

DADOS SÃO FONTES NO JORNALISMO? Contribuições sobre o conceito de Jornalismo de Dados¹

ARE DATA SOURCES IN JOURNALISM? Contributions to the concept of Data Journalism

Vítor Almeida dos Santos ²
Cíntia Xavier³

Resumo: Este artigo traz um percurso sobre o uso de dados no jornalismo na tentativa de responder se dados são fontes de informação no jornalismo. Para isso, utiliza-se de uma revisão na literatura que examina o atravessamento dos dados no jornalismo. A partir desse movimento, percebe-se que as pesquisas se estabelecem em dois segmentos: a utilização dos dados para orientar a atividade jornalística, ou o uso dos dados para compor o material noticioso, como fonte de informação, tal qual reivindica a literatura sobre Jornalismo de Dados (JD). Diante desse quadro, entende-se que o “dado” é um termo ainda em construção, mas que pode ser coletado, manipulado e reproduzido pelo jornalista.

Palavras-Chave: dados. fontes. jornalismo de dados.

Abstract: This article provides an overview of the use of data in journalism, seeking to determine whether data can be considered sources of information in journalism. To this end, it employs a literature review that examines how data intersects with the journalistic field. From this analysis, it is observed that research is established across two segments: the use of data to guide journalistic activity, or the use of data to compose news material as a source of information, as claimed by the literature on Data Journalism (DJ). Given this framework, it is understood that “data” is a term still under construction, yet one that can be collected, manipulated, and reproduced by the journalist.

Keywords: data. source. data journalism.

1. Introdução

Há uma dificuldade conceitual em responder a pergunta do título, haja vista que não se ter uma definição para o termo “dados”, se apresenta como um obstáculo inicial. Por isso,

¹ Trabalho apresentado ao Grupo de Trabalho GT Estudos de Jornalismo. 35º Encontro Anual da Compós, Universidade Federal do Rio Grande do Norte (UFRN). Natal - RN. 9 a 12 de junho de 2026.

² Vítor Almeida dos Santos: UEPG, Mestrando, Bolsista Capes, vitoralmeida.jor@gmail.com

³ Cíntia Xavier: UEPG, Doutora, cxavier@uepg.br.

busca-se, neste artigo, retomar aspectos sócio-culturais, como a dataficação, na tentativa de encontrar caminhos.

O jornalismo é atravessado pelos dados, uma vez que as indústrias de base são substituídas por empresas cuja atividade é armazenar e distribuir informação, tendo os dados como o produto que sustenta esse modelo de negócio (Silveira, 2017).

Essa ruptura para o modelo pós-industrial tem na sua estrutura o que Silveira (2017) chama de sociedades informacionais, também chamadas de sociedade operadas por softwares. Nesse modelo econômico, o valor se encontra na produção de bens que podem ser transferidos por redes digitais (Silveira, 2017). Essa mudança pode ser exemplificada pela fechadura digital que, ao substituir a fechadura de metal, coleta os dados de quantas vezes, quais horários e qual cartão foi utilizado para abrir e fechar a respectiva porta (Silveira, 2017). Também é nesse modelo econômico que se vê o uso desenfreado de redes sociais para socialização, produção e distribuição de conteúdo, inclusive jornalístico.

Para este artigo, considera-se duas modalidades para o uso do conceito de dados. A primeira tem os dados como orientador; a segunda, o dado como fonte, principal interesse desta pesquisa.

O jornalismo se utiliza de dados antes do processo de dataficação. Inclusive, há dois casos que são mencionados como uma proto intersecção entre dados e jornalismo, os protestos de Detroit, 1967, em que Philip Meyer coletou e analisou dados para entender o perfil dos manifestantes (Meyer, 1991). O outro caso, data de mais de 200 anos, quando o *The Guardian*, que na época ainda se chamava *Manchester Guardian*, publicou uma lista com as escolas de *Manchester* e *Salford*, discriminando quantos alunos havia por instituição e qual era o gasto por ano (Rogers, 2011). Isso possibilitou identificar, pela primeira vez, quantas crianças acessaram o ensino gratuito e quantas estavam em situação de pobreza. Inclusive, esses dados desmentiram as estimativas oficiais da época de que haviam cerca de 8000 alunos recebendo educação gratuita, a lista do *The Guardian* apresentava cerca de 25000 crianças (Rogers, 2011).

Contudo, a dataficação da vida (Lemos, 2021), a partir do agenciamento das *big techs* no controle dos dados, somada à institucionalização de políticas de acesso à informação, permitiu ao jornalismo acessar um conjunto de dados que está disponível a partir da digitalização, isto é, dados mediados por computadores.

Ainda que neste percurso olhe-se para o jornalismo no sentido amplo, se faz necessário apropriar-se de textos sobre Jornalismo de Dados (JD), pois se trata de uma prática que reivindica, no próprio termo, o uso de dados, ainda que a partir de bases de dados (Träsel, 2014; Gehrke, 2022).

Para este percurso foram mobilizadas pesquisas, a partir de uma revisão de literatura, que tratam do uso de dados no jornalismo por diferentes entradas. Diante desse quadro, percebe-se que essas pesquisas estão localizadas em dois segmentos conceituais. O primeiro, tem os dados como elemento que atravessa o processo de produção, a partir da relação do jornalista com algoritmos, plataformas e metrificação da atividade laboral. O segundo segmento utiliza os dados para compor o material noticioso, geralmente associado à prática do Jornalismo de Dados.

