

DADOS ABERTOS GOVERNAMENTAIS BRASILEIROS NA CIÊNCIA: UM ESTUDO BIBLIOMÉTRICO NA BASE SCOPUS

Larriza Thurler

Doutora em Ciência da Informação
Universidade Federal do Rio de Janeiro (UFRJ)
Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT)
E-mail: larriza@gmail.com

Fabio Castro Gouveia

Doutor em Química Biológica (Educação, Gestão e Difusão em Biociências)
Fundação Oswaldo Cruz (Fiocruz)
E-mail: fgouveia@gmail.com

INTRODUÇÃO

Diversas iniciativas a nível global estão sendo desenvolvidas para estimular a abertura dos dados, em especial na esfera pública, a fim de promover gestões mais participativas. O Brasil segue algumas normas e compromissos internacionais para o avanço nesta temática. Uma delas é a Parceria para Governo Aberto (OGP, na sigla em inglês). Lançada em 2011, a OGP é uma iniciativa para difundir e incentivar práticas governamentais relacionadas à transparência dos governos, ao acesso à informação pública e à participação social.

Ainda há muito a ser feito, mas já foram dados passos significativos no Brasil, especificamente no que tange à abertura dos dados governamentais, que oferecem um grande potencial para cidadãos, empresas, instituições de pesquisa e ensino, organizações não-governamentais, entre outros. Entretanto, esses dados parecem ser ainda pouco explorados, em grande parte, devido ao gap digital. Segundo a TIC Domicílios 2018, 33% dos domicílios brasileiros ainda não têm acesso à internet. (COMITÊ GESTOR DA INTERNET NO BRASIL, 2019)

Mesmo entre aqueles que têm acesso à internet, há a questão da falta de letramento de dados, ou seja, não há conhecimentos para acesso, uso, tratamento e apresentação de dados por parte dos cidadãos. Letramento de dados implica transdisciplinaridade e facetas técnicas, psicológicas, sociais e discursivas. (BUZATO, TORRES, 2019) Tampouco há competência em informação, que se refere à capacidade de “reconhecer uma necessidade de informação e ter a habilidade de localizar, avaliar e usar efetivamente a informação necessária”. (AMERICAN LIBRARY ASSOCIATION, 1989)

A academia pode desempenhar um papel relevante nesta agenda, contribuindo com o maior engajamento da sociedade nas ações de governo aberto no Brasil. As universidades podem, por exemplo, desenvolver pesquisas bem fundamentadas com os dados abertos a fim de analisar o impacto das políticas públicas na sociedade; avaliar a eficiência e a efetividade dos serviços públicos; monitorar a atuação do governo; usar, cruzar e reusar os dados públicos abertos; e, com isso, auxiliar na percepção de valor de um governo aberto por parte da sociedade, de modo a gerar maior participação política dos cidadãos.

Este artigo se propõe a investigar como as pesquisas científicas têm usado dados abertos governamentais brasileiros. Trata-se de um estudo bibliométrico sobre a produção científica que tenha usado como fonte de dados o Portal Brasileiro de Dados Abertos, com uma abordagem quantitativa e natureza descritiva. Esta análise pode ser relevante para subsidiar políticas de incentivo à pesquisa usando dados abertos, possibilidades de parcerias com instituições privadas ou do terceiro setor e ainda o próprio governo. Além disso, o governo pode avaliar se o Portal está sendo aproveitado plenamente e que ações tomar para que os dados sejam usados de maneira efetiva e de forma crítica.

DADOS ABERTOS GOVERNAMENTAIS

Os dados abertos governamentais são parte de um movimento mais amplo de governo aberto, que teve início no começo do século XXI. De acordo com o artigo 2º do Decreto nº 8.777, de 11 de maio de 2016, que institui a Política de Dados Abertos do Poder Executivo federal no Brasil e estabelece normas para a livre utilização das bases de dados, dados abertos

[...] são 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. Esses dados não podem ter restrição de patentes ou mecanismos de controle. (BRASIL, 2016, p. 21)

Para Tauberer (2014), os dados abertos governamentais impactaram a maneira como o governo opera, envolvendo mais indivíduos para fazer com que o governo seja mais acessível e atraente, por meio da transformação dos dados brutos em algo novo. Pesquisadores podem desempenhar, portanto, um papel relevante de mediador, extraindo valor dos dados abertos governamentais e articulando as interações entre o governo e seus cidadãos.

Atualmente, 75% das bases de dados do governo federal estão abertas. A fim de centralizar a busca e o acesso aos dados e informações públicas, o governo desenvolveu o

Portal Brasileiro de Dados Abertos.¹ O objetivo da plataforma é oferecer um serviço simplificado que organiza e padroniza o acesso aos dados públicos. Ela reúne atualmente mais de sete mil conjuntos de dados, de 170 organizações, divididos em 21 grupos temáticos relacionados à administração pública, tais como governo e política, saúde, educação, meio ambiente, dentre outros. Cada conjunto de dados possui uma descrição, um ou mais recursos, e uma série de outros metadados, como periodicidade de atualização e órgão responsável.

