

---

## TEM #DDJBR AQUI? Mapeando a presença do jornalismo de dados no Brasil<sup>1</sup>

Mathias Felipe DE-LIMA-SANTOS<sup>2</sup>  
Universidade de Navarra, Pamplona, Espanha

### RESUMO

Os dados são o novo petróleo no mundo dos negócios. E na realidade jornalística não poderia ser diferente. Com o surgimento da Internet e digitalização das organizações, novas formas de produção jornalística começaram a surgir. Entender os fenômenos do jornalismo de dados e o surgimento de novas mídias é um dos maiores desafios nos estudos contemporâneos do jornalismo. Partindo de uma revisão bibliográfica do jornalismo de dados no mundo e no Brasil, esse artigo reflete sobre a aplicação da prática através de um mapeamento no país. Combinando os princípios da prática junto a pesquisa, esse trabalho resulta em um estudo teórico e visual para entender até que ponto se pode afirmar que a prática realmente chegou nas organizações e quão concentrado é nas grandes capitais brasileiras. Por fim, apresenta uma análise crítica das 52 organizações mapeadas e aponta sugestões para futuras pesquisas.

**PALAVRAS-CHAVE:** jornalismo de dados; mapeamento; organizações de mídia; nativos digitais; mídia tradicional.

### INTRODUÇÃO

Os dados estão no centro de como nos comunicamos, interagimos e prosperamos em um mundo conectado. Hoje, o mundo é feito de dados e esse é o recurso mais valioso que temos para entender a sociedade. Para saber para onde estamos indo, primeiro precisamos entender onde estamos. As decisões do dia-a-dia precisam de dados e não é fácil entendê-los, por isso, é necessário torná-los visuais e legíveis (GRAY et al., 2012).

O uso de dados nas redações não é algo novo. A difusão de tecnologias digitais nos meios de comunicação ao longo do século XX alterou a prática e incorporou novos aspectos ao trabalho de jornalismo. Dessa forma, surgiram novas formas de informação, como resultado do aumento da digitalização (BRADSHAW, 2014). O jornalista Philip Meyer estabeleceria um marco no campo de RAC – Reportagem com Auxílio de Computador (do inglês, *CAR – Computer-Assisted Reporting*). Foi em 1967, o ano dos protestos dos direitos civis em Detroit, umas das manifestações

---

<sup>1</sup> Trabalho apresentado no GP Gêneros Jornalísticos, XIX Encontro dos Grupos de Pesquisas em Comunicação, evento componente do 42º Congresso Brasileiro de Ciências da Comunicação. Essa pesquisa recebeu financiamento do programa de investigação e inovação Horizonte 2020 da União Europeia no âmbito do acordo de financiamento Marie Skłodowska-Curie n.º 765140.

<sup>2</sup> Doutorando de Comunicação da Universidade de Navarra, e-mail: [mdelimas@unav.es](mailto:mdelimas@unav.es)

mais violentas da história americana, com 43 mortos e mais de 7.000 presos. Meyer com a ajuda de um *mainframe* fez levantamentos estatísticos para mostrar que as pessoas que frequentavam a faculdade eram tão suscetíveis a promoverem revoltas quanto as com o ensino médio (GRAY et al., 2012).

O termo jornalismo de dados se tornou popular desde a última década com a chegada do *Datablog* do jornal britânico *The Guardian*. A prática dados é vista como um novo método de investigar e contar histórias de maneira a informar e engajar o público em maior escala. Além disso, os jornalistas de dados têm papel importante para ajudar a diminuir as barreiras à compreensão e à imersão em dados e aumentar a alfabetização em dados de seu público em larga escala. A prática utiliza da digitalização para lidar com grande quantidade de dados, que jamais poderiam ter sido imaginados, dando nova ênfase ao uso de ferramentas tecnológicas para o processo de produção de notícias. Jornalismo de dados serve como um lembrete de quão importante a análise de *big data* (grandes massas de dados) pode ajudar no processo de trabalho de notícias (HAMMOND, 2017).

Entender os fenômenos do jornalismo de dados e o surgimento de novas mídias é um dos maiores desafios nos estudos contemporâneos do jornalismo. Como a prática amadureceu desde então, os pesquisadores usaram de diferentes métodos para alcançar seus objetivos de pesquisa, como entrevistas qualitativas ou questionários para identificar e compreender quem são os principais atores envolvidos nas principais redações e como a prática está sendo implementada (YOUNG et al., 2018). Porém, os métodos tradicionais não são mais a única forma de compreender quem são os principais envolvidos na atividade do jornalismo de dados do Brasil. Eles não são capazes de dar a dimensão desse campo em expansão, hoje é necessário que ele seja estudado em outros formatos e formas.

