicação possível em diversas áreas, contribuindo para com a implementação da Agenda 2030.

## REFERÊNCIAS

ALBANO, Cláudio Sonaglio. **Dados governamentais abertos**: proposta de um modelo de produção e utilização de informações sob a ótica conceitual da cadeia de valor. 2014. Tese (Doutorado em Administração) – Faculdade de Economia, Administração e Contabilidade, Universidade de São Paulo. São Paulo, 2014. Disponível em: <https://teses.usp.br/teses/disponiveis/12/12139/tde-03062014-170642/publico/ClaudioSnaglioAlbanoVC.pdf>. Acesso em: 20 set. 2020.

ALBANO, Claudio Sonaglio; ARAUJO, Marcelo Henrique de; REINHARD, Nicolau. Fatores motivadores e facilitadores dos relacionamentos em redes: como os gestores públicos reconhecem esses fatores em dados governamentais abertos. **Navus - Revista de Gestão e Tecnologia**, Florianópolis, v. 7, n. 1, p. 73–92, jan. 2017.

ALBUQUERQUE, Almir dos Santos; BASTOS, Rogério Cid; LINO, Manuel Rosa de Oliveira. Qualidade da informação em portais do judiciário. **Encontros Bibli**: revista eletrônica de biblioteconomia e ciência da informação, Florianópolis, v. 14, n. 27, p. 115-134, maio 2009. Disponível em: <https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2009v14n27p115>. Acesso em: 03 set. 2020.

ARAUJO, Marcelo Henrique de *et al.* Dados governamentais abertos. In: ENCONTRO DA ASSOCIAÇÃO NACIONAL DE PÓS-GRADUAÇÃO E PESQUISA EM ADMINISTRAÇÃO, 36., 2012, Rio de Janeiro. **Anais do EnANPAD**. Rio de Janeiro: Enanpad, 2012. p. 1-16.

ÁVILA, Thiago. Uma proposta de modelo de processo para publicação de dados abertos conectados governamentais. **Comunidade Áreas de Integração**, nov. 2015. Disponível em: <http://areasdeintegracao.blogspot.com.br/2017/09/preparacao-de-partes-interessadas-para.html>. Acesso em: 25 ago. 2020.

AYRE, Lori Bowen; CRANER, Jim. Open data: what it is and why you should care. **Public**

**Library Quarterly**, v. 36, n. 2, p. 173–184, 2017. Disponível em: <https://doi.org/10.1080/01616846.2017.1313045>. Acesso em 10 set. 2020.

BARBALHO, Fernando Almeida. A emergência do campo de políticas públicas de dados abertos governamentais no Brasil. **Conhecer: debate entre o público e o privado**, Fortaleza, v. 8, n. 20, p. 118–137, 2018.

BOBROWSKI, Monica; MARRÉ, Martina; YANKELEVICH, Daniel. A homogeneous framework to measure data quality. *In: CONFERENCE ON INFORMATION QUALITY*, 4., 1999, Massachusetts. **Proceedings [...]**. Massachusetts: Mit Sloan School of Management, 1999. p. 115-124.

BRANDT, Mariana Baptista *et al.* Modelo de dados abertos conectados para informação legislativa. **Informação & Sociedade: estudos**, João Pessoa, v. 28, n. 2, p. 149–161, 2018.

BRASIL. **Lei nº 12.965**, de 23 de abril de 2014. Publicada no DOU de 24.4.2014. Diário Oficial, Brasília, DF. 2014. Disponível em: [https://www.planalto.gov.br/ccivil\\_03/\\_ato2011-2014/2014/lei/112965.htm](https://www.planalto.gov.br/ccivil_03/_ato2011-2014/2014/lei/112965.htm). Acesso em: 09 nov. 2020.

BREITMAN, Karin *et al.* Open government data in Brazil. **IEEE Intelligent Systems**, v. 27, n. 3, p. 45–49, 2012.

CALAZANS, Angélica Toffano Seidel. Qualidade da informação: conceitos e aplicações. **Transinformação**, Campinas, v. 20, n. 1, p. 29–45, 2008.

CARDOSO, Paulo Henrique. **Ciência de dados aplicada a dados governamentais abertos sob a ótica da Ciência da Informação**. 2019. 110 f. Dissertação (Mestrado) - Curso de Mestrado em Ciência da Informação, Escola de Comunicação, Universidade Federal do Rio de Janeiro; Instituto Brasileiro de Informação em Ciência e Tecnologia, Rio de Janeiro, 2019. Disponível em: <http://ridi.ibict.br/handle/123456789/1039>. Acesso em: 15 ago. 2020.

CONSELHO NACIONAL DE JUSTIÇA. **Agenda 2030 no Poder Judiciário**. 2018. Disponível em: [http://novo.more.ufsc.br/homepage/inserir\\_homepage](http://novo.more.ufsc.br/homepage/inserir_homepage). Acesso em: 26 ago. 2020.