A PRODUÇÃO CIENTÍFICA A PARTIR DO DADOS.GOV.BR

O universo da pesquisa é a produção científica indexada na base Scopus que tenha usado como fonte de dados o Portal Brasileiro de Dados Abertos. A Scopus é o maior banco de dados de resumos e citações da literatura com revisão por pares: revistas científicas, livros, anais de congressos e publicações, e reúne mais de 60 milhões de registros. A base foi escolhida também por permitir recuperar as publicações cujas referências citam uma URL.

Metodologia

O parâmetro de busca foi o código “WEBSITE”² aplicado à busca avançada da base Scopus no dia 16 de janeiro de 2020. Deste modo, os resultados trouxeram pesquisas que usaram o conteúdo disponível no Portal Brasileiro de Dados Abertos. Considerando que a versão beta do portal foi lançada em dezembro de 2011 e a oficial em maio de 2012, optou-se por delimitar a abrangência temporal de 2012 a 2019.

Resultados: apresentação e discussão

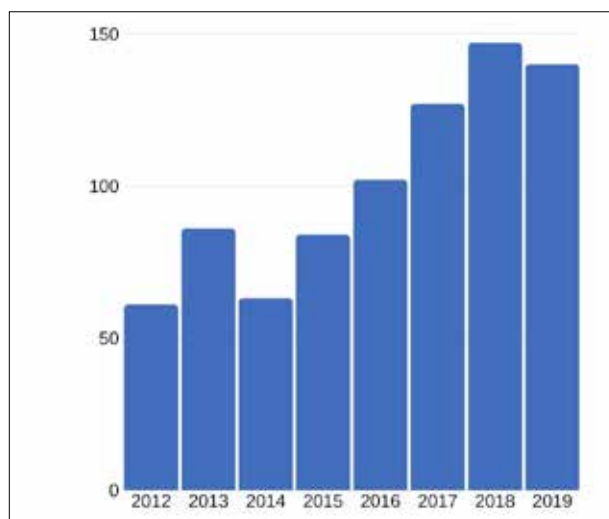
A estratégia de busca resgatou 810 documentos – 664 artigos em periódicos, 47 revisões, 43 livros, 30 capítulos de livros, 25 artigos em conferências e 1 carta ao editor. Os resultados obtidos foram exportados em formato .csv para serem analisados no Open Refine e no software VOSviewer. (VAN ECK; WALTMAN, 2010)

Em relação à evolução temporal, a quantidade de estudos publicados aumentou ao longo dos anos (Gráfico 1). Parte deste aumento pode ser justificado pelo crescimento do volume de documentos indexados na base, mas além disso observou-se uma queda em 2014, ainda por ser estudada. Cabe ressaltar que a busca foi realizada no começo de janeiro de 2020 e que, portanto, o ano de 2019 pode ainda vir a crescer em número de documentos indexados.

¹ Ver: www.dados.gov.br.

² Ver também: www.dados.gov.br.

Gráfico 1 – Evolução temporal dos estudos



Fonte: elaborado pelos autores.

Quanto ao idioma, a maioria dos documentos, como esperado, foi em inglês (505), mas seguida de português (285), provavelmente pelo interesse local, e espanhol (18) e francês (4), ressaltando que há publicações em mais de um idioma. Os cinco periódicos com mais artigos foram: *Revista de Administração Pública*, da Fundação Getúlio Vargas (31); *Ciência e Saúde Coletiva*, da Associação Brasileira de Saúde Coletiva (25); *Revista de Sociologia e Política*, da Universidade Federal do Paraná (24); *Dados*, do Instituto de Estudos Sociais e Políticos da Universidade do Estado do Rio de Janeiro (17) – todas brasileiras – e *Espacios*, da Venezuela (13).

O grafo de redes de coautoria por países é composto por 38 nós, considerando o mínimo de um documento por país. Há destaque para o Brasil, como esperado, seguido dos Estados Unidos e do Reino Unido (Figura 1). Foi utilizado o método *full-counting*, que atribui o mesmo peso aos nós em coocorrência. No grafo podemos observar a diversidade de países que usam dados abertos brasileiros em suas pesquisas, em alguns casos para análises comparativas entre os portais de dados abertos ou uso de dados abertos.