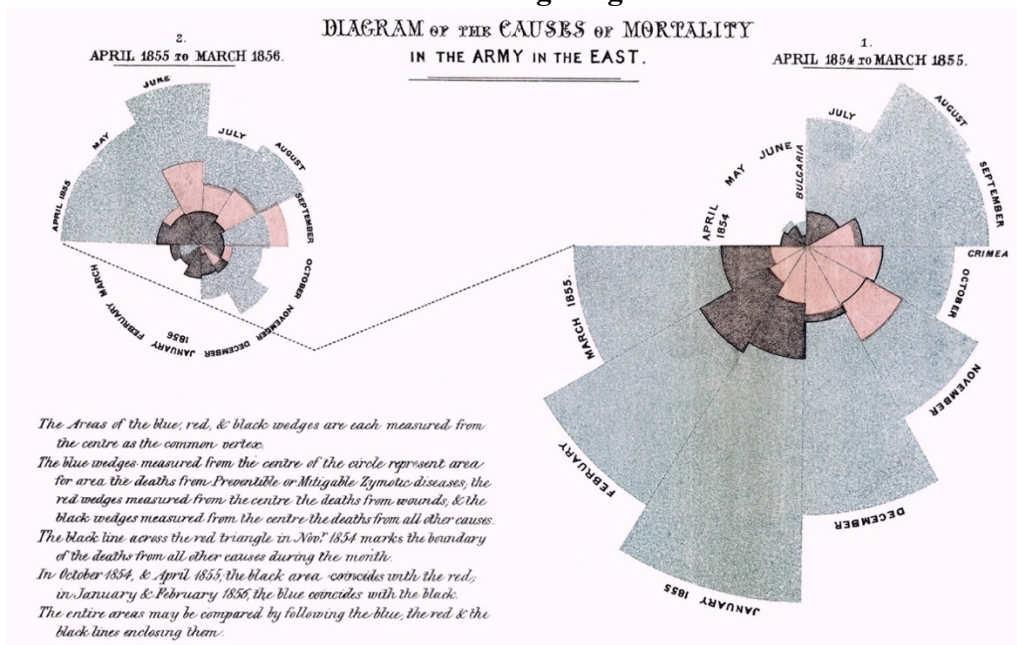
Portanto, este trabalho através da ferramentas de jornalismo de dados mapeia e dá visibilidade a prática no país. Na indústria de notícias brasileiras menos inovadora e com menos recursos (MACHADO, 2006; VENTURA, 2018), até que ponto se pode afirmar que o jornalismo de dados realmente chegou nas organizações de mídia? Quão concentrado é o jornalismo de dados nas grandes capitais brasileiras? Essas são duas questões chaves que nortearão o trabalho. O objetivo é fazer um levantamento nacional de jornalismo de dados e divulgar as instituições que aplicam a prática no país através de um mapa interativo que permite ao leitor buscar por organizações que produzam conteúdos jornalísticos baseado em dados. Através de uma metodologia que combina busca por força bruta, questionários e uso de busca em redes sociais (usando hashtags populares na área, como #jornalismodedados e a que nomeia a esse trabalho, #ddjbr, referente a *data-driven*

*journalism Brazil*), os dados serão coletados para construir o mapa do jornalismo de dados no Brasil. Conclui-se com uma agenda para possíveis desdobramentos do projeto.

## JORNALISMO DE DADOS: EM TORNO DOS CONCEITOS

Apesar da terminologia ‘jornalismo de dados’ ter se popularizado nos últimos anos com o *Datablog* do jornal britânico *The Guardian*, o uso de dados e conhecimentos estatísticos para a produção de notícias é bem anterior. A história do jornalismo de dados começa para os estudiosos em 1850 com Florence Nightingale, uma enfermeira e a primeira mulher a fazer parte da *Royal Statistical Society*. Ela teve um grande impacto, pois foi pioneira na utilização de métodos de representação visual de informações estatísticas para representar as mortes dos soldados britânicos na Guerra da Criméia, conforme mostra a figura 1 (ROGERS, 2010).

**Figura 1 – Diagrama para mostrar as causas da morte na guerra da Criméia por Florence Nightingale**



Fonte: (ROGERS, 2010).

Seu trabalho mostrou claramente que as mortes foram ocasionadas por infecção e que as doenças contraídas pelos soldados ocorreram em hospitais, algo que rotineiramente era subestimado na época. Com esses dados, Nightingale convenceu o parlamento britânico a criar normas de higiene que puderam salvar muitas vidas (SCHACHT, 2017).

Desde então, existem poucos relatos de histórias que aplicaram métodos baseado em dados. Apenas com a chegada dos primeiros computadores ao mercado que foi possível coletar e analisar dados em larga escala. Em 1952, a equipe da rede americana de radiodifusão CBS foi uma das

primeiras a usar esses métodos para prever resultados eleitorais. A equipe alimentava os resultados e pesquisas de antigas eleições presidenciais utilizando um UNIVAC (sigla para *Universal Automatic Computer*, em inglês) da Remington Rand, um dos primeiros computadores comerciais (GRAY et al., 2012).

Nos primeiros anos, até mesmo décadas da tecnologia da informação, ter acesso a uma dessas máquinas não era nada fácil e barato. A maior parte do trabalho ainda era feita à mão. Realizar uma pesquisa eleitoral em 1962 ainda significava ir de porta em porta, colecionar questionários e classificá-los manualmente por partido, candidato, idade, sexo - muitas horas de trabalho para algo que, hoje, levaria apenas alguns cliques no computador. O ano de 1952 foi um marco, pois foi a primeira vez que foi utilizado o RAC para prever o resultado da eleição presidencial (GRAY et al., 2012).

Mais tarde, no ano de 1967, o então jornalista Philip Meyer estabeleceria outro marco no campo. Esse período foi marcado pelos protestos dos direitos civis em Detroit, numa das manifestações mais violentas da história americana, com 43 mortos e mais de 7.000 presos. Meyer com a ajuda de um *mainframe* fez levantamentos estatísticos para mostrar que as pessoas que frequentavam a faculdade eram igualmente suscetíveis a promoverem revoltas, assim como as que haviam chegado ao ensino médio. Ele delineou uma nova forma de produção de conteúdo, conhecido como jornalismo de precisão, modelado a partir das ciências sociais, usando métodos empíricos (particularmente pesquisas e análise de conteúdo) e análise estatística para obter respostas mais definitivas às questões jornalísticas (COX, 2000; GRAY et al., 2012; MEYER, 2002).

Apenas no final dos anos 80 e início dos anos 90 que o jornalismo de precisão, desde a reformulação como RAC, começou a fazer incursões significativas nas redações, liderado por várias histórias vencedoras do Prêmio Pulitzer, que se tornaram um forma de reconhecimento e validação profissional da prática, evoluindo para o que é conhecido hoje como jornalismo de dados (CODDINGTON, 2015; SCHACHT, 2017).

