

**A ABERTURA DE DADOS NAS INSTITUIÇÕES FEDERAIS DE ENSINO SUPERIOR:
UMA ANÁLISE A PARTIR DO PAINEL DE MONITORAMENTO DE DADOS ABERTOS**

ALÉCIO HENRIQUE DANTAS

UNIVERSIDADE FEDERAL DE UBERLÂNDIA (UFU)

JOSÉ EDUARDO FERREIRA LOPES

UNIVERSIDADE FEDERAL DE UBERLÂNDIA (UFU)

VINÍCIUS SILVA PEREIRA

UNIVERSIDADE FEDERAL DE UBERLÂNDIA (UFU)

A ABERTURA DE DADOS NAS INSTITUIÇÕES FEDERAIS DE ENSINO SUPERIOR: UMA ANÁLISE A PARTIR DO PAINEL DE MONITORAMENTO DE DADOS ABERTOS

1. INTRODUÇÃO

Com o avanço das Tecnologias da Informação e Comunicação (TICs), a velocidade com que as informações chegam às pessoas aumenta a cada ano. Estima-se que o crescimento do volume de dados gerados e trafegados na Internet chegará, no ano de 2020, na dimensão de 40 trilhões de gigabytes (ALBUQUERQUE, 2017).

Nesse cenário, os governos vêm passando por uma transformação digital, utilizando cada vez mais as TICs na prestação de serviços aos cidadãos e também como forma de modernizar suas próprias rotinas administrativas (Diniz *et al.*, 2009). Essa digitalização possibilita que os cidadãos acompanhem as ações e resultados alcançados por seus governantes, chegando assim ao que é chamado de controle social e transparência pública.

Uma das formas dos governos promoverem e aumentarem seus níveis de transparência é com a abertura de dados, que são gerados nas mais variadas ações e programas governamentais. Tal iniciativa é denominada de Dados Abertos Governamentais (DAG), sendo uma premissa para que se atinja um maior nível de transparência pública perante os cidadãos e um requisito fundamental para chegar ao denominado Governo Aberto.

Diante disso, uma das iniciativas que visam promover abertura de dados governamentais nos países é a *Open Government Partnership* (OGP) - Parceria para Governo Aberto, fundada em 2011 por 8 países, incluindo o Brasil, a partir da assinatura da Declaração de Governo Aberto. Até o momento da realização do presente estudo, 79 países integravam a parceria.

Visando avançar nessa parceria, o governo brasileiro instituiu, em 2016, a Política de Dados Abertos do Poder Executivo Federal por meio da divulgação do Decreto 8.777/2016. Este, define como objetivos a promoção da publicação de dados, aprimoramento da cultura de transparência pública, fomento do controle social e desenvolvimento de novas tecnologias, integração de serviços públicos, entre outros. Além disso, especifica que a implementação da política se dará por meio da execução de Plano de Dados Abertos (PDA) no âmbito de cada órgão da administração pública federal, direta, autárquica e fundacional, estabelecendo que tais planos deverão ser elaborados e publicados no prazo de 60 dias da data de publicação do decreto, delegando à Controladoria-Geral da União (CGU) a competência de monitorar o cumprimento dos procedimentos e prazos ali dispostos.

Para possibilitar e facilitar o monitoramento da situação dos órgãos quanto à abertura de dados, a CGU lançou, em 2017, o Painel de Monitoramento de Dados Abertos. A ferramenta permite à população verificar e acompanhar a situação dos órgãos do Governo Federal em relação à publicação dos PDAs, assim como a disponibilização das bases de dados de forma aberta, o que aprimora a cultura da transparência pública, fomenta o controle social e promove a geração de novos negócios.

No final de 2018, a Organização para Cooperação e Desenvolvimento Econômico (OCDE) reconheceu os resultados alcançados pela Política de Dados Abertos ao mostrar que o Brasil está bem colocado perante a outros países latino-americanos no estabelecimento de dados abertos, acessíveis e reutilizáveis, o que cria uma grande oportunidade de aceleração da transformação digital no setor público. Apesar disso, é apontado que há necessidade de maior integração e colaboração entre os órgãos, com líderes capacitados a coordenarem e implementarem, estrategicamente, as iniciativas (OECD, 2018). Além disso, de acordo com o projeto *Open Data Barometer* – que mede globalmente como os governos estão publicando e utilizando dados abertos para *accountability*, inovação e impacto social, o Brasil ocupa a 14º

posição no ano de 2019. Dos países latino-americanos, o Brasil está atrás apenas de México, Uruguai e Colômbia.

Nesse contexto, as Instituições Federais de Ensino Superior (IFES) integram os órgãos da administração pública federal autárquicas e fundacionais e, portanto, são atingidas diretamente pelo estabelecido no Decreto 8.777/2016, devendo elaborar e divulgar em seus sítios eletrônicos os respectivos PDAs.

Considerando a contextualização traçada, a pergunta de pesquisa que este estudo pretende responder é "Qual o panorama da adoção da Política de Dados Abertos nas Instituições Federais de Ensino Superior com base no Painel de Monitoramento de Dados Abertos da CGU?".

Desta forma, objetiva-se com esta pesquisa descrever o cenário da adoção da Política de Dados Abertos nas Instituições Federais de Ensino Superior com base no Painel de Monitoramento de Dados Abertos da CGU. De forma adicional, esta pesquisa visa identificar as bases de dados mais frequentes previstas pelas IFES em seus PDAs.

As IFES, cada vez mais, vêm ganhando destaque nos veículos de informação e redes sociais. Ainda, um dos grandes desafios da Administração Pública Brasileira, no século 21, é a promoção de um governo aberto com apoio das TICs (Bueno, Brelàz e Salinas, 2016). Nesse contexto, há uma demanda crescente da população por mais transparência na utilização dos recursos públicos e prestação de contas dos resultados gerados por estas instituições. Com isso, é de fundamental importância investigar como as IFES têm colocado em prática os objetivos de abertura de dados especificados na Política de Dados Abertos, buscando detectar quais bases de dados são mais recorrentes. Optou-se por utilizar os dados contidos no Painel de Monitoramento de Dados Abertos da CGU por ser uma ferramenta centralizadora das informações referentes aos PDAs das instituições e, também, pelo fato da CGU ser o órgão responsável pelo monitoramento da Política de Dados Abertos instituída pelo Decreto nº 8.777/2016.

Este estudo visa contribuir ao preencher uma lacuna teórica referente à Política de Dados Abertos no âmbito das IFES, buscando dar uma visão geral de como se encontra a execução dessa política nas instituições. Adicionalmente, busca-se identificar quais bases de dados as IFES têm dado maior importância na elaboração dos PDAs, visto que não foram encontrados estudos com esse nível de amplitude e que usam como fonte de dados as bases expressas nos PDAs.