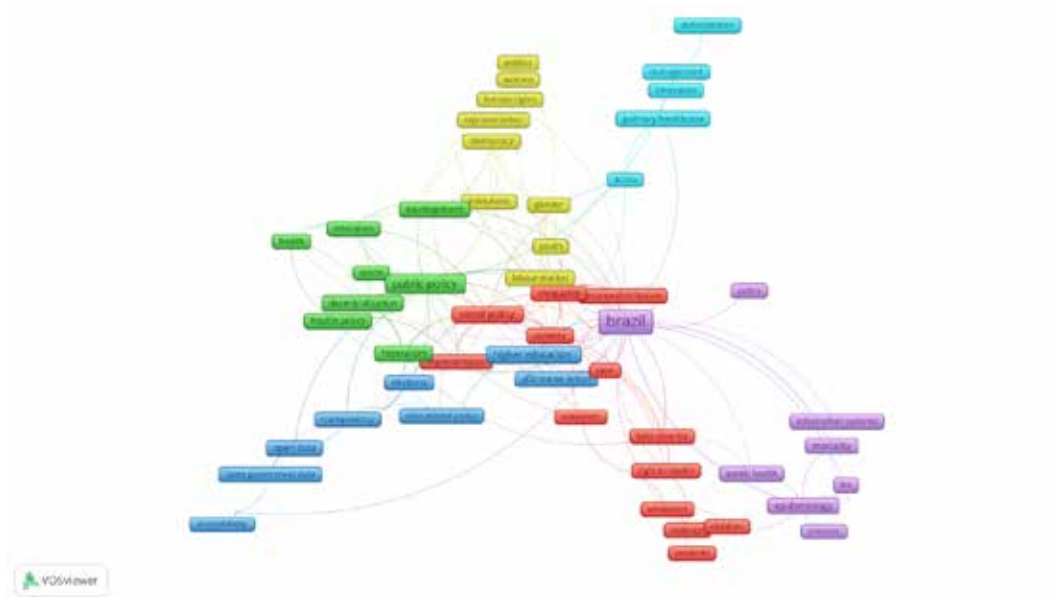
Figura 1 – Redes de coautoria por países



Fonte: elaborado pelos autores no VOSviewer (2010).

No grafo de coocorrência de palavras-chave atribuídas pelos autores (*author keyword*) na visualização “network” evidenciam-se seis clusters temáticos. Foi utilizado o método *fractional-counting*, visando compensar possíveis variações no número de palavras-chave atribuídas pelos autores. Considerou-se quatro como o mínimo de ocorrências de uma palavra-chave, assim, exigindo uma média de ao menos uma vez a cada dois anos do período analisado. A aplicação de um arquivo de thesaurus para tematização resultou em 2393 palavras-chave, sendo que 66 atingiram o critério mínimo de frequência. As palavras-chave “rio de janeiro” e “territory” não estavam conectadas e foram descartadas restando um componente gigante com 64 termos. A Figura 2 contém o grafo onde a resolução foi ajustada para 0.40 para melhor delineamento dos seis agrupamentos formados.

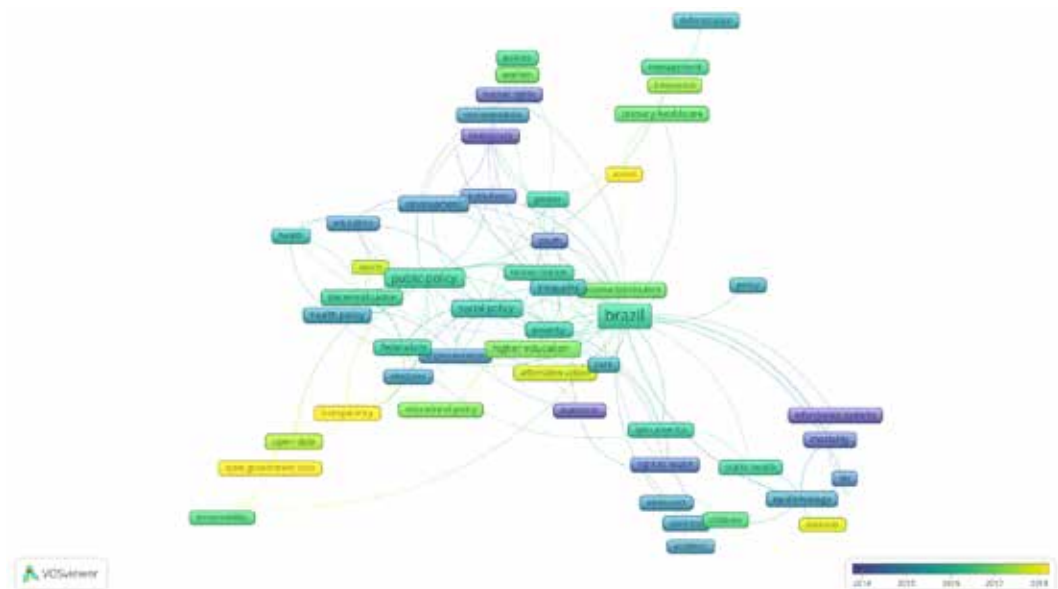
Figura 2 – Redes de coocorrência de termos (palavras-chave de autores)



Fonte: elaborado pelos autores no VOSviewer (2010).