2. Convergência permite o acesso a dados

Os dados atravessam o jornalismo, da pré-produção ao consumo. Ao mesmo tempo, o Jornalismo de Dados (JD) se estabelece como prática que reivindica uma relação mais amigável com este artefato. Por um lado, pesquisas revelam que a atividade jornalística é dependente do serviço dos dados e das suas infraestruturas, como as plataformas e algoritmos (Kalsing, 2021; Silva, 2025; Winques; Longhi, 2022; Lemos, 2020, 2021). Do outro, o JD se constitui como uma prática que utiliza o volume de dados disponíveis para fazer reportagens (Träsel, 2014). Entende-se que, em ambos os casos, a dataficação da vida (Lemos, 2021) é o agente permissor desse atravessamento dos dados no jornalismo.

Não é do escopo deste artigo discutir se estamos na “convergência” (Salaverría; Avilés; Masip, 2010) ou se já migramos para a “pós-convergência” (Fagerjord, 2010; Ramirez, 2013). Até porque, há em ambos os conceitos, características em comum do desenvolvimento tecnológico e de seu impacto na produção e no consumo de mídia. De qualquer forma, entende-se que a indicação de Salaverría, Avilés e Masip (2010) de que a convergência é um conceito amplo também serve para a pós-convergência.

Ainda que a literatura sobre convergência seja dos anos 70, sua popularização veio a partir dos anos 90, quando as empresas jornalísticas passaram a adotar as inovações tecnológicas vigentes (Salaverría, Avilés, Masip, 2010). Antes da virada do milênio, Negroponte, um dos primeiros a cunhar o termo convergência (Ramirez, 2013), afirmou que

a revolução digital havia acabado, argumentando que as “mudanças” já haviam acontecido (Negroponte, 1998).

De fato, Negroponte (1998) neste curto artigo “Beyond Digital” mapeia características que teriam desenvolvimento acentuado apenas nos anos seguintes, como a diminuição do tamanho dos dispositivos, expansão da banda larga e aumento da velocidade dos processadores. Inclusive, é a partir dessa argumentação que Fagerjord (2010) se apoia para escrever sobre o momento após a convergência.

O jornalismo é uma das práticas transformadas por esse desenvolvimento tecnológico que converge na intenção de gerar e manipular dados, também culturalmente chamado de dataficação (Lemos, 2021). Como ainda não se tem claro se o trabalho orientado pelos dados é uma consequência da convergência, podendo ser pensada em um momento “pós”, ou se no próprio desenvolvimento da digitalização já havia a intenção de transformar toda a ação em dados (Lemos, 2021), utiliza-se aqui o conceito de “convergência” (Salaverría; Avilés; Masip, 2010) por abarcar um maior período de tempo e por oferecer uma ideia de *continuum*, principalmente quando se trata da incorporação desse desenvolvimento tecnológico pelos veículos jornalísticos. Para Suzana Barbosa, *continuum* (2013, p. 38):

[...] abrange aspectos relacionados aos desenvolvimentos tecnológicos, à absorção de novos procedimentos para realizar os processos e rotinas de produção do jornalismo, como também os avanços já empreendidos nos estudos para o melhor entendimento do fenômeno da convergência jornalística, suas particularidades, consequências e também divergências. [...] (Suzana Barbosa, 2013, p. 38)

A convergência não afeta apenas os sistemas de mídia, mas a vida das pessoas (Ramirez, 2013). Portanto, o impacto da convergência profissional na produção e na distribuição (Salaverría; Avilés; Masip, 2010) também se orienta pela generalização da tecnologia e sua mediação nas relações sociais (Lemos, 2021).

3. Dataficação do jornalismo

Como já mencionado, não há como falar da relação entre jornalismo e dados sem mencionar o Jornalismo de Dados (JD). Contudo, há uma certa dificuldade teórica em definir o JD. Por enquanto, utiliza-se o conceito de Jornalismo de Dados de Träsel (2014), que entende o JD como um conjunto de práticas que têm em comum a utilização da base de dados. Mtchedlidze (2024) considera o Jornalismo de Dados como um termo polissêmico.

Outros autores, como M Fikri AR (2024) admitem que o estado da arte do JD ainda não é bem conhecido. Por mais que a origem do termo remonte algumas técnicas do Jornalismo de Precisão de Philip Meyer (1991) e do Jornalismo Investigativo, considera-se que para pleno exercício do Jornalismo de Dados, precisa que se tenha o produto dessa prática: os dados. Isso significa que dataficação não é o suficiente. Precisa-se que esses dados estejam em repositórios acessíveis.

Se o Jornalismo de Dados se apresenta como um conceito ainda em construção, os “dados” também. Ao buscar definições para o termo, percebe-se algumas discussões quanto a uma possível equivalência epistemológica entre dados e informação. Lima Junior (2012) traz uma comparação simplificada, enquanto “dados” seriam um artefato binário armazenado por máquina, informação seria o dado estruturado.

O pesquisador israelense Chaim Zins (2007), especialista em ciência da informação, conduziu uma pesquisa a partir do método Delphi com 57 pesquisadores de 16 países para encontrar diferenças entre dados, informação e conhecimento. Charles Ess, da Drury University (Estados Unidos), um dos respondentes da pesquisa, explica que o termo em inglês “*data*” significa códigos binários que podem ser armazenados, processados e transmitidos por computadores, podendo ser “informação” em sentido estrito. Inclusive, o termo “*data*” por si só representa um conjunto de dados. Sua forma singular é “*datum*” do latim. Em comparação, para o professor Charles Ess, informação é o conjunto desses dados, podendo ter sentido na sua origem ou tendo sentido atribuído posteriormente pelo próprio ser humano.

