

GÊNERO E NÚMEROS:

o uso de dados na cobertura do Dia Internacional da Mulher nos sites de três grandes jornais brasileiros

Copyright © 2019
SBPjor / Associação
Brasileira de Pesquisadores em Jornalismo

ANA CAROLINA ARAÚJO
Universidade Federal da Bahia, Salvador – BA – Brasil
ORCID: 0000-0002-3758-0083

DOI: 10.25200/BJR.v15n1.2019.1061

RESUMO – Na última década, o jornalismo de dados evoluiu ao lado das tecnologias de informação e comunicação, espalhando-se cada vez mais pelo ambiente do jornalismo profissional. A proposta deste artigo é compreender se e como o jornalismo impresso tradicional tem reagido às novas ferramentas tecnológicas disponíveis para a prática da reportagem e à crescente disponibilidade de dados abertos, sejam eles de origem pública ou privada. Para investigar estas questões, foram utilizadas como material de análise reportagens publicadas nos sites de três grandes jornais brasileiros – *Folha de S. Paulo*, *O Globo* e *O Estado de São Paulo* – durante a cobertura do Dia Internacional da Mulher em 2017. Os resultados indicam que a presença dos dados ainda ocorre na menor parte do material analisado, que o uso das técnicas digitais de visualização é incipiente e a oferta de dados brutos no processo de produção da notícia não existe. Por outro lado, foi possível perceber que o uso de dados e estatísticas sem indicação ou checagem de fontes é bastante frequente no *corpus* estudado.

Palavras-chave: Jornalismo de dados. Jornalismo guiado por dados. Jornalismo digital. Dia Internacional da Mulher.

GENDER AND NUMBERS:

Using data from International Women's Day coverage on the sites of three major Brazilian newspapers

ABSTRACT – In the last decade, data journalism has evolved alongside information and communication technologies, spreading more and more into the professional journalism environment. The purpose of this article is to understand if and how traditional press has reacted to the new technological tools available for the practice of reporting and the increasing availability of open data, whether public or private. To investigate these

issues, articles published on the websites of three major Brazilian newspapers – *Folha de S. Paulo*, *O Globo* and *O Estado de São Paulo* – were used to analyse the coverage of International Women's Day in 2017. The results indicate that the presence of data is still scarce in the biggest part of the material, that the use of digital visualization techniques is incipient and the raw data supply in the news production process does not exist. On the other hand, it was possible to perceive that the use of data and statistics without reference or source checking is quite frequent in the *corpus*.

Key words: Data Journalism. Data-driven Journalism. Digital Journalism. International Women's Day.

GÊNERO Y NÚMEROS:

El uso de datos en la cobertura del Día Internacional de la Mujer en los sitios de tres grandes periódicos brasileños

RESUMEN – En la última década, el periodismo de datos creció al lado de las tecnologías de información y comunicación, tomando el campo del periodismo profesional. La propuesta de este documento es comprender si y cómo el periodismo impreso tradicional ha reaccionado a las nuevas herramientas tecnológicas disponibles para la práctica del reportaje y la creciente disponibilidad de datos abiertos, ya sean de origen público o privado. Para esta investigación se analizarán reportajes publicados en los sitios de tres grandes periódicos brasileños – *Folha de S. Paulo*, *O Globo* y *O Estado de São Paulo* – durante la cobertura del Día Internacional de la Mujer en 2017. Los resultados indicaron bajo uso de datos, uso incipiente de visualizaciones digitales y ninguna oferta de datos brutos. El uso de datos y estadísticas sin referencias o chequeo de las fuentes fueron frecuentes.

Palabras clave: Periodismo de datos. Periodismo guiado por datos. Periodismo digital. Día Internacional de la Mujer.

1 Introdução

Era novembro de 2008 quando o cientista da computação Joe Hellerstein vaticinou, em um texto num blog sobre inovação tecnológica, que a humanidade estaria vivendo a primeira fase de sua quarta revolução industrial: a “revolução industrial dos dados”. Segundo ele, o processo era incipiente porque a maior parte da informação digital disponível então ainda era “manufaturada” em forma de textos na web e dados de formulários. O texto abordava o início da produção de dados digitais em escala massiva a partir de logs de softwares, GPS, e outros em processos automatizados, acompanhado do aumento da capacidade de armazenamento dos discos rígidos e *flashdrives* a um custo relativamente baixo. O último passo dessa revolução seria a “comoditização” dos softwares de análise de dados e automatização generalizada de processos acessíveis a uma ampla classe de usuários,

conduzindo a um volume de dados que rapidamente superaria a produtividade coletiva de autores de conteúdo em todo no mundo. Se a ideia era essa, a virada já aconteceu. Estamos na era do *big data*¹, uma quantidade imensurável de dados é gerada a cada segundo ao redor do mundo, há discos rígidos de uso doméstico com capacidade de 10 terabytes (sem falar do armazenamento em nuvem), e os dados se tornaram *commodities*.

Dois anos depois da análise de Hellerstein, em 2010, o fundador da Web, Tim Berners-Lee², afirmou que o jornalismo guiado por dados (JGD) era o futuro, e que os jornalistas precisariam, mais do que habilidades gerais com dados, saber como se aprofundar neles, analisar e encontrar o que de importante escondiam (Arthur, 2010). Desde então, aquilo que se convencionou chamar de jornalismo de dados – ou jornalismo guiado por dados (JGD), como alguns autores preferem chamar por motivos que serão discutidos mais à frente – passou a atrair o interesse pelas conexões entre notícias e tecnologia.

Esta pesquisa se baseia nos temas de que trataram Hellerstein e Berners-Lee: a intersecção entre a massa de dados que se produz e oferece a quem tenha nela interesse, e as possibilidades dos usos desses dados como combustível para a produção de notícias. Nossa tentativa é utilizar um recorte de produção jornalística como substrato para a análise do uso dos dados oferecidos na produção de narrativas noticiosas neles baseadas.

Mais do que uma definição sobre se o que se encontrou é ou não jornalismo de dados, a proposta é saber se os dados são utilizados com a finalidade de aprofundar a compreensão sobre um tema, se ela aparece como acessório ou se a digitalização da vida contemporânea está sendo ignorada no processo de produção de notícias. Para isso, serão examinados o uso de dados como elemento da apuração e sua extensão, o nível da análise envolvida, a origem dos dados, o acesso às fontes dos dados e a oferta de visualizações customizadas.

Sendo essa uma pesquisa exploratória para a nomeação do problema e apuração da metodologia de análise para uma pesquisa maior, optou-se por um *corpus* pequeno. Tomamos como referência a cobertura do Dia Internacional da Mulher, com base na observação empírica de que anualmente, da véspera ao dia seguinte da efeméride, a cobertura sobre um mesmo tema cresce, além de haver a divulgação de dados públicos e pesquisas sobre o assunto. Assim, utilizamos como objeto de análise a cobertura, factual ou não, do 8 de março, pelos sites de três dos maiores jornais do Brasil³: *Folha*

de S. Paulo, *O Globo* e *Estado de S. Paulo* (que daqui por diante será referido apenas como Estadão, uma vez que este é o nome dado ao seu portal de notícias na internet).

Sobre isso, é importante reforçar que esta pesquisa poderia ser realizada a partir dos mais diversos conjuntos de matérias jornalísticas que tratassem sobre qualquer tema. Entretanto, compreendendo que entre os potenciais benefícios do uso de dados no jornalismo está a possibilidade de tratar determinados temas enquanto agendas públicas, e não de forma episódica, julgamos que o tema das mulheres proporcionaria um importante substrato de estudo.

2 O potencial dos dados abertos

A disponibilidade de tecnologias e ferramentas para a prática de reportagens apoiadas em dados em todo o seu processo de produção – pauta, apuração, arquitetura da informação, visualização e publicação – não significa, necessariamente, que estes recursos serão utilizados. E, ainda, caso haja a opção pelo caminho dos dados, tampouco significa que eles estarão disponíveis, serão utilizados de modo apropriado e resultarão em material jornalístico de boa qualidade.

O movimento dos dados abertos extrapola o campo das mídias. Na verdade, este se adequou à transformação mundial que nasceu no mercado, como dispositivo de controle e *accountability*, e alcançou governos de diferentes níveis, consolidando o paradigma da Era da Informação (Castells, 1999). O processo se apoia num tripé: abertura, participação e colaboração, (White House, 2009, parágrafo 4) considerando que, através da transparência, compartilhamento e trabalho coletivo, o potencial destes conjuntos de dados poderia se realizar. Promover o acesso a tais dados de forma compreensível enquanto informação e conhecimento seria uma forma de democratizá-los. Vale frisar que o foco do movimento está na publicização de dados públicos, ou gerados com financiamento público, uma vez que o uso de recursos públicos deveria beneficiar a sociedade como um todo. No caso dos dados de empresas privadas, é preciso considerar a propriedade de seus criadores (Kitchin, 2014).