Como contribuição prática, este estudo poderá auxiliar os gestores das IFES a mapearem quais bases de dados devam ser priorizadas para abertura de modo que agregue mais valor para a população.

Desta forma, esta seção buscou oferecer uma visão geral ao tema, particularmente no ambiente brasileiro. Na seção seguinte, são apresentados os principais conceitos e fundamentos teóricos necessários para embasamento do estudo. Em seguida, um detalhamento da metodologia é apresentado. Na quarta parte, é realizada a análise dos dados e resultados obtidos. Por fim, são feitas as considerações finais quanto aos resultados alcançados.

2. REFERENCIAL TEÓRICO

2.1. Governo Aberto e Dados Abertos

Quando se fala de Governo Aberto, é inevitável mencionar os dados abertos. Trata-se de dois assuntos ainda recentes e precoces nas pesquisas científicas que buscam investigar conceitos, benefícios e limitações quanto à sua implementação a partir de iniciativas de governo. De acordo com a OCDE, acadêmicos e sociedade civil ainda não chegaram a um consenso sobre uma definição única de Governo Aberto (OECD, 2016).

Ainda conforme OECD (2016), governo aberto é baseado em três princípios: transparência, *accountability* e participação. Tais princípios compõem a definição dada por aquela de que governo aberto é “uma cultura de governança baseada em políticas públicas inovadoras e sustentáveis e práticas inspiradas nos princípios da transparência, *accountability* e participação que promovam o crescimento da democracia e da inclusão” (OECD, 2016, p. 22).

Posto isso, Albano e Reinhard (2015) indicam que as iniciativas de governo aberto geram oportunidades para a utilização de dados abertos. Por sua vez, a Fundação do Conhecimento Aberto (*Open Knowledge Foundation – OKF*) estabelece como definição que dados abertos são aqueles que podem ser usados e reutilizados por qualquer pessoa, para qualquer propósito, além de poderem ser distribuídos, modificados e compartilhados livremente sob a mesma licença, exigindo-se apenas o crédito de autoria. Complementarmente, quando os dados são gerados e disponibilizados pela administração pública, eles são denominados de Dados Governamentais Abertos (DGA).

Nesse sentido, um conceito de DGA é que são dados do setor público disponibilizados livre e gratuitamente, em formatos abertos, acessados publicamente e com facilidade de serem utilizados (Kalampokis, Tambouris e Tarabanis, 2011). Da mesma forma, Janssen, Charalabidis e Zuiderwijk (2012) definem que dados abertos, no escopo do setor público, são dados sem restrições de privacidade e não confidenciais, gerados com recursos públicos e disponibilizados de forma irrestrita quando ao uso e distribuição.

Um primeiro fator positivo observado, quanto à utilização de dados governamentais pelos cidadãos, é o objetivo de aumentar a transparência e controle dos atos do governo (Albano e Reinhard, 2015). Os DGA podem criar inúmeras oportunidades para o desenvolvimento econômico, pois podem promover a criação de novos negócios, disponibilização de serviços públicos mais eficientes e criação de novos postos de trabalho (OECD, 2016; UBALDI, 2013). Além disso, há vários outros benefícios identificados nessas iniciativas, tais como aprimoramento da governança pública, resgate ou aumento da confiança dos cidadãos no governo, criação de um ambiente propício para a inovação, participação e colaboração entre cidadãos e governo, otimização operacional e técnica e responsabilização dos governos, para citar alguns (Janssen, Charalabidis e Zuiderwijk, 2012; OECD, 2016). No entanto, Helbig et. al. (2012), alertam que iniciativas de governo aberto modificam uma porção de fatores técnicos, políticos e sociais. Estes, conforme o autor, causam perturbações ao ambiente que influenciam diretamente a governança, sendo uma possível consequência a atenuação de quaisquer benefícios esperados quanto a iniciativas de governo aberto.

Em relação à governança pública, Silva e Leite (2013) identificaram que uma das práticas de controle social utilizadas no Brasil é a transparência, que contribui para a governança pública. Seguindo esta linha, Ubaldi (2013) aponta que um aumento da transparência pode fortalecer a prestação de contas do setor público com os cidadãos, destacando que os DGAs são fundamentais para que isso ocorra.

Contudo, para que os dados sejam considerados abertos, eles devem seguir alguns princípios e características. Conforme Eaves (2009), que buscou simplificar esses princípios em três pontos, os DGA devem estar indexados para possibilitar que sejam encontrados, devem estar em formato aberto e legíveis por máquinas (computadores, programas, etc.) para serem utilizados e não possuírem restrições legais para compartilhamento dos mesmos. Resumidamente, os cidadãos devem conseguir achar, usar e compartilhar os dados.

No ano de 2007, houve um encontro de defensores de governo aberto com o intuito de formularem um entendimento mais sólido de como DGAs são essenciais para a democracia. Como resultado, foram definidos 8 princípios para abertura de dados governamentais (OPENGOVDATA.ORG, 2007):

1. Completos: Todos os dados públicos devem ser disponibilizados.
2. Primários: Os dados são coletados na fonte, sem sofrerem modificações e sem agregações, ou seja, alta granularidade.
3. Atuais: Os dados devem estar atuais para terem valor.
4. Acessíveis: Devem estar acessíveis ao maior número de pessoas na Internet, para variados propósitos.
5. Processáveis por máquina: Devem estar estruturados para possibilitar processamento automatizado dos dados.
6. Não discriminatório: Devem estar disponíveis para qualquer pessoa, sem necessidade de registro ou formalização para acesso aos dados.
7. Não proprietários: O formato dos dados deve ser tal que nenhuma organização tenha controle exclusivo.
8. Livres de licenças: Os dados não podem estar sujeitos a qualquer tipo de restrição quanto ao uso, reprodução, compartilhamento ou patentes.

Para o *eGov Interest Group* (Grupo de Interesse em Governo Eletrônico) da *World Wide Web Consortium* (W3C), a publicação de dados governamentais deve seguir os seguintes passos: (i) priorizar a publicação dos dados brutos em formatos abertos e não proprietários, como CSV, XML e RDF; (ii) criar um catálogo online dos dados publicados que possibilitem o seu entendimento e (iii) fazer com que os dados sejam legíveis tanto para humanos quanto para máquinas (W3C, 2009).

Como forma de encorajar as pessoas e, em especial, os governos, Berners-Lee (2010) desenvolveu um esquema de avaliação baseado em 5 estrelas, representando níveis que caracterizam como os dados estão divulgados. Ele considera que os dados são abertos quando estão divulgados sob uma licença aberta e que podem ser reusados livremente. Para ter uma estrela - primeiro nível de avaliação - a informação deve ter sido tornada pública, não importando o formato (pode ser inclusive uma foto ou documento digitalizado). Na medida que as estrelas aumentam, os dados vão se tornando mais fáceis de serem utilizados pelas pessoas. Para propiciar um melhor entendimento e clareza, o Quadro 1 apresenta o esquema de avaliação proposto pelo autor.