Embora o jornalismo de dados tenha sido usado informalmente por profissionais de RAC durante décadas, o primeiro uso registrado do termo por uma grande organização de mídia é do *The Guardian*, que lançou seu *Datablog* em março de 2009. Embora a paternidade do termo seja contestada, é amplamente utilizado desde que os documentos da Guerra do Afeganistão do *Wikileaks* vazaram em julho de 2010 (CODDINGTON, 2015; GRAY et al., 2012; SCHACHT, 2017).

---

## O DESENVOLVIMENTO DO JORNALISMO DE DADOS NO MUNDO

Os debates acadêmicos sobre jornalismo de dados foram implementados por vários pesquisadores ao longo dos últimos anos, em geral com respeito às normas, práticas e culturas de jornalismo de dados a nível nacional. Dentre esses estudos incluem-se os dos Estados Unidos (FINK & ANDERSON, 2015; PARASIE & DAGIRAL, 2013; ROYAL & BLASINGAME, 2015; YOUNG & HERMIDA, 2015), Reino Unido (BORGES-REY, 2016; HANNAFORD, 2015), Canada (HERMIDA & YOUNG, 2017; TABARY et al., 2016), Bélgica (DE MAEYER et al., 2015). Já no Brasil, o estudo da prática baseada em redações ainda é pouco estudada (ARAÚJO, 2018; MACHADO, 2016; MANCINI & VASCONCELLOS, 2016; RODRIGUES, 2014; TRÄSEL, 2013, 2014, 2018; VASCONCELLOS et al., 2015; VENTURA, 2018).

A pesquisa etnográfica de Royal e Blasingame (2015) teve enfoque em como a equipe do The New York Times utilizava dados como parte dos projetos de reportagem baseada em dados. Os autores inferiram que as partes envolvidas os viam como jornalistas afetados pelos valores do mundo digital de colaboração, inovação e criatividade. Em outro trabalho nos Estados Unidos, os autores Parasie e Dagiral (2013) mostram que em Chicago existe a presença dos jornalistas-programadores, também conhecido como *journo-coders*, que surgiram com a utilização de tecnologia nas redações. Já o estudo de Young e Hermida (2015) focou no Los Angeles Times. Suas descobertas refletiram o que Parasie e Dagiral (2013) haviam percebido em Chicago, um movimento de crescente da demanda por jornalistas-programadores pelo veículo. Eles também sugeriram que trazer programadores para a redação ajuda moldar os ideais jornalísticos. A tecnologia ajuda a dar formas e contornos ao jornalismo, assim como o processo inverso também ocorre.

Na Europa, os estudos mostraram diferentes alinhamentos em relação à tecnologia e à identidade do jornalismo. No Reino Unido, muito menos tem sido mencionado sobre os *journo-coders* e sobre sua existência nas organizações de mídia, embora alguns estudos de casos da primeira edição do Manual de Jornalismo de Dados (Data Journalism Handbook, em inglês) mostrem a sua existência (GRAY et al., 2012; HERMIDA & YOUNG, 2017). Nos estudos conduzidos por Hannaford (2015), os dois maiores veículos de mídia do Reino Unido, BBC e Financial Times, foram analisados e os resultados mostraram a presença de equipes híbridas trabalhando na produção de dados, com programadores, jornalistas e designers trabalhando juntos para produzir conteúdo multimídia e interativo (HANNAFORD, 2015; HERMIDA & YOUNG,

2017). Essa abordagem foi amplamente adotada por outras empresas de mídia no mundo, como na Austrália e Brasil (MIELNICZUK & TRÄSEL, 2014; TRÄSEL, 2017).

O jornalismo de dados representa, para muitos autores, a democratização de recursos, ferramentas, técnicas e métodos que antes eram restritos a poucos, como especialistas e governos. Os profissionais que trabalham com a prática têm papel essencial em atenuar as barreiras para a compreensão e imersão nos dados e aumentar a alfabetização de dados em larga escala da audiência (GRAY et al., 2012). Entretanto, organizações tradicionais estão acostumados a uma hierarquia estrita, que tem um impacto negativo na inovação e produção de novas formas e formatos noticiosos, gerando um efeito direto no jornalismo de dados. “Recursos, vontade da liderança e possibilidade para implantar [reportagens baseadas em dados] são fatores determinantes para que as habilidades de jornalistas de dados e computacionais exerçam suas atividades” (HERMIDA & YOUNG, 2017, p.171, tradução nossa).

Além disso, a prática foi capaz de criar uma ‘ponte’ entre tecnólogos e jornalistas nas redações, expandindo seus horizontes nas práticas e nos processos e evoluindo para além das ilhas de conhecimento, permitindo explorar novos tópicos em um nível mais profundo. Dessa forma, há um fortalecimento da capacidade competitiva dos veículos para sobreviver na era digital (CAIRO, 2016; HERMIDA & YOUNG, 2017).

## **ENTENDENDO O JORNALISMO DE DADOS NO BRASIL**

Enquanto no mundo jornalismo de dados está em expansão e consolidação, a prática no Brasil ainda está em desenvolvimento. Um dos pioneiros no trabalho de jornalismo de dados e RAC foi Cláudio Weber Abramo, que morreu em 2018, aos 72 anos de idade. Abramo era bacharel em matemática pela USP e mestre em filosofia da ciência pela Unicamp (ASSIS, 2018; ROCHA, 2018; VENTURA, 2018).

O jornalista foi editor de Economia da Folha de São Paulo (1987) e secretário-executivo da Gazeta Mercantil (1987-88), além de ter passado por outras publicações como Valor Econômico, Abril e IstoÉ. Pioneiro do jornalismo de dados no Brasil com projetos como os ‘Às Claras’ e ‘Excelências’, Abramo era defensor da integridade das instituições públicas, da transparência, do acesso a dados públicos e uma referência no combate à corrupção no Brasil. Como pai da prática decidiu fundar em 2000, junto a um grupo de indivíduos e entidades não-governamentais comprometidos com o combate à corrupção, a Transparência Brasil, organização que se dedica ao acesso à informação e ao combate à corrupção. Entre os vários projetos desenvolvidos pela



organização está o ‘Excelências’, uma base de dados online com informações de processos e do desempenho de parlamentares, ganhador do Prêmio Esso de Melhor Contribuição à Imprensa em 2006.