No Brasil, o arcabouço legal que permitiu a abertura de dados públicos inclui as leis complementares nº 101 (2000) – normas de finanças públicas voltadas para a responsabilidade na gestão fiscal – e nº 131 (2009) – acrescenta dispositivos à LC 101,

com foco na oferta de informações de execução fiscal em tempo real, a lei federal nº 12.527 (2011) – regulamenta o direito constitucional de acesso dos cidadãos às informações públicas – e pela instrução normativa nº 4 (2012), da Secretaria de Logística e Tecnologia da Informação (SLTI) do Ministério do Planejamento, que instituiu a Infraestrutura Nacional de Dados Abertos (INDA). Esse aparato colabora para a intensificação do ciclo de produção, divulgação e consumo dos dados. Num fluxo contínuo, a população, por sua própria presença na sociedade e uso dos serviços públicos, gera dados massivos que são coletados, armazenados e estruturados pelos governos, que em tese, divulga esse material. No cenário aqui analisado, jornalistas usufruem desta abertura de dados para oferecer à população compreensões e interpretações dessas massas de dados, transformando-os de dados brutos em informação concreta e acessível a grupos não especializados.

A lei federal no 12.527, também conhecida como Lei de Acesso à informação (LAI), moveu a transparência ativa prevista na Constituição Federal de 1988 para o campo prático, obrigando órgãos públicos a publicarem dados e informações sobre a instituição e serviços por ela prestados. Exigindo que estes dados fossem divulgados em formatos digitais abertos e não proprietários, foi instaurada a primeira etapa da abertura de dados no Brasil. O marco dessa mudança foi a entrada para a Parceria para o Governo Aberto (Open Government Partnership) como um dos países fundadores em 2011, acompanhado de sete países, que hoje se somam a outros 59, formalmente comprometidos a difundir e incentivar a transparência dos governos, o acesso à informação pública e a participação social. Assim, o Brasil alcançou o cenário de abertura governamental predominante ao redor do mundo (Prado, Ribeiro & Diniz, 2012, p. 16).

A oferta dessas bases de dados complexas e abrangentes abre portas para grandes empreitadas jornalísticas, descoberta de escândalos de corrupção, entendimento de surtos de doenças endêmicas, etc. Mas, é importante pensar que o manejo das bases de dados guarda também grande potencial para a cobertura rotineira da mídia jornalística, como transporte público municipal, saneamento básico e compras governamentais. É possível, portanto, que o maior potencial do uso de dados no jornalismo não esteja nas equipes especializadas, mas no seu uso no cotidiano das redações, como forma de aprofundar e contextualizar questões de interesse público. Mas, como já dito acima: isso tem acontecido?

3 A evolução do uso dos dados no jornalismo

A discussão deste trabalho exige a problematização do próprio conceito em debate, a saber, o jornalismo de dados. Se considerarmos o mero uso de dados na produção de material noticioso como característica central desta linha, a discussão se esvaziará, visto que praticamente qualquer prática jornalística poderia ser assim classificada. O uso de dados numéricos no jornalismo se consolidou a partir do final do século XIX como consequência de uma transformação mais ampla, que foi a “substituição de paradigmas religiosos por explicações científicas [...] [que] atingiram em status de ‘reveladoras de realidade’” (Sponholz, 2009, p. 58). Em todo o campo das ciências sociais, coletar, classificar e analisar dados se tornou um caminho quase indiscutível para alcançar a realidade, o fato.

O uso de dados como forma de aprimorar os procedimentos jornalísticos visando à operacionalização da objetividade não é novo. Remete ao século XX, quando foram formulados os conceitos do jornalismo de precisão por Philip Meyer (1973) e, mais tarde, de reportagem assistida por computador – RAC (Cox, 2000; Gray, Bounegru & Chambers, 2012; Coddington, 2014). É ao longo dos anos 1980 que ambos são disseminados ao redor do mundo, chegando ao Brasil na década seguinte. Esses conjuntos de práticas foram criados para permitir a coleta, o processamento e a extração de significado dos conjuntos de dados disponíveis. Era a transição de uma realidade de poucas bases de dados privadas e, em geral, inacessíveis, para, a partir de um movimento de busca global por transparência e abertura, uma infinidade de bases de dados abertas e acessíveis, principalmente através da web a partir dos anos 2000. A abundância levantou fortes expectativas, especialmente nas empresas jornalísticas, de que estear seu processo produtivo em dados era uma nova forma de afastar a inexistência da narrativa jornalística (Alexandre, 2014, p. 48; Charbonneaux & Gkouskou-Giannakou, 2015, p. 268). O entusiasmo em torno dos dados levou empresas jornalísticas tradicionais à busca por profissionais com a literacia para o trabalho com dados, e ao investimento na formação das equipes existentes (Bertocchi, 2013, Grandin, 2014; Renó & Renó, 2016). Paralelamente, e como parte do processo de transformação do mercado, surgiram também novos produtos que dedicam um investimento especial às reportagens com dados, como a ProPublica nos Estados Unidos, e o Nexso Jornal e a Agência Pública, no Brasil.

De acordo com Megan Knight (2015, p.55), a expressão “*data journalism*” apareceu pela primeira vez num texto de Simon Rogers, no jornal inglês The Guardian. No blog “Guardian Insider”, o jornalista relatou a adoção de um aplicativo para gerar visualização de dados brutos em mapas dinâmicos. Para explicar que jornalistas e desenvolvedores estavam trabalhando juntos para oferecer uma nova leitura dos dados, ele disse: “Isso é jornalismo de dados” (Rogers, 2008, parágrafo 7). Em um vídeo publicado em 2013 (Rogers, 2013), o mesmo autor afirma que em 1821 o jornal vazou uma planilha de custos das escolas locais e reivindica para o Guardian a invenção do jornalismo de dados. Não era a primeira vez que redações manejavam dados, mas a forma, com certeza, era nova.

Uma revisão da literatura mostra que o primeiro registro do uso de dados no jornalismo é de 1952, quando a emissora de TV CBS usou programas de computador para tentar prever os resultados das eleições presidenciais norte-americanas (Gray, Bounegru & Chambers, 2012, p. 9). Contudo, o primeiro registro sistematizado reunindo jornalismo e uso de dados veio mais à frente, com Philip Meyer, que em 1967 decidiu utilizar métodos das ciências sociais com a pretensão de melhorar a qualidade da sua apuração jornalística, trabalho que resultou na publicação de *Precision Journalism*, em 1973. Mirando a objetividade sob um olhar bastante positivista, o autor entendia que o método científico seria capaz de produzir um jornalismo melhor do que aquele baseado em *press releases* e declarações de fontes. Anos depois, Meyer colocou o próprio conceito em perspectiva, pesquisando as novas técnicas de análise e processamento de dados apresentadas em *The New Precision Journalism* (1991). Numa declaração a Bounegru em 2012, mais de 40 anos após propor o jornalismo de precisão, Meyer parece cético: “Quando a informação era escarça, a maior parte de nossos esforços se concentrava em caçá-la e reuni-la. Agora que a informação é abundante, processá-la é mais importante” (Gray, Bounegru & Chambers, 2012, p. 6).

Na década de 1990, a pesquisa sobre o uso e processamento de dados quantitativos no jornalismo foi reunida sob o termo reportagem assistida por computador (RAC). Há indicadores de que o trabalho de Meyer tenha sido, dentro e fora da academia, um dos maiores impulsionadores da consolidação do termo e da prática da RAC, definida por Cox (2000, p.

3) como um conceito amplo que abarca diferentes usos dos computadores no processo de produção das notícias, podendo envolver pesquisa online e bases de dados. A literatura acerca do seu uso e aperfeiçoamento nas redações associa a prática à popularização dos microcomputadores e arquivos digitais, bem como da internet (Machado, 2005, p. 301; Flew, Spurgeon, Daniel & Swift, 2012, p. 157; Crucianelli, 2013; Coddington, 2014, p. 2; Cushion, Lewis & Callaghan, 2016, p. 3).

O uso do computador e das ferramentas web, bem como a oferta de dados abertos, cresceram de forma imensurável na segunda metade do século passado e, exponencialmente, a partir dos anos 2000 (Kitchin, 2014, p. 79; Zuiderwijk, Janssen & Dwivedi, 2015, p. 341).

