
Características, Competências e Processos do Jornalismo de Dados – Um estudo Sobre o Projeto (M)Dados¹

Matheus Costa NUNES²
Luciane Fassarella AGNEZ³
Centro Universitário Iesb

Resumo

Este trabalho pretende analisar notícias publicadas pelo projeto (M)Dados do jornal nativo digital, Metrôpoles. O foco da pesquisa é discernir os processos de apuração de peças jornalísticas que são compreendidas como Jornalismo de Dados, assim como as características e habilidades do profissional que se identifica como jornalista de dados. A análise será feita por meio do estudo de caso, análise de conteúdo e entrevista em profundidade. A fundamentação teórica foi elaborada para compreender como o Jornalismo Digital evoluiu para o Jornalismo de Dados a partir da cultura da convergência e da cibercultura.

Palavras-chave

Jornalismo de Dados; Jornalismo Digital; Convergência; Apuração Jornalística; Banco de Dados.

Introdução

O presente trabalho busca compreender as habilidades profissionais, os processos produtivos e as características de formato do Jornalismo de Dados (JD), compreendido como uma nova técnica de apuração (BRADSHAW, 2014) que agrega aspectos do Jornalismo Digital e do Jornalismo de Precisão. A proposta apresentará uma discussão teórica sobre a evolução do jornalismo no meio digital, as características dos formatos veiculados na web, assim como as novas habilidades e competências necessária para a produção do jornalismo com uma perspectiva guiada por dados. A rotina produtiva dessa área une técnicas de apuração jornalística à programação, técnicas de computação, recursos estatísticos, princípios de *accountability*⁴ e explora múltiplos formatos de visualização da informação.

¹ Trabalho apresentado no GP Conteúdos Digitais e Convergências Tecnológicas, XX Encontro dos Grupos de Pesquisas em Comunicação, evento componente do 43º Congresso Brasileiro de Ciências da Comunicação.

² Recém-graduado em Comunicação Social com habilitação em Jornalismo pelo Centro Universitário Iesb, e-mail: matheuscnx@gmail.com

³ Orientadora do trabalho, professora do curso de Comunicação Social do Centro Universitário Iesb, e-mail: luagnez@gmail.com

⁴ Disponível em: <https://www.oxfordreference.com/view/10.1093/oi/authority.20110803095346383>. Acesso em 24 de jun de 2020.

A fim de aproximar a conceituação da prática, foi escolhida uma iniciativa do jornal brasiliense Metrópolis, chamada (M)Dados, que se propõe a fazer apurações em bancos de dados para produzir reportagens, matérias especiais e para dar apoio de apuração a outras editorias do portal de notícias. A fim de compreender as características do conteúdo, foi feita a análise de peças veiculadas na página do (M)Dados. Para entender as habilidades e a formação de um profissional que pratica essa modalidade de jornalismo, foi realizada uma entrevista em profundidade com Lucas Marchesini, o membro do projeto responsável pela apuração em banco de dados.

Entretanto, antes de chegar ao estudo do caso (M)Dados, é necessário estudar os aspectos que compõem o JD, enquanto um tipo de conteúdo e um conjunto de técnicas, e o Jornalismo Digital, enquanto um meio. Em 2014 foi publicado o *The Data Journalism Handbook* (GRAY et al) a fim de propor quais são as características e boas práticas do JD. A forma simplista de definir tal área, segundo a obra, seria dizer que é “jornalismo feito com dados”. Porém, a definição é insuficiente em razão do jornalismo convencional já usar dados em suas apurações (BRADSHAW, 2014), sendo que os veículos tradicionais também estão presentes em um ambiente de volumosa troca de dados como a internet (SANTOS, 2016).

Segundo Mancini e Vasconcello (2016, p. 76-77), o JD pode ser quebrado em três dimensões: "Dimensão Investigativa, Dimensão Interpretativa e Dimensão Comunicativa". A Dimensão Investigativa diz respeito às técnicas de apuração, a dimensão interpretativa se refere à valorização da precisão dos dados dentro da estrutura textual, enquanto a dimensão comunicativa corresponde aos formatos de apresentação do conteúdo, o que outros autores como GRAY et al (2014) definem como visualização. Dentro dessas dimensões encontra-se a união de princípios como *open government*, *open source* (FLEW et al), Reportagem com o Auxílio de Computadores - RAC (MAYER, 1991), *hacking journalism* (LIMA, 2016), web colaboração (MICHALSKY et al, 2010) e técnicas estatísticas diversas.