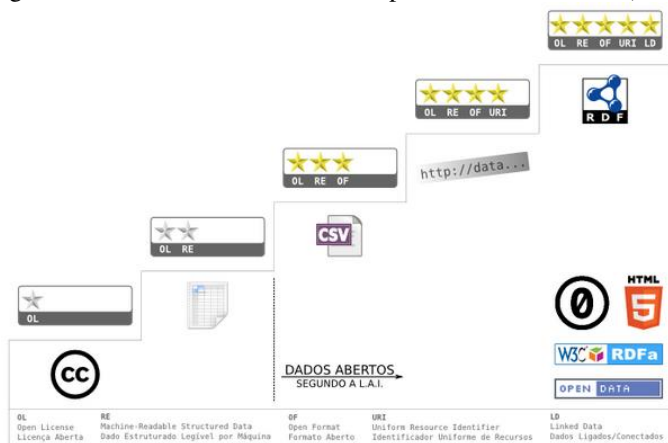
Quadro 1 – Níveis de maturidade de dados abertos

★	Os dados estão na Internet (em qualquer formato) e com licença aberta
★ ★	Dados disponíveis em formatos estruturados e legíveis por máquina (por exemplo, em formato Excel)
★ ★ ★	Igualmente ao nível de 2 estrelas, porém os dados estão em formatos não-proprietários (por exemplo, CSV)
★ ★ ★ ★	Todas as características anteriores (dados na Internet, estruturados e legíveis por máquina, em formato não-proprietário) e utilizando os padrões abertos da W3C (RDF e SPARQL) para identificação das informações
★ ★ ★ ★ ★	Todas as características anteriores e vinculação dos dados para os de outras pessoas (dados linkados)

Fonte: adaptado de Berners-Lee (2010)

Visando aprofundar na proposta de Berners-Lee (2010), Silva (2013) cria uma adaptação do esquema de 5 estrelas - denominado modelo de maturidade - considerando o contexto legal do governo brasileiro, como mostra a Figura 1. O autor frisa que para os dados serem considerados abertos eles devem ter pelo menos 3 estrelas, ou seja, é possível relacionar esse nível do modelo com as características apontadas por Eaves (2009).

Figura 1 – Modelo de maturidade adaptado de Berners-Lee (2010)



Fonte: Silva (2013)

Assim sendo, nota-se que a abertura de dados governamentais perpassa não só dimensões políticas ao promoverem maior transparência, eficiência, aproximação do governo com os cidadãos, aumento do controle social e viabilização de novos negócios, mas também envolve aspectos técnicos e de maturidade na sua operacionalização, que visa disponibilizar os dados para uso efetivo e irrestrito pela população de forma facilitada.

2.2. Legislação Brasileira

No contexto brasileiro, o acesso a informações de órgãos públicos é um direito fundamental expresso no artigo 5º, inciso XXXIII da Constituição Federal Brasileira:

todos têm direito a receber dos órgãos públicos informações de seu interesse particular, ou de interesse coletivo ou geral, que serão prestadas no prazo da lei, sob pena de responsabilidade, ressalvadas aquelas cujo sigilo seja imprescindível à segurança da sociedade e do Estado (BRASIL, 1988).

Para regular o disposto na constituição, foi promulgada em 2011 a Lei nº 12.527, conhecida popularmente como “Lei de Acesso à Informação” (LAI). Em seu artigo 1º, inciso II, é definido que as autarquias e fundações públicas estão subordinadas a essa lei, ou seja, as IFES devem fazer cumprir o estabelecido em seu regimento.

Em relação à natureza das informações a serem acessadas e sua divulgação, o artigo 7º define como direito obter:

- [...]
- IV - informação primária, íntegra, autêntica e atualizada;
 - V - informação sobre atividades exercidas pelos órgãos e entidades, inclusive as relativas à sua política, organização e serviços;
 - VI - informação pertinente à administração do patrimônio público, utilização de recursos públicos, licitação, contratos administrativos (BRASIL, 2011);

E, complementarmente, o artigo 8º que

É dever dos órgãos e entidades públicas promover, independentemente de requerimentos, a divulgação em local de fácil acesso, no âmbito de suas competências, de informações de interesse coletivo ou geral por eles produzidas ou custodiadas.

§ 1º Na divulgação das informações a que se refere o caput, deverão constar, no mínimo:

- I - registro das competências e estrutura organizacional, endereços e telefones das respectivas unidades e horários de atendimento ao público;
- II - registros de quaisquer repasses ou transferências de recursos financeiros;
- III - registros das despesas;
- IV - informações concernentes a procedimentos licitatórios, inclusive os respectivos editais e resultados, bem como a todos os contratos celebrados;
- V - dados gerais para o acompanhamento de programas, ações, projetos e obras de órgãos e entidades; e
- VI - respostas a perguntas mais frequentes da sociedade (BRASIL, 2011).

Portanto, fica evidente que a LAI é um instrumento regulatório que indica o que, como e onde divulgar as informações produzidas pelos órgãos públicos para fomentar a transparência ativa, ou seja, quando não é necessária a solicitação formal por parte da população para a disponibilização de informações.

Quase 5 anos após a divulgação da LAI, em maio de 2016, por meio do Decreto nº 8.777, foi instituída a Política de Dados Abertos, que ofereceu um detalhamento e foco especificamente no que tange a divulgação das bases de dados dos órgãos públicos em formato aberto. O decreto, em seu artigo 2º, entende que

- III - dados abertos - dados acessíveis ao público, representados em meio digital, estruturados em formato aberto, processáveis por máquina, referenciados na internet e disponibilizados sob licença aberta que permita sua livre utilização, consumo ou cruzamento, limitando-se a creditar a autoria ou a fonte;
- IV - formato aberto - formato de arquivo não proprietário, cuja especificação esteja documentada publicamente e seja de livre conhecimento e implementação, livre de patentes ou qualquer outra restrição legal quanto à sua utilização (BRASIL, 2016).

Assim como a LAI, o Decreto nº 8.777/2011 tem como um de seus princípios a publicidade das bases de dados como regra geral e o sigilo como exceção. Além disso, regulamenta que as bases de dados devem ter acesso irrestrito, legíveis por máquina, estarem em formato aberto, permissão para reuso, serem primárias (alta granularidade) e atualizadas periodicamente.

Posteriormente, em julho de 2019, foi divulgado o Decreto nº 9.903, que altera alguns dispositivos do Decreto nº 8.777/2016. As principais alterações se referem à livre utilização pelos Poderes Públicos e pela sociedade dos dados disponibilizados por transparência ativa, permitindo a utilização gratuita das bases de dados quando a União for detentora dos direitos autorais. Além disso, passa para a CGU, exclusivamente, a responsabilidade de divulgação do cronograma de publicação dos Planos de Dados Abertos dos órgãos da administração pública federal direta, autárquica e fundacional. No decreto anterior, esta ação seria realizada por ato conjunto entre o antigo Ministério do Planejamento, Orçamento e Gestão e a CGU (BRASIL, 2019).