No campo do jornalismo, o destaque para esse salto ocorreu em 2010, quando o mundo voltou os olhos para a liberação de milhares de registros da Guerra do Afeganistão e da Guerra do Iraque pelo site Wikileaks⁵. Através do trabalho do fundador do site, o jornalista Julien Assange, e da equipe do *The Guardian*, deu-se acesso a milhares de dados riquíssimos, mas desanimadoramente complexos. A solução encontrada foi o desenvolvimento de um navegador personalizado para buscar palavras-chave, permitindo o acesso concreto ao conhecimento contido na *big data* (Rogers, 2014; Hewett, 2015). Tendo a “revolução industrial dos dados” prevista por Hellerstein em 2008 – na qual o volume de dados gerado por processos automatizados e inteligentes superaria sem chance de retorno sua produção “manufaturada” – se tornado uma realidade, o jornalismo se apropriou dela rapidamente. Assim, definir se os dados se tornaram uma nova categoria de fontes ou um novo conjunto de ferramentas para a prática jornalística parece necessário. Para o jornalismo de dados, uma coisa não se separa da outra, e o papel dos profissionais da área é criar conexões entre dados a princípio incompreensíveis (Bradshaw & Rohumaa, 2011; Bradshaw, 2014). Os cidadãos já não dependem da imprensa para ter acesso aos dados, o que nem sempre acontece com o significado que carregam. Neste cenário, o instituto do “furo jornalístico” como símbolo de excelência de um campo profissional cai por terra. O debate não é necessariamente sobre quem dá a notícia primeiro, mas cada vez mais sobre quem pode fazer as conexões de que os demais não foram capazes, quem enxerga o que os outros não viram através de suas competências jornalísticas e do domínio da técnica.

Para continuar a discussão com o aprofundamento adequado, decidimos analisar brevemente os termos e conceitos usados para designar o conjunto de práticas que, até aqui, tratamos genericamente como jornalismo de dados. Embora alguns jornalistas e professores da área defendam que o jornalismo de dados (JD) é apenas um novo nome para a reportagem assistida por computador (RAC), entendemos que há diferenças em seus objetivos e no produto final. Esta última pode ser definida como uma evolução do jornalismo de precisão (Meyer, 1973), bastante apoiada em metodologias das ciências sociais (*survey*, análise de conteúdo, análise estatística, etc.) para a construção da reportagem investigativa (Coddington, 2014, p. 4). No JD o produto noticioso parte da própria base de dados, dotada de informação em abundância, para a construção de uma narrativa de cunho analítico. Nas palavras de Stray (2011, p. 19), seria o ato de obter, selecionar e reportar dados de interesse público. O foco, mais do que aprimorar a coleta de informações, está na abertura de dados ao público através do seu processamento e visualização, cruzando conhecimento de diferentes campos, especialmente jornalismo, programação e design.

Na tentativa de esvaziar a ideia de que todo jornalismo poderia ser considerado “de dados”, Parasie & Dagiral (2013, p. 854) lançam mão do termo “jornalismo guiado por dados” (JGD) como uma prática mais atualizada, capaz de melhorar a contribuição democrática do jornalismo de três formas: fortalecendo a objetividade jornalística, facilitando a manutenção da *accountability* dos governos pelas empresas jornalísticas e aumentando a participação política dos cidadãos por meio de sua produção e análise de dados. O termo é referencial para Träsel, que em sua tese de doutorado definiu o JGD como “a aplicação da computação e dos saberes das ciências sociais na interpretação de dados, com o objetivo de ampliar a função da imprensa como defensora do interesse público” (Träsel, 2014, p. 90).

Seguindo pelo mesmo caminho está o conceito de jornalismo computacional proposto por Hamilton & Turner (2009, p. 2), que seria a combinação de algoritmos, dados e conhecimento de ciências sociais usados para suplementar a função de *accountability* do jornalismo, unindo a RAC e o JD. Numa perspectiva que nos parece mais adequada, Diakopoulos (2011, parágrafo 2) o descreve como uma vertente tecnologicamente orientada do jornalismo, visando

à aplicação do pensamento computacional às práticas de coleta, produção de sentido e apresentação de informações, usando, por exemplo, processos de automação.

É possível perceber a ausência de uma delimitação clara entre os conceitos apresentados acima. Ao contrário, sob o olhar de alguns autores eles se sobrepõem em vários aspectos (Coddington, 2014; Lima Júnior, 2012; Mancini & Vasconcellos, 2016). A fim de propor alguma harmonia entre eles, indicamos que estas práticas do jornalismo de dados se enquadram nas extensões do paradigma jornalismo digital em base de dados (JDBD), modelo teórico formulado por Barbosa & Torres (2013) para o melhor entendimento sobre o papel das bases de dados no jornalismo contemporâneo. Este modelo, segundo afirmam os autores, expande-se em sucessivas apropriações, tendo em conta a consolidação das bases de dados no contexto atual como estruturantes da atividade jornalística e como agentes singulares no processo de convergência de meios de comunicação, verificando-se assim a sua preponderância (p. 153).

Desta maneira, entendemos o jornalismo guiado por dados como uma das vertentes, um dos aspectos compreendidos pelo Paradigma JDBD (Barbosa & Torres, 2013), uma vez que está no escopo de abrangência do seu conceito, sintetizado como sendo:

[...] o modelo que tem as bases de dados como definidoras da estrutura e da organização, bem como da composição e da apresentação dos conteúdos de natureza jornalística, de acordo com funcionalidades e categorias específicas, que também vão permitir a criação, a manutenção, a atualização, a disponibilização, a publicação e a circulação de cibermeios dinâmicos em multiplataformas. (Barbosa & Torres, 2013, p. 154).

A metodologia de análise utilizada foi desenvolvida a partir de alguns dos trabalhos já citados, como será descrito adiante.

4 Metodologia

O crescimento da oferta de informação através da internet tem sido abordado por diferentes autores a partir da expectativa de transformação social e política dos Estados democráticos (Marques, 2008, p. 50; Pinho, 2008, p. 476; Rothberg, 2008, p. 151; Silva, 2009, p. 41; Bragatto, 2011, p. 133), estando esta abordagem

repetidamente ligada à preservação dos direitos civis, políticos e sociais, a partir de uma mediação do jornalismo. Os meios digitais e, especificamente, a oferta e o uso de bases de dados cada vez mais robustas e intercomunicáveis (massas de dados/*big data* e dados abertos/*open data*), conforme previsto por Berners-Lee (2010), alargaram o campo de trabalho jornalístico, fornecendo o auxílio de um grande volume de dados para sustentar a discussão dos assuntos de interesse público, aí inclusos os campos da saúde, educação, segurança pública, desenvolvimento econômico, diversidade étnica, religiosa e sexual, etc.

A ideia deste trabalho é entender se e de que modo esse aparato de recursos informativos está sendo usado pelo jornalismo. Para esta análise, escolhemos as pautas ligadas a mulher, considerando o recorrente aumento das publicações sobre o tema nas proximidades do 8 de março, Dia Internacional da Mulher. O *corpus* da pesquisa inclui as matérias sobre mulheres publicadas na véspera, na data e no dia seguinte (7, 8 e 9 de março de 2017) por três dos maiores sites jornalísticos de referência do Brasil: *Folha de S. Paulo*, *O Globo* e *Estadão*. Os dados são do Instituto Verificador de Comunicação (IVC)⁶, que investiga a circulação total de cada um, incluindo exemplares impressos e acessos digitais. Vale ressaltar que as três publicações escolhidas figuravam, na ocasião da coleta de dados (10 mar. 2017), no primeiro, terceiro e quarto lugar no ranking do IVC, respectivamente. O segundo lugar era ocupado pelo mineiro *Super Notícia*, diário gratuito do grupo O Tempo, que não possui mecanismo de busca em seu site e por isso não pôde ser incluído nesta pesquisa.

Ao todo foram levantadas 129 matérias para observação, registro e discussão de resultados: 38 do *Estadão*, 41 da *Folha de S. Paulo* e 49 de *O Globo*. Para refinar os dados, foram consideradas apenas as matérias produzidas pelas redações dos veículos, seguindo sua linha editorial e cultura organizacional, desconsiderando material de agências de notícias e de blogs hospedados nos domínios investigados.

A análise reúne variáveis testadas em três diferentes pesquisas, além de quatro critérios propostos pela primeira vez nesta pesquisa, todas reunidas em uma planilha eletrônica. Foram realizadas navegações nos arquivos digitais de cada uma das publicações, selecionando para observação todas as matérias publicadas no período designado que contivessem a palavra mulher

no título ou no corpo do texto. Foi feito o download de todo o conteúdo em arquivos individuais em formato PDF, para garantir que o material pudesse ser usado ao longo de toda a investigação, independentemente de mudanças nos portais de notícias visitados. Após a coleta, cada matéria foi detalhadamente examinada para o preenchimento do instrumento.