Por essa definição, toda vez que se utiliza o termo “informação” para algo encontrado na internet, está se referindo a um estágio posterior do dado, conforme Figura 1, já que a transferência de informação por si só, é codificada e decodificada em dados.

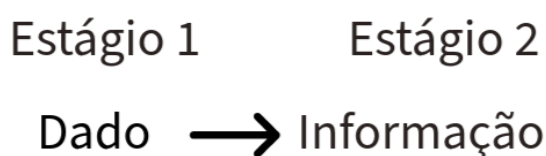


FIGURA 1 - Dado podem ser um estágio anterior a informação

FONTE - elaboração própria (2026)

Estamos mantendo e reproduzindo relações sociais a partir de um gigantesco sistema de controle de informações. As tecnologias da sociedade informacional, baseadas na computação e nas redes digitais cibernéticas, foram constituídas para confirmar se a informação chegou ao seu destino e, na medida do possível, se ela resultou no comando desejado. Assim, o envio de informações quase sempre é acompanhado de seu registro. Dados são comunicados gerando dados sobre a comunicação efetuada, ou seja, metadados são constantemente criados. (Silveira, 2017, p. 18)

Sendo assim, pode-se reconhecer que o jornalismo, principalmente a partir da convergência tecnológica, utiliza dados como fonte de informação. Todavia, esse entendimento generalista de que toda informação que é transmitida por um dispositivo ligado à internet já foi um “dado”, não contribui para a discussão do seu atravessamento no jornalismo. Por isso, neste artigo serão apresentadas duas propostas para tratar da relação entre os dados e seus usos no jornalismo: os dados como orientador e os dados como fonte. Para fazer o debate sobre a relação do jornalismo com os dados se faz necessário apresentar as condições que permitem a incorporação dos dados como atributo natural da vida em sociedade.

O desenvolvimento tecnológico é um atravessador da atividade jornalística. Em um recorte de pelo menos 100 anos, vimos o jornalismo incorporar o rádio, a televisão e a internet, ou seja, não se deve tratar a convergência jornalística como um fenômeno inédito. O que muda, nesse caso, é por onde ancoram essas mudanças. Para este artigo, considera-se que a dataficação é o principal produto da convergência, já que representa uma orientação que afeta os valores do trabalho jornalístico.

No processo de expansão desencadeado a partir do século XIX, a tecnologia sempre foi um fator preponderante para o aprimoramento dos procedimentos da produção jornalística, do trabalho dos profissionais, da oferta informativa, dos modelos dos produtos e dos formatos dos conteúdos, assim como permitiu vencer distâncias para que a velocidade de circulação das notícias pudesse superar barreiras geográficas e temporais e chegar até o público, satisfazendo as necessidades de informação da sociedade (Barbosa, 2013, p. 38)

Entre o final da década de 60 e início de 70, o New York Times estruturou um banco de dados para ser utilizado na rotina produtiva (Lima Junior, 2012). No Brasil, nesse mesmo período, a Folha de S. Paulo começou um projeto para implementar computadores na redação (Lima Junior, 2012). O desenvolvimento desse processo, com a incorporação acentuada de ferramentas computacionais nas redações é que Salaverría, Avilés e Masip (2010, p. 59) chama de convergência jornalística. Para os autores, trata de:

[...] um processo multidimensional que, facilitado pela ampla implementação de tecnologias de telecomunicações digitais, afeta as esferas tecnológica, empresarial, profissional e editorial da mídia, fomentando uma integração de ferramentas, espaços, métodos de trabalho e linguagens antes separados, de modo que os jornalistas produzam conteúdo que é distribuído por meio de múltiplas plataformas, utilizando as linguagens específicas de cada uma (Salaverría; Avilés; Masip, 2010, p. 59, tradução nossa).

Entre as tecnologias de telecomunicações digitais está a possibilidade de transmitir dados por redes sem fio e por cabeamento compartilhado entre internet, telefonia e televisão, ou seja, uma convergência nas redes de conexão (Fagerjord, 2010). É essa convergência tecnológica que facilita a transmissão de conteúdo (Salaverría; Avilés; Masip, 2010) e de dados. Essas tecnologias possibilitam que um número muito maior de indivíduos crie e divulgue informações (Winques; Longui, 2022), produzindo dados sobre si mesmo.

A partir disso, há uma convergência de gêneros e serviços (Fagerjord, 2010), o que não significa que essa integração tenha resultado em menos dispositivos. Não há menos aparelhos digitais do que havia, já que o analógico migrou para o digital (Fagerjord, 2010). O efeito disso é o gerenciamento de toda nossa ação e a transformação em dados, como já mencionado. Lemos (2021) traz o exemplo do livro, que ao migrar para o formato digital, um *e-book*, não significou apenas a transposição da leitura para um *e-reader*, mas a transformação do ato de ler em dados coletados sobre ritmo de leitura, citações destacadas e outras preferências de consumo (Lemos, 2021).

O rastreamento do nosso comportamento através desses dispositivos também ganha uma dimensão cultural, à medida que há ensejo para que se utilize desses recursos. Lima Junior (2012, p. 208) chama essa propensão de “conexão amigável” que faz com que as pessoas sintam que essas tecnologias digitais trazem conforto. Além disso, a diminuição do custo dos dispositivos aliada a alta disponibilidade facilita a aquisição (Lima Junior, 2012).

