

# Avaliando a política de Dados abertos no Legislativo brasileiro: análise qualitativa dos *datasets* da Câmara dos Deputados

*Evaluating the open data policy in the Brazilian Legislative: qualitative analysis of the Chamber of Deputies datasets*

R E V I S T A  
com **política**

revista compolítica

2020, vol. 10(1)

compolitica.org/revista

ISSN: 2236-4781

DOI: 10.21878/compolitica.2020.10.1.351

 Open Access Journal

**Sivaldo Pereira da Silva**

Universidade de Brasília (UnB)

**Ébida Rosa dos Santos**

Universidade de Brasília (UnB)

**Leon Eugênio Monteiro Rabelo**

Universidade de Brasília (UnB)

**Mariah Sampaio F. Luciano**

Universidade de Brasília (UnB)

[University of Brasília]

## Resumo

*Este artigo traz uma proposta de avaliação da política de Dados abertos, utilizando como estudo de caso a Câmara dos Deputados. Em termos metodológicos, o trabalho propõe a aplicação de um conjunto de indicadores baseados em duas frentes: (a) análise qualitativa externa (acerca das informações que estão no entorno da publicação) e (b) análise qualitativa interna (que se concentra na avaliação interior dos arquivos). Os resultados demonstram que as bases de dados da Câmara dos Deputados possuem baixa ocorrência de links inoperantes; apresenta bom feedback informativo, porém, na análise interna dos arquivos, apenas cerca de 12% dos recursos avaliados estavam em excelente estruturação, não apresentando erros ou disfunções.*

*Palavras-chave: dados abertos, governo aberto, parlamento aberto, Câmara dos Deputados.*

## Abstract

*This article presents a proposal for evaluating the open data policy, using the Brazilian Chamber of Deputies as a case study. In methodological terms, the work proposes the application of a set of indicators based on two fronts: (a) external qualitative analysis (about the information surrounding the publication) and (b) internal qualitative analysis (evaluating the functionality inside the files). The results demonstrate that the databases of the Chamber of Deputies have low occurrence of inoperative links; it presents good informative feedback, however, about the internal analysis of the archives, only about 12% of the resources evaluated were in excellent structure, without errors or dysfunctions.*

*Keywords: open data, open government, open parliament, Chamber of Deputies.*

## *Avaliando a política de Dados abertos no Legislativo brasileiro: análise qualitativa dos datasets da Câmara dos Deputados*

Sivaldo Pereira da SILVA  
Ébida Rosa dos SANTOS  
Leon Eugênio Monteiro RABELO  
Mariah Sampaio F. LUCIANO

**N**as últimas duas décadas, a transparência tem se tornado uma demanda crescente e ganhou ainda mais notoriedade com as inovações do ambiente digital, principalmente a partir da virada do século. Um dos recursos mais promissores nesse sentido são as políticas de Dados abertos, isto é, a publicação de dados em arquivos estruturados, legíveis por máquinas (algoritmos) e com livre acesso. Isso vem sendo implementado como uma política pública em diversos países, na forma de decretos, leis e outras normativas que reforçam o papel do Estado e seus agentes como curadores de informação de interesse público, tornando-as abertas e on-line de forma perene e regular para que possam ser apropriadas por diferentes finalidades.

Todo esse conjunto converge aos princípios do movimento que vem sendo chamado de Governo aberto. Um Governo aberto é aquele que cria mecanismos institucionais, informacionais e comunicacionais capazes de dirimir constrangimentos e potencializar a relação com seus cidadãos, em questões como transparência, *accountability*, participação e inovação (Chun, 2010; McDermott, 2010; Meijer *et al.*, 2012; Janssen *et al.*, 2012; Wirtza; Birkmeyera, 2015; Attard, 2015; Possamai, 2016; Safarov *et al.*, 2017; Almada; Dominguez, 2017; Máchová; Lnenicka, 2017).

Cumprir assinalar também, em caráter preventivo, que é preciso se fazer uma distinção conceitual entre ‘Governo aberto’ e ‘Dados abertos’. Mesmo sendo os Dados abertos um protocolo tecnológico fundamental para o escrutínio público das instituições públicas, eles não são, evidentemente, sinônimo de Governo aberto. Como bem aponta Yu & Robinson (2012), por mais que sejam áreas interligadas, “a política de governo aberto não

pode ser confundida com a tecnologia de dados abertos”. A tecnologia não trará necessariamente soluções democratizantes, coisa que requer ainda todo um conjunto de ações e encaminhamentos de ordem propriamente política. Deve ser ainda lembrado que, nos desdobramentos contemporâneos das iniciativas de dados abertos, para além do uso democratizante, constata-se com frequência sua apropriação também corporativa, articulando-os de acordo com as lógicas do mercado e com a busca de lucro na nova economia de dados (Birchall, 2018). Há também hoje o debate em curso sobre as implicações adversas da transparência gerada pelos dados abertos, nas quais irrompem questões ligadas a políticas de vigilância e invasão de privacidade, decorrentes do acesso aberto a dados digitais. (Tusikov, 2015; Tyler, 2017; Harcourt, 2018).

De qualquer modo, como aponta Albert Meijer (2013) por mais que seja interessante o “debate normativo” e as muitas definições conceituais a serem estabelecidas a priori, eles não são suficientes para elucidar a dinâmica própria dos fenômenos concretos, que precisam cada um ser estudados em suas particularidades e idiosincrasias empíricas. Muitas vezes, está se falando de problemáticas que, uma vez instauradas dentro de uma determinada arena social, frequentemente mudam as próprias regras dessa mesma arena (Meijer, 2013). A abertura de dados, portanto, se constitui como parte integrante de um amplo movimento de abertura informacional, com as mais diversas consequências, a serem estudadas de perto.

No Legislativo, a concepção de Parlamento aberto soma-se a esse movimento. Além de maior abertura no sentido de mais participação, o adjetivo “aberto” também significa a exigência de publicação de dados aptos para a apropriação social de forma automatizada (levando em conta os avanços tecnológicos disponíveis), sem constrangimentos ou barreiras (Mandelbaum & Swislow, 2013; Faria & Rehbein, 2016).

Tendo em mente este cenário, a questão central que guia este artigo consiste em indagar sobre a qualidade das publicações dos Dados abertos pelo Legislativo brasileiro, especificamente pela Câmara dos Deputados. A pesquisa pretende trazer uma análise sobre indicadores qualitativos aplicáveis à política de Dados abertos, especificamente na

avaliação dos *datasets*<sup>1</sup> de uma das principais casas legislativas brasileiras, na forma de um estudo de caso.

Para alcançar tais objetivos, o trabalho prossegue em duas etapas. Na próxima seção será realizada uma breve discussão conceitual e normativa sobre a concepção de Dados abertos, Governo aberto e Parlamento aberto. Ainda nessa seção serão caracterizados alguns dos principais índices de avaliação da política de Dados abertos hoje em funcionamento e as tendências do campo, para termos assim uma visão mais completa das lacunas e avanços neste cenário. Na segunda seção, o artigo se concentrará em descrever parâmetros e indicadores para a avaliação qualitativa dos *datasets*, tendo como estudo de caso a análise dos Dados abertos publicados pela Câmara dos Deputados.