O primeiro conjunto de variáveis é uma adaptação da tipologia proposta por Megan Knight (2015, p. 59) ao analisar matérias com dados publicadas em jornais britânicos, a fim de observar suportes e formatos do jornalismo de dados enquanto práxis da mídia tradicional (N=106). A autora classificou as matérias coletadas nas categorias grupo de mídia (aqui substituída por veículo), editoria, categoria de uso dos dados e origem dos dados.

A essas categorias foram somadas as de Cushion, Lewis & Callaghan (2016, p. 7), que analisaram uma extensa base de dados de uma pesquisa encomendada pela BBC Trust sobre a relação entre jornalismo e dados na cobertura jornalística do Reino Unido, incluindo aí programas de rádio e televisão, além de mídia online (N=4.285). Daí foram selecionadas as seguintes variáveis: uso de dados, fonte dos dados e problematização dos dados.

A fim de aprimorar a análise foram adotadas mais três categorias, desta vez destacadas do Paradigma JDBD (Barbosa & Torres, 2013, p. 155): inter-relacionamento/hiperlinkagem, visualização e convergência. Considerando aspectos específicos da prática jornalística, achamos por bem incluir três categorias ausentes de toda a bibliografia revisitada durante o desenho desta pesquisa: diversidade de fonte (uma; duas; três ou mais fontes de dados), acesso às fontes de dados numéricos (não cita a fonte dos dados; cita a fontes dos dados; cita a fonte dos dados e oferece link de acesso) e oferta dos dados brutos utilizados. A seguir, será feita a descrição de cada variável analisada, bem como apresentados os operadores de classificação:

Quadro 1 – Categorias de operadores de análise

Categoria	Operadores
Veículo – empresa responsável pela notícia	<ol style="list-style-type: none"> 1. <i>Folha de S. Paulo</i> 2. <i>O Globo</i> 3. <i>Estadão</i>
Editoria – enquadramento temático da notícia	<ol style="list-style-type: none"> 1. Brasil 2. Cidade 3. Comportamento 4. Cultura 5. Economia 6. Esporte 7. Infográficos 8. Internacional 9. Mulher 10. Opinião 11. Política 12. Saúde 13. Sociedade
Uso de dados – indicar a presença de qualquer dado numérico e seu grau de aprofundamento	<ol style="list-style-type: none"> 1. Não há 2. Referência vaga, sem suporte empírico 3. Referência curta, mas clara 4. Referência detalhada, com contexto
Tipo do uso dos dados – quando verificado, classificar o uso de dados	<ol style="list-style-type: none"> 1. Mapa estático 2. Mapa dinâmico 3. Gráfico ou quadro 4. Infográfico 5. Lista 6. Olho/destaque numérico 7. Tabela 8. Análise textual dos dados 9. Linha do tempo 10. Citação dos dados
Diversidade de fontes – quantas fontes de dados foram utilizadas	<ol style="list-style-type: none"> 1. Uma fonte 2. Duas fontes 3. Três fontes 4. Quatro fontes;
Acesso às fontes – checar a possibilidade de acesso à fonte original dos dados	<ol style="list-style-type: none"> 1. Não cita a fonte dos dados 2. Cita a fontes dos dados 3. Cita a fonte dos dados e oferece link de acesso
Oferta de dados brutos – checar a oferta dos dados brutos utilizados na análise	<ol style="list-style-type: none"> 1. Sim 2. Não

Fonte dos dados – informar a partir de quem os dados foram inseridos na matéria	<ol style="list-style-type: none"> 1. Jornalista 2. Entrevistado
Origem dos dados – por quem foram produzidos os dados citados	<ol style="list-style-type: none"> 1. Governo 2. Instituições públicas ou privadas 3. Fonte desconhecida
Problematização dos dados – verificar se o jornalista contrapõe os dados informados com informações contrárias, questionando um ponto de vista	<ol style="list-style-type: none"> 1. Sim 2. Não
Inter-relacionamento/ hiperlinkagem – informar se há incorporação de páginas web para oferecer contexto e favorecer a análise crítica	<ol style="list-style-type: none"> 1. Não há links 2. Há links internos (para a base de dados do próprio site) 3. Há links externos (para outras bases de dados)
Visualização – indicar a presença de modo de visualização exclusivamente digital	<ol style="list-style-type: none"> 1. Presente 2. Ausente
Convergência – considerando a produção convergente de conteúdo nas redações, registrar se e quantos formatos ou suportes foram utilizados na matéria	<ol style="list-style-type: none"> 1. Um suporte 2. Dois suportes 3. Três suportes 4. Não ocorreu

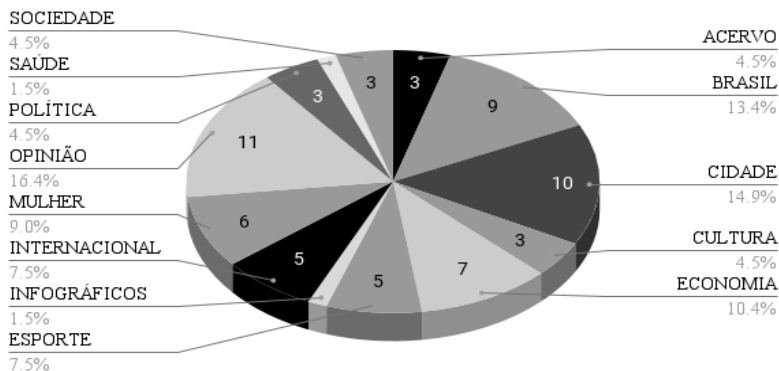
Fonte: Knight (2015), Cushion, Lewis & Callaghan (2016), Barbosa (2007) e a autora.

A etapa seguinte será a apresentação dos resultados coletados e discussão dos dados.

5 Resultados e discussão

Foram encontradas 129 matérias, tratando de temas ligados à mulher entre os dias 7 e 9 de março de 2017, nos sites dos jornais Estadão (38), O Globo (49) e Folha de S. Paulo (42), então submetidas à análise de acordo com os critérios apresentados na seção anterior.

Gráfico 1 – Matérias que fazem alguma referência a dados, por editoria



Fonte: elaboração própria.

Conforme o Gráfico 1, descartadas as matérias que não utilizam dados (62 matérias ou 48,1% do total da amostra), a editoria de Opinião é a que mais usa o recurso, seguida por Cidade e Brasil. Há algumas matérias sobre saúde e violência contra a mulher em diferentes editorias que não apresentam qualquer dado ou estatística. Ao contrário do que poderia dizer o senso comum, a editoria de economia só aparece em quarto lugar. A editoria Folha Mulher, que de acordo com o site do jornal foi criada para ampliar a cobertura sobre as questões de gênero, teve apenas quatro das 42 matérias temáticas do jornal no período.

A segunda categoria de análise tratou sobre o uso de dados numéricos e seu grau de aprofundamento. Conforme o gráfico acima, quase a metade das matérias continha uma informação curta, porém clara, em relação aos dados, como abaixo:

“Menos de 10% do Congresso são formados por mulheres, número muito menor do que em muitos países da América Latina, infinitamente menor do que em países desenvolvidos”. (Costa, 2017, on-line).

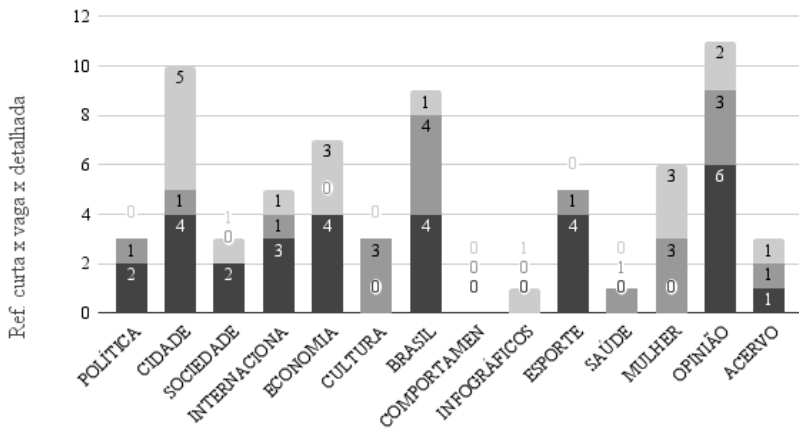
Por outro lado, a menor parte da amostra continha dados explicados detalhadamente, com informações contextuais que favorecessem a compreensão do leitor, como neste caso:

“E, a cada 12 minutos em 2016, uma mulher sofreu agressão física no Rio. Os números, divulgados ontem, Dia Internacional da Mulher, pela Polícia Militar e pela Secretaria estadual de Segurança, respectivamente, dão um quadro da gravidade da violência de gênero no estado”. (Galdo, 2017, on-line).

