

Jornalismo de dados e a qualidade dos dados abertos governamentais no Brasil: problemas, desafios e obstáculos nos processos de apropriação profissional [*Data Journalism and Quality of Open Government Data in Brazil: Issues, Challenges and Obstacles in The Professional Appropriation Process*]

REVISTA
compolítica

Revista Compolítica
2022, 12(2), p. 5-32
compolitica.org/revista
ISSN: 2236-4781

DOI: 10.21878/compolitica.2022.12.2.598

Sivaldo Pereira da Silva

Universidade de Brasília (UnB)
[University of Brasilia]

Maria Dominguez

Instituto Nacional de Ciência e Tecnologia em Democracia Digital (INCT.DD)
[National Institute of Science and Technology in Digital Democracy]

Ébida Santos

Universidade de Brasília (UnB)
[University of Brasilia]

Patrícia Lima

Universidade Federal de Santa Catarina (UnB)
[Federal University of Santa Catarina]

Resumo

Este artigo tem como objetivo caracterizar a experiência de jornalistas no reuso de dados abertos governamentais para a produção de conteúdos. O trabalho traz uma pesquisa qualificada com 51 profissionais com dados coletados a partir da aplicação de survey e entrevista estruturada, com análise exploratória do tipo quantiquantitativa. Os resultados demonstram que “arquivos desatualizados”, “falta de inteligibilidade” e “informação incompleta” são as disfunções mais recorrentes e destacadas como um grande problema para mais de 60% dos entrevistados. Cerca de 94% dos entrevistados afirmaram que já desistiram de uma pauta devido a problemas na qualidade dos datasets. O estudo conclui haver obstáculos reais no nível prático que afeta a qualidade das informações publicadas e seu efetivo reuso.

Palavras-chave: Jornalismo de Dados; Dados Abertos Governamentais; Transparência.

Abstract

This article has the objective to characterize the journalists' experience reusing open government data to produce journalistic content. In order to understand this interface, this paper proposes a research based on a survey and semi-structured interviews with 51 professionals, and an quantitative and qualitative exploratory analysis. The results show that “outdated files,” “lack of intelligibility” and “incomplete information” are the most frequent dysfunctions, mentioned as problems to more than 60% of the journalists interviewed. Around 94% have already gave up on a journalistic agenda due to the datasets' quality problems. This study concludes that there are significant practical obstacles that interferes on the quality of the published information and their effective reuse.

Keywords: Data Journalism; Open Government Data; Transparency.

Jornalismo de dados e a qualidade dos dados abertos governamentais no Brasil: problemas, desafios e obstáculos nos processos de apropriação profissional

Sivaldo Pereira da SILVA
Ébida SANTOS
Maria DOMINGUEZ
Patrícia LIMA

Historicamente, o Estado sempre foi um polo produtor de dados e isso está grafado na própria origem etimológica da palavra “estatística”, inicialmente vinculada ao estudo das informações produzidas pelo ente estatal (*statisticum collegium*). Por outro lado, o jornalismo se firmou como um importante setor que consome, filtra e processa dados governamentais, sendo que o próprio fortalecimento deste campo profissional passa pela posição de intermediário cognitivo entre a máquina estatal e o público (CODDINGTON, 2015).

Com o desenvolvimento do ambiente digital e a expansão do uso generalizado de algoritmos e bases de dados permeando praticamente todos os setores na vida contemporânea, direta e indiretamente, a produção de informações governamentais ganhou uma nova dimensão. Publicar bases de dados estruturados em formatos abertos e legíveis máquinas, aptas para serem coletadas e tratadas por diferentes atores e intermediários através de sistemas automatizados, tornou-se uma política de Estado.

No Brasil e em diversos países, houve esforços e avanços significativos — principalmente a partir da segunda década deste século — na criação de leis e de infraestruturas como portais da transparência e de dados abertos. Conforme o Painel de Monitoramento de Dados Abertos da Controladoria Geral da União¹, cerca de 3.388 bases de dados foram abertas desde o início da implementação dos Planos de Dados Abertos dos órgãos públicos federais, em 2016. O Portal Brasileiro de Dados Abertos (dados.gov), por exemplo, é responsável pela oferta de mais de 13.000 bases de dados². Embora esses números sejam relevantes quando observados no todo, ao olharmos de forma mais específica para

¹ O Painel de Monitoramento de Dados Abertos está disponível em: <<http://governoaberto.cgu.gov.br/noticias/2017/cgu-lanca-painel-de-monitoramento-de-dados-abertos-do-governo-federal>>. Acesso em: 30 abr. 2020.

² Este número de mais de 13.000 conjuntos de dados encontrados no dados.gov foi identificado em setembro de 2022.

esses dados, pode-se afirmar que há uma série de fragilidades no processo de publicação dos *datasets* governamentais (SILVA et al. 2020). O que implica, em última instância, a existência de barreiras para a apropriação social e política dos usuários de dados abertos governamentais, como jornalistas, programadores, pesquisadores, empresas e organizações não governamentais que fazem o reuso desses recursos para diversos fins. Na vida real, os problemas técnicos podem surtir diferentes impactos no processo de apropriação de dados abertos, a depender do perfil e da expertise de cada ator.

Diante desse cenário, o principal objetivo deste artigo é caracterizar e compreender a experiência de jornalistas de dados no processo de apropriação de dados abertos governamentais para a produção de narrativas. Espera-se, aqui, lançar luz sobre os tipos de problemas e desafios que esses profissionais enfrentam e o que consideram mais relevante nesse sentido. Para responder a essas questões, o artigo foi organizado em duas partes subsequentes. A próxima seção buscará relacionar as concepções de dados abertos e jornalismo de dados demonstrando seus entrelaçamentos e horizontes normativos, o que nos permitirá delimitar alguns parâmetros conceituais e analíticos. Na seção seguinte, traremos o resultado de pesquisa de percepção, aplicada a jornalistas, sobre a experiência de uso de dados abertos governamentais no exercício profissional (com foco no jornalismo de dados). Neste estudo, buscou-se produzir uma análise qualitativa sobre os principais elementos que norteiam o uso de dados abertos por esses profissionais, apontando possíveis problemas e soluções para a melhoria de políticas públicas neste campo.

Dados abertos e jornalismo de dados: faces e interfaces

O uso de informação produzida por entes estatais nas narrativas jornalísticas não é uma novidade e sempre esteve, em maior ou menor grau, presente na configuração do *ethos* da imprensa. De certo modo, falar em “jornalismo de dados” pode soar redundante, pois não há jornalismo sem dados, embora possa haver versões ou gêneros mais literários com menor foco na informação, mas que dificilmente se abstenham totalmente desta. Diversos autores têm discutido qual seria o conceito mais apropriado para o jornalismo de dados e o que o diferencia das formas anteriores tanto no âmbito de pesquisas internacionais, como nacionais (MEYER, 1973; GRAY; CHAMBERS; BOUNEGRU, 2012; HAMILTON; TUNER, 2009; ANDERSON, 2012; BARBOSA 2007; LIMA JÚNIOR, 2011; TRÄSEL, 2014; MANCINI; VASCONCELLOS, 2016). Para efeito desta pesquisa, consideramos que qualquer definição mais eficaz passa pelo desenvolvimento de novas *expertises*:

Talvez o ponto central no debate sobre o que venha a ser o JD seja a incorporação de novas competências pelo campo jornalístico. Esse tema está presente nas inúmeras tentativas de definir o JD, contudo, a nosso ver, ainda de forma tangencial. O exame dessas competências, seja a partir de pesquisas do tipo etnográficas, seja do exame do conteúdo e da forma do JD praticado no Brasil, pode ajudar a sair de algumas dificuldades teóricas que ainda persistem. Definir o JD pela dimensão tecnológica somente, por exemplo, implicaria em aceitar um argumento de natureza determinística: uma vez adotado o aparato técnico, estaríamos diante de uma necessária mudança no fazer jornalístico, no caso, uma mudança inexorável rumo ao JD [Jornalismo de Dados] (MANCINI; VASCONCELLOS, 2016, p. 71).

É preciso considerar ainda os problemas decorrentes da junção entre jornalismo e dados, como aponta Bradshaw (2014), pois no ambiente digital os dados não são apenas uma reunião de números numa planilha. Fotos, vídeos, documentos confidenciais e áudios podem também ser retratados em números e especificamente em leitura binária. Para o autor, a limitação que cerca a definição do jornalismo de dados certamente se encontra na diferenciação do jornalismo feito e acostumado apenas com tabelas numéricas para o jornalismo que combina o tradicional faro jornalístico e a habilidade de contar uma história envolvente, com o alcance e escala absolutas da informação digital agora disponível.