Com esse levantamento de legislações brasileiras referentes à transparência, acesso à informação e divulgação de bases de dados, percebe-se que o escopo legal buscou seguir os princípios e abranger as recomendações mencionadas na seção anterior, sendo uma condição necessária para a construção de um ambiente que promova a cultura da transparência, fomente o controle social, desenvolvimento de novas tecnologias, pesquisas científicas e facilite o compartilhamento de informações entre os órgãos públicos.

3. METODOLOGIA

Esta pesquisa é caracterizada como descritiva utilizando-se de dados secundários. De acordo com Gil (2008), uma pesquisa descritiva tem como objetivo estudar e descrever as

características de uma população ou grupo. Como estratégia metodológica, adotou-se a pesquisa de natureza quantitativa. A coleta de dados das IFES foi realizada pela ferramenta Painel de Monitoramento de Dados Abertos da CGU.

Primeiramente, foi realizada uma consulta das informações organizacionais da Administração Pública Federal no Sistema de Informações Organizacionais (SIORG) do Governo Federal, que possui como gestor responsável o Ministério da Economia. A consulta teve como objetivo catalogar todas as Instituições Federais de Ensino Superior vinculadas ao Ministério da Educação, registrando a denominação e sigla de cada unidade organizacional. Nesta etapa, foram catalogadas, ao todo, 101 instituições, sendo 63 Universidades Federais e 38 Institutos Federais de Educação, Ciência e Tecnologia.

Em seguida, para cada IFES, foi realizada uma consulta na ferramenta Painel de Monitoramento de Dados Abertos da CGU, registrando em uma planilha eletrônica os seguintes dados apresentados pela ferramenta: Situação do PDA, ano referente ao cronograma de abertura da instituição, bases de dados previstas para abertura, bases em atraso, bases a serem abertas e bases abertas. Adicionalmente, foram registradas, manualmente, as bases de dados referentes à cada instituição e o indicador de sua situação quanto à abertura (aberta, atrasada ou prevista). Essas coletas de dados foram realizadas de forma manual pois a ferramenta não oferecia um meio de copiar as informações diretamente para outro programa. Além disso, apesar de ser disponibilizada uma funcionalidade para baixar a base de dados do painel, não se optou por utilizá-la, pois foi verificado que os dados baixados estavam desatualizados perante às informações apresentadas pela ferramenta.

Os dados foram coletados no mês de abril de 2019 e são referentes à situação dos PDAs no mês de março de 2019, conforme apresentado pela ferramenta durante a coleta. Finalizada esta etapa, foram registradas, ao todo, 912 bases de dados distribuídas entre as 101 IFES.

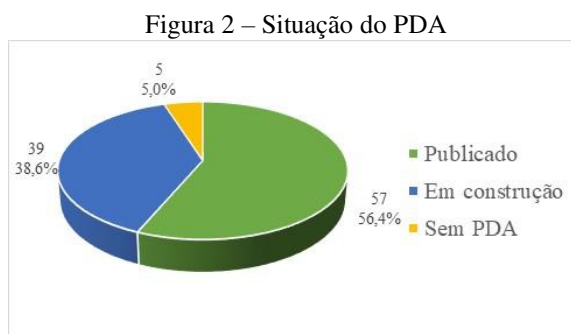
Para a análise dos dados, foram utilizadas análises estatísticas, medidas descritivas de dispersão e tendência central, representação gráfica e técnicas de mineração de texto.

4. DISCUSSÃO E RESULTADOS

A análise dos dados foi realizada em 2 partes. A primeira se refere ao objetivo de descrever o panorama da adoção da Política de Dados Abertos nas IFES, a partir dos dados coletados. A segunda parte busca atingir o objetivo de identificar as bases de dados mais frequentes constantes nos PDAs das IFES.

4.1. Panorama das IFES na adoção da Política de Dados Abertos

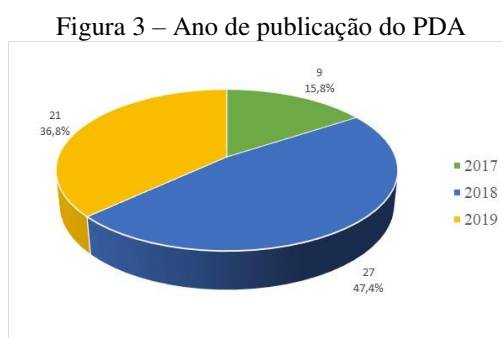
Em relação à situação do PDA das instituições analisadas, a Figura 2 apresenta a quantidade de PDAs quanto à situação, ou seja, se o PDA da IFES é inexistente, está em construção ou está publicado.



Fonte: Dos autores

Nota-se, portanto, que pouco mais da metade das IFES elaboraram seus PDAs e estão publicados em seus sítios eletrônicos, representando 57 instituições. Apenas 5% das IFES ainda não possuem PDA, ou seja, 95% das instituições ou possuem PDA elaborado e publicado ou estão em processo de construção do plano. A partir deste ponto, a análise se dará considerando apenas as 57 IFES que possuem o PDA divulgado.

Em relação ao ano de publicação do PDA, a Figura 3 apresenta a distribuição da quantidade de PDAs das IFES referentes aos anos de 2017, 2018 e 2019.

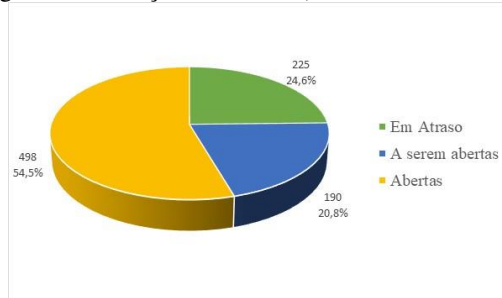


Fonte: Dos autores

Como a Política de Dados Abertos foi instituída, por meio do Decreto nº 8.777/2016, no ano de 2016, é natural que os primeiros PDAs tenham sido publicados no ano seguinte. Percebe-se que poucas IFES (15,8%) possuem seus PDAs vinculados ao ano de 2017 e quase metade delas publicaram seus PDAs no ano de 2018. Para o ano do presente estudo, ou seja, 2019, aproximadamente 37% das IFES possuem PDAs publicados. Pode-se especular, portanto, que aquelas IFES que estão com PDAs datados em 2017, elaboraram e publicaram o plano no ano posterior à legislação e não promoveram sua revisão e atualização. Visto que, passado o ano de 2018, há IFES com PDAs de 2017, pode-se considerar que tais instituições estão atrasadas em relação à revisão e atualização de seus planos. Deste modo, conclui-se que, por volta de 84% das IFES estão com seus PDAs atualizados ou em processo de revisão e atualização.