### *Parlamento e Dados abertos: configurações conceituais e parâmetros normativos*

Para compreendermos o sentido da ideia de “Parlamento aberto” e “Dados abertos”, precisamos antes contextualizar a concepção de “Governo aberto”. A ideia de Governo aberto ganhou força institucional na segunda década deste século, quando surgiu a Parceria para Governo Aberto ou OGP (*Open Government Partnership*). Trata-se de uma iniciativa de âmbito internacional, lançada em 2011 por oito países fundadores (incluindo Brasil e Estados Unidos), cujo objetivo declarado era "difundir e incentivar globalmente práticas governamentais relacionadas à transparência dos governos, ao acesso à informação pública e à participação social"<sup>2</sup>.

Neste mesmo espírito, em 2012, foi criado o movimento Parlamento aberto, inicialmente com o *OpeningParliament.org*, uma articulação transnacional composta principalmente

---

<sup>1</sup> Um *dataset* é um “conjunto de arquivos” que estão relacionados a uma temática ou ênfase similar.

<sup>2</sup> Disponível em: < [governoaberto.cgu.gov.br/](http://governoaberto.cgu.gov.br/)> Acesso, 18 de outubro de 2018. Especificamente, os oito países fundadores (em ordem alfabética): África do Sul, Brasil, Estados Unidos, Filipinas, Indonésia, México, Noruega e Reino Unido. Até o final da segunda década deste século, o número de países signatários já chegava próximo a 80.

por parlamentares de diversos países e organizações civis. A Câmara dos Deputados brasileira estava entre os seus signatários.

Um documento importante que sintetiza essa iniciativa é a *Declaration on Parliamentary Openness*<sup>3</sup> aprovada ainda em 2012, numa conferência que contou com a participação de representantes de 38 países. Um dos objetivos ressaltados pelo documento é a “Promoção de uma cultura de abertura”, onde consideram que:

A informação parlamentar pertence ao público. Toda a informação parlamentar deve poder ser reutilizada ou republicada pelos cidadãos com as exceções estritamente definidas pela lei. Para assegurar uma cultura de abertura, o Parlamento deve adotar medidas para garantir a participação cidadã e uma sociedade civil livre, reforçar a observação efetiva dos trabalhos parlamentares, e proteger vigorosamente estes direitos através do escrutínio efetivo da sua parte (p. i-ii).

A declaração elenca ainda um conjunto de 44 princípios que devem ser seguidos para garantir a abertura e a transparência parlamentar. A relevância da publicação de *datasets* devidamente estruturados, na ideia de Parlamento aberto, está nitidamente visível neste conjunto principiológico: 29 dos 44 princípios estão vinculados aos efeitos advindos de uma boa política de Dados abertos. Dentre estes podemos citar: (a) reconhecer a informação parlamentar como um bem público; (b) fortalecer uma cultura de abertura através de legislação; (d) envolver cidadãos e sociedade civil; (e) assegurar um efetivo controle parlamentar; (f) difundir informação completa; (g) fornecer informação atualizada; (h) assegurar a exatidão da informação; (i) fornecer informações relativas aos membros do Parlamento; (j) informar os cidadãos sobre a agenda parlamentar; (l) fornecer as informações relativas ao orçamento e despesas; (m) conceder livre acesso; (n) fornecer a informação em formatos abertos e estruturados; (o) garantir sua exploração técnica; (p) utilizar formatos não proprietários e *software* livre; (q) permitir recolher a informação para reutilização; (r) utilizar mecanismos de pesquisa simples e estáveis.

Fortalecendo as diretrizes previstas na Declaração, a Parceria para Governo Aberto (OGP) também incorporou o parlamento, em 2013, ao criar em sua estrutura o Grupo de

---

<sup>3</sup> Disponível em diversos idiomas, inclusive em português em <<https://openingparliament.org/declaration/>>. Acesso: 2 de junho de 2019.

Trabalho para Abertura Legislativa (*Legislative Openness Working Group - LOWG*), adotando em 2016 o Guia de Políticas para o Engajamento Parlamentar (*Parliamentary Engagement Policy Guidance*)<sup>4</sup>.

Todas essas iniciativas convergem para uma maior exigência de abertura das casas legislativas em sua crescente complexidade, o que por sua vez requer instrumentos capazes de dar conta do volume e diversidade de dados gerados. Ao ser normativamente guiada por determinados princípios que buscam viabilizar o acesso e o livre reuso de informações públicas, estruturadas para o processamento automático de máquinas, a política de Dados abertos (seja no âmbito do Governo ou Parlamento aberto) é, naturalmente, objeto de avaliações. Surgiram, assim, diversos estudos sobre o desempenho de bons indicadores, geralmente na forma de índices comparativos, com algum tipo de classificação (*ranking*) entre as políticas de Dados abertos, *em sua maioria focada nos poderes executivos nacionais*.

A título de caracterização das formas de avaliação de Dados abertos, podemos sintetizar quatro índices produzidos por diferentes instituições e que podem fornecer um bom panorama de avaliações hoje vigentes: Open Data Barometer (ODB), Global Open Data Index (GODI), Open Data Inventory (ODIN) e OurData Index (Open-Useful-Reusable Data Index).

O Open Data Barometer (ODB)<sup>5</sup> é um índice produzido pela World Wide Web Foundation que elabora e publica um *ranking* de países quanto à performance de indicadores de Governo e Dados abertos. A metodologia é composta basicamente por *survey* com especialistas; com gestores governamentais; e pesquisa aplicada a uma amostra de *datasets*.

O *Global Open Data Index (GODI)*<sup>6</sup> é uma avaliação com publicação anual, executado pela *Open Knowledge Network*, cuja primeira edição ocorreu em 2013. A pergunta principal, que guiou o último relatório (2016/2017), concentrou-se em entender como os

---

<sup>4</sup> Ver em <https://www.opengovpartnership.org/resources/parliamentary-engagement-policy-guidance-2017>

<sup>5</sup> Ver em < [www.opendatabarometer.org](http://www.opendatabarometer.org) > acesso em 9 de junho 2018.

<sup>6</sup> Ver em < [www.index.okfn.org](http://www.index.okfn.org) > acesso em 16 de maio 2018.

governos em todo o mundo publicam dados abertos. Uma característica do GODI é que a avaliação não recai necessariamente sobre "países" e sim sobre "lugares" (*places*)<sup>7</sup>, isto é, opera-se também com algum foco subnacional ou regional.

O *Open Data Inventory* (ODIN)<sup>8</sup> é um índice produzido pela *Open Data Watch* (ODW): uma Organização não-governamental, fundada por especialistas em dados, que visa monitorar e produzir informações e normatizações capazes de guiar o melhor modo de implementação de dados abertos com foco em diversos países. A primeira edição do relatório foi publicada em 2016. Este índice trabalha basicamente com indicadores que buscam avaliar se as ofertas de informação nos portais de dados abertos são de fato abertas e completas. Buscam identificar as áreas temáticas nas quais os dados são publicados e os elementos que apontam maior ou menor abertura dos dados (como existência de termos de uso; opções de *download*; se estão em arquivos legíveis por máquina; se os dados estão publicados em arquivos não-proprietários etc.).