Por isso, o conhecimento em Ciência de Dados; a desenvoltura em coletar, automatizar e processar dados digitais estruturados e não estruturados; o uso de *softwares* e algoritmos para automatizar processos e explorar estruturas de bases de dados; a capacidade de interpretar um *dataframe* e perceber suas topografias e padrões tácitos; a aptidão em driblar as barreiras técnicas; a capacidade de cruzar variáveis e transformar dados isolados em conhecimento; a possibilidade de gerar formas dinâmicas de visualização e cognição de informação, etc. são alguns dos muitos elementos que caracterizam o jornalismo de dados. Entretanto, como aponta Bradshaw (2014), a *expertise* dos jornalistas que lidam com dados também se configura como um desafio.

Naturalmente, nem todos os jornalistas de dados terão o mesmo nível de conhecimento em todas as dimensões possíveis, assim como nem todo profissional no telejornalismo domina todas as dimensões deste ofício. Mas a combinação, em grau maior ou menor, dessas diversas habilidades certamente são as bases importantes para se definir se uma prática profissional é ou não jornalismo de dados. Para compreendermos os gargalos desta *expertise*, devemos observar as dificuldades e desafios que jornalistas de dados enfrentam desde a configuração de um ambiente de trabalho propício, disponibilidade de ferramentas condizentes e formação adequada até a existência de matéria-prima

qualificada deste campo. Neste sentido, a qualidade das bases de dados governamentais são insumos importantes e como jornalistas lidam com esses dados merecem especial atenção.

Obviamente, seria reducionista definir o jornalismo de dados apenas pelo uso de dados abertos governamentais. A investigação em bases de dados de empresas, a raspagem e o processamento de informação dispersa na *web* (por meio de técnicas de *webscrapping*, *linked data* etc.) também compõem insumos desta prática jornalística. Porém, devido à própria natureza do jornalismo em suas origens políticas e deontológicas (BARNET; GABER, 2001; KOVACH; ROSENTIEL, 2003; MCNAIR, 2009; COOK, 2011; BENNET, 2012), o trato de dados governamentais é, sem dúvida, uma de suas faces mais importantes.

Nesse cenário, vemos surgir o movimento que advoga pelos Dados Abertos Governamentais, guiado pelos princípios do Governo Aberto, com grandes expectativas sobre os seus benefícios políticos, sociais e econômicos, servido como matéria-prima para a atividade de diversos atores como jornalistas, programadores, pesquisadores e organizações não governamentais. O volume de dados abertos governamentais acessíveis para consulta e uso por quaisquer partes interessadas, como os jornalistas, cresceu nos últimos anos. Isso aconteceu na esteira das políticas de dados abertos, que vêm se consolidando como uma das mais importantes frentes de insumos para o exercício, desenvolvimento e fortalecimento desta modalidade jornalística e do próprio jornalismo típico deste século. Como aponta Stoneman (2015):

When the Open Data movement began, and governments started opening their books, and publishing data, several journalists and editors, saw Open Data as a potential saviour of an industry in decline, with splintered audiences, rising staff costs, and increasing access to all of sources of information, there was a need for people with expertise to acquire, analyse and explain the data which governments were opening (STONEMAN, 2015, p. 15).

O movimento de dados abertos nasceu na forma de uma mobilização de ativistas que contou com o apoio de gestores governamentais para se consolidar como políticas públicas em diversos países³. Isso deve ser compreendido como algo que vai muito além da tradicional publicação de informações

³ Uma das iniciativas mais importantes do movimento de dados abertos é Parceria para Governo Aberto (*Open Government Partnership*) lançada em 2011 inicialmente por oito países fundadores e que ao final da mesma década já contava com mais de 80 nações signatárias. O Brasil foi um dos principais atores fundadores da OGP, juntamente com os EUA.

públicas em portais governamentais e não deve ser confundido com dados públicos divulgados de qualquer natureza e formatos. Para que um conjunto de dados seja considerado “aberto”, uma série de pré-requisitos precisa ser contemplada (ATTARD *et al.*, 2015).

Essa padronização é regida por um conjunto de princípios normativos e técnicos voltados para garantir que a publicação esteja apta para ser apropriada por diversos atores, incluindo jornalistas, tendo em vista o uso de algoritmos e *softwares* para a coleta e processamento dinâmico das informações. Podemos resumir esses princípios⁴ nos seguintes termos:

- (a) *Completos* — os dados não podem ser parciais ou com lacunas de informação. *Dataframes* com *missing* ou com uma parte da informação omitida prejudica o reuso e inviabiliza a compreensão do todo;
- (b) *Primários* — devem ser de primeira mão, isto é, produzidos originalmente pelo ente governamental que os publica. Isso garante a confiabilidade, *accountability* e responsabilização da curadoria dos dados.
- (c) *Atuais* — os dados devem ser publicados o quanto antes possível, pois o atraso na publicação pode significar perda de valor e de contexto, diminuindo as possibilidades de reuso);
- (d) *Processáveis por máquina* — não bastam ser digitalizado. Precisam ser, além de digitais, estruturados em formatos legíveis por algoritmos e *softwares* de análises estatísticas. Significam publicações em formatos de *dataframe*, árvore ou outras formas estruturadas processáveis de forma automatizada. Arquivos como imagens (JPG, GIF etc.) ou de editores de texto (DOCX, RTF, PDF etc.) são formatos impróprios, pois, embora sejam digitais, não são estruturados para bases de dados.
- (e) *Acessíveis* — os dados precisam ser disponíveis para o público mais amplo, sem restrições de qualquer tipo para os propósitos mais variados, não devendo pressupor cadastro, registro ou identificação de quem usa, ou qualquer tipo de mecanismo restritivo;

⁴ A lista é a síntese do debate normativo e deontológico sobre o que são dados abertos. Consolidada principalmente a partir da aprovação em 2007 do documento “*The 8 Principles of Open Government Data*”, elaborado por um grupo de trabalho multidisciplinar constituído por especialistas, ativistas, gestores e representantes de organizações civis e empresariais reunidos na cidade de Sebastopol, na Califórnia (EUA).

- (f) *Formatos livres* — devem ser publicados por meio de licenças livres e em formatos não proprietários. Se o formato do arquivo tem *copyright* e pertence a uma empresa, isso gera insegurança no uso, por isso, devem ser publicados em formatos abertos de domínio público.
- (g) *Gratuitos* — não pode haver qualquer pagamento de taxa de uso, taxa de acesso ou tributação no reuso dos dados. Devem ser disponibilizados sem onerar o usuário pela apropriação. O pressuposto é os dados serem públicos e o Estado ser apenas o seu curador.
- (h) *Permanentes* — uma vez publicados, os dados não podem ser retirados do acesso público e o seu endereço eletrônico não pode ser alterado. Qualquer oscilação neste sentido inviabiliza aplicações e pesquisas baseadas em algoritmos.
- (i) *Documentados* — para se compreender *datasets* complexos é preciso que haja documentação explicando as variáveis, como foram coletadas, que métricas são utilizadas na sua mensuração e outras informações. Isso está diretamente vinculado à transparência das publicações, pois aquilo que não podemos compreender é, por natureza, opaco.
- (j) *Interativos* — no nível cotidiano da apropriação pelos usuários, surgem dúvidas sobre a natureza dos dados, mesmo quando são documentados. Por isso, os dados requerem uma dimensão comunicativa que possibilita a interação entre a curadoria dos dados e aqueles que farão o seu reuso no âmbito social.

Observando mais atentamente, perceberemos que todos esses princípios que guiam a padronização do conceito de dados abertos estão preocupados, principalmente, com a universalidade e a garantia de mecanismos para a apropriação social dos dados, publicados em um contexto de uso cada vez mais intenso de algoritmos e ferramentas de processamento automatizado. Isso nos serve como indicadores de qualidade que afetam diretamente a experiência do usuário. Silva et al. (2020) propõem que a qualidade de uma boa política de dados abertos deve ter como base três pilares normativos fundamentais, sendo eles:

[1] A inteligibilidade consiste na propriedade dos recursos em serem compreensíveis, onde o “aberto” pressupõe que o dado deve ser acessível cognitivamente para o usuário, se aproximando da noção de transparência, ou seja, aberto para o seu entendimento e reuso. [...]

[2] A operacionalidade diz respeito à propriedade dos recursos em termos de funcionalidade no seu processo de apropriação, livrando o usuário de barreiras, constrangimentos ou custos desnecessários de reuso. Em termos práticos, significa analisar o quanto um recurso (arquivo) está operacional para a livre apropriação. [...] Por fim, a [3] interatividade se refere à dimensão

mais comunicacional que envolve os recursos. Significa avaliar o quanto um *dataset* pode ser objeto de diálogo, de feedback informativo, interação social ou discussão pública (SILVA et al., 2020, p. 8).

No caso do jornalismo de dados, a ausência dessas dimensões e princípios pode inviabilizar uma pauta ou dificultar a elaboração de uma reportagem. No processo de produção de uma narrativa jornalística guiada por dados, os problemas de falta de qualidade na publicação de Dados Abertos podem afetar os diversos estágios do processo de produção, desde a coleta dos dados (arquivos inacessíveis, links quebrados etc.) até o tratamento das informações.