Quanto ao indicador de abertura das bases, a Figura 4 apresenta a distribuição das bases previstas para abertura nos PDAs quanto à situação, ou seja, se estão atrasadas (prevista, mas não publicada), a serem abertas (prevista e dentro do planejamento para abertura) ou abertas (previstas e publicadas). Vale destacar que uma inconsistência nos dados foi detectada em uma das instituições, resultando em um total de 913 bases, diferente das 912 bases que foram coletadas manualmente no painel.

Figura 4 – Situação das bases (indicador de abertura)



Fonte: Dos autores

A partir da Figura 4, verificou-se que por volta de 54% das bases previstas para abertura encontram-se abertas, ou seja, publicadas. As bases previstas e que estão atrasadas quanto à

abertura representam 24% e as bases previstas que ainda não foram publicadas, contudo estão conforme o planejado, somam quase 21% das IFES.

Com o intuito de contribuir nesta análise, a Tabela 1 apresenta algumas características gerais das 57 IFES que possuem PDAs publicados quanto à variáveis estatísticas descritivas. É possível constatar que, em média, as IFES preveem em seus PDAs 16 bases de dados para abertura, variando entre o mínimo de 2 e máximo de 58. O valor 13 para a mediana aproxima-se da média, indicando uma tendência de distribuição normal (uniforme) dos dados.

Tabela 1 - Estatística descritiva das bases previstas

Medida	Valor
Média por IFES	16,02
Desvio Padrão	10,61
Mediana	13
Mínimo	2
Máximo	58

Fonte: Dos autores

Portanto, de forma geral, o panorama da adoção da Política de Dados Abertos nas IFES caracteriza-se por pouco mais da metade (56%) das instituições possuem seus PDAs elaborados e publicados, sendo que quase 37% destas possuem PDAs divulgados no ano condizente com o presente estudo. Além disso, praticamente a metade das bases de dados previstas para abertura nos PDAs estão abertas e publicadas. E por fim, as variáveis estatísticas descritivas mostram que há uma tendência de distribuição uniforme referente à quantidade de bases previstas para abertura das IFES, com média aproximada de 16 bases por instituição.

4.2. Identificação das bases de dados mais frequentes

Após aplicação de um filtro nas 912 bases coletadas para resultar apenas em registros exclusivos, ou seja, bases de dados com nomes distintos, constatou-se um total de 674 bases. Esta quantidade limita a realização de uma análise mais apurada para identificar as bases de dados mais frequentes e presentes nos PDAs das IFES, pois, como cada instituição elabora o seu próprio PDA, de forma independente e autônoma, mesmo que todas tenham em comum pertencerem à área da Educação Superior, incorre-se no risco de despadroneamento da nomenclatura dada às bases de dados a serem divulgadas, comprovada pela alta quantidade de nomes distintos.

Ademais, durante a coleta, notou-se com frequência a ocorrência de bases referentes aos mesmos conjuntos de dados, mas com denominações ligeiramente diferentes (por exemplo, “Bolsistas” e “Bolsistas (desenvolvimento acadêmico e assistência estudantil)”, “Contratos” e “Contratos Administrativos”). Também houve ocorrências de bases com erros de digitação (por exemplo, “Alucos trancados” no lugar de “Alunos trancados”). Tais fatores colaboraram para que fosse encontrado um alto número de bases com diferentes nomeações.

Diante disso, para viabilizar a análise quantitativa, foi necessário realizar uma categorização das bases de dados de forma a agrupá-las, eliminando, assim, a ocorrência de uma alta quantidade de nomes distintos e também aquelas que possuem erro de digitação.

Essa categorização foi elaborada em dois níveis. O primeiro nível se refere às categorias “Administração”, “Ensino”, “Pesquisa” e “Extensão” e foi motivada por estarem expressas na Constituição Federal do Brasil, em seu Art. 207, onde é definido que “as universidades gozam de autonomia didático-científica, administrativa e de gestão financeira e patrimonial, e obedecerão ao princípio de indissociabilidade entre ensino, pesquisa e extensão.” (BRASIL, 1988). Já para o segundo nível, com o intuito de nortear a definição de subcategorias dentro de

cada uma das categorias do primeiro nível, foi efetuada uma contagem das palavras mais frequentes nas 912 bases de dados das IFES. Para isso, foi utilizado o *software* R com a biblioteca *tm* que possui funções para mineração de texto (*text mining*) e a biblioteca *wordcloud* para gerar uma nuvem de palavras, deixando as mais frequentes no centro da nuvem. Este resultado pode ser conferido na Figura 5.

Figura 5 – Nuvem de palavras contidas nas bases de dados



Fonte: Dos autores

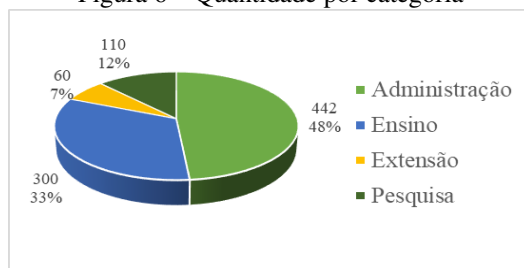
Com isso, é possível constatar, visualmente, que nas bases de dados aparecem, com mais frequência, as seguintes palavras: cursos, dados, servidores, pesquisa, projetos, alunos, extensão, pós-graduação, graduação, relação, contratos e ensino. Outras palavras relevantes que cabe destacar são: licitações, patrimônio, concursos, disciplinas, curriculares, orçamentária, convênios, cargos, terceirizados, bolsas, pessoas, bolsistas, indicadores e docentes. As palavras mais ao extremo da nuvem, apesar de serem menos frequentes, também colaboram para que sejam criadas as subcategorias. Cabe frisar que algumas palavras, apesar de diferentes, são parecidas e se referem ao mesmo assunto, por exemplo “pessoas” e “pessoal”. Portanto, é necessário avaliar com parcimônia todas as ocorrências registradas na nuvem de palavras objetivando a identificação de subcategorias.

Desta forma, usando como referência a nuvem de palavras em conjunto com as características das bases de dados analisadas, chegou-se às categorias e subcategorias apresentadas no Quadro 2.

Assim, a partir das categorias estabelecidas, foi realizada uma análise quantitativa com o objetivo de identificar a frequência de ocorrência em cada uma delas.

Deste modo, a Figura 6 apresenta a quantidade de bases de dados referentes ao primeiro nível de categorização.