Por fim, um quarto índice que nos serve como exemplo dos diversos mecanismos de acompanhamento e avaliação de Dados abertos é o *OurData Index (Open-Useful-Reusable Data Index)*<sup>9</sup>. É produzido pela OCDE (Organização para a Cooperação e Desenvolvimento Econômico) no âmbito das edições do "*Government at a Glance*". O índice é elaborado a partir de uma metodologia de *survey* (questionário) aplicado aos gestores dos países avaliados.

Em linhas gerais, os índices tratam de indicadores sobre questões como: os temas sobre os quais os dados versam (saúde, segurança, finanças governamentais, poluição etc.); os parâmetros técnicos gerais que garantem a abertura dos dados para leitura automática por máquinas (se estão em arquivos com formatos condizentes); se não existem barreiras para o livre reuso (licenças livres; não restrições ao acesso); se os países possuem ações concretas de incentivo à apropriação social de dados abertos; se possuem legislação

---

<sup>7</sup> Significa dizer que há casos em que um país ou união geopolítica com *status* de Estado podem ser analisados em seus níveis sub-nacionais separadamente para se tentar captar diferenças consideradas relevantes (por exemplo, avaliam a Irlanda do Norte separada da Grã-Bretanha).

<sup>8</sup> Ver em < <http://odin.opendatawatch.com> > acesso em 17 de junho 2018.

<sup>9</sup> <http://www.oecd.org/gov/digital-government/open-government-data.htm>

condizente; se produzem dados com qualidade para o livre reúso; se desenvolvem ações de participação, etc.

Ainda que devam ser considerados importantes por tratarem de avaliações longitudinais e comparativas acerca da evolução dos Dados abertos no mundo, o conjunto de índices hoje vigentes tem duas incompletudes ou fragilidades relevantes: (a) carece de uma efetiva análise qualitativa dos *datasets* em sua estrutura interna e; (b) carece de análises específicas em relação aos Dados abertos do Poder Legislativo (geralmente, privilegiam a avaliação do Poder Executivo em detrimento dos demais Poderes).

Para tentar avançar sobre essas duas lacunas, a próxima seção traz os resultados de um estudo sobre a qualidade dos *datasets* publicados pela Câmara dos Deputados brasileira. O objetivo é contribuir para o amadurecimento de indicadores e análises mais qualitativas das políticas de Dados abertos bem como para o avanço de estudos e índices sobre Parlamento aberto e Dados abertos no Legislativo.

### *Análise qualitativa dos datasets da Câmara dos Deputados*

Do ponto de vista normativo, a avaliação da Política de Dados abertos envolve diferentes dimensões que vão desde a publicação de dados até o processo de interação com o usuário. Ciente deste quadro maior, a proposta do presente estudo não é produzir uma avaliação *in totum* e sim contribuir para compreender parte desse amplo processo avaliativo. Nesta direção, duas frentes de análise qualitativa foram priorizadas em relação à publicação dos *datasets*: (a) *análise qualitativa externa*, isto é, quando a análise recai sobre as informações que estão no entorno da publicação (metadados, significados, interação etc.) e (b) *análise qualitativa interna*, quando a análise se concentra na estrutura da publicação (sobretudo na operacionalidade e adequação técnica dos arquivos em sua estrutura interna)



### 1. Parâmetros metodológicos: indicadores e *corpus*

Para efetivar essas duas frentes de análises qualitativas, foi preciso desenvolver um conjunto de parâmetros capazes de guiar a avaliação proposta de modo sistemático. Por isso, uma das contribuições da pesquisa é também metodológica, pois busca colaborar para a construção de indicadores de avaliação da política de Dados abertos, aplicada ao caso da Câmara dos Deputados, mas que também seja replicável em outras instituições governamentais, tanto no Legislativo quanto em outros Poderes. Neste sentido, convém esclarecer o *corpus* estudado bem como os instrumentos de análise e a forma de aplicação de indicadores nessas duas frentes analíticas.

Em relação ao *corpus* de análise, o estudo trabalhou com o universo total dos *datasets* publicados pela Câmara dos Deputados em seu portal próprio de Dados abertos até julho de 2018 (período de coleta), com organização original dividida em três grandes conjuntos de dados (*datasets*) intitulados: (a) despesas pela cota para exercício da atividade Parlamentar; (b) proposições por ano de apresentação; (c) autores das proposições por ano de apresentação.

Tendo em vista que um *dataset* pode ser constituído por diversos arquivos (recursos), por uma questão de recorte e estratégia metodológica, a análise recaiu sobre o que chamamos de “conteúdo-base” dos *datasets*: trata-se de um pacote de informação específica que pode aparecer - a mesma informação - em diferentes arquivos. Por exemplo, imaginemos um *dataset* que tenha três recursos nos formatos CSV, XML, JSON. Ao analisarmos cada um desses arquivos descobrimos que se trata do mesmo conteúdo publicado em diferentes formatos (por exemplo: a lista com os telefones e endereços de postos de saúde de uma cidade é publicada no formato CSV e também XML). Chamamos esse conteúdo replicado em diferentes tipos de arquivos de *conteúdo-base*. Por esta razão, analisar o mesmo conteúdo repetido em diferentes tipos de arquivo seria inócuo uma vez que os erros dos conteúdos são replicados em todos os arquivos, já que *estes* têm a mesma fonte do original.

Como premissa conceitual, a definição de Dados abertos prevê que sejam arquivos estruturados<sup>10</sup> para leitura automática de máquinas. Por isso, tomamos como procedimento de análise empírica o arquivo CSV como arquivo-padrão,<sup>11</sup> capaz de representar o conteúdo-base pelo fato de ser o padrão de arquivo estruturado mais comum dentro dos padrões de Dados abertos (geralmente, todos os dados estão publicados em CSV). Caso o arquivo CSV não existisse, a análise recairia para o arquivo JSON e, caso este não exista, a terceira opção seria o arquivo XML. Caso não exista CSV, JSON nem mesmo XML, seria analisado o próximo arquivo estruturado existente por ordem alfabética. Nestes termos, o *corpus* a ser analisado somou 162 arquivos (conteúdos-bases).

Quanto aos parâmetros de análise dos *datasets* em relação ao entorno da publicação, os indicadores foram desenhados com base em questões qualitativas que versam sobre o significado dos recursos<sup>12</sup>, seu contexto, orientações para o seu reuso bem como sua capacidade de ser objeto de discussão ou de gerar algum grau de responsividade. Isso foi sintetizado na forma de cinco indicadores conforme detalhamento no Quadro 1:

Quadro 1: Indicadores qualitativos externos

Indicador	O que identifica ou verifica?
Rotulação	Verifica o nível de inteligibilidade do <i>dataset</i> em termos de clareza sobre seu conteúdo, isto é, identifica se possui um rótulo; algo breve, porém capaz de servir como uma etiqueta informativa de entrada que traga detalhes suficientes para o usuário antever o que o arquivo contém sem precisar abri-lo.