Com a dataficação, não se trata apenas da conversão de um objeto analógico em digital, mas da modificação de ações, comportamentos e conhecimentos baseados na performance dos dados elaborada por sistemas de inteligência algorítmica. Esta deve ser pensada como um conjunto de métodos de coleta, processamento e tratamento de dados para realizar previsões [...] (Lemos, 2021, p. 194)

Para o jornalismo, o resultado é um ecossistema midiático complexo (Motta, 2024), resultando em um sistema em que um volume gigantesco de dados que promovem a lógica algorítmica e que põe em disputa por atenção em plataformas a mídia tradicional, veículos

independentes, influenciadores e pessoas comuns. Nesse caso, toda a infraestrutura é pensada para que ações transformadas em dados sirvam como orientador do trabalho jornalístico.

Chaim Zins (2007), que pesquisou a origem dos termos dados, informação e conhecimento, utiliza a palavra “artefato” como um dos possíveis sinônimos para “dados”. Portanto, neste artigo também será utilizado “artefato” para se referir aos “dados”.

3.1 Dado como orientador

As empresas jornalísticas também tiveram a incumbência de compreender a audiência, de modo que se possa vender mais espaços publicitários, refinar os horários de circulação do jornal ou dos programas de maior audiência no rádio e televisão. Esse processo ocorre pela transformação dos dados coletados sobre a audiência em algoritmos, que por sua vez viram estratégias para ranqueamento, título e corpo do texto noticioso, ou seja, metrifica o trabalho jornalístico (Kalsing, 2021).

A partir da convergência tecnológica, com a digitalização da produção e do consumo (Salaverria; Avilés; Masip, 2010) e da ação transformada em dado armazenável (Lemos, 2021) e disponível, organizações que vendem a atenção, entre elas, as empresas jornalísticas, dispuseram de mais métricas para identificar o comportamento da audiência.

Ao se observar para além do jornalismo, identifica-se que as métricas se fazem presentes na maioria das profissões, sobretudo naquelas diretamente impactadas pela indústria 4.0. Elas permeiam as mais distintas instâncias, desde a vida acadêmica – com quantificações que vão do sistema de avaliação da Capes ao Google Scholar, sem contar as bonificações por citações – até a área da saúde. Algumas empresas, para manter os trabalhadores motivados (e, de quebra, para aumentar a produtividade e o lucro), utilizam a bonificação por resultados. E esse sistema de recompensa não é necessariamente monetário (Kalsing, 2021, p. 45-46).

Para além de tratar de um jornalismo de dados ou de um jornalismo orientado por dados, o que se tem é a vida orientada por dados. Para quantizar todas nossas ações, precisa-se de uma infraestrutura que, primeiro, dê conta de armazenar e distribuir esses dados a terceiros e, segundo, que tenha um potencial de atração para que as ações estejam mediadas por essa própria infraestrutura. No caso atual, são as plataformas.

[...] Enquanto existem mais vozes na esfera da produção, a concentração é cada vez maior na esfera de circulação, com plataformas e toda forma de intermediários modulando e estabelecendo limites, regras e padrões de controle ao manancial de informação criado. Essa contradição não é estática, e sim resultado da disputa entre os agentes diversos da sociedade (mercado, Estado e sociedade civil) que se

estabelece tanto na concorrência de mercado quanto na instituição de regramentos que disciplinam as atividades (Valente, 2023, p. 147).

A plataforma capta os dados de audiência, os organiza a partir dos seus próprios interesses e reescreve-os por meio de algoritmos que vão orientar as estratégias de busca da atenção. Essa sequência de processos contamina a atividade jornalística, que permanece em disputa por oferecer conteúdos noticiosos orientados por valores jornalísticos e não apenas pela lógica das plataformas.

Nathalia Lopes da Silva (2025), percebe que a noticiabilidade ainda que permaneça sendo socialmente construída, é influenciada por três fatores: maior uso das plataformas nas redações, e por consequência a disponibilidade de dados sobre acesso, compartilhamento e buscas, via algoritmo; utilização de inteligência artificial para produção do conteúdo; por último, o uso do *WhatsApp* para contato com as fontes. Desses três fatores, os dois primeiros se relacionam ao produto da dataficação, ou seja, o jornalista acessa o dado após a coleta e mediação algorítmica. Já o último fator, é o primeiro estágio de uso de uma infraestrutura de coleta e armazenamento de dados pelo jornalista.

Por mais que pareça que a decisão de aderir às plataformas não tenha benefícios para o trabalho jornalístico, há atrativos para o setor comercial da empresa. O primeiro são dados (transformados em informação) sobre o comportamento da audiência (Kalsing, 2021). O outro atrativo é a alta adesão do público a essas plataformas, o que aumenta, potencialmente, o número de acessos. Para Winkes e Longhi (2022), as plataformas se tornaram espaços de debate público. Em contrapartida, essas empresas jornalísticas não controlam a “circulação do conteúdo”, a “lógica dos algoritmos” nem a “monetização” (Kalsing, 2021, p. 37).

Lemos (2020; 2021) organiza esses ecossistema em “Plataformização, dataficação e performatividade algorítmica (PDPA)” que operam a partir da dataficação das ações, gerenciamento dos dados a partir de grandes infraestruturas e mediação dos algoritmos a serviço dessas mesmas estruturas. Para o jornalismo, sobra uma relação mediada com o público, tendo o resultado fabricado pela própria mediação (Winkes; Longui, 2022).

Os jornalistas escrevem notícias em computadores, gravam vídeos com celulares, contatam pessoas para falar ao jornal por meio de redes sociais (plataformas) e publicam conteúdos nessas mesmas plataformas. Do outro lado, o resultado desse acesso orienta a rotina, incidindo sobre o que vai ser publicado, como já detalhou Janaína Kalsing (2021). Enquanto a autora chama essa orientação de jornalistas metrificados, nossa preocupação é

com um estágio anterior: o jornalismo “datafocado”, onde toda “ação é transformada em dados” (Lemos, 2021), pois não há métricas se não há dados. No momento em que os dados são utilizados para embasar decisões durante o processo de produção, abra-se para que também seja utilizado como fonte de informação.