Embora no Brasil existam leis, decretos, normas infralegais e uma estrutura relativamente robusta de publicação de dados abertos, é preciso investigar que tipo de problemas e obstáculos os jornalistas de dados têm enfrentado na apropriação e uso desses dados. Isto é, quando lida diretamente com a manipulação de arquivos e informações publicadas, quais as dificuldades e entraves que encontram nestes processos e quais desafios hoje despontam como os mais proeminentes.

Políticas de dados abertos no Brasil e a experiência de jornalistas de dados

No intuito de responder às questões levantadas até aqui, esta seção traz um conjunto de dados e análises sobre o modo como estes profissionais lidam com o reuso de dados abertos governamentais. São identificados os desafios aos quais gestores e curadores de dados abertos devem estar atentos para haver políticas mais efetivas neste campo. Neste caminho, foi realizado um estudo de recepção com profissionais das cinco regiões do país, levantando informações e identificando padrões sobre os principais elementos que marcam as suas experiências. Em termos metodológicos, os procedimentos de coleta e análise foram executados considerando os seguintes parâmetros e instrumentos:

a) *Corpus* — o *corpus* analisado foi composto por 51 profissionais, identificados como jornalistas de dados, seja pelo seu trabalho publicamente reconhecido, seja pela sua autodenominação ou indicação

utilizando a técnica de *snowball*⁵. Trabalhou-se com uma amostragem não probabilística, mas garantindo um nível adequado de representatividade regional para se captar as diferentes realidades vividas. Por se tratar de um estudo qualitativo sobre a percepção de profissionais acerca do uso de dados abertos, e tendo em vista que o universo deste tipo de profissional no Brasil ainda é bastante incipiente, considerou-se a amostra significativa e adequada para os fins propostos, possibilitando segurança para afirmações e generalizações no nível global da amostra.

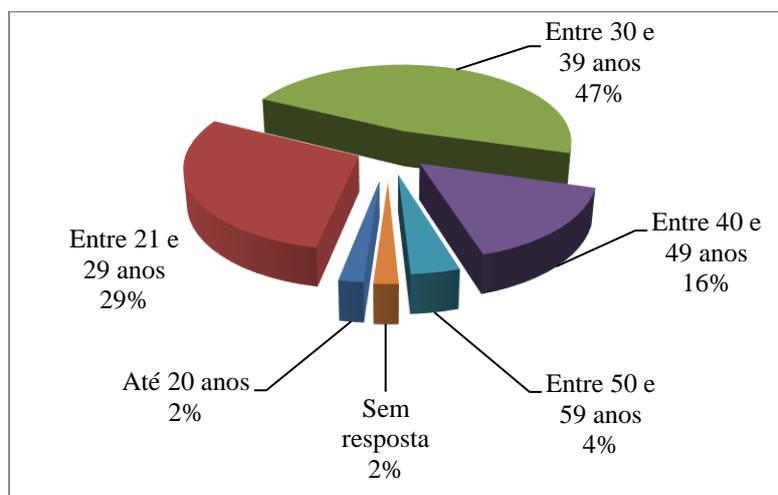
b) Survey e entrevista estruturada — foi aplicado um instrumento de coleta de dados composto por duas partes: (a) *survey* com perguntas fechadas desenhadas considerando os princípios, dimensões e indicadores normativos que configuram a qualidade de publicação de Dados Abertos e; (b) entrevista em profundidade do tipo estruturada, com perguntas abertas para livre resposta do entrevistado. Na *survey*, o intuito foi captar indicadores quantiquantitativos e na entrevista estruturada, o objetivo foi captar informações mais qualitativas sobre a experiência dos entrevistados; (c) *análise* — os dados foram analisados utilizando estatística de frequência, uma vez que a amostragem e o próprio objetivo mais qualitativo da pesquisa não requerem o uso de estatística inferencial. A tabulação dos dados textuais oriundos da entrevista estruturada contou com a utilização do *software* Iramuteq, que produz análises lexicais com visualizações do discurso dos entrevistados, mostrando correlações e tendências.

Análise

Quanto à caracterização dos entrevistados, quase a metade (47%) tem entre 30 e 39 anos e cerca de 29% da amostra têm entre 21 e 29 anos. Se separarmos os grupos com idade inferior a 40 anos, veremos que os respondentes que possuem até essa faixa etária representam 78%. Isso demonstra que prepondera dentre os entrevistados uma faixa etária relativamente jovem. Jornalistas com 40 anos ou mais somam apenas 20% da amostra (não houve registro de pessoas declaradas acima de 60 anos), conforme detalha o gráfico 1):

⁵ O *snowball* é um método aplicável a pesquisas qualitativas quando o público-alvo é de difícil acesso. Assim, a partir de uma primeira seleção de entrevistados e após coletar suas respostas, solicita-se a indicação de outra pessoa que se enquadre na amostra da pesquisa.

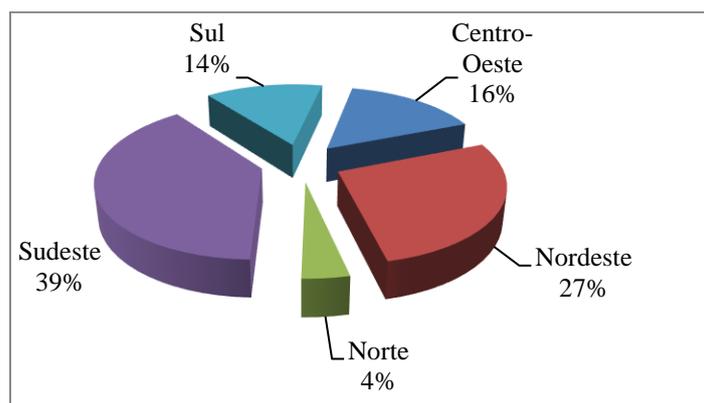
Gráfico 1 – Faixa etária da amostra (N=51)



Fonte: Elaboração própria a partir de *survey* com profissionais

Em termos regionais, no gráfico 2, a amostra traz profissionais das cinco regiões do país. Na amostra, as regiões Sudeste e Nordeste possuem respectivamente 39% e 27% dos entrevistados. A região Centro-Oeste aparece com 16%, o Sul com 14% e o Norte, com 4%.

Gráfico 2 – Proporcionalidade da amostra por região do país (N=51)

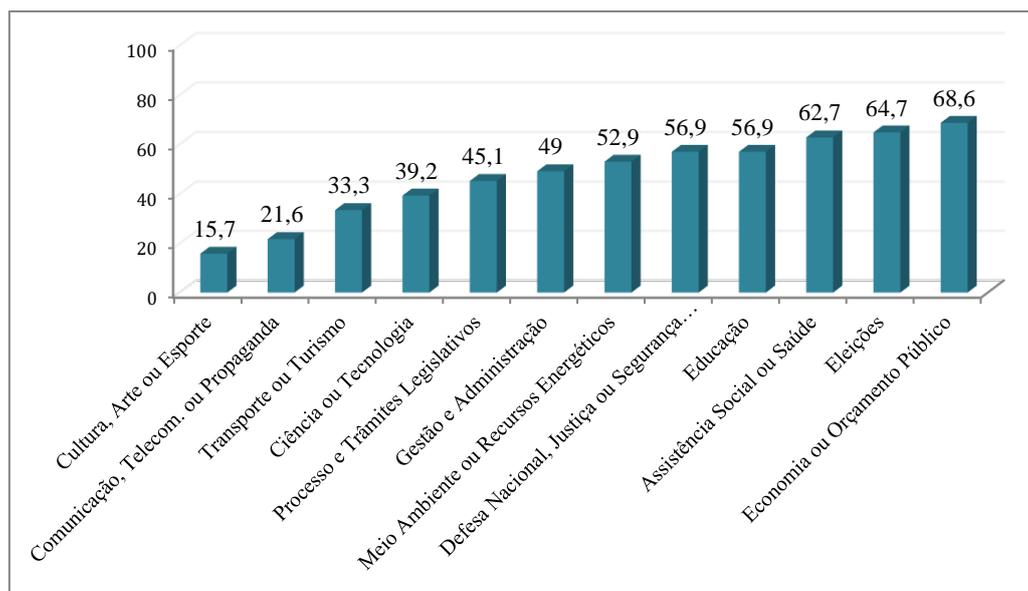


Fonte: Elaboração própria a partir de *survey* com profissionais

No que se refere às áreas temáticas mais exploradas pelos profissionais no processo de produção de conteúdos baseados em dados abertos, duas áreas se destacam: “Economia ou Orçamento Público” e “Eleições”. Outras também merecem ser observadas, como “Assistência Social ou Saúde”,

“Educação”, “Defesa Nacional, Justiça ou Segurança Pública” e “Meio Ambiente ou Recursos Energéticos”, mencionadas por mais da metade dos entrevistados, conforme detalha o gráfico 3.