<sup>10</sup> Arquivos não-estruturados são recursos publicados sem uma estrutura adequada para leitura automática de algoritmos, isto é, leitura automática por máquinas (exemplo: uma página de HTML, um PDF, um DOCX, um ODT, uma imagem como JPG... são arquivos digitais, mas não estão estruturados para leitura automática. Mesmo que um PDF ou um arquivo de Word traga uma tabela ela não é legível de modo automático por um algoritmo). Quando chamamos de “estruturado” significa que é um arquivo legível por *softwares* de bases de dados. Exemplo: CSV, JSON, XLS, XML.

<sup>11</sup> O arquivo CSV é considerado por diversas normatizações, guias e organizações como W3C um formato-base no qual o dado deve ser disponibilizado por ser um formato leve, simples, universal e não proprietário. Isso não significa que o *dataset* precisa ter apenas formatos CSV: significa que este é considerado um formato-base exigido em qualquer *dataset*, ainda que ele seja enriquecido com outros formatos mais avançados como por exemplo arquivos com notações RDF para websemântica.

<sup>12</sup> Um recurso é uma unidade de arquivo que compõe um *dataset*. Há vários tipos de arquivos por meio dos quais um recurso pode ser publicado (como HTML, CSV, JSON, XML etc).

Parametrização	Verifica se o <i>dataset</i> possui alguma informação autoinstrutiva capaz de explicar seus parâmetros viabilizando assim o seu uso na prática. Isto inclui explicação sobre suas variáveis, metodologias ou outras informações técnicas que vão além da rotulação. Exemplos: manual de dados, cartilha ou arquivo similar com esta função.
Temporalidade	Identifica o <i>timing</i> do <i>dataset</i> ; data de criação e o tempo-base de atualização dos dados (atualização captada no momento da aplicação da planilha).
Discursividade	Verifica se o <i>dataset</i> /recurso é apoiado por algum fórum ou espaço público para discussão e compartilhamento de experiências e usos.
Comunicação	Verifica se o <i>dataset</i> /recurso possui algum mecanismo ativo de <i>feedback</i> informativo entre a curadoria e o usuário e a efetividade de resposta em caso de demanda específica do usuário.

Fonte: próprios autores.

Já a análise qualitativa interna diz respeito à avaliação do interior dos arquivos que compõem os *datasets*. Isso implica em se observar como os arquivos (recursos) são operacionais em suas condições estruturais, bem como a disponibilidade efetiva do arquivo; o tipo de recurso (se de fato está em formato aberto); se o arquivo possui vazios de dados (*missing*), má organização ou anomalias funcionais. O quadro 2 traz uma síntese descritiva destas questões em 5 indicadores:

Quadro 2: Indicadores qualitativos internos

Indicador	O que identifica ou verifica?
Disponibilidade	Verifica se o <i>dataset</i> /recurso possui algum erro que inviabilize a sua acessibilidade como <i>link</i> quebrado; arquivo corrompido, isto é, se o arquivo de fato está disponível para <i>download</i> de modo concreto.
Formatação	Identifica se o <i>dataset</i> está publicado em formatos estruturados legíveis por máquinas; que tipos de formatos; se são formatos não-proprietários que garantam a devida abertura para a apropriação social sem restrições de uso.
Padronização	Verifica se o conteúdo do <i>dataset</i> apresenta qualidade em termos de padrão quanto à morfologia das informações, homogeneidade quanto às nomenclaturas evitando grafias divergentes para uma mesma informação; verifica se o arquivo não possui “sujeiras de dados” (isto é, resquícios de informação que atrapalha a leitura padronizada de informação); trechos

	ilegíveis ou erro de estruturação que compromete o padrão esperado de leitura por máquinas.
Integridade	Verifica se o recurso respeita o princípio da completude dos dados, isto é, se não apresenta lacuna ( <i>missing</i> ) de informação em suas células internas.
Opacidade	Verifica se o arquivo possui variáveis opacas, isto é, colunas devidamente nomeadas e transparentes quanto ao seu real significado para que o usuário possa fazer o correto uso dos dados a que se refere.

*Fonte: próprios autores.*

Todos esses indicadores elencados nos Quadros 1 e 2 foram construídos levando em conta os aspectos normativos que estão nas bases da literatura sobre Dados abertos e da observação empírica acerca dos problemas que podem facilitar ou dificultar a apropriação do ponto de vista do usuário. Neste sentido, uma boa síntese da base normativa nas quais os indicadores estão apoiados são os oito princípios que os Dados abertos devem seguir, perspectiva do Governo Aberto. Os dados devem ser: (1) completos; (2) primários; (3) atuais; (4) acessíveis; (5) processáveis por máquina; (6) com acesso não discriminatório; (7) com formatos não proprietários; (8) livres de licenças. Outros princípios também foram posteriormente adicionados a esta lista: (9) *on-line* e gratuitos; (10) permanentes; (11) confiáveis; (12) com presunção de abertura; (13) documentados; (14) seguros para abrir e; (15) projetado com participação pública<sup>13</sup>. Os indicadores utilizados no presente estudo atravessam boa parte desses princípios.

Para guiar a análise, foi aplicada uma planilha que condensou o conjunto de indicadores dispostos nos Quadros 1 e 2 com o objetivo de orientar a coleta sistemática de dados e organizá-los para tabulação subsequente. Na prática, a planilha reflete uma série de elementos normativos e operacionais, considerando o atual estado de desenvolvimento e

<sup>13</sup> “*The 8 Principles of Open Government Data*” são uma das primeiras iniciativas institucionais que buscou pactuar diretrizes qualitativas para dados abertos governamentais. Foi aprovada em 2007 por um grupo de trabalho multidisciplinar (composto por estudiosos, ativistas, gestores e representantes de organizações civis e empresariais) que se reuniu na Califórnia (EUA) e elaborou inicialmente oito princípios que são hoje amplamente aceitos como atributos basilares sobre o que são Dados abertos. Maiores informações em: <https://opengovdata.org/>

tendências do debate sobre Dados abertos e os padrões internacionais hoje adotados neste campo.

A planilha foi elaborada para ser aplicada em dois passos. No primeiro passo (chamada de Parte I) foi dedicada a analisar o entorno de publicação dos recursos; coletar meta-informações sobre os *datasets*; também coletar algumas informações disponíveis no portal e Dados Abertos da Câmara (conforme indicadores do Quadro 1). O processo de coleta, nesta primeira fase, se deu basicamente a partir de informações publicadas em páginas *web* (HTML) do portal de Dados Abertos da Câmara. Já na segunda etapa de aplicação da Planilha, a coleta de informação passou a ser nos próprios arquivos, que foram abertos e analisados pelos pesquisadores e pesquisadoras. O objetivo foi examinar por dentro, testando-os, observando suas estruturas, suas formas e buscando identificar possíveis erros. Nesta segunda etapa, a análise incidiu especificamente sobre os conteúdos-bases (recursos-bases) que compõem o *dataset*, conforme explicado anteriormente.

## 2. Resultados e discussão

Primeiramente, na análise qualitativa do entorno da publicação dos *datasets* os resultados demonstram que há alguns indicadores razoáveis, mas também algumas fragilidades. Em relação ao indicador de atualização, não há informações claras e coesas no portal de Dados abertos da Câmara sobre a data exata da atualização de cada arquivo publicado. Há uma menção genérica de que os *datasets* têm atualização diária, porém, não existem informações adicionais sobre isso ou sobre quando um determinado arquivo deixou de ser atualizado. Para que haja segurança no reuso dos dados, a publicação precisa informar textualmente o dia e o horário em que houve o último *upgrade* na base de dados. Tendo em vista que nem todos os arquivos precisam necessariamente ser atualizados diariamente (a depender da natureza do conteúdo) é imprescindível que esta descrição de temporalidade esteja prevista em todos os dados publicados.