Figura 6 – Quantidade por categoria



Fonte: Dos autores

Quadro 2 – Categorização das bases de dados

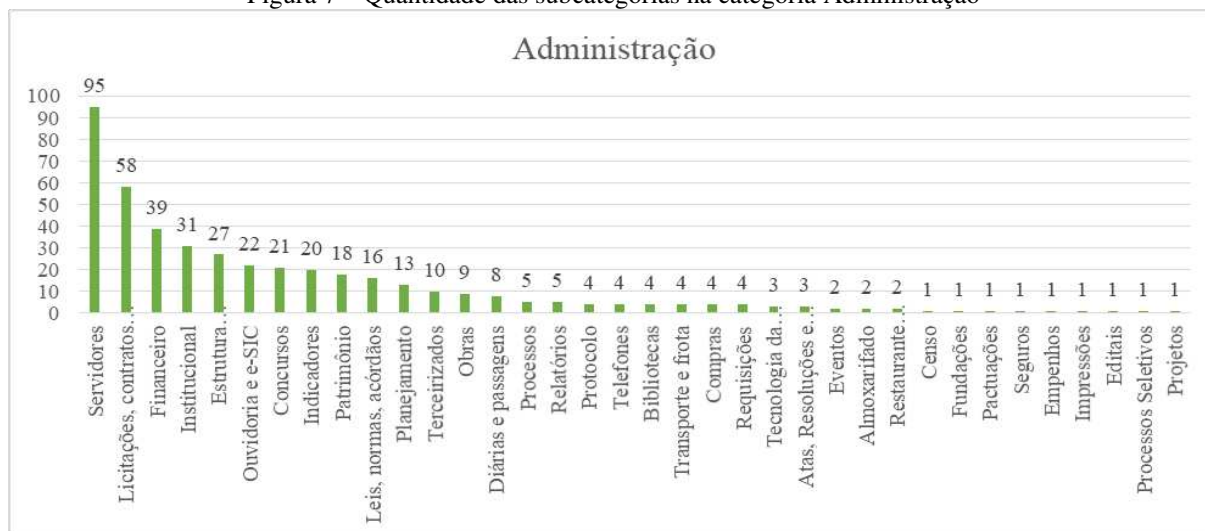
Categoria	Subcategorias
Administração	Servidores; Licitações, contratos e convênios; Financeiro; Institucional; Estrutura; Organizacional; Ouvidoria e e-SIC; Concursos; Indicadores; Patrimônio; Leis, normas, acórdãos; Planejamento; Terceirizados; Obras; Diárias e passagens; Processos; Relatórios; Protocolo; Telefones; Bibliotecas; Transporte e frota; Compras; Requisições; Tecnologia da Informação; Atas, Resoluções e Instruções Normativas; Eventos; Almoxarifado; Restaurante Universitário; Censo; Fundações; Pactuações; Seguros; Empenhos; Impressões; Editais; Processos Seletivos; Projetos
Ensino	Cursos; Alunos; Indicadores; Assistência Estudantil; Docentes; Bolsas; Disciplinas; Processos Seletivos; Egressos; Matrículas; Turmas; Ensino; Ingressantes; Matrizes curriculares; Projetos; Estágio; Bibliotecas; EAD; Mobilidade; Relatórios; Dados acadêmicos sobre o semestre atual; Visitas técnicas; Eventos; Candidatos; Ausências; Empresas juniores; Regulação e Supervisão da Educação Superior; Evadidos; Atividades Acadêmicas; Desligados; Espaço Físico
Pesquisa	Projetos; Grupos; Pesquisadores; Bolsas; Iniciação Científica; Pesquisa; TCCs, dissertações e teses; Produção Acadêmica; Cursos; Patentes; Indicadores; Incubadoras; Defesas; ISSN e ISBN; Relatórios; Eventos; Financeiro
Extensão	Projetos; Extensão; Cursos; Bolsas; Indicadores; Certificados; Produtos; Revista; Eventos; Coordenadores; Financeiro

Fonte: Dos autores

Nota-se que quase metade das bases de dados se referem à categoria Administração (48%), e, em seguida, Ensino (33%), Pesquisa (12%) e Extensão (7%). As categorias Administração e Ensino, quando juntas, representam mais de 80% das bases de dados.

Em relação à categoria Administração, a Figura 7 apresenta a frequência de cada uma de suas subcategorias.

Figura 7 – Quantidade das subcategorias na categoria Administração



Fonte: Dos autores

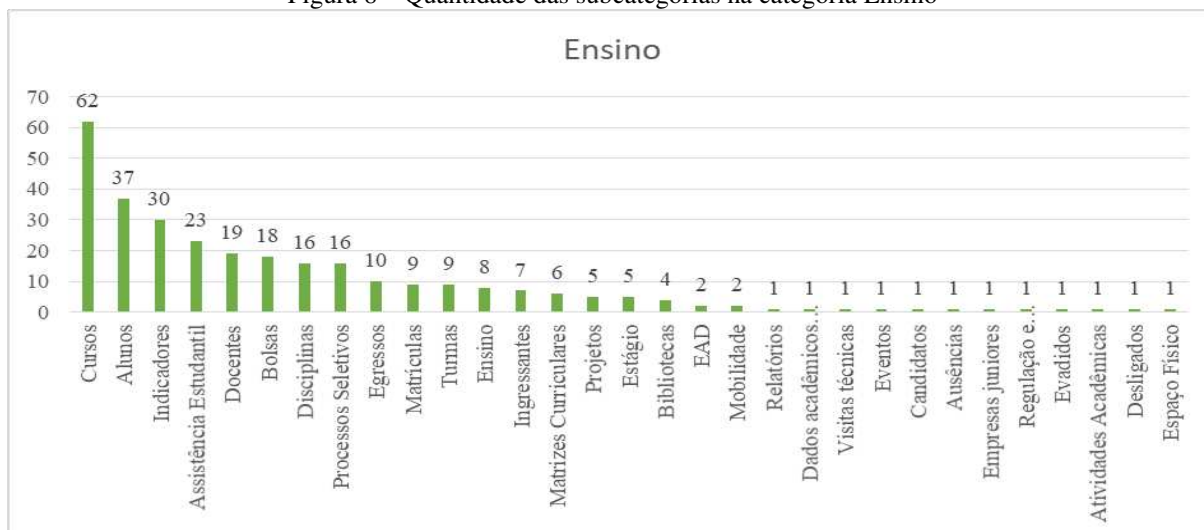
Percebe-se que a maior ocorrência de bases de dados na categoria Administração refere-se a dados de servidores que, no caso das IFES, são compostos por técnicos-administrativos e docentes. Na segunda posição mais frequente, estão dados sobre licitações, contratos e convênios. Além destes, vale destacar uma frequência significativa em bases com dados financeiros, estruturas organizacionais, Ouvidoria e e-SIC e patrimoniais.

Este resultado está em consonância com o recomendado pela legislação, especificamente à Lei de Acesso à Informação (BRASIL, 2011), que indica como informações

mínimas a serem divulgadas aquelas referentes à estrutura organizacional, dados financeiros, licitações, contratos, patrimônio e perguntas mais frequentes da sociedade.