Gráfico 3 – Quantidade de reportagens produzidas sobre diferentes áreas governamentais



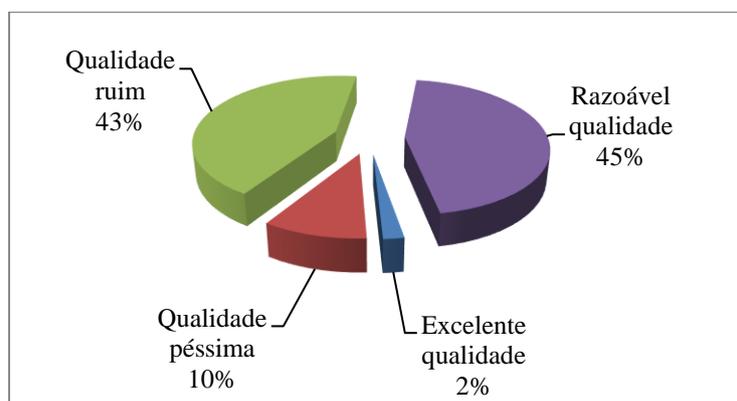
Fonte: Elaboração própria a partir de *survey* com profissionais

Na análise sobre a percepção de qualidade das bases de dados governamentais na experiência dos profissionais, os resultados demonstraram não haver uma percepção plenamente positiva. Pouco mais da metade dos profissionais (53%) consideraram que os dados abertos governamentais no Brasil têm qualidade ruim ou péssima. Embora haja uma parcela significativa de avaliações pessimistas neste indicador, há também outra parcela importante de profissionais (45%,) que consideraram as bases de dados disponíveis razoáveis, ou seja, com uma qualidade mediana a ponto de reconhecerem elementos positivos nas suas experiências de trabalho, ainda que não sejam ideais. Para apenas 2% dos entrevistados, a qualidade dos dados tem uma excelente performance, conforme podemos visualizar no gráfico 4.

Esta percepção dúbia, de uma metade que considera mediana e outra que considera ruim, provavelmente, reflete um processo ainda em construção, resultante do movimento de implementação gradativa da política de dados abertos dos últimos anos. No nível federal, esse movimento acontece principalmente a partir de 2016, após a publicação do Decreto 8.777, que regulamentou a Política

Nacional de Dados Abertos e forçou a criação de Planos de Dados Abertos por órgãos públicos federais, delimitando responsáveis, prazos e as bases de dados que deverão ser abertas por cada ente. Desde a emissão desta política, de fato, houve um grande volume de dados públicos que passaram a estar abertos e disponíveis ao público, entretanto, nem sempre nos formatos e qualidade adequados para o reuso e apropriação por jornalistas. Soma-se a isso uma série de retrocessos vistos no campo da transparência pública nos últimos anos no Brasil, notadamente desde 2019, quando se passou a perceber ameaças e ataques à Lei de Acesso à Informação e à transparência on-line do governo federal brasileiro (ALMADA et al, 2022).

Gráfico 4 – Percepção sobre qualidade dos dados abertos governamentais (N=51)



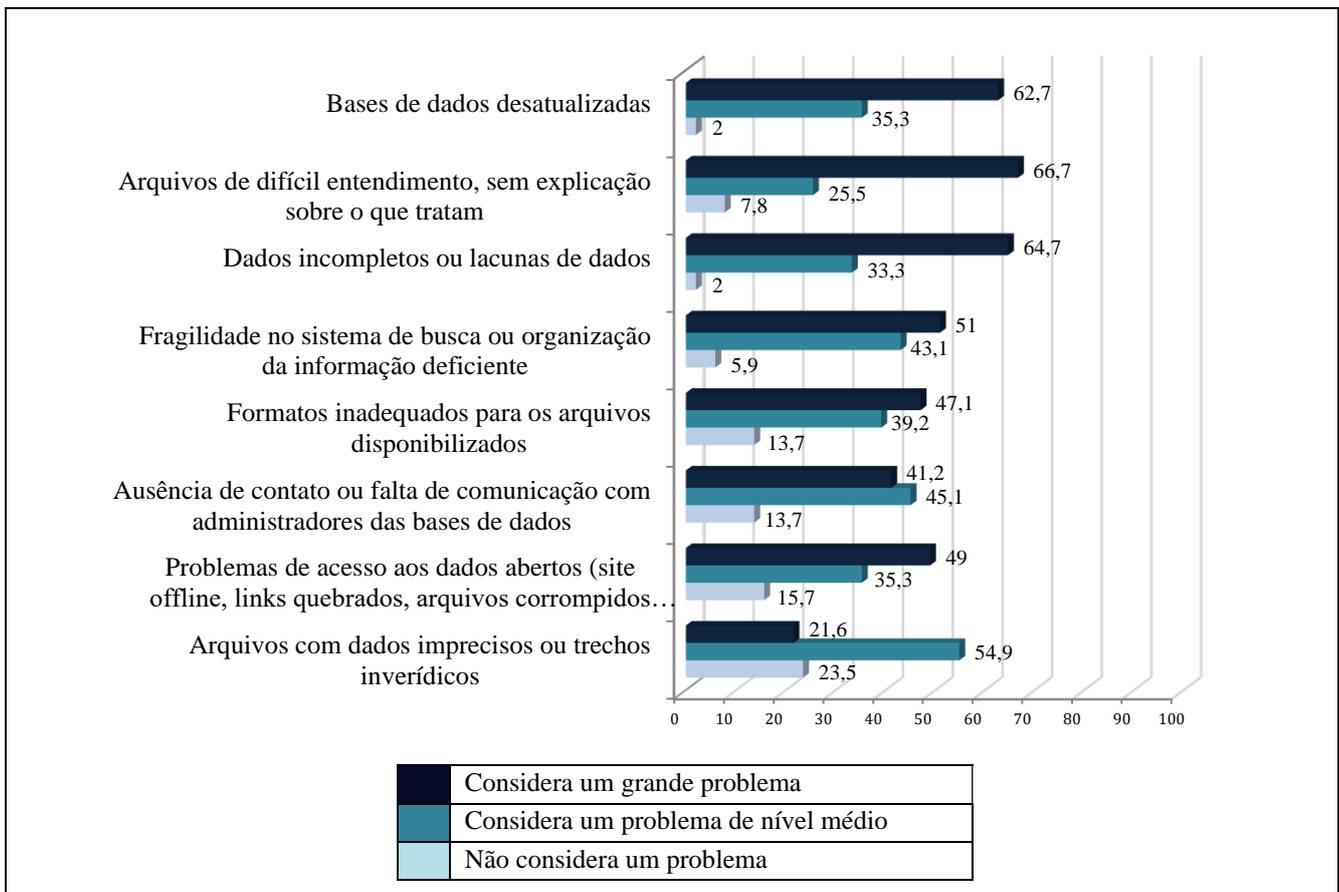
Fonte: Elaboração própria a partir de *survey* com profissionais

Analisamos mais a fundo que tipo de problemas são considerados mais relevantes, em termos de impacto, no processo de produção: coleta, tratamento e processamento dos dados. O gráfico 5 traz um detalhamento apontando a intensidade do problema na percepção dos entrevistados, em uma escala que avalia o problema como “um grande problema”, “um problema secundário (mediano)” e “um não problema”. Esta gradação nos permite detectar a existência de disfunções no processo de reuso dos dados pelos profissionais e, simultaneamente, possibilita dimensionar o grau de sua importância como elemento limitador do trabalho jornalístico.

De tal modo, "Bases desatualizadas", "Arquivos de difícil entendimento ou não explicados", "Dados incompletos ou com lacunas" são os três tipos de disfunções consideradas “um grande problema” por mais de 60% dos entrevistados. O primeiro está vinculado a problemas que atingem o próprio cerne da atividade jornalística, a função de reportar aquilo que é atual. Quando dados estão desatualizados,

perdem o valor-notícia (SILVA, 2014; FRANCISCATO, 2014; TRAQUINA, 2001; CORREIA, 1997). O segundo remete-nos a problemas de inteligibilidade nos arquivos, isto é, mostra a existência de problemas na formatação e documentação das bases de dados, remetendo à opacidade nos arquivos ou à falta de explicação sobre os elementos contidos nas colunas e linhas de uma base de dados, por exemplo. Essa falta de explicação sobre a formatação das bases de dados acaba sendo contraditória com a própria função e princípios que regem uma boa política de dados abertos, dificultando a compreensão e reuso dos dados. Já o terceiro problema aparece com algum grau de unanimidade: nenhum profissional considerou que arquivos com lacunas de dados não sejam um problema (todos citaram como um grande problema ou problema mediano). Isso pode significar, na prática, a inviabilidade de uso de um arquivo, a depender do nível de *missing* que um *dataframe* contém e do grau de relevância dos dados faltantes.