Outro projeto de destaque de Abramo foi o ‘Às Claras’, uma plataforma que organizava e disponibilizava gastos das eleições (ASSIS, 2018; ROCHA, 2018). O matemático-jornalista também foi um dos mais importantes articuladores para a elaboração da Lei de Acesso à Informação (LAI), aprovada em 2011, que assegurava o direito de acesso aos dados produzidas por órgãos e entidades da União, Estados e Municípios. O trabalho de Abramo abriu portas para outras organizações começarem a trabalhar com dados abertos e jornalismo de dados. Em 2012, o Estado de São Paulo criou o Estadão Dados, a primeira equipe voltada para a produção de reportagens baseadas em dados em uma redação. Considerado hoje um dos casos mais bem-sucedidos, a equipe já criou reportagens baseadas em dados do Twitter, como “O que 15 mil tweets revelam sobre seu candidato” para as eleições presidenciais de 2018 (ASSIS, 2018; ROCHA, 2018).

Outra equipe que vem se destacando é a G1 Dados, do portal G1.com parte do grupo Globo. Com base na cidade de São Paulo, eles foram um dos primeiros brasileiros premiados no maior prêmio de jornalismo de dados global, o *Data Journalism Awards*. O projeto “Monitor da Violência” foi premiado na categoria escolha pública de 2018 e prometia investigar os assassinatos no Brasil ao longo de uma semana. O objetivo era entender o que está impulsionando a violência, humanizar os dados e apelar ao governo que tome uma ação. O grosso da investigação ocorreu entre 21 e 27 de agosto de 2017 e envolveu 230 jornalistas de 55 redações afiliadas da emissora em todo o país. O projeto foi tão grande que reuniu não só jornalistas do grupo, como também uma universidade e uma ONG para fazer uma investigação a nível nacional sobre as mortes pela violência. Entre os resultados assustadores obtidos, eles descobriram que houve um assassinato a cada oito minutos (BOUCHART, 2017).

Um ano antes, a ABRAJI (Associação Brasileira de Jornalistas Investigativos) ganhou o mesmo prêmio na edição de 2017, porém na categoria de pequenas redações jornalísticas. Competindo com 99 trabalhos de 30 países, o projeto CTRL+X mapeava as tentativas de censura a veículos de mídia no Brasil. O portal apresentava na época do prêmio cerca de 2500 processos judiciais de políticos contra empresas de mídia. Segundo Tiago Mali, coordenador do projeto, a ferramenta foi útil para acompanhar a liberdade de expressão no país, principalmente durante o período eleitoral de 2016 (NAHSER, 2018).

Foi tão importante o trabalho de Abramo para o jornalismo de dados, que abriu a portas para esses reconhecidos projetos. Hoje, o jornalista-matemático dá nome ao recém-criado prêmio

---

para reconhecer a excelência na produção de trabalhos jornalísticos guiados por dados no Brasil, o prêmio Cláudio Weber Abramo de Jornalismo de Dados<sup>3</sup>, que terá sua primeira edição ao final de 2019. Com esse reconhecimento, espera-se que o Brasil consiga se desenvolver ainda mais na produção de jornalismo de dados e atinja a maturidade alcançada por outros países.

## **METODOLOGIA: DESVENDADO OS CAMINHOS DA PESQUISA**

Nessa seção esmiuçamos os caminhos da pesquisa a partir da descrição da metodologia adotada, por entender que a escolha utiliza algo inovador, combinando o uso da prática à pesquisa. Portanto, o objetivo central é apresentar as etapas de construção do trabalho, ao mesmo tempo em que descreve o caminho de produção de uma história baseada em dados.

O trabalho utilizou, portanto, de diferentes práticas para a produção do produto jornalístico resultante, baseando na pirâmide invertida desenvolvida por Bradshaw (2007), conforme mostra figura 2, para descrever o *pipeline* de produção de uma história baseado em dados. Segundo o autor existem duas formas de como começar uma reportagem de dados: ou você tem uma pergunta que precisa de dados para ser respondida ou possui um conjunto de dados que precisa ser questionado. Seja qual for, a primeira fase é a compilação de dados. O autor defende que esse é o estágio mais importante, não apenas porque todo o resto depende dessa parte, mas porque é provavelmente o estágio que será retornado mais vezes em cada um dos estágios subsequentes.

Para criar uma base de dados com os meios que produzem histórias baseadas em dados foi utilizado um questionário disponibilizado em grupos de WhatsApp de dados abertos, de Facebook de jornalismo de dados e dados abertos, além de coletar dados do Twitter com as hashtags #ddjbr e #jornalismodedados. Foram coletados dados do Twitter durante o início de 2019, de março até meados de abril, através da plataforma Workbench<sup>4</sup>. Apesar da comunidade de jornalismo de dados brasileira produzir pouco conteúdo na rede social comparado a mundial, através das hashtags #ddjbr e #jornalismodedados foi possível encontrar outros veículos que não haviam sido mapeados pelos informantes, assim como um perfil @ddj\_br, que produzia newsletter até o fim de 2018 sobre reportagens baseadas em dados. Esse perfil foi essencial para mapear exemplos de histórias das grandes organizações.