Quanto aos indicadores de rotulação, para cada *dataset* existe uma descrição rápida, capaz de proporcionar uma noção geral sobre o que se trata, porém, os termos utilizados ainda podem suscitar várias dúvidas, como podemos ver no exemplo da Figura 1:

Figura 1 - Reprodução de tela com rotulação de *dataset*



CÂMARA DOS DEPUTADOS

## Proposições

Proposições por ano de apresentação

Arquivos em que cada registro contém as informações sobre uma proposição apresentada à Câmara dos Deputados para deliberação.

Atualização diária.

Selecione o ano

Selecione o tipo de arquivo

- proposicoes-2009.xml
- proposicoes-2008.csv
- proposicoes-2008.xlsx
- proposicoes-2008.xml
- proposicoes-2007.csv

Fonte: website da Câmara dos Deputados

Neste exemplo, é possível notar que não há um detalhamento suficiente a ponto de caracterizar uma boa rotulação, pois são utilizados termos genéricos como “proposições”, “apresentação”, “deliberação”, não esclarecendo se isso envolve todos os tipos de “proposições” (como requerimentos; projetos de Lei; pedido de abertura de CPI etc.), que tipo de “apresentação” e que tipo de “deliberação” se trata (votação em Comissões? Votação em plenário? Deliberação pela mesa diretora?). Embora os textos sejam curtos e carentes de algumas informações mais gerais e complementares, foi considerado que há o cumprimento mínimo do indicador de rotulação dos recursos, possibilitando assim uma inteligibilidade inicial, isto é, um conhecimento por parte do usuário sobre o que o *dataset* trata ou contém.

Na análise de indicadores de parametrização, isto é, de elementos de inteligibilidade mais aprofundados que tragam parâmetros de variáveis e características técnicas sobre os dados, os resultados apontam para a inexistência deste tipo de referência. O usuário não é capaz de saber como os dados foram coletados, que tipo de mensuração ou de unidades de medida pode encontrar nas variáveis e se estas trazem variáveis do tipo “numérica”, “string”, “float” etc. Ou seja, há uma lacuna de parâmetros técnicos mais detalhados acerca dos dados, que possibilitaria ao usuário compreender melhor como trabalhar o conjunto de dados publicados.

Já os indicadores que envolvem algum nível de interatividade tiveram melhor desempenho. Em termos de discursividade, foi identificada a existência de um fórum para compartilhamento de códigos e experiências sobre os Dados abertos da Câmara no portal *GitHub*,<sup>14</sup> apontando positivamente para este item, embora o volume, o alcance e o impacto do fórum não tenham sido analisados, o que caberia outras pesquisas específicas. Para avaliar o indicador de comunicação, realizou-se um teste de responsividade com o envio de uma mensagem solicitando esclarecimento sobre a variável de uma tabela (em dia útil e horário comercial). Mediu-se o tempo de resposta e se o conteúdo respondia de fato à indagação. O resultado do teste foi avaliado como positivo, com um tempo de resposta considerado rápido (3 horas e 45 minutos após o envio) e o conteúdo recebido foi condizente ao solicitado.

Na análise qualitativa interna dos recursos, isto é, na análise dos arquivos em si, o desempenho dos indicadores variou, tendo alguns itens com boa exequibilidade e outros com performance aquém do esperado. Com desempenho positivo, tivemos o indicador que trata da disponibilidade operacional dos arquivos (funcionamento do acesso ao dado de forma livre sem constrangimentos). Para este aspecto, foi realizado teste de navegação para verificar se os *links* publicados de fato possibilitavam o *download* do arquivo que prometiam. O resultado demonstrou que 100% dos *links* para arquivos CSV (arquivo base para análise, conforme estipulado nos parâmetros metodológicos) estavam operacionais no momento da análise. *Links* quebrados ou inoperantes para arquivos foram detectados,

---

<sup>14</sup> Em <https://github.com/labhackercd/dados-abertos/issues>

porém, isso ocorreu de modo isolado e não expressivo: em apenas 2,4% dos recursos totais testados e apenas vinculados à reprodução de conteúdo no formato de arquivo ODS<sup>15</sup> (sendo que o mesmo conteúdo estava disponível e ativo em arquivo CSV).

Quanto ao indicador de formatação dos arquivos, os dados publicados pela Câmara dos Deputados seguem um mesmo padrão em relação à quantidade de arquivos e aos formatos. Cerca de 95% dos conteúdos-bases são publicados em três formatos: CSV, XLSX, XML (apenas 5% foi publicado em mais de três formatos e nenhum acima de seis formatos). Outros formatos como JSON, GeoJSON não foram identificados na amostra estudada, embora seja possível localizar arquivos em JSON na versão anterior do portal<sup>16</sup> e também na API que o portal disponibiliza. Em linhas gerais, todos os arquivos publicados no portal têm uma versão em formato estruturado, condizente com o mínimo exigido pela concepção de Dados abertos. Não foram identificados arquivos com notações para websemântica (como RDF) o que seria um indicador de avanço em direção ao horizonte dos Dados abertos conectados<sup>17</sup>.

Já um dos indicadores que medem a existência de variáveis com clareza sobre seus significados teve um baixo desempenho. A ocorrência de variáveis opacas (isto é, categorias de dados cujo “nome” não está transparente) foi bastante alta: cerca de 86,4% dos recursos-bases analisados apresentaram algum tipo de variável não inteligível ou que levantava dúvidas sobre o seu real significado<sup>18</sup>. Isso inviabiliza que o usuário possa fazer uso efetivo do recurso, pois em alguns casos não terá certeza sobre o real sentido de uma

---

<sup>15</sup> De modo mais específico, isso foi encontrado em recursos do *dataset* “Despesas pela Cota para Exercício da Atividade Parlamentar” (Exemplo: o recurso <<http://www.camara.leg.br/cotas/Ano-2010.ods>>, relativo a um ano de 2010, leva a uma página informada como “não encontrada” pelo servidor do portal).

<sup>16</sup> Ver em <<http://www2.camara.leg.br/transparencia/dados-abertos/dados-abertos-legislativo>> acesso 25 de julho 2018.

<sup>17</sup> Arquivos com notações para websemântica são hoje considerados a nova fronteira para a web pois avançam em direção a uma web dados (a web tradicional é basicamente uma web de páginas). Trata-se de arquivos com metadados que explicam os próprios dados, levando-os para a possibilidade de leituras semânticas mais complexas.