Gráfico 5 – Tipos de problemas mais enfrentados no processo de reuso (N=51)

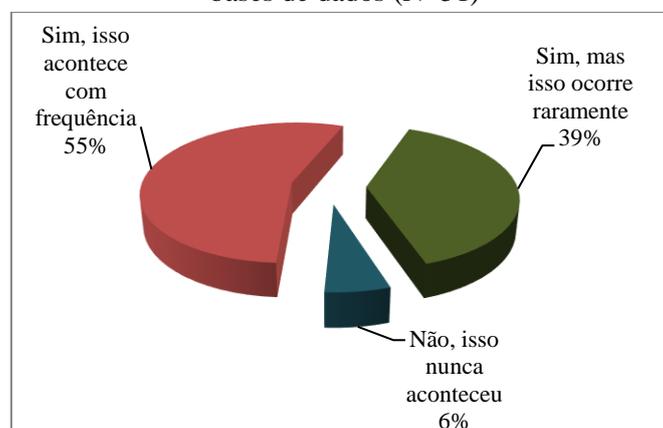


Fonte: Elaboração própria a partir de *survey* com profissionais

Problemas de indexação (busca e organização da informação) e formatos inadequados de arquivos aparecem como um problema relevante para quase metade dos profissionais. Por outro lado, vale também destacar que a questão da confiabilidade dos dados não aparece como um problema central, mas desponta como um problema secundário de destaque para quase 55% dos entrevistados, apontando os dados imprecisos ou inverídicos como um problema. Importante também dimensionar e relativizar o real impacto de alguns desses problemas em termos práticos. Quando algo é relatado como “um grande problema”, significa que há a percepção de maior falha em termos de uma boa política de dados abertos. Por exemplo, para 44% dos entrevistados, problemas primários de acesso aos arquivos (*links* quebrados ou arquivos corrompidos) foram considerados um grande problema. Isso pode significar a inviabilidade de muitas pautas antes mesmo de seu início, e indica também, em termos técnicos, que muitos arquivos “publicados”, na prática, não existem, não são acessíveis ou não são adequados para o reúso.

Ao serem combinados, os problemas de qualidade das bases de dados podem ter efeitos importantes no cotidiano dos jornalistas de dados, pois significam repercussões práticas no próprio desenvolvimento de matérias e reportagens e na sobrevivência de uma pauta baseada em dados abertos. Conforme detalha o gráfico 6, a grande maioria dos entrevistados, cerca de 94% dos respondentes, declarou que já chegou a desistir de uma pauta na tentativa de utilizar dados abertos governamentais. Para 55%, inclusive, isso ocorre com frequência. Experiências plenamente satisfatórias foram raras, já que os relatos apontaram que para apenas um grupo muito pequeno de profissionais (6%) não houve problemas no processo de produção de matérias e reportagens, que sempre conseguiram finalizar pautas baseadas em tais recursos.

Gráfico 6 – Desistência na produção de pauta/matéria vinculada a Dados Abertos, devido a problemas nas bases de dados (N=51)



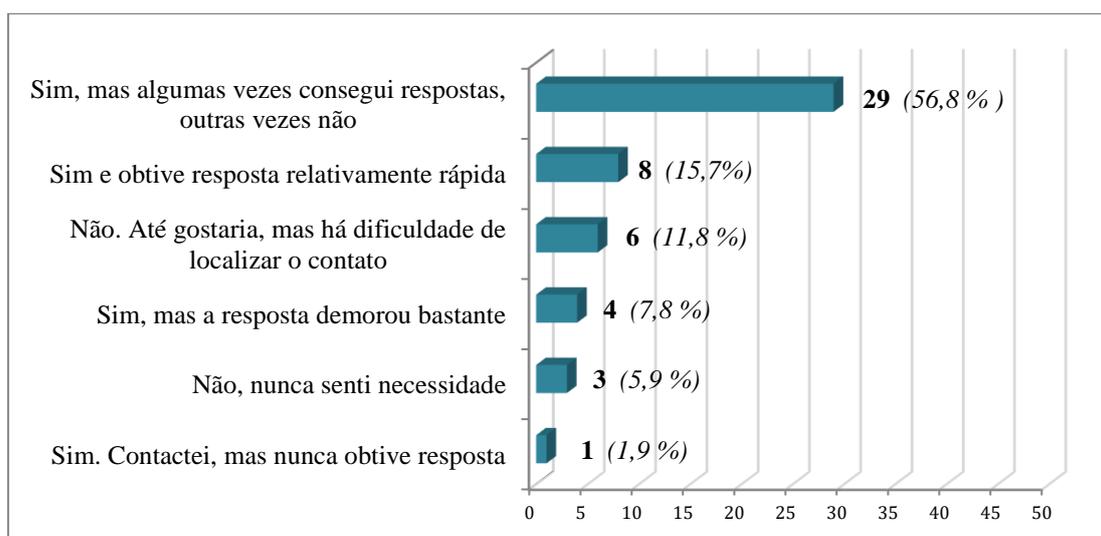
Fonte: Elaboração própria a partir de *survey* com profissionais

A desistência de pautas é exemplificada pelo caso do Respondente 8 que, após utilizar da Lei de Acesso à Informação para obter dados da Polícia Federal, recebeu arquivos incompletos e de difícil compreensão. Em seu relato conta que:

Ao recorrer ao pedido, deixando as perguntas mais claras ainda, novos dados me foram enviados, no entanto, ainda sem esclarecimento sobre o que eu tinha pedido. Tentei contato com a pessoa que tinha me respondido, mas me orientaram a fazer um novo pedido de acesso às informações, que, depois, vieram com informações imprecisas e tive que optar por derrubar a pauta (RESPONDENTE_8, entrevista, 2020).

Como discutimos nos parâmetros de uma boa política de dados abertos, além da qualidade técnica dos arquivos (em sua operacionalidade), eles também precisam ser inteligíveis, ponto diretamente vinculado à noção de transparência. Quando há problemas de inteligibilidade ou de operacionalidade, a perspectiva da interação é um elemento importante nos processos de efetiva apropriação social dos *datasets*. Quando jornalistas e outros usuários de dados abertos enfrentam problemas de operacionalidade ou inteligibilidade, é preciso haver a possibilidade de contato e *feedback* informativo com os gestores públicos responsáveis pelas bases de dados acessadas para que o processo de reuso tenha continuidade. O gráfico 7 traz uma síntese da percepção dos entrevistados quanto à necessidade de comunicação com gestores/curadores dos dados e a sua efetividade quando isso ocorre.

Gráfico 7 – Experiência de interatividade e *feedback* informativo – N (%)



Fonte: Elaboração própria a partir de *survey* com profissionais

Primeiramente, cerca de 94% dos profissionais relataram que, de algum modo, sentiram necessidade de entrar em contato com os gestores responsáveis pela publicação dos *datasets* (curadores ou mantenedores). Isso demonstra que este é um elemento importante na política de dados abertos, seja devido aos diversos problemas qualitativos enfrentados pelos profissionais, que tende a aumentar a pressão por demanda comunicativa; seja devido à própria natureza deste ofício, que requer plena segurança e entendimento do conteúdo (com temas cada vez mais complexos, com processamentos quantitativos em crescente volume de dados) que será “traduzido” e reportado, uma vez que a atividade deste tipo de intermediário requer clareza e segurança para repassar informações corretas ao público mais amplo.

Segundo, chama a atenção o baixo índice de profissionais que contataram e obtiveram uma resposta relativamente rápida dos gestores responsáveis pelos dados utilizados: apenas 15,7%. Também é importante destacar que quase 57% relataram ter realizado contato, mas nunca obtiveram resposta. Todos esses números indicam um evidente problema de comunicação e interação na política de dados abertos no Brasil. Essa problemática também se relaciona com a própria noção de dados que os gestores têm. Foi possível observar, nas falas dos entrevistados, a dificuldades de comunicação, pois o que é disponibilizado por vezes são dados desagregados, de difícil leitura e desatualizados. A morosidade nas respostas por parte dos gestores também é evidenciada pelos entrevistados, dificultando a produção jornalística.

A dificuldade no contato para sanar dúvidas, para a obtenção de dados não publicados ou ainda para a solicitação de novas bases de dados pode ser verificada também nas palavras dos respondentes quando apontam: “Os pedidos via LAI (Lei de Acesso à Informação) demoram e nem sempre são atendidos, o que pode impedir a realização de determinado trabalho (R_35)”; “Então as minhas principais dificuldades são dados desatualizados, em PDF ou outro formato que dificulte o uso, e demora na resposta de Lei de Acesso [à Informação — complemento nosso] (R_41)”; “O órgão demora até três meses para liberar os dados de estações meteorológicas manuais (R_3)”.