3.2 Dado como fonte

As fontes, pessoas ou documentos, são utilizadas pelos jornalistas, durante o processo de apuração para informar sobre determinado fato, aparecendo de forma explícita ao longo do texto (Gehrke, 2018), ou em *off*. No começo dos anos 2000, antes do termo “data journalism”, Elias Machado (2002) já havia identificado que o ciberespaço, ou seja, a comunicação em rede, ofereceria ao jornalismo a possibilidade de expandir suas fontes, já que o repórter poderia levantar os dados disponíveis sobre determinada pauta antes de ir em busca da declaração de alguma personalidade. Ainda que Machado (2002) não esclareça a natureza desse dado, indica uma possibilidade de superar o jornalismo declaratório.

Não são só as plataformas que utilizam dados para seus interesses, o jornalismo também. Além da dataficação mencionada, a política de dados abertos possibilita que o jornalismo utilize dados como fonte de informação.

Gehrke (2018) elabora três categorias para pensar a fonte documental no Jornalismo de Dados: estatística, reprodução e arquivo documental. Reconhece que dentre essas categorias, a fonte documental estatística é aquela utilizada para descrever a dimensão generalista quando se pensa no termo “dados”, que é associada ao número. Gehrke (2018, p. 93) inclui entre as fontes estatísticas as bases de dados, número extraídos de pesquisas, portanto, inclui uma “abordagem quantitativa”.

Erkmen (2023) entende que diferentemente da Reportagem Assistida por Computador (RAC) e do Jornalismo Computacional⁴, o Jornalismo de Dados estaria associado a um uso desses dados como fonte primária, ainda que o autor não explique o que seria uma fonte primária. Francesca Morini (2025) na sua incursão etnográfica pelo jornal alemão *Die*

⁴ Conceito criado por Sarah Cohen, Fred Turner e James Hamilton (2011) que dá uma ideia do uso de técnicas de ciência da computação no jornalismo. A amplitude do conceito também acaba dificultando a aproximação ou afastamento do conceito de Jornalismo de Dados.

Tageszeitung (taz) também percebeu uma centralidade no uso dos dados, passando de “mais uma fonte” para a fonte principal da reportagem.

O termo “dados”, nesse contexto, é genérico, porque Erkmen (2023) considera que o JD é dependente da disponibilidade de dados públicos. Träsel (2014) traz uma característica adicional a esses dados que são as “bases de dados”, que podem ser públicas, e que para Gehrke (2018) seria uma fonte documental estatística. Inclusive, é dessas bases de dados públicas que trata o Jornalismo de Precisão, de Meyer (1991) um dos precursores do Jornalismo de Dados (Träsel, 2014).

Suzana Barbosa (2013) coloca as bases de dados como estruturantes do trabalho jornalístico, já que estaria incorporado desde a pré-produção até a circulação. Esse paradigma, entende-se, estaria mais ligada a dataficação do que o uso dos dados como fontes. Mesmo assim, a entrada “dados como orientador” contribui para que o jornalista incorpore a lógica de que se pode confiar nos dados. Se são esses mesmos dados que determinam estratégias de busca de atenção e compartilhamento (Kalsing, 2021), por que não poderia determinar a confiabilidade da apuração jornalística?

A rotinização dos microcomputadores e do acesso à Internet no final do século XX leva a um aumento na oferta de bases de dados públicos e de software simples e de baixo custo capazes de analisá-las. Com isso, ressurgiu o interesse pelo Jornalismo de Precisão e pela RAC, em especial a partir da metade dos anos 2000, quando Holovaty (2006) publica um artigo influente a respeito da coleta e manutenção de bancos de dados pelas redações (Träsel, 2014, p. 23).

Se na prática jornalística que não se autodenomina “de dados”, a busca ocorre em dados estruturados e apresentados ao público como uma pesquisa acadêmica, em bibliotecas do Instituto Brasileiro de Geografia e Estatística (IBGE) ou levantamentos de órgãos públicos, o Jornalismo de Dados se distinguiria apenas por buscar acessar bases de dados em estado “bruto”⁵, ou seja, que ainda não estão integradas. Por serem arquivos extensos não podem ser manipulados em formato de planilha, portanto pode-se utilizar linguagens de programação, como o *Python*.

Lima Junior (2012, p. 211) chama de *big data* “conjunto de dados (*dataset*) cujo tamanho está além da habilidade de ferramentas típicas de banco de dados em capturar, gerenciar e analisar”. O autor (2012) também questiona a indefinição sobre qual o tamanho necessário para que um conjunto de dados seja considerado *big data*. Sendo assim, enquanto

⁵ Do inglês “*raw data*”.

o Jornalismo de Dados trabalha com *big data*, outras formas de jornalismo trabalham com “dados”.

Essa pergunta é que divide, por uma linha muito frágil, o uso de dados no jornalismo tradicional e no Jornalismo de Dados. Se o jornalista utiliza a ferramenta de planilhas para analisar os dados, *big data* será as bases de dados a partir de 10 milhões de células se o *software* utilizado for o Google planilhas, se a opção for o *software Excel*, *big data* será as bases de dados superiores a 1.048.576 linhas por 16.384 colunas. Pois ao ultrapassar os limites de uso comum de dados, o jornalista teria que recorrer a linguagem de programação voltada ao gerenciamento de dados.