Já na Figura 3 temos a visualização com cores atribuídas pelo ano médio de aparição da palavra-chave. Por se tratar do mesmo *layout* da Figura 2 podemos estabelecer quais as palavras-chave/temáticas são mais recentes para cada um dos agrupamentos observados.

Figura 3 – Redes de coocorrência de termos ao longo dos anos (palavras-chave de autores)



Fonte: elaborado pelos autores no VOSviewer (2010).

Os grafos revelam que as pesquisas encontradas estão centradas nos dados do Brasil, como esperado, e relacionados a políticas públicas. Há seis agrupamentos mais definidos de pesquisas: (a) em azul escuro estão as pesquisas relacionadas às questões governamentais, abordando temáticas como eleições, políticas educacionais, transparência, dados abertos governamentais, *accountability* e ação afirmativa; (b) em vermelho, às questões sociais, como raça, políticas sociais, pobreza, distribuição de renda, adolescentes, crianças, violência e direito à saúde; (c) em roxo, às questões de saúde, como por exemplo políticas de saúde, mortalidade, HIV, saúde pública, epidemiologia e zoonoses; (d) em verde, à gestão pública, envolvendo temas como educação, saúde, esportes, desenvolvimento, federalismo; (e) em amarelo, aos direitos humanos, tais como direitos humanos, democracia, gênero, juventude, mulheres, mercado de trabalho, juventude, representação; e (f) em azul claro, relacionados à gestão e inovação. As palavras-chave mais recentes, de 2017 a 2019, estão relacionadas à transparência, dados abertos governamentais e acesso, e entre 2016 e 2017, a temas relativos a saúde, educação e gestão pública.

CONSIDERAÇÕES FINAIS

A abertura de dados governamentais está baseada na ideia de envolvimento de todos os setores da sociedade – cidadãos, iniciativa privada, academia e o próprio governo – para possibilitar uma maior apropriação destes dados. Nesse contexto, as universidades desempenham um papel fundamental para fazer com o que cidadão perceba o valor dos dados governamentais abertos e o impacto no seu cotidiano.

A busca na base Scopus revelou que o número de estudos utilizando dados do Portal Brasileiro de Dados Abertos vem aumentando ao longo dos anos. Além disso, tais pesquisas estão centradas especialmente em temas relacionados à administração pública e saúde. Mais recentemente, temas como governo aberto, transparência, políticas públicas e saúde vêm sendo mais pesquisados. Um estudo complementar relevante seria avaliar como esses estudos têm chegado à população e impactado a percepção da sociedade, para que seja estimulado o uso dos dados tanto para desenvolvimento de aplicativos ou sites quanto para pressionar o governo para melhoria de serviços prioritários.

AGRADECIMENTOS

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (Capes) – Código de Financiamento 001 e Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Processo 430982/2018-6, chamada universal MCTIC/CNPQ nº 28/2018.

REFERÊNCIAS

AMERICAN LIBRARY ASSOCIATION. Presidential committee on information literacy: final report. ALA, Washington, DC, 1989. Disponível em: <http://www.ala.org/acrl/publications/whitepapers/presidential>. Acesso em: 26 fev. 2020.

BRASIL. Decreto nº 8777, de 11 de maio de 2016. Institui a Política de Dados Abertos do Poder Executivo federal. *Diário Oficial da União*: seção 1, Brasília, DF, n. 90, p. 21, 12 maio 2016. Disponível em: http://www.in.gov.br/materia/-/asset_publisher/KujrwoT7ZC2Mb/content/id/21520099/doi-2016-05-12-decreto-n-8-777-de-11-de-maio-de-2016-21520002. Acesso em: 26 fev. 2020.

BUZATO, M. E. K.; TORRES, C. C. Dadificação e transdisciplinaridade nos estudos do letramento: o jornalismo de dados como contexto de pesquisa. *Revista da Anpoll*, Brasília, DF, v. 1, n. 49, p. 128-141, 2019.

COMITÊ GESTOR DA INTERNET NO BRASIL. *Pesquisa sobre o uso das tecnologias de informação e comunicação nos domicílios brasileiros: TIC domicílios 2018*. São Paulo: Comitê Gestor da Internet no Brasil, 2019. Disponível em <https://www.cetic.br/publicacao/pesquisa-sobre-o-uso-das-tecnologias-de-informacao-e-comunicacao-nos-domicilios-brasileiros-tic-domicilios-2018/>. Acesso em: 18 nov. 2019.

TAUBERER, J. *Open Government Data*. [S. l.]: Amazon Digital, 2014.

VAN ECK, N. J.; WALTMAN, L. Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics*, London, v. 84, n. 2, p. 523-538, 2010.