O achado mais interessante é intermediário entre estas categorias. Considerando técnica jornalística, é estranho pensar que 28,4% das matérias utilizavam referências vagas e sem qualquer referência à fonte ou base de dados para referendar a citação, como aqui:

“A despeito de todas as mulheres estarem potencialmente submetidas à proibição penal do aborto, a criminalização afeta desproporcionalmente mulheres negras e indígenas, pobres, de baixa escolaridade e que vivem distante de centros urbanos”. (Cardoso, 2017, on-line).

Gráfico 2 – Aprofundamento da referência, por editoria



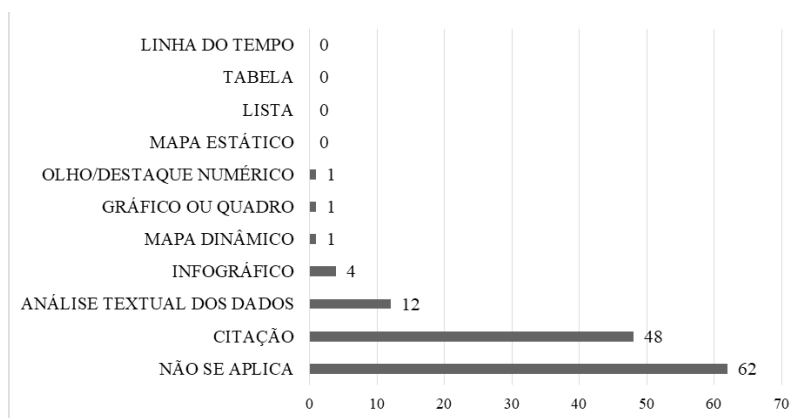
Fonte: elaboração própria.

Quando essa variação de aprofundamento é dividida por editoria, nota-se que a maior parte dos dados sem suporte é utilizado em Brasil e Opinião. Já a maioria das matérias com referência detalhada e contextualizada está na editoria de Cidade, seguida por Economia e Mulher, que dividem a segunda posição. Embora a metodologia utilizada não permita essa avaliação, entendemos ser premente a busca de explicação para este achado, incluindo a disponibilidade de dados sobre certos temas e a rotina produtiva das equipes envolvidas.

Em relação aos tipos de usos de dados, a Gráfico 3 indica que, excluídas as matérias que não os utilizam, mais de 70% das ocorrências são citações simples. Em seguida aparece a análise textual, que foi encontrada em 12 casos. Os recursos visuais foram

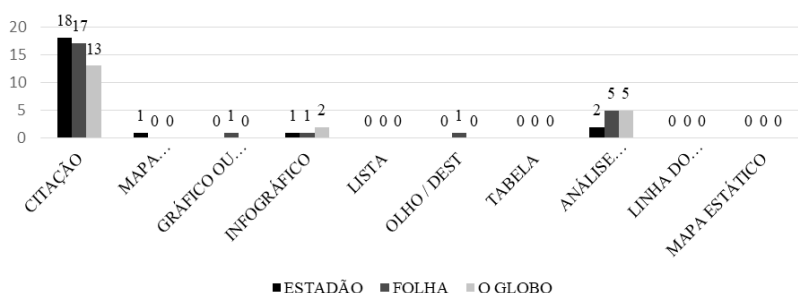
pouco utilizados, com ausência de mapas e linhas do tempo. Mesmo considerando os diferentes veículos, a citação simples segue predominante. É interessante destacar que o Estadão, apesar de ter sido o jornal com mais matérias sobre o tema, foi o que menos utilizou dados.

Gráfico 3 – Tipos de usos de dados



Fonte: elaboração própria.

Gráfico 4 – Tipo de uso dos dados por veículo



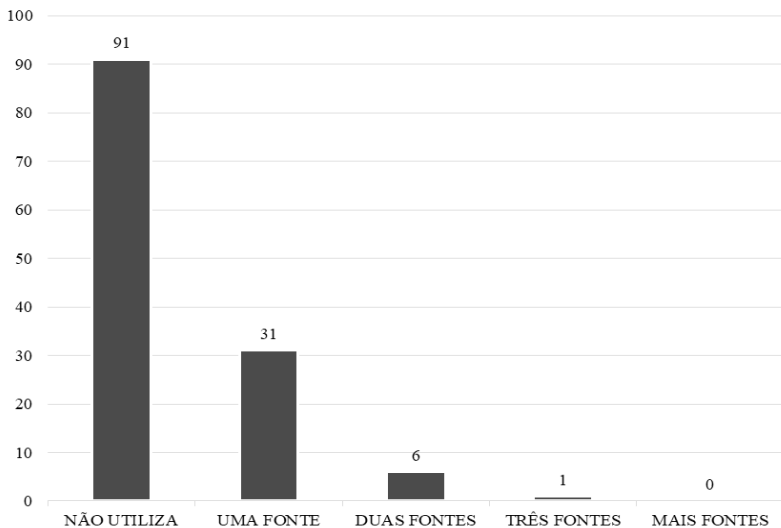
Fonte: elaboração própria.

Considerando que o jornalismo digital, comparado ao suporte impresso, seria menos atingido pela limitação de espaço, bem como o caráter informativo do material aqui analisado, resolvemos analisar a diversidade das fontes de dados utilizadas na construção do texto.

Vale ressaltar que diversidade de fontes aqui se refere a fontes de dados numéricos, não à quantidade de pessoas entrevistadas no processo de reportagem. De acordo com os resultados levantados, a diversidade foi bastante limitada, conforme o Gráfico 5.

Entre as matérias que usam dados, mais de 80% trazem apenas uma fonte, 15% duas fontes e somente uma matéria confrontou três fontes de dados. Um resultado atípico revela que em 24 das 67 matérias da amostra usam referências numéricas (35,8%), os dados foram usados sem a indicação da fonte, contrariando os princípios básicos de apuração e verificação no jornalismo. No total da amostra, apenas uma reportagem da *Folha* sobre violência contra a mulher fornecia um link para o site que era a fonte da informação. Também não foi encontrada nenhuma ocorrência de oferta de dados brutos utilizados na produção da reportagem, como indicam manuais de boas práticas do jornalismo de dados, a exemplo do *The Data Journalism Handbook* (2012).

Gráfico 5 – Diversidade de fontes



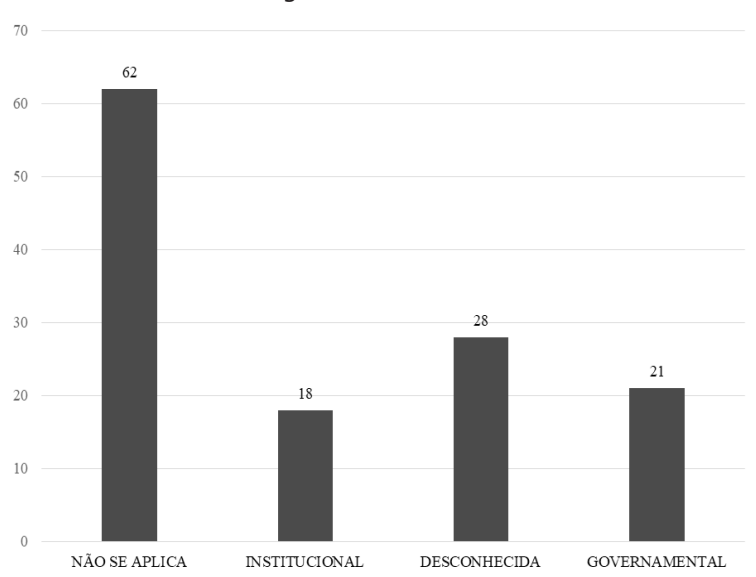
Fonte: elaboração própria.

Em seguida foi analisada a fonte dos dados nas matérias, a fim de registrar se fora o jornalista ou algum entrevistado a lançar mão de dados para conferir validade ao próprio discurso. O número de jornalistas que invocaram dados foi quase duas vezes maior do que o

número de entrevistados que o fizeram. Como já dito, não é possível definir a causa deste resultado a partir da metodologia utilizada, mas sugerimos dois caminhos dedutivos: primeiro, parece que os jornalistas sentem maior necessidade de legitimação dos discursos a partir dos números, visto que a escolha das fontes como tal já pressupõe alguma autorização; em segundo lugar, podemos considerar que os jornalistas acham suficientes as declarações de seus entrevistados, não cobrando deles dados que sustentem suas falas.