<sup>18</sup> Uma variável é opaca quando o recurso possui pelo menos uma de suas variáveis não-inteligíveis para o usuário, ou seja, quando não é possível ter certeza sobre o que significa uma variável por não ser intuitivo ou por não haver explicações sobre o seu significado. Em todos esses casos, representa um vazio de informação para o usuário que pode inutilizar o recurso por comprometer a sua integridade ou completude.



parte do *dataset*. Na figura 2 (grifado e enumerado como item “4”), podemos visualizar um exemplo de variável opaca (dentre outros indicadores que veremos adiante):

Figura 2 - Exemplos de lacunas de dados, despadronizantes e variáveis opacas (Onde 1 e 2 são lacunas de dados; 3 é despadronizante; e 4 é variável opaca)

txtCNPJCPF	txtNumero	indTipoDocumento	datEmissao	vlrDocumento	vlrGlosa	vlrLiquido	numMes	numAno
		0	2015-10-17 00:00:00	85,93	0	85,93	9	2015
		0	2015-10-23 00:00:00	40,53	0	40,53	10	2015
		0	2015-11-13 00:00:00	105,46	0	105,46	10	2015
		0	2015-11-13 00:00:00	40,16	0	40,16	11	2015
		0	2015-12-15 00:00:00	99,96	0	99,96	11	2015
		0	2015-12-15 00:00:00	40,44	0	40,44	12	2015
		0	2016-01-12 00:00:00	138,78	0	138,78	12	2015
		0	2016-03-17 00:00:00	2,8	0	2,8	12	2015
		0		168,7	0	168,7	1	2015
		0		153,93	0	153,93	4	2015
		0		125,03	0	125,03	9	2015
		0		115,23	0	115,23	3	2015
		0		107,75	0	107,75	8	2015
		0		80,82	0	80,82	5	2015
		0		80,74	0	80,74	2	2015
		0		65,84	0	65,84	7	2015
		0		64,1	0	64,1	6	2015
		0		36,45	0	36,45	10	2015
33000118000926	00000017224	0	2016-01-01 00:00:00	330,05	2,9	327,15	12	2015
33000118000926	00000017227	0	2016-01-01 00:00:00	74	0	74	12	2015
33000118000926	00000017879	0	2015-11-01 00:00:00	489,38	19,74	469,64	10	2015

Fonte: arquivo CSV do portal de Dados Abertos da Câmara dos Deputados<sup>19</sup>

Indagamos ao setor de curadoria dos dados da Câmara dos Deputados sobre este problema de não-clareza de determinadas variáveis e se havia algum manual ou glossário que esclarecesse tal questão. A resposta obtida explicou que o problema se devia ao processo de migração do portal para a atual versão e que, no passado, “havia uma página no Portal da Câmara que explicava os campos desses arquivos. Essa página acabou se tornando inacessível externamente após um rearranjo do Portal”[...]<sup>20</sup>.

<sup>19</sup> Tela do *dataset* "Despesas pela Cota para Exercício da Atividade Parlamentar", arquivo CSV analisado disponível em < <http://www.camara.leg.br/cotas/Ano-2015.csv.zip> > Acesso em 3 de agosto 2018.

<sup>20</sup> Recebido através de e-mail oficial do órgão gestor responsável pela publicação de Dados abertos da Câmara dos Deputados, em resposta a teste de responsividade para a seguinte indagação: “O que significam as siglas "vlrglosa" que nomeia a coluna R do arquivo do Excel e a sigla "numlote" na coluna Y do arquivo deste link: <<http://www.camara.leg.br/cotas/Ano-2018.xlsx>>. Na verdade, várias siglas dos arquivos da página de Dados abertos não estão claras. Você tem algum manual ou glossário que explique o que cada sigla significa? Se tem, poderiam me passar o link?” Na resposta, o e-mail também informou o *link* (situado fora da página do Portal) onde estão as informações sobre as variáveis: < <https://github.com/labhackercd/dados-abertos/issues/106>>.

A republicação deste “glossário” na página do portal certamente impactaria na ocorrência de variáveis opacas em diversos arquivos. Pelo fato de o glossário não estar publicado no *site* (na atual versão) e, na prática, invisível para o usuário comum em seu processo normal de navegação pelo portal de Dados abertos da Casa, ele foi assim considerado como “inexistente”, uma vez que a análise recai sobre o *dataset* e seu entorno de publicação. Ainda sobre “parametrização”, vale destacar o mérito da Câmara em ter publicado uma API<sup>21</sup> e também uma página<sup>22</sup> com alguns tutoriais com dicas gerais de reuso.

Um outro indicador que demonstrou problemas em sua performance foi a integralidade dos arquivos, isto é, a ocorrência de células com lacuna de dados. Em 89,8% dos recursos-bases estudados detectou-se *missing* (quando o recurso possui claramente ausências de informações), principalmente em casos de tabelas com células vazias em alguma variável. Um exemplo deste tipo de problema pode também ser visualizado na Figura 1, nos itens grifados 1 e 2.

No que se refere ao indicador que analisa a existência de padronização interna dos dados, o estudo detectou que 55,8% dos conteúdos-bases contêm elementos despadronizantes ou grafias divergentes. Como já mencionado no Quadro 2, despadronizantes são resquícios de notações que não têm valor efetivo em tabelas estruturadas cuja existência dificulta que *softwares* e algoritmos possam fazer leitura de forma automatizada<sup>23</sup>. Podemos ver exemplos ilustrativos deste problema na Figura 2, grifo número 3 (onde há sujeiras de dados com valor vazio, despadronizando o formato mais recorrente de datas). De modo

---

<sup>21</sup>Entregando dados em formatos JSON e XML. Disponível em <<https://dadosabertos.camara.leg.br/swagger/api.html>> Acesso 27 de julho 2018.

<sup>22</sup>Especificamente, dois vídeos ensinando como fazer gráficos dinâmicos e dois tutoriais explicando como usar a API e como corrigir a acentuação UTF-8 nos arquivos CSV. Estes tutoriais trazem informações tanto para leigos como para programadores, porém sem aprofundamento a título de ser considerado uma parametrização na forma de cartilha ou manual técnico efetivo para o reuso. Mesmo no caso do texto sobre API, seria o que mais se aproxima de um manual, porém é específico para a API e não para o portal e os recursos como um todo. Disponível em <<https://dadosabertos.camara.leg.br/howtouse/central-tutoriais.html>> Acesso: 28 de julho 2018.

<sup>23</sup> Trata-se de palavras ou números que estão “sobrando” e que não trazem nenhum tipo de informação relevante e despadronizam os dados. Exemplo: um valor de data que tem excessos de zeros os quais deveriam ter valores de horas, minutos, segundos. Porém, são despadronizantes pois tem valor vazio e dificultam a leitura automática do padrão de data. Exemplo: 05/ 10/ 2018 00: 00: 00: 00.

semelhante, mas com foco na falta de uniformidade, a grafia divergente ocorre quando dentro do mesmo arquivo (recurso) os dados (de uma mesma informação) são registrados de formas distintas, quando deveriam ser homogeneizados, conforme exemplos da Figura 3 (ver grifos em círculo, nos quais as mesmas empresas aparecem com nomes grafados “com acento” *versus* “sem acento” ou em formas “por extenso” *versus* “abreviada”):