De toda forma, o jornalismo está tomado pelo uso do dado como fonte de informação. A apuração mediada por um computador, celular ou outro dispositivo conectado à internet, faz com que a informação utilizada em uma notícia ou reportagem já tenha sido um dado armazenado. A diferença ficaria na forma de coleta, que depende da habilidade do jornalista e na estrutura de apresentação desses dados, se em estado bruto ou pronto para a publicação;

3.3 Dados são fonte no jornalismo?

Há duas respostas para essa pergunta. A primeira é que se tomar “dados” no sentido geral, como um artefato anterior à informação, toda informação mediada a partir da digitalização é um dado, ou seja, no trabalho jornalístico ficaria de fora apenas consultas a fontes que não estivessem mediadas por dispositivos tecnológicos. Um exemplo seria a interação face a face com registro em dispositivo analógico.

Como neste percurso de pesquisa, considera um sentido mais restrito o termo “dados”, como um objeto quantizado, geralmente apresentado como um número, entende-se que o dado pode ser uma fonte jornalística, ainda que nem todo dado utilizado na construção de um material noticioso seja uma fonte. Portanto, a segunda resposta para a pergunta é: depende do acesso ao dado.

De modo geral, o dado é coletado, manipulado e reproduzido. A coleta é a captura de evidências a partir da realidade; a manipulação é a sistematização, organização desses dados, acrescido de cruzamento com outros dados, a depender do caso; a reprodução é a circulação desse dado a partir de um sentido atribuído. Cada uma dessas etapas tem especificidades de acordo com o tipo ou volume de dados. O jornalista portanto pode estar envolvido em cada

uma dessas etapas, na coleta dos dados, na manipulação de dados previamente construídos ou na reprodução de um dado que já foi coletado e construído. A TABELA 1 apresenta três eixos de acesso do jornalista aos dados e sua relação com as etapas de construção dos dados: o dado construído por jornalistas, dado produzido por terceiros e o dado produzido por terceiros mas transformado pelo jornalista.

TABELA 1
Eixos de acesso aos dados pelos jornalistas

Eixo de acesso aos dados	Coleta	Manipulação	Reprodução
Dado construído por jornalistas	X	X	X
Dado produzido por terceiros			X
Dado produzido por terceiros mas transformado pelo jornalista		X	X

FONTE – elaboração própria (2026)

Eixo 1 - Dado construído por jornalistas: é quando o jornalista apreende um conjunto de evidências sobre determinado fato e transforma em um banco de dados. A coleta desse dado por parte do jornalista pode ser feita por meio da observação ou de um *survey*, por exemplo. É quando o jornalista constrói o dado, vide Institutos de Pesquisas e organizações sem fins lucrativos. No livro “Rota 66: A História da Polícia que Mata” (Caco Barcellos, 2018), o autor pede a seu ajudante Sidney que fique de plantão no IML observando as entradas dos cadáveres (Barcellos, 2018), a fim de ampliar sua base de dados estruturada em 1975. Neste tipo de acesso, considera-se que os dados não são fontes, mas uma informação ou conjunto de informações construídos pelo próprio jornalista.

Eixo 2 - Dado produzido por terceiros: é quando se consulta fontes documentais, como relatórios e levantamentos, para se extrair os dados. Nesse caso, o dado já tem um sentido atribuído, estando pronto para uso. Um exemplo é o caso dos dados produzidos pelo Instituto Brasileiro de Geografia e Estatística (IBGE) que são utilizados na íntegra como fonte no jornalismo. Entende-se aqui que o dado extraído é uma fonte documental.

Eixo 3 - Dado produzido por terceiros mas transformado pelo jornalista: é quando o jornalista acessa os dados para gerar uma nova informação, a partir do cruzamento com outros dados. Nesse caso, é dependente de bases de dados disponíveis e de requerimentos à Lei de Acesso à Informação. Nesta categoria, o jornalista acessa os “dados brutos” ou “*raw data*”, geralmente categorizadas como *big data*, por ultrapassarem o volume que pode ser gerenciado em *softwares* de planilhas. Aqui, o dado não é utilizado na íntegra, a partir do seu sentido original, como no exemplo anterior, mas a partir da manipulação do jornalista. Como resultado, o jornalista constroi uma nova informação que não estaria presente caso esse dado fosse apenas reproduzido. Entende-se que o dado manipulado é uma fonte de informação, ainda que se apresente no conteúdo noticioso para compor uma nova informação.

Nos Eixos 2 e 3, o jornalista é totalmente dependente da disponibilidade de dados, ou seja, se beneficia do fenômeno da dataficação e da institucionalização das políticas de acesso à informação, utilizando relatórios de empresas privadas, de governos, e de bases de dados públicas. A confiabilidade dos dados produzidos por terceiros depende da transparência na metodologia dessas organizações. Caso não seja informado, quem assegura a confiabilidade na origem desse dado é a própria instituição, que atuaria justificando o dado (Ramsälv; Ekström, 2024).

Há uma outra forma de apresentação dos dados no jornalismo que diz respeito às fontes humanas que informam dados. Em ano de eleição, vê-se veículos jornalísticos, como G1, Estadão, GZH, com cobertura específica de checagem de informações em debates eleitorais. Esse tipo de relato não é categorizado no Eixo 3 por não se poder averiguar se seria a reprodução de um dado já disponível, se é um levantamento da própria fonte humana, ou se é uma informação a partir do cruzamento de dados. Além disso, entende-se que nesse caso, o dado compõe o relato da fonte, sendo tratado pelo jornalista a partir dessa perspectiva.