Ainda tratando sobre as características dos dados, foi levantada a origem das informações citadas pelos autores dos textos e por suas fontes, que poderiam variar entre fontes governamentais, fontes institucionais (empresas, sindicatos, associações, organizações sem fins lucrativos, etc.) ou fontes desconhecidas (quando um dado era citado sem indicação de quem o produzira ou divulgara), como pode ser visto no Gráfico 6 a seguir. A maior parte dos dados citados em textos publicados pelos três jornais é proveniente de fonte desconhecida, ou seja, predominam afirmações que podem não ter sustentação na realidade, reiterando a pouca atenção dada à checagem durante o processo de apuração. É preciso considerar três ângulos desta ocorrência. No primeiro, os jornalistas citam dados sem revelar suas fontes, o que pode nos fazer perguntar até mesmo se elas existem.

Gráfico 6 – Origem dos dados



Fonte: elaboração própria.

No segundo, os entrevistados citam dados sem referência alguma e não são questionados pelos repórteres, que inclusive publicam tal informação. Isso aconteceu em 65% das vezes em que as fontes utilizaram dados em suas declarações. E por fim, produtos com essa falha de checagem e verificação passaram incólumes por todo o processo produtivo da notícia, passando pelo processo de edição e fechamento, e isso aconteceu muitas vezes.

Acerca da categoria problematização de dados, verificou-se que em nenhuma das matérias analisadas houve presença de dados colocados pelo jornalista e pela fonte simultaneamente. Em outras palavras, em nenhuma das 17 vezes em que as fontes lançaram mão de dados em seus discursos, independentemente da origem destes dados, os jornalistas complementaram seus textos com outros dados que confirmassem ou refutassem os pontos de vista apresentados, cumprindo seu papel de questionar posições e oferecer contexto e variedade de perspectivas sobre um tema ou acontecimento.

A categoria de inter-relacionamento/hiperlinkagem foi adotada do modelo de Barbosa (2008) não apenas para verificar o uso da ferramenta disponível, mas considerando o recurso uma possibilidade de oferta de contextualização. Assim, ainda que houvesse limitadores para a produção de texto e material gráfico de contextualização, seria possível conduzir o leitor a outros ambientes de abordagem daquela pauta, o que, de acordo com nossos resultados, não é uma abordagem recorrente. Os links estavam ausentes em 70% das matérias verificadas, 20% contavam com links para outras matérias no próprio veículo, e apenas sete matérias (5%) levavam o leitor a outros sites com mais informações de interesse. Ainda que os três jornais possuam um vasto arquivo digitalizado que poderia ser conectado e ampliar a informação para o leitor através de mecanismos extremamente simples, isso não tem sido uma prática nos sites destes jornais.

Assim como dissemos que o jornalismo online supera, de alguma forma, as limitações espaciais, podemos sugerir que ele oferece mais ferramentas de visualização de dados, em alguns casos exclusivos do meio digital. Inclusive não apenas os programas de computadores que ampliam o leque de recurso dos designers, mas também uma série de ferramentas gratuitas de geração de visualização e de fácil manuseio. Ainda assim, na amostra analisada a importância dada à visualização foi pequena, identificada em seis das 129 matérias coletadas, sendo quatro infográficos, um quadro e uma visualização interativa (mapa dinâmico numa matéria do *Estadão*).

Mesmo se consideradas apenas as 67 matérias que utilizaram dados, a opção pela geração de visualizações foi feita em menos de 10% dos casos. Numa outra correlação, é possível pensar que as visualizações só foram apresentadas em um terço das 18 matérias que ofereciam referência detalhada e com contexto.

Figura 1 – Exemplo de matéria com infografia



Fonte: Estadão (2017). Recuperado de infograficos.estadao.com.br/cidades/o-mundo-segundo-as-mulheres/.

A última categoria apresentada foi a da convergência, que buscou analisar o uso multimídia na apresentação de conteúdos. Nesta categoria, foram encontradas apenas cinco ocorrências do uso de duas mídias na apresentação de uma matéria, incluindo texto e vídeo. Não foram inseridos na análise de matérias onde estava disponível texto e foto ou texto e visualizações estáticas, uma vez que estas alternativas também poderiam ser utilizadas em jornais impressos, e nem por isso seriam consideradas conteúdo multimídia.

6 Considerações finais

Este estudo examina e discute se algumas alegações feitas pelos defensores do jornalismo de dados são verdadeiras, buscando compreender em que nível elas têm sido alvo de apropriação nas redações da mídia tradicional. Este processo é endereçado especialmente aos casos em que a disponibilidade de dados abertos pode contribuir para a cobertura temática (e não episódica) de uma

agenda, neste caso, as mulheres. Ao fim e ao cabo, não há sinais de que o uso dos dados tenha de fato invadido as redações da mídia tradicional brasileira, ou de que os jornalistas estejam de fato lançando mão da massa de dados disponível para qualificar a sua produção jornalística. Entretanto, é importante dizer que todos os veículos analisados utilizaram, em algum momento, o conjunto de práticas que chamamos de jornalismo guiado por dados.

Em vista da existência de uma equipe especializada na área, e apesar de ter produzido menos matérias no período e sobre o tema analisado, o *Estadão* foi, proporcionalmente, o que mais utilizou dados. Além disso, o portal foi o único a propor uma visualização digital mais avançada, um mapa dinâmico. Os outros dois jornais, de um modo geral, utilizaram os dados de modo menos frequente. Em termos de visualizações, o *Estadão* produziu três, enquanto *O Globo* e *Folha* apresentaram, cada um, duas visualizações, apesar de terem publicado uma quantidade significativamente maior de matérias sobre mulheres no período estudado.

Considerando a literatura levantada para o estudo, foi confirmada a expectativa de que a maior parte dos dados utilizados seria proveniente de fontes governamentais, favorecendo assim o discurso de atores sociais que já ocupam posições de poder e têm acesso privilegiado aos canais de produção de notícias. Um achado inesperado, entretanto, foi que o uso de dados de fontes desconhecidas foi maior do que aqueles provenientes de fontes governamentais e institucionais somados, e que essa prática é frequentemente atribuída aos jornalistas mais do que a seus entrevistados. Diferentemente do que tem se verificado em países como a Inglaterra e os Estados Unidos, no Brasil, ainda são minoria as matérias apoiadas em massas de dados divulgados pelos governos ou outras instituições, o que caracterizaria objetivamente a prática de um jornalismo guiado por dados. Esse fato é mais marcante no que se refere à comunicação de temas ligados a saúde e segurança pública, por exemplo. A respeito deles, o jornalismo praticado em países mais bem posicionados no ranking do índice de desenvolvimento humano (IDH) tem se baseado com frequência nesse tipo de fonte, que inclui resultados de pesquisas e conjuntos de dados abertos governamentais.

A ideia defendida pelos entusiastas do jornalismo de dados, de que revelações advindas de pedidos via Lei de Acesso à Informação e vazamentos de informações estatais se tornariam importantes fontes para o trabalho jornalístico, também pareceu não se confirmar,

ao menos no período e tema selecionados para esta pesquisa. Apesar de tantos acontecimentos identificados nos últimos três anos que têm configurado nos meios digitais o fenômeno popularmente conhecido como Primavera das Mulheres, tudo indica que as pautas ligadas a elas continuam sendo cobertas a partir de ocorrências e não enquanto um tema relevante da agenda pública.

Também é importante pensar que, mesmo quando os jornalistas estão interessados nos dados, sua habilidade com eles, cuja necessidade premente foi sugerida por Berners-Lee em 2010, ainda está no campo da idealização. Em outras palavras, mesmo quando utilizam dados, a maioria dos jornalistas ainda o fazem de forma superficial, sem o necessário aprofundamento e contextualização da informação que apresentam.

Entendemos que a metodologia aqui proposta, modificada e testada várias vezes antes da aplicação final para a coleta de dados, se revelou adequada e eficaz para os fins esperados. A investigação possibilitou o debate em torno do alargamento do campo jornalístico possibilitado pelo uso de grandes volumes de dados como alicerces da discussão dos assuntos de interesse público.

Conforme citado anteriormente, optamos pela reunião de variáveis de análise utilizadas em três estudos anteriores, acrescidos, mais tarde, de novos quatro critérios considerados necessários para uma análise detalhada do conteúdo selecionado. Ressaltamos aqui o caráter inovador desta proposta, visto que a combinação dos métodos permite uma nova matriz de investigação que preenche lacunas existentes em cada um deles. Assim, o material coletado foi analisado no que diz respeito a suportes, formatos, origem e fonte dos dados, recursos tecnológicos e de design, além de um olhar mais específico sobre a prática jornalística e seus resultados discursivos, quando tratamos de problematização dos dados apresentados, diversidade de fontes ouvidas e oferta dos dados brutos.