Já para Coddington (2014), o JD deve ser compreendido em quatro dimensões: a experiência jornalística, somada a técnicas de apuração de reportagens, análise de amostras quantitativas com uso de software, similares às técnicas oriundas do Big Data, mais a compreensão do jornalista de um público ativo e que participa do processo de produção e interpretação da notícia.

Na quarta dimensão proposta por Coddington (2014) pode ser enquadrado o exemplo do portal de notícias americano *Five Thirty Eight*, que hospeda uma página chamada *Data* onde os leitores podem fazer *downloads* de todos os dados brutos e apurados pela redação, além de poderem visualizar o código dos softwares desenvolvidos pela equipe. Não só as informações estão abertas para a checagem, como os leitores e colaboradores são convidados a analisarem as fontes dos dados para criarem artes e novas histórias. Essa cultura conhecida como *open source*, ou seja, disponibilizar o código de softwares desenvolvidos, combinada com web colaboração (MICHALSKY *et al*, 2010) são marcas fundamentais do público ativo que Coddington (2014) define como a quarta dimensão do JD.

Em suma, a convergência de diversos princípios institucionais e de práticas da web-cultura culminou no advento de uma certa economia de dados. Portanto, resta compreender: como o Jornalismo de Dados é feito, quem é o profissional apto a executá-lo e como ele evoluiu a partir do Jornalismo Digital e do Jornalismo em Base de Dados?

1.1. A tecnologia e o profissional

Novas ferramentas e técnicas de transmissão de informação alteram as formas de comunicação e de produção de cultura. Por exemplo, o advento da tecnologia elétrica é para McLuhan (1998) a prolongação do sistema nervoso central. A experiência individual e social em torno de novas plataformas de disseminação de informação, age como uma violenta mudança para o sistema nervoso central. A força que impulsiona essa mudança é, segundo Sibilla (2012), o sentimento do homem pós-orgânico de insuficiência perante a realidade, que resulta em uma busca para se complementar com a técnica.

Já Haraway (2016) vai mais adiante e afirma que o ser humano é parte de um subsistema, uma simbiose entre orgânico e inorgânico que estabelece uma relação de codependência. A relação entre controle, homem e máquina se mantém pelo uso da linguagem, mas não há um patamar mais elevado para nenhum dos três, pois todos se encontram juntos e concatenados. Segundo a teoria de Haraway (2016), os problemas na comunicação, ou na ciência da comunicação, podem ser compreendidos hoje como problemas de codificação da linguagem. Assim como para o Jornalismo de Dados a organização das informações passa primeiro pela programação, para depois ser interpretada pelo jornalista na forma de uma história (GRAY *et al*, 2014).

Para entender o homem que vive, e trabalha, dentro de um paradigma cibernético, é importante compreender o conceito. Cibernético vem do grego *kubernetikós* e significa aquele que pilota, ou próprio de quem conduz. A ideia foi revisitada em 1948 pelo matemático Norbert

Wiener (apud KIM, 2004) que a redefiniu como um modelo de inputs e outputs que são semelhantes a seres vivos e máquinas. De acordo com Kim (2004), Wiener acreditava que a cibernética abarca uma teoria de mensagens (ou seja, câmbio de informações), que carrega modelos matemáticos similares entre seres orgânicos, tecidos sociais e máquinas.

Tanto homem quanto máquina partilham do conjunto das representações que construíram no ambiente digital segundo a lógica da linguagem no ciberespaço (KIM, 2004). Para a máquina, a forma é o conteúdo (BARTHES, 2012) e para o homem os símbolos ali presentes carregam também uma camada de interpretação e significação.

1.2. Características e recursos do Jornalismo de Dados

É no ambiente digital, onde a linguagem é transformada em código e a colaboração homem-máquina é feita, que o jornalista de dados encontra a sua fonte mais valiosa de histórias. O Jornalismo de Dados, ou Jornalismo Guiado por Dados se caracteriza pela “produção, tratamento e cruzamento de grandes quantidades de dados, de modo a permitir maior eficiência na recuperação de informações, na apuração de reportagens a partir de conjuntos de dados, na circulação em diferentes plataformas” (TRÄSEL, 2014, p.292).