4. Conclusões

A proposta apresentada tenta desenvolver um percurso que possa indicar a utilização dos dados como fonte no jornalismo. A primeira dificuldade é não se ter uma definição para o termo “dados”, tendo como pistas a pesquisa de Lima Junior (2012) e de Chaim Zins (2007).

Por isso recorreu-se ao fenômeno que implica na proliferação dos dados: a dataficação (Lemos, 2021).

Para analisar a incorporação dos dados no jornalismo, também se faz necessário recuperar as condições tecnológicas que possibilitaram tal uso. A pesquisa considera a dataficação como um dos eixos da convergência, ainda que seja necessário maior investigação para colocá-la como uma característica da convergência ou como uma consequência. Salaverría, Avilés e Masip (2010) já haviam apontado certas incongruências quanto ao encaixe de um fenômeno como parte ou como fruto da convergência.

Pode-se dizer que há um caminho percorrido até a dataficação do jornalismo, que inclui as redes de conexão sem fio (Fagerjord, 2010) associado ao barateamento do custo dos dispositivos (Lima Junior, 2012) e portabilidade (diminuição do tamanho) (Negroponte, 1998). Ao mesmo tempo, há a convergência dos serviços (Fagerjord, 2010) e a transformação para o modelo de plataformas e sua infraestrutura de coleta e armazenamento de dados. Esses dados passam pelo processo de mediação algorítmica e são absorvidos como orientador da atividade jornalística ou como fonte de informação.

Neste último caso, o dado como fonte é considerado mais acessível quando pode ser coletado diretamente ou manipulado por *softwares* de uso generalizado como o *Google* planilhas e o *Excel*, e menos acessível quando o ultrapassa o limite de análise desses softwares, sendo chamado de *big data*, necessitando de habilidades da computação, como uso de linguagem de programação. Tal habilidade estaria associada a uma abordagem do Jornalismo de Dados.

No entanto, o jornalista acessa os dados em estágios diferentes (na manipulação ou reprodução), também podendo construí-los a partir de sua própria coleta. O exame para uso desses dados ainda permanece circunscrito ao processo de produção, portanto, a pauta, o tempo para apuração e a habilidade em tratar desses dados influenciam qual dos três eixos será aplicado.

Entre as limitações do nosso percurso, fica como desdobramentos futuros o ensejo em abordar outras variáveis menos tecnicistas, mas igualmente importantes. Entre elas, entende-se que a adoção dos “dados” como fonte traz uma noção de objetividade ligada à tradição positivista do Jornalismo de Precisão (Meyer, 1991), pois superaria a declaração de fontes humanas ao oferecer métodos claros e reprodutíveis (Träsel, 2014). A partir disso, pode-se discorrer sobre a natureza do dado acessado pelo jornalismo e qual interesse político

e econômico por trás da infraestrutura que armazena esses dados e do algoritmo que manipula.

REFERÊNCIAS

AR, M Fikri. Trends in examining data journalism research: a systematic literature review in 2010-2023. **Masyarakat, Kebudayaan Dan Politik**, [S.L.], v. 37, n. 1, p. 74-93, 29 jul. 2024. Universitas Airlangga. <http://dx.doi.org/10.20473/mkp.v37i12024.74-93>.

BARBOSA, Suzana. Jornalismo convergente e continuum multimídia na quinta geração do jornalismo nas redes digitais. p 33-54. In: CANAVILHAS, João. **Notícias e Mobilidade: O Jornalismo na Era dos Dispositivos Móveis**. 2013

BARCELLOS, Caco. **Rota 66: a história da polícia que mata**. 19. ed. Rio de Janeiro: Record, 2018.

COHEN, Sarah; HAMILTON, James T.; TURNER, Fred. Computational journalism. **Communications Of The Acm**, [S.L.], v. 54, n. 10, p. 66-71, out. 2011. Association for Computing Machinery (ACM). <http://dx.doi.org/10.1145/2001269.2001288>.

ERKMEN, Özlem. Data Journalism: a systematic literature review. **Journalism Studies**, [S.L.], v. 25, n. 1, p. 58-79, 6 dez. 2023. Informa UK Limited. <http://dx.doi.org/10.1080/1461670x.2023.2289885>.

FAGERJORD, Anders. After Convergence: YouTube and Remix Culture. In J. Hunsinger, L. Klastrup, & M. Allen (Eds.), **International Handbook of Internet Research**: Springer, 2010, p.187-200. Disponível em: https://www.researchgate.net/publication226034248_After_Convergence_YouTube_and_Remix_Culture

GEHRKE, Marília. **O uso de fontes documentais no jornalismo guiado por dados**. 2018. 130 f. Dissertação (Mestrado) - Curso de Faculdade de Biblioteconomia e Comunicação. Programa de Pós-Graduação em Comunicação e Informação, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2018.

GEHRKE, Marília. **Os elementos da transparência no Jornalismo Guiado por Dados**. Florianópolis: Insular, 2022.

GEHRKE, Marília; BENETTI, Marcia. O TWITTER COMO FONTE NO JORNALISMO GUIADO POR DADOS. **Brazilian Journalism Research**, Brasília, v. 16, n. 3, p. 1-24, dez. 2020.

KALSING, Janaína. **Jornalistas metrificados e a plataformação do Jornalismo". Programa de Pós-graduação em Comunicação (Dissertação)**, Porto Alegre: Universidade Federal do Rio Grande do Sul, 2021.