Assim, atendendo às demandas de uma análise crítica do discurso da mídia jornalística tradicional, o instrumento apresentado oferece não apenas chaves para a compreensão do que é dito, mas também, e em alguns momentos mais importante, do não dito. Por exemplo: quando um jornalista ou uma autoridade governamental cita um dado sobre a violência contra a mulher que não tem qualquer referência objetiva – como verificado repetidamente durante a pesquisa – abrem-se portas para a interpretação desse silêncio sobre a fonte da informação que talvez possa nem existir. Por que tal agente

social opta por um discurso vazio de sentido para tratar de um tema tão importante e no qual sua contribuição foi considerada relevante por um jornalista? Mais do que as questões normativas e de forma jornalística, é possível aqui observar aqui algumas características da prática do jornalista profissional e da sua postura individual ante a deontologia do jornalismo.

Levando em conta os resultados descritos e discutidos, entendemos que, apesar de ter sido moldada para a análise de temas femininos publicados no Dia Internacional da Mulher, a metodologia proposta pode ser replicada no exame de outros objetos. Na própria agenda dos direitos da mulher, a metodologia possibilitaria o estudo de outras coberturas jornalísticas, como a situação legal do aborto no Brasil, a violência doméstica contra mulheres, ou as barreiras de gênero no mercado de trabalho. Uma sugestão para a melhora da qualidade dos resultados seria aumento do *corpus* para permitir maior precisão na caracterização das tendências do uso de dados na mídia jornalística.

Os achados desta pesquisa indicam que a conexão entre o uso de dados e a cobertura de pautas ligadas não só à mulher, mas ao desenvolvimento democrático e de modo amplo, é, sim, um caminho de pesquisa promissor. No contexto deste artigo, contudo, ficou demonstrado que o uso de dados nas redações tradicionais não tem recebido maior atenção do que qualquer outra técnica jornalística rotineira.

NOTAS

- 1 Mayer-Schönberger & Cukier (2013) relatam que a explosão na quantidade de dados começou em ciências como a astronomia e a genética nos anos 2000, quando foi cunhado o termo. De acordo com eles, o conceito inicial de big data estava relacionado à incapacidade técnica de lidar com uma quantidade tão grande de informações. O significado se expandiu e hoje os autores utilizam o termo para se referirem a coisas feitas em larga escala que não poderiam ser concretizadas numa escala menor, como extrair insights ou criar formas de valor, de maneira que alteram mercados, organizações, e o relacionamento entre cidadãos e governos etc. (Mayer-Schönberger & Cukier, 2013, p. 6).
- 2 Físico britânico, criador da World Wide Web (WWW ou Internet), fundador da World Wide Consortium (W3C), o fórum de desenvolvimento técnico da Web, e cofundador do Open Data Institute, em

- Londres, que trabalha com o objetivo de manter a privacidade, liberdade e abertura da rede.
- 3 De acordo com o Relatório de Posição e Participação no mercado editorial em março de 2017 fornecido pelo Instituto de Verificação de Circulação (IVC), entidade sem fins lucrativos que realiza auditoria multiplataforma de mídia e fornece ao mercado dados isentos e detalhados sobre comunicação, incluindo tráfego web, tanto de desktops quanto de smartphones, tablets e aplicativos, bem como circulação, eventos, e inventário e campanhas de mídia out of home.
 - 4 No original: “When information was scarce, most of our efforts were devoted to hunting and gathering. Now that information is abundant, processing is more important” (tradução da autora).
 - 5 Fundada pelo jornalista australiano Julian Assange em 2006, a WikiLeaks se autodenomina uma organização multinacional de mídia e biblioteca coletiva e foi. A entidade é especializada na análise e publicação de grandes conjuntos de dados oficiais censurados ou restritos a respeito de guerras, espionagem e corrupção, já tendo publicado mais de 10 milhões de documentos e análises associadas com autenticidade verificada.
 - 6 Relatório de Posição e Participação no mercado editorial em março de 2017 fornecido pelo Instituto de Verificação de Circulação (IVC).

REFERÊNCIAS

Alexandre, I. A. R. (2014). *Jornalismo de Dados: o estado da arte nos jornais generalistas diários em Portugal* (Dissertação de mestrado). Recuperado do Repositório Universidade Nova (hdl.handle.net/10362/13615).

Arthur, C. (2010) Analysing data is the future for journalists, says Tim Berners-Lee. Recuperado de www.theguardian.com/media/2010/nov/22/data-analysis-tim-berners-lee.

Barbosa, S. (2007). *Jornalismo Digital em Base de Dados (JDBD). Um paradigma para produtos jornalísticos digitais dinâmicos* (Tese de Doutorado). Recuperado do Repositório Institucional UFBA (repositorio.ufba.br/ri/handle/ri/11299).

Barbosa, S. (2008). Modelo Jornalismo Digital em Base de Dados (JDBD) em interação com a convergência jornalística. *Textual & Visual Media*, 1, pp. 87–106.

Barbosa, S., & Torres, V. (2013). O paradigma “Jornalismo Digital em Base de Dados”: modos de narrar, formatos e visualizações para conteúdos. *Galáxia*, 25, pp. 152–164. <http://dx.doi.org/10.1590/S1982-25532013000200013>.

Berners-Lee, T. (2010, mai. 26). *Open, Linked Data for a Global Community* [Arquivo de Vídeo]. Recuperado de www.youtube.com/watch?v=gala5jXCFe0.

Bertocchi, D. (2013) *Dos dados aos formatos – Um modelo teórico para o design do sistema narrativo no jornalismo digital*. (Tese de Doutorado). Recuperado de Biblioteca Digital USP. doi: 10.11606/T.27.2014.tde-21092015-122011

Bradshaw, P. & Rohumaa, L. (2011). *The Online Journalism Handbook. Skills to survive in the digital age*. Harlow, England: Pearson.

Bradshaw, P. (2014). *O que é Jornalismo de Dados? Manual de Jornalismo de Dados*. Recuperado de datajournalismhandbook.org/pt/introducao_0.html.

Bragatto, R. C. (2011). Democracia e internet: apontamentos para a sistematização dos estudos da área. *Compólitica*, 2(1), pp. 132-163. <https://doi.org/10.21878/compolitica.2011.1.2> <https://doi.org/10.21878/compolitica.2011.1.2.14>

Brasil (2012, abr. 12). *Instrução Normativa n. 4, de 12 de abril de 2012*: Institui a Infraestrutura Nacional de Dados Abertos – INDA. Brasília. Recuperado de dados.gov.br/pagina/instrucao-normativa-da-inda.

Brasil (2000, maio 4). *Lei n. 101, de 04 de maio de 2000*: Estabelece normas de finanças públicas voltadas para a responsabilidade na gestão fiscal e dá outras providências. Brasília. Recuperado de www.planalto.gov.br/ccivil_03/Leis/LCP/Lcp101.htm.

Brasil (2009, maio 27). *Lei n. 131, de 27 de maio de 2009*: Acrescenta dispositivos à Lei Complementar no 101, de 4 de maio de 2000, que estabelece normas de finanças públicas voltadas para a responsabilidade na gestão fiscal e dá outras providências, a fim de determinar a disponibilização, em tempo real, de informações pormenorizadas sobre a execução orçamentária e financeira da União, dos Estados, do Distrito Federal e dos Municípios. Brasília. Recuperado de www.planalto.gov.br/Ccivil_03/leis/LCP/Lcp131.htm

Brasil (2011, nov. 18). *Lei n. 12.527, de 18 de novembro de 2011*: Regula o acesso a informações previsto no inciso XXXIII do art. 5º, no inciso II do § 3º do art. 37 e no § 2º do art. 216 da Constituição Federal; altera

a Lei no 8.112, de 11 de dezembro de 1990; revoga a Lei no 11.111, de 5 de maio de 2005, e dispositivos da Lei no 8.159, de 8 de janeiro de 1991; e dá outras providências. Brasília. Recuperado de www.planalto.gov.br/ccivil_03/_ato2011-2014/2011/lei/l12527.htm.

Cardoso, D. (2017, mar. 7). PSOL entra com ação no STF para descriminalizar aborto até 12ª semana. *Estadão*. Recuperado de saude.estadao.com.br/noticias/geral,psol-entra-com-acao-no-stf-para-descriminalizar-o-aborto-ate-12-semana-de-gestacao,70001690371

Castells, M. (1999). A era da informação: economia, sociedade e cultura. v. 1. *A Sociedade em Rede*. São Paulo: Paz e Terra.