Agregando os dados de todas essas fontes foi permitido gerar um banco de dados com cinquenta e dois ( $n=52$ ) organizações que produzem histórias baseadas em dados, descrito com

---

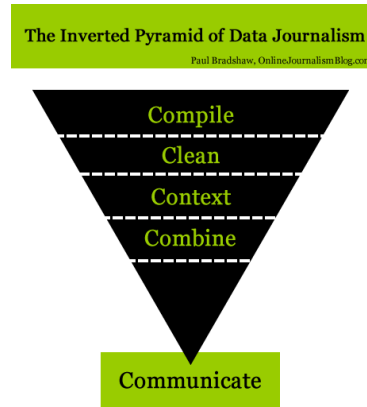
<sup>3</sup> <https://jornalismodedados.org>, Acesso em: 1º de maio de 2019.

<sup>4</sup> Workbench - <http://workbenchdata.com/>



mais detalhes na análise. Manualmente foram adicionados os endereços corretos e as informações de descrição de cada meio.

**Figura 2 – Pirâmide Invertida**



Fonte: (BRADSHAW, 2007).

Com os dados obtidos, a próxima fase é a limpeza. Ter confiança nas histórias escondidas atrás dos dados, significa poder confiar na qualidade dos mesmos, e isso é possível através de um processo de limpeza cuidadoso. Para Bradshaw (2007), essa etapa tem dois objetivos básicos: remover o erro humano e converter os dados em um formato consistente com outros dados que estão sendo utilizados. Nesse produto jornalístico, essa etapa foi fundamental para padronização de dados e limpar erros humanos que foram adicionados através do formulário. A ferramenta *open source* OpenRefine foi usada para auxiliar no processo de limpeza. Essa plataforma é utilizada para ajudar na limpeza através da execução de quatro tarefas essenciais: remoção de registros duplicados, separação de valores contidos no mesmo campo, análise da distribuição de valores ao longo de um conjunto de dados e agrupamento de diferentes representações do mesmo tipo.

A próxima etapa na pirâmide é a contextualização. Como qualquer em qualquer fonte jornalística, os dados nem sempre podem ser confiáveis. Eles podem vir com suas próprias histórias, preconceitos e objetivos. Então Bradshaw (2007) pede para questionar seus dados: quem coletou, quando e com que propósito? Como foi reunido? O que exatamente eles querem dizer? Nesse caso, como os dados foram coletados com a ajuda de terceiros, foi preciso fazer a checagem de cada *link* para ver a conformidade com o que foi definido como jornalismo de dados no questionário. Seguindo a definição de Vasconcellos et al. (2015), o jornalismo de dados é a prática jornalística que exige a coleta, organização e exploração de dados para se obter relações que podem ser significativas entre os dados. Com base nisso, foram checadas todas as histórias e portais das organizações para verificar o alinhamento com o jornalismo de dados. Após cuidadosa análise, foram considerados para o mapeamento organizações que não são apenas de mídia, mas que

---

produzem conteúdo jornalístico feito por atores com formação na área, como a Sociedade Brasileira de Pediatria e o Conselho Federal de Medicina.

A penúltima fase dessa pirâmide invertida é a combinação. O autor, Bradshaw (2007), acredita que boas reportagens podem ser encontradas em um único conjunto de dados, mas geralmente é necessário combinar duas ou mais bases de dados. Assim como no jornalismo, é melhor uma história de múltiplas fontes do que uma única fonte. A combinação também pode vir através de outras formas, com o uso de dados e entrevistas. Seguindo os preceitos de Bradshaw (2017), esse trabalho utilizou de dois canais diferentes para mapeamento de dados para a geração de uma base de dados. Os dados foram coletados através de um questionário e através do Twitter. Essas duas fontes permitiram gerar uma tabela com as organizações que trabalham com jornalismo de dados no Brasil.

A última fase do jornalismo de dados é a ‘comunicação’. Considerada a mais relevante, pois visualizar os resultados – em geral, através de um mapa, gráfico ou infográfico, ou ainda em uma animação ou produto interativo – é o que diferencia a prática do uso de dados pelo jornalismo (BRADSHAW, 2007). Como esse trabalho busca utilizar preceitos do jornalismo de dados, foi utilizado um mapa interativo que permite ao usuário selecionar cada ponto no mapa e obter mais informação sobre cada organização, assim como acessar um *link* para uma história baseada em dados. Para ajudar o trabalho de jornalistas, *startups* estão desenvolvendo soluções prontas que eliminam a necessidade do jornalista de programar e oferecem formas de visualização dos dados. Essas ferramentas são conhecidas como *out-of-the-box* e ajudam na produção de reportagens principalmente para pequenas e médias organizações de mídia que não tem equipes suficientes. Nesse caso, a ferramenta Carto<sup>5</sup> foi escolhida para apresentação dos dados através de um mapa utilizando em associação com programação para web.

## **DADOS SOBRE DADOS: QUEM TRABALHA COM JORNALISMO DE DADOS NO BRASIL?**

Dada a necessidade de estudar o jornalismo de dados e entender sua presença no país e, dessa forma, descobrir quem são os principais atores que estão desenvolvendo essa prática no país utilizamos do preceitos da prática para esse estudo. A primeira tarefa desafiadora era como mapear a maior parte dessas organizações. No mundo, o jornalismo de dados é conhecido por ser uma prática jornalística que tem muitos valores do mundo *hacker*, como colaboração e tecnofilia (TRÄSEL,

---

<sup>5</sup> <https://carto.com>

2018). Acreditando nesses valores, um questionário foi enviado para grupos de pessoas que trabalham com dados no Brasil.

O questionário foi dirigido a dois públicos alvos: pessoas que trabalham com jornalismo de dados e pessoas que se interessam sobre o assunto. Conforme o perfil do informante, as perguntas eram diferentes. As questões iniciais focavam em entender se a pessoa trabalhava em alguma organização com jornalismo de dados ou ela tinha interesse na prática. Aos informantes que trabalham com jornalismo de dados foram, então, direcionadas questões que perguntavam o nome, cidade, estado, CEP da organização, assim como link para uma história baseada em dados e contato da equipe de jornalismo de dados. Já aqueles que responderam que não trabalham em organizações de mídia foram pedidos os dados similares, porém com enfoque de cadastro de organizações de mídia. Ao todo foram coletados vinte ( $n=19$ ) entradas com dados de organizações.