CORREA, Andreiuid Sheffer; SOUZA, Raul Mendes de; SILVA, Flavio Soares Correa da. Towards an automated method to assess data portals in the deep web. **Government Information Quarterly**, v. 36, n. 3, p. 412–426, 2019. Disponível em: <https://doi.org/10.1016/j.giq.2019.03.004>. Acesso em: 05 set. 2020.

CORSAR, David; EDWARDS, Peter. Challenges of open data quality: More than just license, format, and customer support. **Journal of Data and Information Quality**, v. 9, n. 1, p. 10–12, 2017.

DAVENPORT, Thomas H. **Ecologia da informação**. São Paulo: Campus, 2001.

DAWES, Sharon S.; HELBIG, Natalie. Information Strategies for Open Government: challenges and prospects for deriving public value from government transparency. *In: INTERNATIONAL CONFERENCE ON ELECTRONIC GOVERNMENT*, 9., 2010,

Lausanne, Suíça. **Proceedings [...]**. Lausanne, Suíça: Ifip Wg, 2010. p. 50-60. Disponível em: <https://hal.inria.fr/hal-01056566/document>. Acesso em: 05 set. 2020.

DZIKRULLAH, Fahmi; RINJANI, Muhammad Angga. A framework design to develop integrated data system for smart e-government based on big data technology. **Bulletin of Social Informatics Theory and Application**, v. 1, n. 2, p. 41–51, 2017.

EPPLER, Martin; WITTIG, Dörte. Conceptualizing information quality: a review of information quality frameworks from the last ten years. *In: CONFERENCE ON INFORMATION QUALITY*, 5., 2000, Cambridge. **Proceedings [...]**. Massachusetts: Mit, 2000. p. 1-14.

FIGUEIREDO, Ana Sofia. Data Sharing : convert challenges into opportunities. **Front Public Health**, Londres, v. 5, p. 1–6, dec. 2017. Disponível em : <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5723929/>. Acesso em: 15 set. 2020.

FINK, Katherine. Opening the government’s black boxes: freedom of information and algorithmic accountability. **Information Communication and Society**, v. 21, n. 10, p. 1453–1471, 2018.

FOWLER, Dan; BARRATT, Jo; WALSH, Paul. Frictionless Data: making research data quality visible. **International Journal of Digital Curation**, v. 12, n. 2, p. 274–285, 2018.

FRICIONLESS DATA (ed.). **Center for Data Science and Public Policy, Workforce Data Initiative**. 2017. Disponível em: <https://frictionlessdata.io/blog/2017/08/15/center-for-data-science-and-public-policy-workforce-data-initiative/>. Acesso em: 25 ago. 2020.

FRICIONLESS DATA (ed.). **Data package pipelines**: a framework for processing data packages in pipelines of modular components. [s.d.]. Disponível em: <https://frictionlessdata.io/tooling/data-package-pipelines/#check-it-out>. Acesso em: 25 ago. 2020.

GALLOPÍN, Gilberto; VESSURI, Hebe. Science for sustainable development. **Interfaces between Science and Society**, v. 336, n. June 2012, p. 35–51, 2019.

GERALDO, Lucas; DALENOGARE, Campos. Abordagens teóricas em dados governamentais abertos. **Revista Gestão e Tecnologia**, v. 19, n. 5, p. 296-314, 2019.

HASNAIN, Ali; REBHOLZ-SCHUHMAN, Dietrich. Assessing FAIR data principles against the 5-star open data principles. *In: EXTENDED SEMANTIC WEB CONFERENCE*, 15., 2018, Grécia. **The Semantic Web: ESWC 2018 Satellite Events**. Grécia: Eswc, 2018. p. 469-477. Disponível em: <https://www.researchgate.net/publication/326757298>. Acesso em: 07 set. 2020.

HANSEN, Karsten Kryger; BUSS, Mareike; HAAHR, Lea Sztuk. **A fairy tale**: a fake story in a trustworthy guide to the fair principles for research data. [s.l.] FAIR Project, 2018. Disponível em: <http://doi.org/10.5281/zenodo.2248200>. Acesso em: 03 set. 2020.

HIGMAN, Rosie; BANGERT, Daniel; JONES, Sarah. Three camps , one destination : the

intersections of research data management, FAIR and open. **Insights**, v. 32, n. 1, maio 2019. Disponível em: <http://doi.org/10.1629/uksg.468>. Acesso em: 23 ago. 2020.

INSTITUTO DE PESQUISA ECONÔMICA APLICADA. **Agenda 2030**: ODS - Metas Nacionais dos Objetivos de Desenvolvimento Sustentável. Brasília, DF: IPEA, 2018.