Charbonneaux, J.; Gkouskou-Giannakou, P. (2015). O Jornalismo de “Dados”, uma Prática de Investigação? Um olhar sobre os casos alemão e grego. *Brazilian Journalism Research*, 11(2), pp. 266–291. <https://doi.org/10.25200/BJR.v11n2.2015.592>

Coddington, M. (2014). Clarifying Journalism’s Quantitative Turn: A Typology for Evaluating Data Journalism, Computational Journalism, and Computer-assisted Reporting. *Digital Journalism*, 3(3), pp. 331–348. <https://doi.org/10.1080/21670811.2014.976400>

Costa, M. T. (2017, mar. 8). ‘Mulheres têm menos voz, participação, e não têm poder de decisão’. *O Globo*. Recuperado de oglobo.globo.com/economia/mulheres-tem-menos-voz-participacao-nao-tem-poder-de-decisao-diz-miriam-muller-21028867

Cox, M. (2000). The Development of Computer-Assisted Reporting. Newspaper Division, Association for Education in Journalism and Mass Communication, *Southeast Colloquium*, v. 2030, n. 305, p. 22. Recuperado de citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.631.6220

Crucianelli, S. (Org.). (2013). *Ferramentas Digitais para Jornalistas 2.0*. Recuperado de knightcenter.utexas.edu/books/FerramentasDigitaisparaJornalistas.pdf.

Cushion, S, Lewis, J. & Callaghan, R. (2016) Data Journalism, Impartiality and Statistical Claims. *Journalism Practice*, 11(10), pp. 1198-1215. <https://doi.org/10.1080/17512786.2016.1256789>

Diakopoulos, N. (2011) *A Functional Roadmap for Innovation in Computational Journalism*. Recuperado de www.nickdiakopoulos.com/2011/04/22/a-functional-roadmap-for-innovation-in-computational-journalism/

Flew, T., Spurgeon, C., Daniel, A. & Swift, A. (2012) The Promise of Computational Journalism. *Journalism Practice*, 6 (2), pp. 157–171. <https://doi.org/10.1080/17512786.2011.616655>

Galdo, R. (2017, mar. 6). Crimes contra mulheres no topo de queixas

à PM. *O Globo*. Recuperado de oglobo.globo.com/rio/crimes-contramulheres-no-topo-de-queixas-pm-21034906

Gray, J., Bounegru, L. & Chambers, L. (Orgs.). (2012) *The Data Journalism Handbook – How Journalists Can Use Data to Improve the News*. Recuperado de datajournalismhandbook.org/pt/index.html.

Grandin, F. R. (2014) Criação de valor a partir do Jornalismo Guiado por Dados. In *11th World Media Economics and Management Conference*. pp. 1–30. Rio de Janeiro: UFRJ. Recuperado de www.academia.edu/7240997/Criação_de_valor_a_partir_do_Jornalismo_Guiado_por_Dados

Hamilton, J. & Turner, F. (2009) *Accountability through Algorithm: Developing the Field of Computational Journalism*, Stanford, CA, Center For Advanced Study in the Behavioral Sciences Summer Workshop, Universidade de Stanford. Recuperado de web.stanford.edu/~fturner/Hamilton%20Turner%20Acc%20by%20Alg%20Final.pdf

Hellerstein, J. (2008) *The Commoditization of Massive Data Analysis*. Recuperado de radar.oreilly.com/2008/11/the-commoditization-of-massive.html.

Hewett, J. (2015) Learning to teach data journalism: Innovation, influence and constraints. *Journalism*, 17 (1), pp. 119-137. <https://doi.org/10.1177/1464884915612681>

Kitchin, R. (2014). *The Data Revolution – Big Data, Open Data, Data Infrastructures & Their Consequences* (R. Rojek) London: SAGE Publication Ltd.

Knight, M. (2015). Data journalism in the UK: a preliminary analysis of form and content. *Journal of Media Practice*, 16 (1), pp. 55–72. <https://doi.org/10.1080/14682753.2015.1015801>

Lima Júnior, W. T. (2012) Big Data, Jornalismo Computacional e Data Journalism: estrutura, pensamento e prática profissional na Web de dados. *Estudos em Comunicação*, 12, pp. 207–222. Recuperado de www.ec.ubi.pt/ec/12/pdf/EC12-2012Dez-11.pdf

Machado, E. (2005) A Base de Dados como formato no Jornalismo Digital. *Actas do VII Lusocom*. Covilhã: Universidade Beira Interior, pp. 301–310. Recuperado de www.bocc.ubi.pt/pag/machado-elias-base-dados-formato-jornalismo-digital.pdf

Mancini, L. & Vasconcellos, F. (2016). Jornalismo de Dados: conceito e categorias. *Fronteiras – estudos midiáticos*, 18(1), pp. 69–82. Recuperado de revistas.unisinos.br/index.php/fronteiras/article/view/fem.2016.181.07

Marques, F. P. J. A. (2008) *Participação política e internet: meios e oportunidades digitais de participação civil na democracia contemporânea, com um estudo do caso do estado brasileiro*. (Tese

de Doutorado). Recuperado do Repositório UFBA (<https://repositorio.ufba.br/ri/handle/ri/11303>)

Mayer-Schönberger, V. & Cukier, K. (2013). *Big data: A revolution that will transform how we live, work, and think*. New York: Houghton Mifflin.

Meyer, P. (1973) *Precision Journalism: A Reporter's Introduction to Social Science Methods*. Bloomington: Indiana University Press.

Meyer, P. (1991) *The new precision journalism*. Bloomington, Indiana: Indiana University Press.

Parasie, S. & Dagiral, E. (2013) Data-driven journalism and the public good: "Computer-assisted-reporters" and "programmer-journalists" in Chicago. *New media & Society*, 15(6), pp. 853–871. <https://doi.org/10.1177/1461444812463345>

Pinho, J. A. G. de. (2008). Investigando portais de governo eletrônico de estados no Brasil: muita tecnologia, pouca democracia. *Revista de Administração Pública*, 42(3), pp. 471–493. <http://dx.doi.org/10.1590/S0034-76122008000300003>.

Prado, O., Ribeiro, M. & Diniz, E. (2012) Governo eletrônico e transparência: olhar crítico sobre os portais do governo federal brasileiro. In J. A. G. Pinho (Org.), *Estado, sociedade e interações digitais: expectativas democráticas*. (pp. 13 – 39) Salvador: Edufba.

Renó, L., & Renó, D. (2016). Jornalismo de dados e tecnologia: algoritmo na produção da notícia transmídia | Data Journalism and Technology: Algorithm in the Production of Transmedia News. *Razón Y Palabra*, 20 (1-92), pp. 1186-1203. Recuperado de www.revistarazonypalabra.org/index.php/ryp/article/view/295

Rogers, S. (2008) "Turning Official Figures into Understandable Graphics, at the Press of a Button." *Guardian Insider Blog*. Recuperado de www.Gguardian.co.uk/help/insideGuardian/2008/dec/18/unemploymentdata.

Rogers, S. (2014) Data journalism is the new punk. *British Journalism Review*, 25(2), pp. 31–34. <https://doi.org/10.1177/0956474814538181>

Rogers, S., Gallagher, A. (2013, April 4) *What is data journalism at the Guardian?* [Arquivo de Vídeo] Recuperado de www.theguardian.com/news/datablog/video/2013/apr/04/what-is-data-journalism-video.

Rothberg, D. (2008). Por uma agenda de pesquisa em democracia eletrônica. *Opinião Pública*, 14(1), pp. 149–172. <http://dx.doi.org/10.1590/S0104-62762008000100006>.

Silva, J. A. B. e (2009) Transformações no processo de produção da

notícia. *Bibliocom*, 2(1), pp. 38–41. Recuperado de portcom.intercom.org.br/revistas/index.php/bibliocom/article/download/1524/1502

Sponholz, L. (2009). *Jornalismo, Conhecimento e Objetividade: Além do Espelho e das Construções*. Florianópolis: Insular.

Stray, J. (2011) *A Computational Journalism Reading List*. Recuperado de jonathanstray.com/a-computational-journalism-reading-list

Träsel, M. (2014). *Entrevistando planilhas: estudo das crenças e do ethos de um grupo de profissionais de jornalismo guiado por dados no Brasil* (Tese de Doutorado) Recuperado do Repositório Institucional PUCRS (repositorio.pucrs.br/dspace/handle/10923/6841)

White House. (2009). *Open Government Directive*. Recuperado de www.digitalgov.gov/open-government-directive/

Zuiderwijk, A., Janssen, M. & Dwivedi, Y. K. (2015) Acceptance and use predictors of open data technologies: Drawing upon the unified theory of acceptance and use of technology. *Government Information Quarterly*. 32(4), pp. 429-440. <https://doi.org/10.1016/j.giq.2015.09.005>

ANA CAROLINA ARAÚJO. Professora substituta do Departamento de Jornalismo da Faculdade de Comunicação da Universidade Federal da Bahia e estudante de doutorado do Programa de Pós-graduação em Comunicação e Cultura Contemporânea. Pesquisa nas áreas de Transparência Pública, Dados Abertos e Jornalismo Identitários. E-mail: contatoacaraujo@gmail.com

RECEBIDO EM: 01/11/2017 | ACEITO EM: 16/06/2018