Uma possibilidade que é frequentemente explorada em produtos do JD é o uso de números e estatísticas no lide, para dar credibilidade à informação. O trabalho do jornalista com base de dados permite que estimativas inéditas sejam geradas e que informações qualitativas recebam um respaldo quantitativo (GRAY, et al, 2014).

Os números agem na notícia para simular um sentido de verdade ao perseguir a objetividade (BULLA, 2015). A facilidade que esses signos possuem de representar quantidades os tornam úteis em dialogar com a língua e trazer embasamento para discursos. Para Meyer (2002), a nossa atenção tende a ser atraída pelos aspectos dos problemas que se expressam em números e que, portanto, eles têm a capacidade de criar uma aparência de importância sobre a questão. O uso de números na narrativa jornalística também é visto como uma tentativa de se apropriar do respaldo da ciência, que com frequência usa teorias probabilísticas nas argumentações, a fim de adicionar mais credibilidade e objetividade às reportagens.

O JD também herda algumas de suas características do Jornalismo Digital, a partir da cibercultura (MACHADO, 2005), sendo elas quatro facetas que se encontram interligadas (BARDOEL, 2001): Interatividade, customização do conteúdo, hipertextualidade e multimídia. A interatividade entre usuário e jornalista é essencial, por sua vez, para garantir a customização do conteúdo consumido. Já a hipertextualidade, de acordo com Mielniczuk (2003), permite que uma mídia se mantenha conectada a uma outra mídia, ou a um repositório de arquivos, a um banco de dados e assim por diante. Enquanto a multimídia permeia não só o formato do conteúdo, mas também a apuração feita online e em banco de dados.

Segundo Machado (2005), no início do Jornalismo Digital as bases de dados reuniam informações textuais sobre os usuários, mas hoje acumulam conexões entre imagens, texto e georreferenciamento, abrindo espaço para uma convergência que contribui para apurações mais complexas. Não só as formas de pesquisa se tornam, mais complexas, mas as formas visualização da informação também ganham requinte e diversidade (GRAY et al, 2014).

Os recursos de apuração do Jornalismo de Dados, segundo Marcelo Trasel (2014), munem o profissional de imprensa de dois recursos sem precedentes. Primeiro, o jornalista adquire a habilidade de encontrar informações com valor notícia em meio a milhares, ou até mesmo milhões, de registros. Tal trabalho sem a assistência de computadores seria impossível, ou iria requerer muito tempo. Segundo, o apurador de histórias possui a capacidade de cruzar diferentes bases de dados, para a produção de um conhecimento original e inédito. Segundo a obra *The Data Journalism Handbook* (GRAY et al, 2014), o banco de dados usado para gerar a peças jornalística pode ser moldado partir das necessidades da apuração, mas sem retirar a oficialidade delas.

O banco de dados pode ser visto pelo jornalista contemporâneo como a matéria prima para a geração de novas histórias. Dentro da base de dados, o jornalista pode definir padrões orientados a objetos (MACHADO, 2005) e considerar ou eliminar as conexões estabelecidas dentro desses microcosmos. O ponto de partida para essas apurações é encontrar e ter acesso de bases já existentes de governos e instituições privadas.

Segundo Coddington (2015, p. 334), a nova habilidade que o jornalista deve adquirir para significar o ciberespaço “é a expertise de encontrar elementos narrativos em meio a volumes de dados”. Outrora essa expertise era relativamente limitada a analistas de dados, estatísticos e cientistas sociais, mas a democratização dessas técnicas e métodos permite aos profissionais de imprensa “cavar” dentro desses espaços em busca de insumos para reportagens. A prática do Jornalismo de Dados se configura como uma união desses métodos ao jornalismo investigativo profissional, a cultura *open source*, e as características formais do Jornalismo Digital, mas sem excluir a importância de destrezas do jornalismo tradicional, como entrevistar, cruzar fontes e contar histórias.