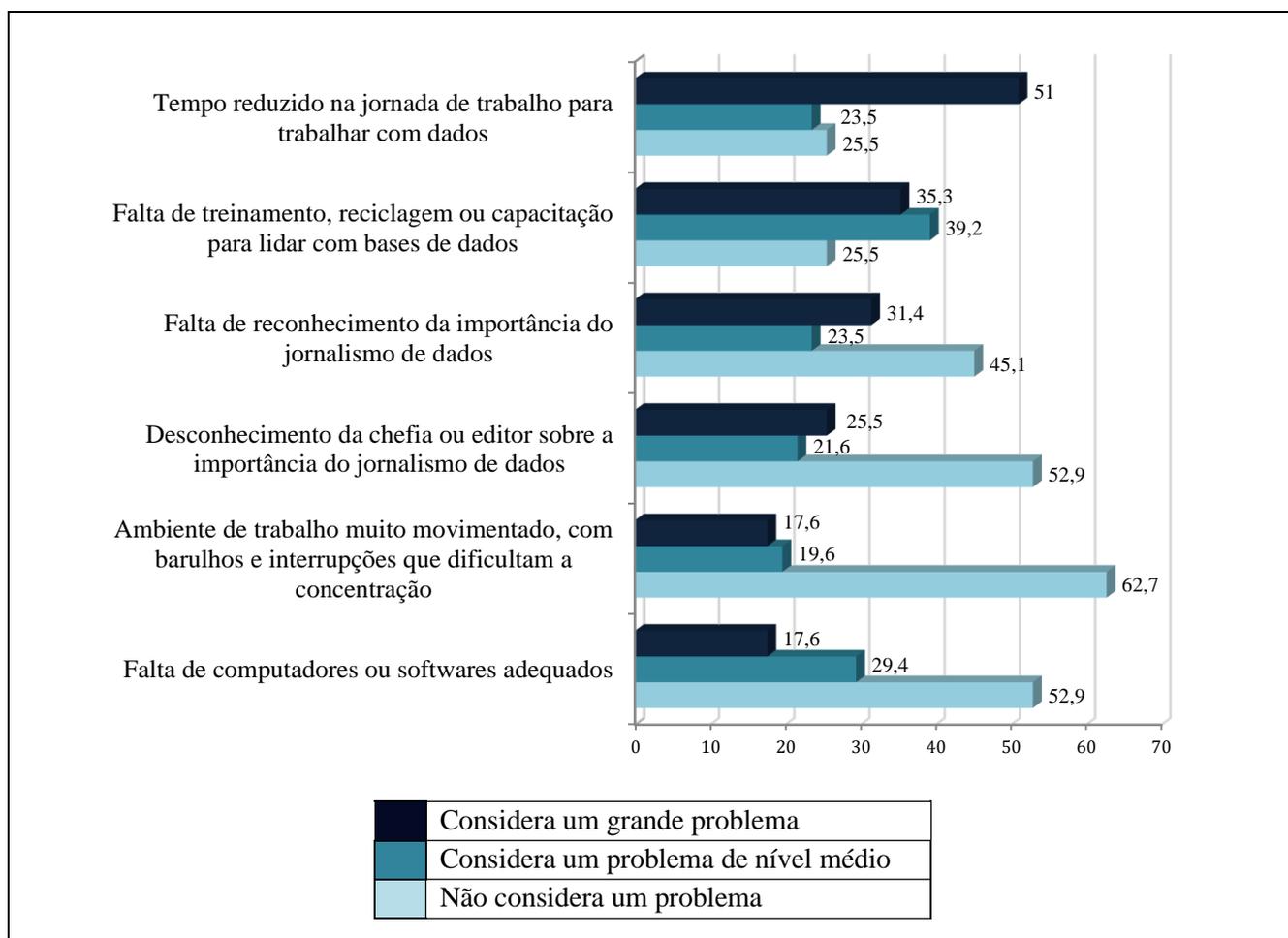
Apesar de a LAI ter completado dez anos desde a sua entrada em vigor, estes resultados apontam que ainda existem importantes lacunas em sua implementação⁶. O Brasil entra na lista de países com

⁶ Para mais informações sobre os avanços e desafios para a implementação da LAI no Brasil, ver relatório ‘10 anos da Lei de Acesso à Informação: De onde viemos e para onde vamos’, da ONG Artigo 19. Disponível em:

política sistematizada para a consulta de dados de interesse público quando mais de cem nações já haviam investido na transparência pública. A LAI foi regulamentada tardiamente no Brasil - a primeira lei de acesso à informação do mundo data do século XVIII, a sueca - mas ainda assim apresenta avanços importantes para a transparência, como a possibilidade de solicitação de informações públicas em formatos abertos.

Além dos problemas de inteligibilidade, operacionalidade e falta de interatividade que analisamos até aqui — diretamente vinculados às políticas de dados abertos —, é importante também observar outros obstáculos que jornalistas de dados enfrentam na execução de seus ofícios, vinculados ao seu ambiente de trabalho. O gráfico 8 traz um panorama neste sentido, também com uma escala de percepção em três níveis: “um grande problema”, “um problema secundário (mediano)” e “um não problema”. Para 51% dos profissionais, “O tempo reduzido na sua jornada de trabalho para coletar dados, processar e produzir narrativas” é considerado um grande problema, como sintetiza o Respondente 21: “O acesso à informação é difícil, as equipes são pequenas, é quase impossível ter o tempo necessário para realizar a análise necessária dentro das redações”. O tempo dedicado à produção é relatado pelos respondentes como um dos desafios enfrentados na execução das suas atividades. A falta de treinamento, capacitação ou reciclagem" também é um problema relatado por cerca de 1/3 dos entrevistados (um grande problema para 35,3%; um problema mediano para 39,5% e não é um problema para 20,9%).

Gráfico 8 – Tipos de problemas mais enfrentados no processo de reúso (N=51)

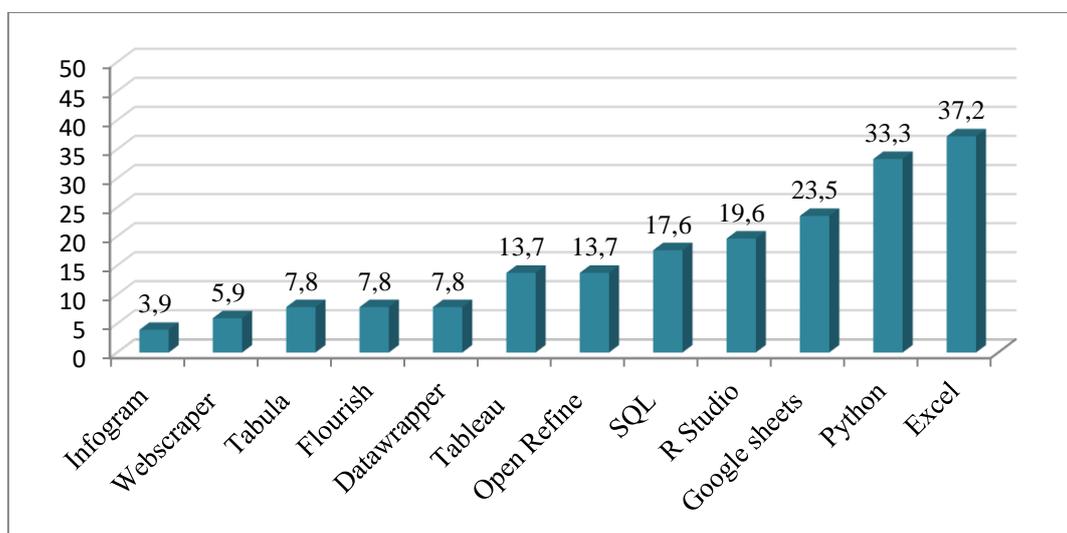


Fonte: Elaboração própria a partir de *survey* com profissionais

Problemas como “Falta de reconhecimento da importância do jornalismo de dados”, “Desconhecimento da chefia sobre a importância do jornalismo de dados”, “Falta de equipamentos (computadores e *software*) e “Ambiente de trabalho muito movimentado” tendem a ser considerados um grande obstáculo na percepção de boa parte dos profissionais entrevistados (entre 45% e 62% relataram que estes itens não são um problema).

Quanto às ferramentas mais utilizadas, o gráfico 9 traz uma síntese dos instrumentos mais mencionados. Em pergunta de resposta espontânea, cada entrevistado foi indagado a listar as ferramentas que mais utiliza no seu dia a dia (ficando livres para listar quantos itens achassem necessário, por isso, os percentuais no referido gráfico não são absolutos, pois se cruzam).

Gráfico 9 – Ferramentas e *softwares* mais citados pelos profissionais (N = 51)



Fonte: Elaboração própria a partir de entrevista estruturada com profissionais.

Notemos que o *software* Excel, Python, Google Sheets e R Studio são as ferramentas mais citadas na resposta espontânea. Vale destacar que o primeiro e o terceiro (Excel e Google Sheets) são *softwares* de duas das maiores empresas privadas na área de tecnologia (respectivamente Microsoft e Google). Já o segundo e o terceiro (Python e R Studio) não são *softwares* fechados e sim linguagens de programação não proprietárias, ou seja, códigos de fonte aberta mantidos por comunidades de programadores. Principalmente o uso dessas duas linguagens requer um grau mais avançado no conhecimento de técnicas de Ciência de Dados e de computação, uma vez que funcionam a partir de linhas de comando de códigos⁷.