Disponível em:

[http://www.ipea.gov.br/portal/index.php?option=com\\_content&view=article&id=33895&Itemid=433](http://www.ipea.gov.br/portal/index.php?option=com_content&view=article&id=33895&Itemid=433). Acesso em 02 set. 2020.

IZDEBSKI, Krzysztof. **Transparency and open data principles**: why they are important and how they increase public participation and tackle corruption. Varsóvia, Polônia: [s.n.], 2015. Disponível em: <https://transparencee.org/wp-content/uploads/2015/12/open-data-principles-by-krzysztof-izdebski.pdf>. Acesso em 15 set. 2020.

JANSSEN, Marijn; CHARALABIDIS, Yannis; ZUIDERWIJK, Anneke. Benefits, adoption barriers and myths of open data and open government. **Information Systems Management**, v. 29, n. 4, p. 258–268, 2012.

KALAMPOKIS, Evangelos; TAMBOURIS, Efthimios; TARABANIS, Konstantinos. Open Government Data: a stage model. In: INTERNATIONAL CONFERENCE, EGOV, 10., 2011, Delft, Holanda. **Electronic Government, Proceedings**. Berlim, Alemanha: Springer, 2011. p. 235-246. Disponível em: [https://doi.org/10.1007/978-3-642-22878-0\\_20](https://doi.org/10.1007/978-3-642-22878-0_20). Acesso em: 10 set. 2020.

KASSEN, Maxat. A promising phenomenon of open data: a case study of the Chicago open data project. **Government Information Quarterly**, v. 30, n. 4, p. 508–513, 2013. Disponível em: <http://dx.doi.org/10.1016/j.giq.2013.05.012>. Acesso em 25 ago. 2020.

KLEIN, Rodrigo Hickmann; KLEIN, Deisy Barbiero; LUCIANO, Edimara Mezzomo. Open Government Data: concepts, approaches and dimensions over time. **Revista Economia & Gestão**, Belo Horizonte, v. 18, n. 49, p. 4–24, 2018.

KUMAR, V. Open Government Data Initiatives: a comparison of India and China. **Research Gate**, n. March, 2016. Disponível em: [https://www.researchgate.net/profile/Vinit\\_Kumar9/publication/305768855\\_Open\\_Government\\_Data\\_Initiatives\\_a\\_comparison\\_of\\_India\\_and\\_China/links/57a05a6408aece1c72171d69/Open-Government-Data-Initiatives-a-comparison-of-India-and-China.pdf](https://www.researchgate.net/profile/Vinit_Kumar9/publication/305768855_Open_Government_Data_Initiatives_a_comparison_of_India_and_China/links/57a05a6408aece1c72171d69/Open-Government-Data-Initiatives-a-comparison-of-India-and-China.pdf).

MANUAL dos dados abertos: governo. [s.l.]: Laboratório Brasileiro de Cultura Digital; W3C; Cgi.Br, 2011. (Dados Abertos Governamentais). Traduzido e adaptado de opendatamanual.org. Disponível em:

[https://www.w3c.br/pub/Materiais/PublicacoesW3C/Manual\\_Dados\\_Abertos\\_WEB.pdf](https://www.w3c.br/pub/Materiais/PublicacoesW3C/Manual_Dados_Abertos_WEB.pdf).

Acesso em: 25 ago. 2020.

NEHMY, Rosa Maria Quadros; PAIM, Isis. A desconstrução do conceito de “qualidade da informação”. **Ciência da Informação**, Brasília, DF, v. 27, n. 1, p. 36–45, 1998.

OLIVEIRA, Lairson E. Rodrigues de Alencar; LÓSCIO, Bernadette Farias. Uma abordagem para captura de informações sobre aplicações que fazem uso de dados abertos. **Revista**

Brasileira de Administração Científica, Aracaju, v. 5, n. 2, p. 127–140, 2014.

OPEN KNOWLEDGE FOUNDATION. What is open data? **Opend data handbook**, [2011]. Disponível em: <https://opendatahandbook.org/guide/en/what-is-open-data/>. Acesso em: 25 ago. 2020.

PAIM, Isis. Problematização do conceito "Qualidade" da Informação. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 1, n. 1, p. 111–119, 1996. Disponível em: <http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/8>. Acesso em: 28 ago. 2020.

PAIM, Isis; NEHMY, Rosa Maria Quadros; GUIMARÃES, Cesar Geraldo. Problematização do conceito "Qualidade" da informação. **Perspectivas em Ciência da Informação**, [s.l.], v. 1, n. 1, nov. 2007. ISSN 19815344. Disponível em: <http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/8/27>. Acesso em: 22 set. 2020.

PORTER, Michael Eugene. Towards a dynamic theory of strategy. **Strategic Management Journal**, v. 12, p. 95-117, 1991.