Figura 3 - Exemplo de grafias divergentes

HP COMERCIO E DERIVADOS DE PETROLEO LTDA
HP COMÉRCIO E DERIVADOS DE PETRÓLEO LTDA
HP COMERCIO E DERIVADOS DE PETROLEO LTDA
HP COMERCIO E DERIVADOS DE PETROLEO LTDA
HP COMERCIO E DERIVADOS DE PETROLEO LTDA.
HP COMERCIO E DERIVADOS DE PETRÓLEO LTDA.
ITAPOAN DIST DE DER DE PET LTDA
J A SOBRAL & CIA LTDA
J A SOBRAL & CIA LTDA
J A DERIVADOS DE PETROLEO LTDA
J A DERIVADOS DE PETROLEO LTDA
J A DERIVADOS DE PETRÓLEO LTDA.
JMF COMERCIO DE COMBUSTIVEIS LTDA
JOÃO CESAR FARJALA GUSMÃO
JSM COMBUSTÍVEIS LTDA
JSM COMBUSTIVEIS LTDA
LUBRIFICANTES GASCOL INDÚSTRIA E COMÉRCIO LTDA
LUBRIFICANTES GASOL INDUSTRIA E CM LTDA
LUBRIFICANTES GASOL INDUSTRIA E COMERCIO LTDA
LUBRIFICANTES GASOL INDÚSTRIA E COMÉRCIO LTDA
MEGA POSTO BROTAS
MEGA POSTO BROTAS LTDA

Fonte: reprodução de coluna de arquivo de Dados abertos da Câmara dos Deputados

Esses elementos afetam a qualidade do *dataset*, uma vez que o recurso com informações despadronizadas ou com grafias de um mesmo dado, que é inscrito em variadas formas, requer maior energia e ferramentas por parte do usuário para limpá-los, tratá-los e uniformizá-los, para somente após isso fazer um efetivo processo de reúso. Em alguns casos, a depender do tipo de informação ou do volume de despadronização, pode inviabilizar o uso de uma base de dados. Outros elementos quanto ao indicador de

padronização - como trechos ilegíveis ou problemas de estrutura no recurso - também foram avaliados, mas não houve identificação de ocorrência deste tipo de disfunção.

### *Considerações finais*

Este artigo teve como principal objetivo avaliar parte da política de Dados abertos da Câmara dos Deputados, mais concentradamente, tratou da qualidade dos *datasets* e suas propriedades voltadas para a apropriação social. Buscou-se contribuir para o desenvolvimento de indicadores e análises mais qualitativas dos *datasets* bem como para o avanço de estudos e índices sobre Parlamento aberto e Dados abertos no Legislativo brasileiro.

Inicialmente, como aporte conceitual e normativo, destacamos que as concepções de Governo aberto e Parlamento aberto são parte de um movimento maior, envolvendo princípios de participação, abertura e transparência. Ao ser normativamente guiada por determinados princípios, que buscam viabilizar o acesso e o livre reuso de informações públicas estruturadas para o processamento automático de máquinas, a política de Dados abertos (seja no âmbito do Governo ou Parlamento aberto) é, naturalmente, objeto de avaliações. Sintetizamos alguns índices que avaliam tal política e demonstramos que eles são, em geral, comparativos e com foco no poder Executivo, apontando assim a necessidade de se produzir mais estudos e avaliações também no âmbito Legislativo.

Em linhas gerais, os resultados demonstraram que os recursos publicados não apresentam erros de *links* inoperantes ou quebrados (detectado apenas em percentual ínfimo). Também positivamente, as análises sobre interatividade demonstraram um rápido *feedback* informativo por parte do gestor no teste de responsividade e a existência de fórum de discussão sobre os *datasets*. Porém, a avaliação qualitativa interna dos *datasets* também demonstrou que apenas cerca de 12% dos recursos avaliados estavam em excelente estruturação, não apresentaram erros ou disfunções. Em todo o resto foi detectado algum tipo de opacidade, lacuna de dados ou outras disfunções. Este quadro aponta para um panorama no qual o usuário terá dificuldades em se apropriar dos dados

publicados de forma qualitativa, havendo problemas e barreiras que impedem ou dificultam o fenômeno da apropriação.

Naturalmente, a apropriação de Dados abertos em todo seu potencial está no uso de *softwares* ou algoritmos para gerar conhecimento a partir do reuso de dados estruturados e isso tem relação direta com o tipo de usuário sobre o qual estamos falando. Não se trata do cidadão comum. O usuário-padrão (tomado como base para esta análise) é aquele que tem habilidades mínimas para lidar com processamento de dados digitais usando ferramentas de Ciência de Dados. Em termos práticos, estamos falando de cientistas de dados, jornalistas de dados, programadores e outros especialistas que lidam com dados digitais através de *softwares* ou algoritmos (que podem trabalhar em diversas instituições como centros de pesquisas, meios de comunicação, empresas, governo, *startups*). Então, digamos, quando um programador, jornalista de dados ou pesquisador se depara com um *dataset* com problemas isso gera uma barreira de apropriação que poderá ou não ser ultrapassada a depender dos recursos que o usuário detém (recursos humanos, expertise, tempo etc.). Por isso, um *dataset* com baixa qualidade para reuso tende a impactar mais no usuário que detém pouco recurso ou expertise para driblar tais problemas. Por exemplo, enquanto elementos não padronizados aumentam o custo de reuso (porque demandam mais energia do usuário em uniformizar dados), outras disfunções como variáveis opacas e lacuna de dado podem aumentar este custo, já que a ausência de inteligibilidade de variáveis torna parte dos dados potencialmente inutilizáveis e a ausência de informação com células vazias pode impossibilitar cruzamentos de informação de modo integral e seguro. Elementos aparentemente técnicos têm impactos reais no processo de apropriação social, pois geram barreiras que só poderão ser movidas por aqueles usuários que detêm expertise ou recursos para tal.

Do ponto de vista metodológico, os indicadores e a análise qualitativa dos *datasets* aplicados neste estudo se mostraram operacionais e podem evoluir em melhor calibragem, especificidade, gradações e abrangência. No horizonte de pesquisa, estudos comparativos (entre casas legislativas ou entre países) e longitudinais (na forma de indicadores específicos para Parlamento Aberto) são caminhos a trilhar.

Importante notar ainda que uma característica da pesquisa é a detecção de disfunções ou erros nos *datasets* baseada em uma análise “boleana”, isto é, na verificação se o “erro” ocorre (*true*) ou não ocorre (*false*) no recurso. Significa dizer que tal método é útil ao apontar se o problema existe ou não no recurso, mas não consegue determinar a sua intensidade, ou seja, quantas vezes o problema aparece dentro do mesmo recurso. Neste sentido, análises qualitativas mais específicas e granulares precisam ser também desenvolvidas no futuro.