Com o objetivo de obter o maior número possível de empresas que trabalham com jornalismo de dados, recorreremos ao Twitter para encontrar outros atores que não chegaram até o questionário. Foram agregados os dados de todas essas fontes, que, então, foram limpados para eliminar informações duplicadas ou faltantes. Ao todo foi possível gerar um banco de dados com cinquenta e duas ( $n=52$ ) organizações. Das entidades mapeadas é possível ver que grande parte delas são organizações de mídia independentes ou nativas digitais (51,92%), desde as mais populares como UOL, Nexo, G1, Poder360 e Intercept, até as menores e mais desconhecidas da grande população como Colaboradados, Datalab, Filtro Fact-Checking e Volt Data Lab. Logo atrás estão as mídias tradicionais que representam 36,5% do total, a lista inclui empresas tradicionais brasileiras, como TV Globo, Globonews, Estado de S. Paulo, Folha de S. Paulo, assim como oriundas de fora do país, como BBC e El País. A menor parte é composta de organizações que não são de mídia, porém contam com jornalistas de em seu corpo de funcionários produzindo histórias baseada em dados.

**Tabela 1 – Número de Organizações por cidade**

<i>Tipo de Organização</i>	<i>Organização de mídia</i>
<b>Mídia Tradicional</b>	19 (36,5%)
<b>Nativa Digital</b>	27 (51,92%)
<b>Organização não-midiática</b>	4 (7,69%)
<b>Outros</b>	2 (4%)
<i>Total</i>	<b>52</b>

Fonte: Elaborada pelo autor com base nos dados coletados.

Com os dados obtidos pelos informantes e através do Twitter, pode-se verificar que existiam informações que não estavam completas, devido a falta de conhecimento sobre esse dado por parte do informante. Por isso, as organizações que possuíam apenas as cidades, e não possuíam

endereço disponíveis na *web*, foram colocadas em pontos centrais da cidade onde se localizam. As demais foram adicionadas nos endereços corretos buscando no Google Maps. Através de um *script* foi possível coletar a latitude e longitude através da códigos da API do Google Maps de cada um dos endereços disponíveis e colocar com precisão onde se localiza cada meio.

**Tabela 2 – Número de Organizações por cidade**

<i>Cidades</i>	<i>Total de organizações</i>
<b>Brasília</b>	3
<b>Campo Grande</b>	1
<b>Curitiba</b>	2
<b>Florianópolis</b>	2
<b>Fortaleza</b>	1
<b>Porto Alegre</b>	1
<b>Recife</b>	2
<b>Rio Claro</b>	1
<b>Rio de Janeiro</b>	13
<b>Salvador</b>	1
<b>São José dos Campos</b>	1
<b>São Paulo</b>	23
<b>Vitória</b>	1
<i>Total de organizações</i>	<i>52</i>

Fonte: Elaborada pelo autor.

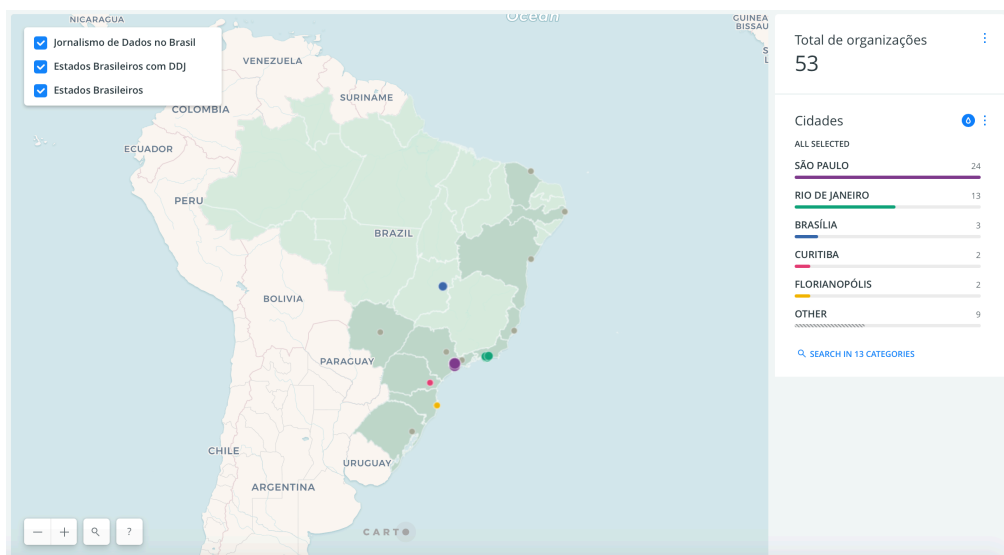
Com os dados coletados, compilados e limpos, foi necessário voltar a pergunta proposta para entender a melhor forma de visualizá-los. O que esse projeto propõe é responder onde há jornalismo de dados no Brasil. Tendo essa proposta em mente, a melhor forma para fazer isso seria a utilização de uma ferramenta para mapeamento. Como mencionado anteriormente, Carto foi escolhida para apresentação dos dados, pois permite a visualização e análise de dados geográficos, além de um mapa utilizando em associação com programação para web. Antes conhecida como CartoDB, é uma plataforma de computação em nuvem de Software como Serviço (SaaS) que fornece ferramentas de sistema de informação geográfica (SIG) e ferramentas de mapeamento na web para exibição em um navegador.

Apesar de ser uma ferramenta simples, o Carto possui interface para programação onde é possível alterar *layout, design*, assim como a leitura dos dados com código SQL. Para o mapeamento foi preciso utilizar codificação CSS e HTML para ajustar detalhes da visualização, como o número de cores a serem exibidas. A primeira etapa foi a escolha do tipo de mapa de base que seria utilizado para exibição dos dados. Foi escolhido o *Voyage* com base nos princípios de clareza e evitar uma poluição visual com rodovias e informações hidrográficas que não serão úteis

nesse caso (CAIRO, 2016). O mapa base apresenta as divisões de estados e países, facilitando a localização visual dos dados apresentados, conforme mostra a figura 3.