QUARATI, Alfonso; DE MARTINO, Monica. Open government data usage: a brief overview. *In*: INTERNATIONAL DATABASE APPLICATIONS & ENGINEERING SYMPOSIUM, 23., 2019, Atenas, Grécia. **Proceedings [...]**. Atenas, Grécia: Ideas, 2019. p. 1-8. Disponível em: <https://doi.org/10.1145/3331076.3331115>. Acesso em: 28 ago. 2020.

REICHHELD, F. Frederick. The one number you need to grow. **Harvard Business Review**, v.81, n.12, 46–57, 2013.

REN, Guang-Jie; GLISSMANN, Susanne. Identifying information assets for open data: the role of business architecture and information quality. *In*: INTERNATIONAL CONFERENCE ON COMMERCE AND ENTERPRISE COMPUTING, 14., 2012, Washington, Dc. **Proceedings [...]**. Washington, Dc: Ieee Computer Society, 2012. p. 94-100.

RUIJER, Erna; GRIMMELIKHUIJSEN, Stephan; MEIJER, Albert. Open data for democracy: developing a theoretical framework for open data use. **Government Information Quarterly**, v. 34, n. 1, p. 45–52, 2017. Disponível em: <http://dx.doi.org/10.1016/j.giq.2017.01.001>. Acesso em: 25 ago. 2020.

SAXENA, Stuti. Open Government Data (OGD) usage in India: a conceptual framework using TOE & UTAUT frameworks. **Socrates**, Kanpur City, India, v. 4, n. 3, p. 124–144, 2016.

TAMBOURIS, Efthimios. Multidimensional Open Government Data. **JeDEM - eJournal of eDemocracy and Open Government**, Krems, Áustria, v. 8, n. 3, p. 1–11, 2016.

TRINDADE, Ana Lucia Batista; OLIVEIRA, Mírian. Atributos Para Avaliação da Qualidade da Informação em Sistemas de Gestão do Conhecimento. *In*: ENCONTRO DE ADMINISTRAÇÃO DA INFORMAÇÃO, 1., 2007, Florianópolis. **Anais [...]**. Maringá, Pr: Anpad, 2007. p. 1-14. Disponível em: <http://www.anpad.org.br/admin/pdf/enadi318.pdf>. Acesso em: 15 set. 2020.



VALENTIM, Marta. **Gestão, mediação e uso da informação**. São Paulo: Unesp, 2010.

VETRÒ, Antonio et al. Open data quality measurement framework: Definition and application to Open Government Data. **Government Information Quarterly**, v. 33, n. 2, p. 325–337, 2016.

VILELA, Allyson Bruno Campos Barros. **OpenData Processor**: uma ferramenta para a automatização do processo de extração e publicação de dados abertos. 2018. 97 f. Dissertação (Mestrado) - Curso de Engenharia de Software, Programa de Pós-Graduação em Engenharia de Software, Universidade Federal do Rio Grande do Norte, Natal, 2018. Disponível em: [https://repositorio.ufrn.br/jspui/bitstream/123456789/26014/1/OpenDataprocessorferramenta\\_Vilela\\_2018.pdf](https://repositorio.ufrn.br/jspui/bitstream/123456789/26014/1/OpenDataprocessorferramenta_Vilela_2018.pdf). Acesso em: 28 ago. 2020.

WIESE, Frauke *et al.* Open power system data – frictionless data for electricity system modelling. **Applied Energy**, v. 236, p. 401-409, fev. 2019. DOI:10.1016/j.apenergy.2018.11.097. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0306261918318130>. Acesso em: 20 ago. 2020.

WILKINSON, Mark D. et al. Comment: The FAIR Guiding Principles for scientific data management and stewardship. **Scientific Data**, v. 3, 2016.

W3C. CSV on the Web: a primer. W3C Working Group Note, 25 feb. 2016. Disponível em: <https://www.w3.org/TR/tabular-data-primer/>. Acesso em: 25 ago. 2020.

ZHU, Xiaohua; FREEMAN, Mark Antony. An evaluation of U.S. municipal open data portals: A user interaction framework. **Journal of the Association for Information Science and Technology**, v. 70, n. 1, p. 27–37, 2019.

ZUIDERWIJK, Anneke et al. Socio-technical Impediments of Open Data. **Electronic Journal of e-Government**, v. 10, n. 2, p. 156–172, 2012. Disponível em: <http://www.ejeg.com/issue/download.html?idArticle=255>. Acesso em: 25 ago. 2020.

ZUIDERWIJK, Anneke; JANSSEN, Marijn; DAVIS, Chris. Innovation with open data: Essential elements of open data ecosystems. **Information Polity**, v. 19, n. 1–2, p. 17–33, 2014



Artigo submetido ao sistema de similaridade

Submetido em: 09/11/2020 – Aprovado em: 13/11/2020 – Publicado em: 20/11/2020

---