LANOSGA, Gerry; CAMAJ, Lindita; MARTIN, Jason. Testing a Typology of Data Journalism: professional and epistemological orientations across and between global regions. **Journalism Practice**, [S.L.], p. 1-20, 28 ago. 2024. Informa UK Limited. <http://dx.doi.org/10.1080/17512786.2024.2394564>.

LEMONS, André. Dataficação da vida. **Civitas - Revista de Ciências Sociais**, [S.L.], v. 21, n. 2, p. 193-202, 24 ago. 2021. EDIPUCRS. <http://dx.doi.org/10.15448/1984-7289.2021.2.39638>.

LEMONS, A. Plataformação, dataficação e performatividade algorítmica (PDPA): Desafios atuais da cibercultura. In Prata, Nair; Pessoa, Sonia C. (Orgs). **Fluxos Comunicacionais e Crise da Democracia**. São Paulo: Intercom, 2020, pp 117-126.

LIMA JUNIOR, Walter Teixeira. Big Data, Jornalismo Computacional e Data Journalism: estrutura, pensamento e prática profissional na Web de dados. **Estudos em Comunicação**, n. 12. Covilhã, Portugal, 2012

MACHADO, Elias. O ciberespaço como fonte para os jornalistas. **Biblioteca Online de Ciências da Comunicação**, Covilhã, p. 1-11, jun. 2002.

MEYER, Philip. **The New Precision Journalism**. United States Of America: Indiana University Press
Bloomington & Indianapolis, 1991.

MORINI, Francesca (2025) Data Journalism as “Terra Incognita”: Newcomers’ Tensions in Shifting Towards
Data Journalism Epistemology, **Journalism Practice**, 19:1, 76-92, DOI: 10.1080/17512786.2023.2185656

MOTTA, Suelen Homrich. O ecossistema midiático em convergência: hashtags e memes no discurso televisivo.
Tese, UFPR, 2024. Capítulo 2- O ecossistema midiático na pós-convergência. Disponível em:
<https://acervodigital.ufpr.br/handle/1884/94054>

MTCHEDLIDZE, Junai. Technical Expertise in Newsrooms: understanding data journalists’ roles and
practices. **Journalism And Media**, [S.L.], v. 5, n. 3, p. 1316-1328, 13 set. 2024. MDPI AG.
<http://dx.doi.org/10.3390/journalmedia5030083>.

NEGROPONTE, N. Beyond digital. **Wired**, 6.12. 1998. Acesso em: 1 fev 2026.

RAMÍREZ, Rubén. Post-Convergent Mediatization: Toward a Media Typology Beyond Web 2.0. **Mediatization
Studies**, v. 4, 2013, p. 9-23. Disponível em: <https://journals.umcs.pl/ms/article/view/11449>

RAMSÄLV, Amanda; EKSTRÖM, Mats; WESTLUND, Oscar. The epistemologies of data
journalism. *New Media & Society*, [S.L.], v. 26, n. 11, p. 6307-6324, 22 jan. 2023. SAGE
Publications. <http://dx.doi.org/10.1177/14614448221150439>.

ROGERS, Simon. **Data journalism at the Guardian**: what is it and how do we do it? London: The Guardian,
26 set. 2011. Disponível em:
<https://www.theguardian.com/news/datablog/2011/sep/26/data-journalism-guardian>. Acesso em: 10 fev. 2026.

SALAVERRÍA Ramón; AVILÉS, José Alberto García; MASIP, Pere. Concepto de convergencia periodística. In:
GARCÍA, Xosé López; FARIÑA, Xosé Pereira. (Eds.). **Convergência digital**: reconfiguración de los medios de
comunicación en España. Santiago de Compostela: Universidade, Servizo de Publicacións e Intercambio
Científico, 2010. p.41-64.

SILVA, Nathalia Lopes Da. Reconfigurações da noticiabilidade na pós-convergência: manifestações residuais no
Jornal Correio do Estado (MS). Tese, UFMS, 2025.

SILVEIRA, S.A. Tudo sobre todos: redes digitais, privacidade e venda de dados pessoais. São Paulo: Edições
Sesc São Paulo, 2017.

TRÄSEL, Marcelo Ruschel. **ENTREVISTANDO PLANILHAS**: estudo das crenças e do ethos de um grupo de
profissionais de jornalismo guiado por dados no Brasil. 2014. 315 f. Tese (Doutorado) - Curso de Programa de
Pós-Graduação em Comunicação Social, Pontifícia Universidade Católica do Rio Grande do Sul., Porto Alegre,
2014.

VALENTE, J. A emergência dos monopólios digitais: concentração e diversidade na internet no Brasil. In:
Bandeira, O.; Pasti, A.; Mendes, G. (Orgs.). Quem controla a mídia? Dos velhos oligopólios aos monopólios
digitais. São Paulo: Veneta; Coletivo Intervezes, 2023

WINQUES, Kérley; LONGHI, Raquel Ritter. Dos meios às mediações (algorítmicas): mediação, recepção e
consumo em plataformas digitais. **Matrizes**, [S.L.], v. 16, n. 2, p. 151-172, 2 set. 2022. Universidade de São
Paulo. Agência de Bibliotecas e Coleções Digitais. <http://dx.doi.org/10.11606/issn.1982-8160.v16i2p151-172>.

ZINS, Chaim. What is the meaning of "data", "information", and "knowledge"?, 2007



ZINS, Chaim. Conceptual approaches for defining data, information, and knowledge. **Journal Of The American Society For Information Science And Technology**, [S.L.], v. 58, n. 4, p. 479-493, 22 jan. 2007. Wiley. <http://dx.doi.org/10.1002/asi.20508>.