Quanto à categoria Ensino, a Figura 8 apresenta a frequência de cada uma de suas subcategorias.

Figura 8 – Quantidade das subcategorias na categoria Ensino

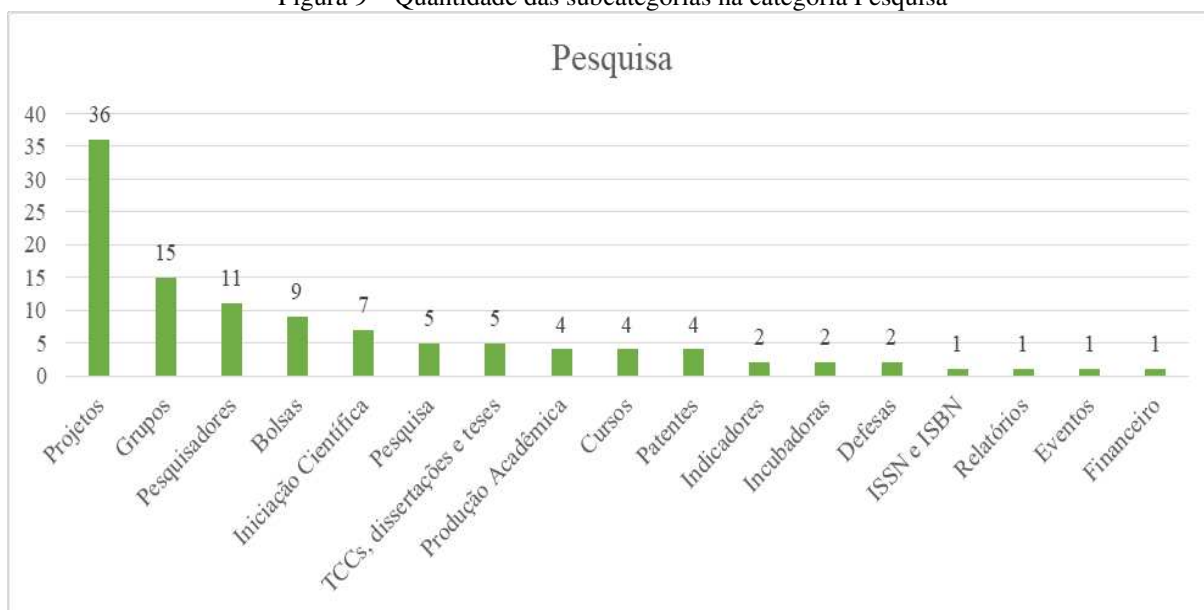


Fonte: Dos autores

A maior frequência registrada dentro da categoria Ensino é a subcategoria Cursos, ou seja, bases de dados relativas a informações sobre os cursos das IFES são as que mais aparecem nos PDAs para abertura. Seguidamente, bases de dados referentes a alunos, indicadores (relacionados à ensino), assistência estudantil, docentes, bolsas, disciplinas e processos seletivos são as que possuem mais ocorrências dentro desta categoria.

Seguindo a análise, em relação à categoria Pesquisa, a Figura 9 apresenta a frequência de cada uma de suas subcategorias.

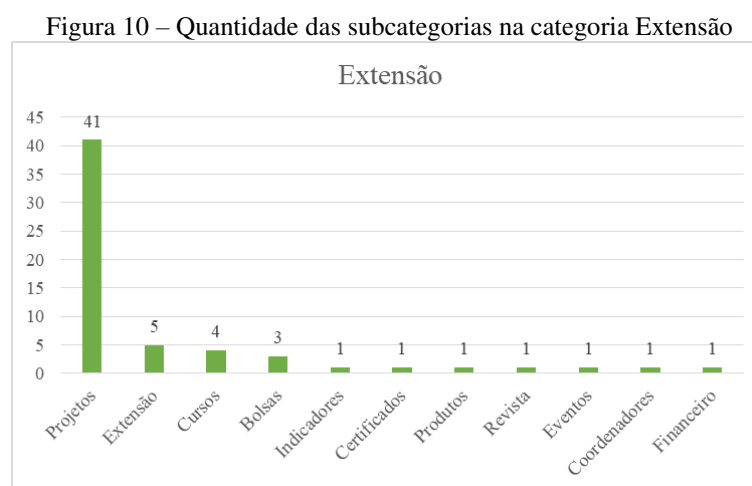
Figura 9 – Quantidade das subcategorias na categoria Pesquisa



Fonte: Dos autores

É evidente, pela Figura 9, que o maior foco das bases de dados da categoria Pesquisa são referentes a informações sobre os projetos de pesquisa das IFES. Dados sobre grupos de pesquisa, pesquisadores e bolsas também possuem certa significância apesar de apresentarem frequência bem inferior a maior registrada.

E finalmente, quanto à categoria Extensão, a Figura 10 apresenta a frequência de cada uma de suas subcategorias.



Fonte: Dos autores

Os dados de projetos de extensão lideram essa categoria, representando aproximadamente dois terços dentre todas as subcategorias registradas. Importante frisar que a subcategoria Projetos contempla tanto projetos quanto ações ou atividades de extensão. Tais nomenclaturas, que se referem a dados de mesma natureza, foram encontradas nas bases de dados analisadas.

Portanto, os resultados indicam que, dentro do tripé ensino, pesquisa e extensão, as IFES oferecem maior atenção para abertura de dados da categoria Ensino, prevalecendo informações relativas aos cursos, alunos, assistência estudantil e docentes. Nas categorias Pesquisa e Extensão, a predominância são de dados referentes aos projetos desenvolvidos em suas respectivas áreas.

5. CONSIDERAÇÕES FINAIS

A pesquisa realizada procurou descrever como está o cenário da adoção da Política de Dados Abertos por parte das IFES. O estudo constatou que pouco mais da metade das IFES elaboraram e divulgaram seus PDAs, sendo que, destas, aproximadamente 84% estão com seus planos atualizados, sendo um indicativo positivo quanto à aderência de parte das IFES em relação à política instituída. Por outro lado, há ainda a necessidade de um significativo avanço para que a totalidade das IFES adotem o estabelecido na política.

Outro importante achado mostra que aproximadamente metade das bases de dados previstas pelas IFES nos PDAs encontram-se abertas e apenas um quarto delas estão atrasadas. Este é outro indicativo que as IFES alinhadas à política têm procurado cumprir o estabelecido nos PDAs, mas que, devido à quantidade de bases em atraso ser significativa, pode representar que as instituições possuem dificuldades ou limitações para a efetiva abertura de dados.