Foram utilizadas treze (13) cores para representar as cidades onde existe pelo menos uma organização atuando com jornalismo de dados. O tamanho da circunferência é maior conforme a maior presença de organizações dentro da mesma cidade, São Paulo e Rio de Janeiro lideram, respectivamente, como mostra a tabela 2. Para visualizar essas informações em forma numérica foi decidido colar um *widget* lateral que apresenta o total de organizações e o total por cada cidade. Conforme o usuário aproxima o mapa de uma cidade e as demais não aparecem, tanto o número total de organizações quanto as cidades atualizam para apresentar apenas os dados correspondentes das cidades que estão enquadradas no momento. Esse mapa foi produzido de forma que possa ser facilmente embutido em um website, permitindo que assim outras pessoas ou organizações possam adicionar em seus portais disseminando o jornalismo de dados para mais brasileiros.

**Figura 3 – Mapa do jornalismo de dados no Brasil.**



Fonte: Captura realizada pelo autor.

Seguindo os preceitos de transparência também presentes no jornalismo de dados, o resultado desse estudo pode ser acessado através do link abaixo: <http://bit.ly/MapaDDJBr>

## CONCLUSÃO

O projeto não só coloca em prática o jornalismo de dados, mas permite fazer análise críticas sobre a relevância dessa disciplina no país. A presença massiva de organizações nas duas maiores capitais, São Paulo (23 empresas) e Rio de Janeiro (13), mostram que o jornalismo de dados ainda não está consolidado no país. A limitação dessa prática nas capitais pode ter reflexo no papel fundamental do jornalismo na atualidade. No nosso entendimento o papel da imprensa como cão

---

de guarda para sociedade pode ser influenciado por essa falta de desenvolvimento do jornalismo de dados.

O país tem avanços significativos na política de dados abertos com a Lei da Transparência. Essa lei brasileira, sancionada em 2009, obriga a União, os estados e os municípios a divulgar seus gastos na Internet em tempo real. As investigações jornalísticas tradicionais estão longe de abraçar dados como fonte de informação, por isso o jornalismo de dados é fundamental para a atuação do jornalismo.

O presente projeto não aborda as razões que levam as organizações de mídia brasileira a não adotarem a prática. Mas pesquisas no mundo mostraram que a limitação de tempo ou recursos influenciam adoção do jornalismo de dados. A crise da indústria jornalística afeta a decisão por reportagens que demandam tempo, pois cada recurso investido necessita ser bem aproveitado. O erro está exatamente nessa concepção, jornais tradicionais que investiram em novas formas de conteúdo para engajar a audiência e conseguir criar novos modelos de negócios sustentáveis, como The Guardian e The New York Times, são exemplos de sucesso.

Além da concentração de jornalismo de dados nas cidades do Rio e de São Paulo, o mapa mostra que veículos de imprensa locais que trabalham com jornalismo de dados ainda são poucos ou não tem presença destacada na rede mundial de computadores. Isso reflete que menos pessoas serão expostas e alfabetizadas no uso de dados.

Esse projeto também mostrou que não são apenas organizações de mídias que estão produzindo conteúdo jornalístico com dados. Com a crise, jornalistas estão sendo contratados para trabalhar em outras organizações e passam a fazer conteúdo para os canais dessas empresas. Esses profissionais aproveitam do jornalismo de dados para criarem histórias que permitam envolver suas audiências. As organizações Sociedade Brasileira de Pediatria, Conselho Federal de Medicina e APCEF/SP são apenas algumas das organizações em que se pode identificar a presença de jornalistas de dados.

Outro ponto a se destacar é a grande presença de mídias nativas digitais na produção de conteúdo baseado em dados. Apesar da prática existir em mídias tradicionais, a presença dos nativos digitais é mais massiva. Além disso, existem organizações de mídia que se apresentam como focadas apenas em dados, como Volt Lab, Gênero e Número, Fiquem Sabendo, entre outras.

Vale ressaltar, que esse projeto possui limitações que não permite analisar o nível de maturidade de cada mídia na produção de jornalismo de dados. Além dessas limitações já mencionadas, esse projeto não apresenta um widget para busca por macrorregiões, como bairros e ruas. Infelizmente nem todos as organizações puderam ser contatadas ou possuíam informações



que garantiam a precisão do endereço, por isso um busca por macrorregião não seria tão fidedigna. Outro ponto, que foi testado com a audiência e não foi possível melhorar devido as limitações da plataforma Carto e de seu uso em aplicações móveis, foi as janelas de informações sobre as organizações. Torná-las mais colorida ou atrativa para o usuário aumentaria o espaço requisitado da tela, dificultando a utilização do mapa por um usuário de celular.

Com base em todos os pontos mencionados acima, pode-se ver que o jornalismo de dados tem grande potencial de exploração no país e que precisa de atores espalhados pelas diversas capitais. Só com uma presença massiva no país, o jornalismo de dados pode ser apreciado e, assim criar de um único nome e imagem para a audiência e para as redações brasileiras. Assim, criaremos uma consciência sobre a importância de estabelecer um significado do jornalismo de dados como produto de qualidade para atrair e reter clientes leais.