Na análise lexical das entrevistas estruturadas, em que a resposta é aberta e espontânea, podemos perceber que muitos dos dados descritos também se consolidam no discurso dos profissionais. Nas Figuras 1 e 2, temos a visualização da nuvem⁸ de palavras resultantes de duas perguntas abertas: a)

⁷ Power BI, Dax, QGIS e Google (em geral) também foram mencionados, porém, apenas uma vez cada.

⁸A nuvem de palavra é uma forma de visualização de dados que realiza uma aglomeração de palavras destacadas conforme a sua ocorrência em termos quantitativos. Ou seja, trabalha com a noção básica de frequência: quanto mais a palavra aparece (quanto mais ela é frequente), mais ela é destacada no texto (em termos de tamanho). As palavras mais importantes também são posicionadas no centro da nuvem. A técnica, neste caso executada através do *software* Iramuteq, possibilita a identificação das palavras-chave em um corpus textual e nos permite fazer interpretações qualitativas.

Sobre a importância de dados abertos governamentais para o jornalismo e (b) o que deveria ser feito por gestores para melhorar tal política. Na Figura 1, o conteúdo das respostas das duas questões unificadas (pergunta “a” e “b”). Já na Figura 2, separamos apenas o discurso dos entrevistados para a questão vinculada ao papel dos gestores (pergunta “b”). Analisemos ambas, a seguir:

Figura 1 — Nuvem de palavras sobre a importância de dados abertos para o jornalismo e o que os gestores deveriam fazer para melhorar a política de dados abertos



Fonte: Elaboração própria a partir de entrevista estruturada com profissionais

Ao observar a importância de dados abertos para o jornalismo e ações de gestores para melhorar as políticas de abertura, as falas dos respondentes se evidenciam na transparência, dados atualizados, formatos e disponibilização. Os jornalistas defendem em seus discursos a eficiência das bases de dados e a facilidade de acesso para o exercício do jornalismo de dados, que para os profissionais possibilita à sociedade outras fontes jornalísticas de denúncia, acompanhamento e transparência da gestão pública. Sobre a padronização dos formatos, por exemplo, o Respondente 31 destaca: “Acredito que o trabalho fundamental é a padronização de formatos abertos e de dicionários de dados, mesmo dentro de uma mesma base, como o TSE, por exemplo”.

Figura 2 — Nuvem de palavras com respostas sobre o que gestores deveriam fazer para melhorar a política de dados abertos

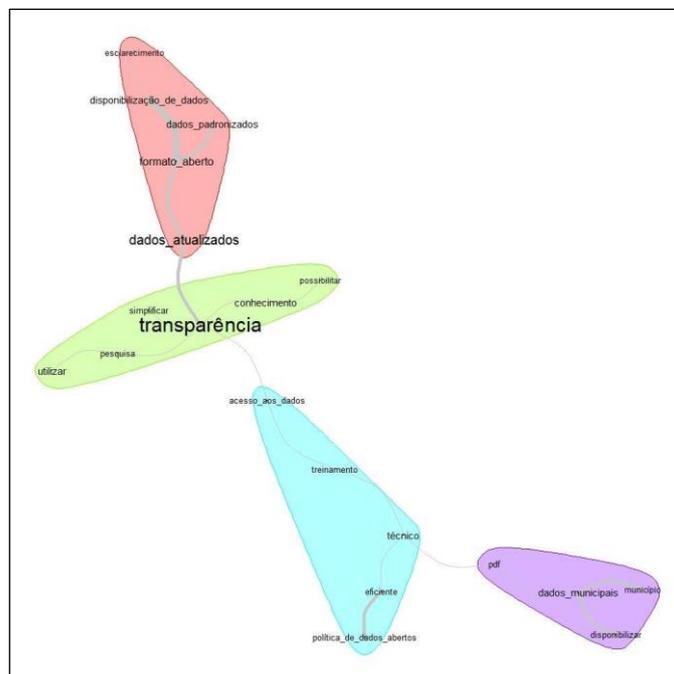


Fonte: Elaboração própria a partir de entrevista estruturada com profissionais

Primeiramente, notamos que há uma forte frequência da palavra “transparência”, que remete à vinculação das políticas de dados abertos como uma política inserida na concepção de transparência pelos profissionais. Isso também é reforçado com frequentes menções à LAI. No caso da LAI, os entrevistados geralmente mencionam que utilizam desta legislação para requerer informações que não conseguem obter nos portais governamentais. Ou seja, diante dos problemas de transparência ativa, recorrem com frequência a um mecanismo de transparência passiva para obter dados que necessitam. Há uma defesa, na fala dos entrevistados, do acesso aos serviços de transparência e de uma maior oferta de bases de dados abertos, disponibilização e cumprimento da LAI. Na análise de similitude, a Figura 3 traz a árvore de palavras que demonstra a relação (baseada na Teoria dos Grafos⁹) entre palavras e os respectivos *clusters* que formam:

⁹ Análise de Similitude é técnica baseada na teoria dos grafos, neste caso, aplicada através do software Iramuteq. A análise de Similitude aposta na relação entre objetos traçando uma linha a cada conexão entre eles. Os Grafos representam conexão entre palavras e ajudam a visualizar a estrutura em que palavras são mencionadas e suas ligações semânticas.

Figura 3 — Análise de Similitude sobre a importância de dados abertos para o jornalismo e o que os gestores deveriam fazer para melhorar a política de dados abertos



Fonte: Elaboração própria a partir de entrevista estruturada com profissionais

O *cluster* central, baseado no termo "transparência", vincula esta concepção à necessidade de "simplificar" as publicações para que se apresentem mais inteligíveis; de "possibilitar" a extração de "conhecimento" de forma mais sólida; de otimizar a "pesquisa" realizada pelo jornalista. Num segundo *cluster*, cuja palavra central é "dados atualizados" (com forte ligação em "transparência" o discurso está voltado para os problemas quanto ao formato dos arquivos (necessidade de "formatos abertos" uma vez que encontram muitos arquivos fechados, como PDF), à "disponibilização de dados", à necessidade de maior padronização ("dados padronizados") e maior "esclarecimento" no sentido de existência de documentação no processo de publicação dos *datasets*. Um terceiro *cluster*, formado pelo discurso dos profissionais (abaixo do cluster "transparência"), trata da "política de dados abertos" como uma política que envolve melhorias "técnicas", maior "eficiência" e melhor "acesso aos dados". Ele aponta para a necessidade de mais capacitação ("treinamento") para servidores públicos e gestores que criam e mantêm os *datasets*. Por fim, um último *cluster* trata dos problemas na qualidade de *datasets* nos municípios brasileiros. Diversas falas apontaram para a inexistência ou má qualidade (quando existe) da política de dados abertos no nível municipal.

Considerações finais

Este artigo teve como objetivo caracterizar a experiência de jornalistas no processo de apropriação e reuso de dados abertos governamentais. Primeiramente, buscamos lançar luz sobre a relação entre dados abertos e jornalismo de dados, apontando que esta face profissional se define na combinação de expertises. Ou seja, o tradicional conhecimento e técnicas de apuração e habilidade para produzir narrativas típicas do jornalismo se somam às novas aptidões oriundas de campos como Ciência de Dados, Ciências da Computação, Estatística etc.

Porém, mesmo que haja uma devida expertise desenvolvida, é necessário que haja a disponibilidade de um grande volume de matéria-prima digital informacional (principalmente dados estruturados) e que isso esteja aberto e acessível, com a qualidade requerida.

Uma importante fonte de insumos para este tipo de jornalismo são as bases de dados abertas governamentais. O surgimento do movimento de governo aberto e a difusão de políticas de dados abertos tem sido uma realidade que vem se expandindo principalmente a partir da segunda década deste século. Iniciativas como a *Open Government Partnership* (OGP), Lei de Acesso à Informação (LAI) e o próprio Decreto 8.777/2016 refletem e consolidam uma série de princípios que definem a qualificação de uma boa política de dados abertos. Mas nem sempre os princípios são de fato cumpridos. A simples publicação on-line de informação governamental em um *website* não constitui um dado aberto. Elencamos uma série de elementos qualitativos neste sentido que precisam ser cumpridos e ressaltamos que isso tem repercussões diretas na real possibilidade de apropriação social e reuso das informações publicadas. A partir desta normatividade, buscamos analisar como jornalistas de dados têm lidado com a qualidade dos dados abertos no Brasil, quais dificuldades e desafios neste horizonte.