## Referências

- ALMADA, M. P.; DOMINGUEZ, M. *E-Transparência e dados abertos governamentais: o ambiente digital como lócus dos avanços do governo brasileiro nas políticas, leis e parcerias internacionais*. Trabalho apresentado no VII Compolítica - Congresso da Associação Brasileira de Pesquisadores em Comunicação e Política. Porto Alegre, 2017.
- ATTARD, J.; ORLANDI, F.; SCERRI, S.; AUER, S. A systematic review of open government data initiatives. *Government Information Quarterly*, v. 32, n. 4, 399-418, 2015.
- BIRCHALL, C. Interrupting Transparency. In: ALLOA, Emmanuel & THOMÄ, Dieter (org). *Transparency, Society and Subjectivity: Critical Perspectives*. Cham: Palgrave-Macmillan, 2018.
- CHUN, Soon Ae *et al.* Government 2.0: Making connections between citizens, data and government. *Information Polity*, 15, 1-9, 2010.
- FARIA, C.; REHBEIN, M. Open parliament policy applied to the Brazilian Chamber of Deputies, *The Journal of Legislative Studies*, 1-20, 2016.
- HARCOURT, B. E. Virtual Transparency: From the Panopticon to the Expository Society and Beyond. In: ALLOA, Emmanuel & THOMÄ, Dieter (org). *Transparency, Society and Subjectivity: Critical Perspectives*. Cham: Palgrave-Macmillan, 2018.
- JANSSEN, M.; CHARALABIDIS, Y.; ZUDERWIJK, A. Benefits, Adoption Barriers and Myths of Open Data and Open Government. *Information Systems Management*, 29 (4), 258-268, 2012.
- MÁCHOVÁ, R.; LNENICKA, M. Evaluating the Quality of Open Data Portals on the National Level. *Journal of Theoretical and Applied Electronic Commerce Research*, 12 (1), 21-41, 2017.
- MANDELBAUM, A.; SWISLOW, D. Definiendo el Parlamento abierto. In: Vielba, I.R.; Gonzalo, M.A.; DOMÍNGUEZ, E. C. (Org.) *Parlamentos abiertos a la sociedad: participación y monitorización*. Madrid: Fundación Ideas, Friedrich Ebert Stiftung, 2013, p. 27-34.
- MCDERMOTT, P. Building open government. *Government*. In: *Information Quarterly*, 27, 401-413, 2010.

MEIJER, A. J.; CURTIN, D.; HILLEBRANDT, M. Open government: connecting vision and voice. *International Review of Administrative Sciences*, v. 78, n. 1, 10–29, 2012.

MEIJER, A. J. Understanding the Complex Dynamics of Transparency. *Public Administration Review*, v. 73, n. 3, p. 429-439, 2013.

OECD. *Government at a Glance 2017*. Paris: OECD Publishing, 2017. Disponível em <[http://dx.doi.org/10.1787/gov\\_glance-2017-en](http://dx.doi.org/10.1787/gov_glance-2017-en)> Acesso. 03 julho 2018.

OECD. *Open Government: The global context and the way forward 2016*. Paris: OECD, 2016. Disponível em <<http://www.oecd.org/gov/open-government.htm>> Acesso. 03 julho 2018.

ONU (Organização das Nações Unidas). *Guide on Lessons for Open Government Data Action Planning for Sustainable Development*. Nova York: ONU, Department of Economic and Social Affairs, 2017. Disponível em <<http://workspace.unpan.org/sites/Internet/Documents/UNPAN97913.pdf>> acesso 13 de abril 2018.

POSSAMAI, A. J. Portal brasileiro de dados abertos: Novas práticas para o fortalecimento da democracia e da gestão pública na Era Digital. In: SILVA, S. P. da; BRAGATTO, R. C.; SAMPAIO, R. C. *Democracia digital, comunicação política e redes: Teoria e prática*. Rio de Janeiro: Letra e Imagem, 2016.

ROBERTS, A. *Blacked Out: government secrecy in the information age*. Nova York: Cambridge University Press, 2006.

SAFAROV, I.; MEIJER, A.; GRIMMELIKHUIJSEN, S. Utilization of open government data: A systematic literature review of types, conditions, effects and users. *Information Polity*, v. 22, n.1, p. 1-24, 2017.

TUSIKOV, N. *Transparency in the Post-Snowden Age of Mass State and Corporate Surveillance*. State University of New York, Buffalo, 2015. Disponível em: [https://www.academia.edu/10619820/Transparency\\_in\\_the\\_Post-Snowden\\_Age\\_of\\_Mass\\_State\\_and\\_Corporate\\_Surveillance](https://www.academia.edu/10619820/Transparency_in_the_Post-Snowden_Age_of_Mass_State_and_Corporate_Surveillance) consultado em 09/02/2020.

TYLER, C. Jeremy Bentham on open government and privacy. In: *Journal of Information Ethics*, v. 26, n.1, Spring, 2017, pg. 112-29. Disponível em [https://www.academia.edu/36586027/Bentham\\_privacy\\_and\\_open\\_government\\_Open\\_Access\\_version.pdf](https://www.academia.edu/36586027/Bentham_privacy_and_open_government_Open_Access_version.pdf) consultado em 09/02/2020.

WIRTZA, B. W.; BIRKMEYERA, S. Open Government: Origin, Development, and Conceptual Perspectives. *International Journal of Public Administration*. 1–16, 2015.

YU, H.; ROBINSON, D. The New Ambiguity of “Open Government” In: *UCLA Law Review Discourse*, 2012. Disponível em <https://www.uclalawreview.org/pdf/discourse/59-11.pdf> consultado em 09/02/2020.

## Agradecimentos

Gostaríamos de agradecer aos colegas integrantes do CTPol pela contribuição e apoio no processo de discussão e testes metodológicos que precederam este estudo (em



ordem alfabética): Ana Thereza Nogueira Soares, Daniel Jorge Teixeira Cesar, Diogo Ataíde Leite Campos, Eduardo de Lima Rodrigues, Luana Ferreira Alves, Maria Eduarda Gomes de Souza, Nicole de Faria Bartolini Mattiello, Pedro Ivo de Sá Guimaraes e Vivian Peron

Uma versão com dados preliminares deste estudo foi apresentada no I Congresso do Instituto Nacional de Ciência e Tecnologia em Democracia Digital, na cidade de Salvador, em setembro de 2018.

### *Os Autores*

Sivaldo Pereira da Silva é professor da Faculdade de Comunicação (FAC) e do Programa de Pós-Graduação em Comunicação da Universidade de Brasília (PPG-FAC-UnB), é doutor em Comunicação e Cultura Contemporânea, coordenador do Centro de Estudos em Comunicação, Tecnologia e Política (CTPol-UnB). E-mail: [sivaldop@unb.br](mailto:sivaldop@unb.br).

Ébida Rosa dos Santos é doutora em Comunicação pela Universidade de Brasília (UnB) e pesquisadora do Centro de Estudos em Comunicação, Tecnologia e Política (CTPol-UnB). E-mail: [ebidasantos@gmail.com](mailto:ebidasantos@gmail.com).

Leon Eugênio Monteiro Rabelo é doutorando do Programa de Pós-Graduação em Comunicação da Universidade de Brasília (PPG-FAC-UnB) e pesquisador do Centro de Estudos em Comunicação, Tecnologia e Política (CTPol-UnB). E-mail: [leon.rabelo@gmail.com](mailto:leon.rabelo@gmail.com).

Mariah Sampaio F. Luciano é mestranda do Programa de Pós-Graduação em Comunicação da Universidade de Brasília (PPG-FAC-UnB) e pesquisadora do Centro de Estudos em Comunicação, Tecnologia e Política (CTPol-UnB). E-mail: [mariah.sfl@gmail.com](mailto:mariah.sfl@gmail.com).

Data de submissão: 10/09/2019

Data de aprovação: 17/03/2020