## REFERÊNCIAS

- ARAÚJO, L. V. A web e o jornalismo de dados: mapeamento de conceitos chave. **Dispositiva**, v. 5, n. 1, p. 144–163, 2018.
- ASSIS, C. Morre Cláudio Weber Abramo, defensor da transparência e pioneiro do jornalismo de dados no Brasil. Disponível em: <<https://knightcenter.utexas.edu/pt-br/blog/00-20034-morre-claudio-weber-abramo-defensor-da-transparencia-e-pioneiro-do-jornalismo-de-dados>>. Acesso em: 4 de maio de 2019.
- BORGES-REY, E. Unravelling Data Journalism. **Journalism Practice**, v. 10, n. 7, p. 833–843, 2 out. 2016.
- BOUCHART, M. Holding the powerful accountable, using data. Disponível em: <<https://medium.com/data-journalism-awards/holding-the-powerful-accountable-using-data-f0bd5ebb24e3>>. Acesso em: 7 maio. 2019.
- BRADSHAW, P. The inverted pyramid of data journalism. p. 1–7, 2007.
- BRADSHAW, P. **Stories and Streams - Teaching collaborative journalism with peer to peer learning**. 2017.
- BRADSHAW, P. Data Journalism. In: **Ethics for Digital Journalists**. [s.l.] Routledge, 2014. p. 214–232.
- CAIRO, A. **The Functional Art**. 1. ed. Berkeley: New Riders Publishing, 2016.
- CODDINGTON, M. Clarifying Journalism’s Quantitative Turn. **Digital Journalism**, v. 3, n. 3, p. 331–348, 4 maio 2015.
- COX, M. The Development of Computer-Assisted Reporting. **Newspaper Division, Association for Education in Journalism and Mass Communication, Southeast Colloquium**, v. 2030, n. 305, p. 22, 2000.
- DE MAEYER, J. et al. Waiting for Data Journalism. **Digital Journalism**, v. 3, n. 3, p. 432–446, 4 maio 2015.
- FINK, K.; ANDERSON, C. W. Data Journalism in the United States. **Journalism Studies**, v. 16, n. 4, p. 467–481, 4 jul. 2015.
- GRAY, J.; BOUNEGRU, L.; CHAMBERS, L. **The Data Journalism Handbook**. 1st. ed. Sebastopol: O’Reilly Media, 2012.
- HAMMOND, P. From computer-assisted to data-driven: Journalism and Big Data. **Journalism**, v. 18, n. 4, p. 408–424, 2017.

- HANNAFORD, L. Computational journalism in the UK newsroom. **Journalism Education**, v. 4, n. 1, p. 6–21, 2015.
- HERMIDA, A.; YOUNG, M. L. Finding the Data Unicorn. **Digital Journalism**, v. 5, n. 2, p. 159–176, 7 fev. 2017.
- MACHADO, E. **O jornalismo digital em base de dados**. [s.l.] Calandra, 2006.
- 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**. [s.l.] Rowman & Littlefield Publishers, 2002.
- MIELNICZUK, L.; TRÄSEL, M. Data-driven journalism as professional innovation and its challenges for education. **Contemporanea Comunicação e Cultura**, v. 15, n. 2, p. 609–629, 2017.
- NAHSER, F. Behind the scenes of a murder investigation that mobilised 230 journalists all over Brazil. Disponível em: <<https://medium.com/data-journalism-awards/behind-the-scenes-of-a-murder-investigation-that-mobilised-230-journalists-all-over-brazil-e5136e2ea1ea>>. Acesso em: 7 maio. 2019.
- PARASIE, S.; DAGIRAL, E. Data-driven journalism and the public good: “Computer-assisted-reporters” and “programmer-journalists” in Chicago. **New Media & Society**, v. 15, n. 6, p. 853–871, 18 set. 2013.
- ROCHA, C. Como Claudio Weber Abramo trabalhou pela transparência. **Nexo**, v. 2018, p. 19–22, 2018.
- RODRIGUES, Kelly de Conti. Jornalismo de dados na web: estudo da produção de sentido na infografia do blog do Estadão Dados e do La Nación Data Blog. In: **10º Interprogramas de Mestrado em Comunicação da Faculdade Cásper Líbero**. [s.l.]: Faculdade Cásper Líbero, 2014.
- ROGERS, S. Florence Nightingale, datajournalist: information has always been beautiful. **The Guardian**, 13 ago. 2010.
- ROYAL, C.; BLASINGAME, D. Data journalism: An explication. **International Symposium in Online Journalism**, v. 5, n. 1, p. 24–46, 2015.
- SCHACHT, K. The history of data driven journalism –. Disponível em: <<https://datentaeter.de/the-history-of-data-driven-journalism/>>. Acesso em: 1 maio. 2019.
- TABARY, C.; PROVOST, A. M.; TROTTIER, A. Data journalism’s actors, practices and skills: A case study from Quebec. **Journalism**, v. 17, n. 1, p. 66–84, 2016.
- TRÄSEL, M. JORNALISMO GUIADO POR DADOS: relações da cultura hacker com a cultura jornalística. **Anais do XXII Encontro Anual da Compós, Universidade Federal da Bahia**, p. 1–16, 2013.
- 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, 2014.
- TRÄSEL, M. Hacks and hackers: the ethos and beliefs of a group of Data-Driven Journalism professionals in Brazil. **Repórteres e hackers: o ethos e as crenças de um grupo de profissionais de Jornalismo Guiado por Dados no Brasil**, v. 25, n. 1, p. 1–14, 2018.
- VASCONCELLOS, F. et al. The five Data Journalism categories, or how to understand Data Journalism in Brazil. p. 1–24, 2015.
- VENTURA, M. P. Jornalismo de Dados como diferencial: o caso do Nexa. **Pauta Geral - Estudos em Jornalismo**, v. 5, n. 2, p. 240–254, 2018.
- YOUNG, M. L.; HERMIDA, A. From Mr. and Mrs. Outlier To Central Tendencies. **Digital Journalism**, v. 3, n. 3, p. 381–397, 2015.
- YOUNG, M. L.; HERMIDA, A.; FULDA, J. What Makes for Great Data Journalism?: A content analysis of data journalism awards finalists 2012–2015. **Journalism Practice**, v. 12, n. 1, p. 115–135, 2018.