Os resultados demonstraram que metade dos profissionais entrevistados (53%) consideraram que os dados abertos governamentais no Brasil têm qualidade ruim ou péssima. Apenas 2% avaliaram essas bases como excelentes. Isso aponta claramente que há um grande *gap* qualitativo a ser resolvido pela gestão e curadoria dos *datasets*, no âmbito governamental. Na análise mais minuciosa das dificuldades, a pesquisa identificou que problemas de atualização, completude, opacidade e indexação (organização da informação) aparecem como grandes entraves para o jornalista. Não devemos pensar

nesse conjunto de disfunções apenas como meros problemas técnicos. Significam, na prática, barreiras que dificultam a apropriação dos dados e seu reuso e devem ser observados com atenção pelos gestores. Outra problemática que devemos dar especial atenção diz respeito aos relatos de 49% dos entrevistados que consideraram a existência de arquivos inacessíveis (links quebrados ou arquivos corrompidos) como um grande problema. O que se pode observar a partir do que apontam os entrevistados é que os desafios se concentram na qualidade do que é disponibilizado, mas também no processo de resposta dos pedidos realizados. Tratando-se de dados públicos, devemos pontuar que nem sempre algumas áreas governamentais disponibilizam dados de forma acessível como, por exemplo, saúde e educação. Pautas que muitas vezes não se desenvolvem pela falta de resposta ou, quando se tem devolutiva, a base de dados requer tratamento refinado e exaustivo por parte do profissional.

Isso impacta diretamente o processo inicial da coleta de dados e significa que dados aparentemente “publicados” na prática não existem, um indicador que corrobora outras pesquisas que demonstraram, por exemplo, que 21% dos *datasets* do Governo Federal (em nível ministerial) publicado no Portal Brasileiro de Dados Abertos não possuem os arquivos que prometem para download (SILVA et al 2020). O impacto desse conjunto de disfunções é um claro problema para o desenvolvimento do jornalismo de dados no Brasil e isso é reforçado pelo fato de 94% dos entrevistados declararem que já chegaram a desistir de uma pauta na tentativa de utilizar dados abertos governamentais brasileiros sendo que 55% afirmaram que isso ocorre com frequência.

Se por um lado temos problemas de inteligibilidade e operacionalidade que os dados demonstram, também há dificuldades de comunicação e interação entre jornalistas e curadores dos dados. Quando apenas 15,7% dos entrevistados tiveram *feedback* informativo adequado e de forma rápida e 94% relataram que já sentiram necessidade de comunicação ou já se comunicaram (mas não obteve uma experiência adequada de *feedback*) isso demonstra que há demanda de comunicação e ao mesmo tempo um baixo desempenho do gestor neste sentido. Isso nos chama atenção para pensarmos na política de dados abertos como um processo que envolve bem mais que a publicação de arquivos. A análise lexical dos discursos dos entrevistados nas entrevistas estruturadas também reforçou os dados mais quantitativos da *survey*. A formação dos *clusters* em torno de terminologias como “transparência”, dados atualizados”, “formatos abertos”, “disponibilização de dados”, “dados padronizados”, “treinamento”, “eficiência” e “dados municipais” (ou “municípios”) refletem as

demandas e os problemas enfrentados pelos jornalistas em seu processo de apropriação profissional e reúso dos dados abertos.

Por fim, para que haja avanços tanto nas políticas de transparência quanto no fortalecimento do jornalismo de dados no Brasil, é preciso que mais estudos longitudinais sejam desenvolvidos para que possamos ter uma perspectiva mais real dos problemas hoje existentes e da sua evolução ou solução.

No sentido da solução, a educação para dados pode desempenhar um papel importante. Entretanto, não basta que as escolas de jornalismo trabalhem para ampliar as expertises de futuros profissionais para lidar com a profusão de dados e os desafios decorrentes desse novo cenário. Gestores públicos e curadores de dados abertos precisam produzir melhor acompanhamento sobre a qualidade dos dados publicados e observar esses indicadores buscando uma maior efetividade neste campo, hoje fundamental para maior transparência do Estado, melhores insumos informativos para a opinião pública e, conseqüentemente, fortalecimento de processos democráticos.

Referências

ALMADA, M. P. et al. A transparência do Executivo Federal brasileiro: uma comparação entre os governos Dilma Rousseff e Jair Bolsonaro. **Opinião Pública**, v. 28, p. 169-199, 2022.

ANDERSON, C. W. Notes Towards an Analysis of Computational Journalism. **HIIG Discussion Paper Series**, n. 1, 2012. Disponível em: <http://ssrn.com/abstract=2009292>.

ATTARD, J., et al. A systematic review of open government data initiatives, **Government Information Quarterly**, 32(4), p. 399-418, 2015.

BARNETT, S.; GABER, I. **Westminster Tales: The 21st Century Crisis in British Political Journalism**. London: Continuum, 2001.

BENNETT, W. L. **News: The politics of illusion**. 9 ed. University of Washington, 2012

BRADSHAW, P. O que é o Jornalismo de Dados? In GRAY, J., BOUNEGRU, L., Chambers, L. (Org.). **Manual de Jornalismo de Dados**. Abraji: São Paulo, 2014, p. 8-10.

CODDINGTON, M. Clarifying journalism's quantitative turn: A typology for evaluating data journalism, computational journalism, and computer-assisted reporting. **Digital Journalism**, v.3, n.3, p. 331-348, 2015

COOK, T. O jornalismo político. **Revista Brasileira de Ciência Política**, n. 6, p. 203-247, 2011.

CORREIA, F. **Os jornalistas e as notícias**. Lisboa: Editorial Caminho, 1997.

HAMILTON, J., TURNER, F. Accountability Through Algorithm: Developing the Field of Computational Journalism. In **Report from the Center for Advanced Study in the Behavioral Sciences**, Summer Workshop, p. 27-41, 2009.

KOVACH, B.; ROSENSTIEL, T. **Os elementos do jornalismo**: o que os profissionais do jornalismo devem saber e o público deve exigir. São Paulo: Geração Editorial, 2003.

VIANNA JUNIOR, A. J. et al. Contribuições da Lei de Acesso à Informação para o avanço da transparência no Brasil. **Revista Gestão e Desenvolvimento**, v. 16, n. 3, p. 134-152, 2019

LIMA, W. T. Jornalismo computacional em função da "Era do Big Data". *Revista Líbero*, 14(28), 45-52, 2011.

MCNAIR, B. Journalism and democracy. In WAHL-JORGENSEN, K., HANITZSCH, T. (Org.). **The Handbook of Journalism Studies**. Londres: Routledge, 2009, p. 347-349.

MANCINI, L.; VASCONCELLOS, F. Jornalismo de Dados: conceito e categorias. **Revista Fronteiras: estudos Midiáticos**, v. 18, n. 1, p. 69-82, 2016.

MEYER, P. **Precision Journalism**: a reporter's introduction to social science methods. Bloomington: Indiana University Press, 1973.

NADERIFAR, M. Goli, H. GHALJAEI, F. Snowball Sampling: A Purposeful Method of Sampling in Qualitative Research. **Strides in Development of Medical Education**, v. 14, n. 3, 2017.

NEVEU, E. Four generations of political journalism. In Kuhn, R., Neveu, E. (Org.), **Political journalism: new challenges, new practices**. Londres: Routledge, 2002, p 22-44.

STONEMAN, J. **Does Open Data Need Journalism?** Oxford: Reuters Institute for the Study of Journalism, p. 1-32, 2015.

SILVA, S.; SOARES, A.; CESAR, D.; RABELO, L. Indicadores para avaliação qualitativa de Dados Abertos: inteligibilidade, operacionalidade e interatividade nos datasets do Governo Federal no Portal Brasileiro de Dados Abertos. **Informação & Sociedade**, v. 30, n. 1, p. 1-20, 2020.

SILVA, G. Para pensar os critérios de noticiabilidade. In: SILVA, G.; SILVA, Marcos P.; FERNANTES, M. L. (Org.). **Crériterios de noticiabilidade**: problemas conceituais e aplicações. Florianópolis: Insular, p. 51-67, 2021.

TRAQUINA, N. **O estudo do jornalismo no século XX**. São Leopoldo: Editora Unisinos, 2001

TRÄSEL, M. Jornalismo guiado por dados: aproximações entre a identidade jornalística e a cultura hacker. **Estudos em Jornalismo e Mídia**, v. 11, n. 1, p. 291-304, 2014.

Sobre os(as) autores(as)

Sivaldo Pereira da Silva é professor da Universidade de Brasília (UnB). Doutor em Comunicação e Cultura Contemporâneas pelo PósCom/UFBA, pesquisador associado ao Instituto Nacional de Ciência e Tecnologia em Democracia Digital (INCT.DD). E-mail: sivaldop@unb.br.

Ébida Santos é Doutora em Comunicação pela Universidade de Brasília (UnB). É pesquisadora do CTPol-UnB. E-mail: ebidasantos@gmail.com.

Maria Dominguez é pesquisadora associada ao Instituto Nacional de Ciência e Tecnologia em Democracia Digital (INCT.DD). Mestre em Comunicação e Cultura Contemporâneas pelo PósCom/UFBA. E-mail: mariadominguezcp@gmail.com.

Patrícia Lima é professora e coordenadora dos cursos de Jornalismo e Publicidade e Propaganda do Centro Universitário Estácio de Brasília. Doutora pela Universidade Federal de Santa Catarina (UFSC) E-mail: patricialimajornalista@gmail.com.

Data de submissão: 03/02/2022

Data de aprovação: 04/10/2022