Acredita-se que a identificação das bases de dados mais frequentes presentes nos PDAs das IFES seja a maior contribuição desta pesquisa ao dar um passo inicial das bases mais predominantes nessas instituições, podendo auxiliar os gestores nas decisões estratégicas e

priorização de informações para abertura visando gerar maior valor público diante dos dados disponibilizados. Desta forma, mostrou-se que dados referentes aos cursos, alunos, servidores, licitações, contratos, convênios e financeiros são os mais comuns. Contudo, como foi exposto pelo estudo, há uma notória despadronização da nomenclatura das bases de dados. Uma possível solução para este problema pode vir de recomendações feitas por fóruns ou associações de dirigentes das IFES para seguirem um mesmo vocabulário na definição de seus PDAs.

Considerando que a presente pesquisa não adentrou no quesito da qualidade e maturidade das bases de dados abertas pelas IFES, conforme fundamentos apresentados no referencial teórico, sugere-se como campo para futuros estudos a verificação da utilização dos princípios para abertura de dados governamentais, observando os formatos, estruturação, atualização e níveis de maturidade que os dados disponibilizados representam.

Portanto, as IFES têm caminhado no sentido de atender aos objetivos definidos na Política de Dados Abertos instituída pelo Decreto nº 8.777/2016, viabilizando um ambiente que aprimora a cultura da transparência, favorece a criação e oferta de novos serviços públicos digitais integrados, fomenta a abertura de novos negócios gerando mais empregos e renda, desenvolvimento de inovações tecnológicas e possibilita que a sociedade participe democraticamente da gestão pública através do controle social que a abertura de dados governamentais proporciona. As IFES possuem um papel fundamental no alcance desses benefícios tendo em vista seu papel na sociedade, e devem buscar uma melhora de sua *accountability*, aproximando-se dos cidadãos que, a cada dia, vêm buscando aumentar seus níveis de participação social nos governos, facilitado por meio do uso das TICs.

REFERÊNCIAS

ALBANO, C. S.; REINHARD, N. **Desafios para Governos e Sociedade no Ecossistema Brasileiro de Dados Governamentais Abertos**. Cadernos Gestão Pública e Cidadania, v. 20, n. 67, p. 214–234, 2015.

ALBUQUERQUE, E. M. **O que faremos com os 40 trilhões de gigabytes de dados disponíveis em 2020?**, 2017. Disponível em: <<https://br.okfn.org/2017/09/29/o-que-faremos-com-os-40-trilhoes-de-gigabytes-de-dados-disponiveis-em-2020/>>. Acesso em: 12 jul. 2019.

BERNERS-LEE, T. *Linked Data: Design Issues*. [S.l.]: W3C, 2010. Disponível em: <<http://www.w3.org/DesignIssues/LinkedData.html>>. Acesso em: 10 jul. 2019.

BRASIL. **CONSTITUIÇÃO DA REPÚBLICA FEDERATIVA DO BRASIL DE 1988**, Brasília, DF, out 1988. Disponível em: <http://www.planalto.gov.br/ccivil_03/constituicao/constituicao.htm>. Acesso em: 27 jun. 2019.

BRASIL. LEI Nº 12.527, DE 18 DE NOVEMBRO DE 2011. **Lei de Acesso à Informação**, Brasília, DF, nov 2011. Disponível em: <http://www.planalto.gov.br/ccivil_03/ato2011-2014/2011/lei/l12527.htm>. Acesso em: 27 jun. 2019.

BRASIL. DECRETO Nº 8.777, DE 11 DE MAIO DE 2016. **Institui a Política de Dados Abertos do Poder Executivo federal**, Brasília, DF, mai 2016. Disponível em: <http://www.planalto.gov.br/ccivil_03/ato2015-2018/2016/decreto/d8777.htm>. Acesso em: 27 jun. 2019.

BRASIL. DECRETO Nº 9.903, DE 8 DE JULHO DE 2019. **Altera o Decreto nº 8.777, de 11 de maio de 2016, que institui a Política de Dados Abertos do Poder Executivo federal, para dispor sobre a gestão e os direitos de uso de dados abertos**, Brasília, DF, jul 2019. Disponível em: <http://www.planalto.gov.br/ccivil_03/_Ato2019-2022/2019/Decreto/D9903.htm>. Acesso em: 12 jul. 2019.

BUENO, R. L. P.; BRELÀZ, G.; SALINAS, N. S. C. **Administração Pública Brasileira no Século 21: Seis Grandes Desafios**. Revista do Serviço Público, v. 67, n. Ed. Especial, p. 7–28, 2016.

DINIZ, E. H. *et al.* **O governo eletrônico no Brasil: perspectiva histórica a partir de um modelo estruturado de análise**. Revista de Administração Pública, v. 43, n. 1, p. 23–48, 2009.

EAVES, D. *The three laws of open government data*, 2009. Disponível em <<http://eaves.ca/2009/09/30/three-law-of-open-government-data/>>. Acesso em: 27 jun. 2019.

GIL, A. C. Métodos e técnicas de pesquisa social. 6. ed. São Paulo: Atlas, 2008.

HELBIG, N. *et al.* **Modeling the Informational Relationships between Government and Society: A Pre-Workshop White Paper**. p. 17, 2012.

JANSSEN, M.; CHARALABIDIS, Y.; ZUIDERWIJK, A. *Benefits, Adoption Barriers and Myths of Open Data and Open Government*. *Information Systems Management*, v. 29, n. 4, p. 258–268, 2012.

KALAMPOKIS, E.; TAMBOURIS, E.; TARABANIS, K. *A classification scheme for open government data: towards linking decentralised data*. *International Journal of Web Engineering and Technology*, v. 6, n. 3, p. 266, 2011.

OECD. *Open Government: The Global Context and the Way Forward*, OECD Publishing, Paris. <http://dx.doi.org/10.1787/9789264268104-en>, 2016.

OECD. *Digital Government Review of Brazil: Towards the Digital Transformation of the Public Sector*, *OECD Digital Government Studies*, OECD Publishing, Paris, <https://doi.org/10.1787/9789264307636-en>, 2018.

OPENGOVDATA.ORG, *Eight principles of open government data*, 2007. Disponível em: <<https://opengovdata.org/>>. Acesso em: 13 jul. 2019.

SILVA, F.; LEITE, H. **Práticas de controle social na percepção do poder executivo: um estudo dos entes federativos brasileiros**. Revista Ciências Administrativas, v. 19, n. 2, p. 419–445, 20 dez. 2013.

SILVA, N. **Maturidade em Dados Abertos: Entenda as 5 Estrelas**. 2013. Disponível em: <<https://br.okfn.org/2013/01/17/maturidade-em-dados-abertos-entenda-as-5-estrelas/>>. Acesso em: 11 jul. 2019.

UBALDI, B. *Open Government Data: Towards Empirical Analysis of Open Government Data Initiatives*, *OECD Working Papers on Public Governance*, No. 22, OECD Publishing, Paris, <https://doi.org/10.1787/5k46bj4f03s7-en>, 2013.

W3C, *Publishing open government data*, *W3C Working draft*, 2009. Disponível em: <http://www.w3.org/TR/gov-data/>. Acesso em 4 jul. 2019.