De forma prática, o JD envolve a execução quatro etapas, segundo Bradshaw (2014): encontrar os dados, interrogar os dados, *data mashing* (método que envolve tornar os dados mais heterogêneos) e visualizar os dados. As três primeiras etapas envolvem aplicar um filtro sobre o volume de dados, enquanto o último diz respeito a representação das informações, por meio da plethora de formatos que a multimídia digital nos dispõe.

É notável, portanto, a característica do JD como um conjunto de técnicas, cultura web, recursos computacionais e valores norteadores da profissão jornalística. A aplicabilidade desse

conjunto de métodos visa gerar *insights* e, possivelmente, pautas, em cima de múltiplas fontes de informações em conjuntos densos de dados.

2. Metodologia

O estudo de caso foi escolhido como primeiro recurso metodológico; antes de tudo em razão da abrangência de análise que ele comporta; segundo pela possibilidade de tratar do projeto (M)Dados (ou Núcleo de Processamento de Grande Volumes de Dados do Metrôpoles) e do trabalho de Lucas Marchesini simultaneamente; terceiro por permitir análises quantitativas e qualitativas; quarto por possibilitar o estudo de um caso sem isolá-lo do contexto mais amplo (YIN, 2010), do advento jornalismo de dados; e por último, pela amplitude temática que o (M)Dados possui ser compatível com a amplitude temática que o método suporta. Todos esse processos serão necessários para observar que a missão que o próprio Metrôpoles impõe ao (M)Dados está sendo alcançada efetivamente.

O segundo recurso metodológico eleito para o objeto de estudo foi a análise de conteúdo por causa dos seguintes motivos: a possibilidade de aplicar análises quantitativas e qualitativas; a flexibilidade de analisar um mesmo conteúdo com diferentes critérios de seleção; a possibilidade de observar o conteúdo em aspectos temporais, temáticos e técnicos; a capacidade de fazer uma leitura qualitativa profunda das reportagens selecionadas, a perspectiva de contabilizar aspectos quantitativos das peças jornalísticas, interatividade, número de fontes, base de dados e personagens usados nas peças escolhidas; e por último a chance de distinguir o que é que marca o conteúdo do jornalismo de dados, daquele que não é.

Os aspectos analisados de cada peça do conjunto foram: tema, recursos visuais, número de dados, número de hyperlinks, banco de dados usados, fontes ouvidas e base de dados disponível no corpo da reportagem. A temática foi observada a fim de compreender com quais editoriais o JD consegue dialogar melhor e também para avaliar se a perspectiva dos dados se aplica em temáticas diversas.

Os recursos visuais que estão presentes nas reportagens foram analisados a fim de observar a multimidialidade dos produtos jornalísticos, uma das características centrais do jornalismo (MIELNICZUK, 2003) e que foi incorporada pelo JD (GRAY et al, 2014). Nessa categoria de análise também é possível observar quais são os formatos que mais se adequam à representação de dados.

O número de hyperlinks em cada reportagem foi contabilizado a fim de observar o nível de interatividade das peças. Esse recurso é frequentemente usado para criar uma ponte entre informações reportadas e suas fontes, ou para contextualizar a reportagem dentro de outra. Portanto, um amplo uso de hyperlink pode indicar o quanto o produto jornalístico busca engajar o leitor com o assunto.

Foi contemplado o número de fontes e personagens usados em todas as publicações por dois motivos: para compreender a importância que uma fonte humana possui em apurações que se guiam pelo uso de fontes documentais e para avaliar o grau de polifonia, ou seja, de pluralidade de vozes.

Os bancos de dados que foram usados foram também observados a fim de se ter uma noção de quais são esses recursos e suas origens (público ou privado). Como o banco de dados se constitui como a matéria prima para o JD, é importante notar quais são as instituições que os disponibilizam.

Por ultimo, o terceiro recurso metodológico usado foi a entrevista em profundidade, com o intuito de investigar as habilidades (DUARTE, 2005) que Lucas Marchesini possui como jornalista de dados; conseguir relatos sobre a rotina produtiva do jornal e quais são as ferramentas operadas e como elas são operadas.

Tabela I - Conjunto para análise de conteúdo

Título	Data de Publicação
Mega-Sena: veja um raio-X do prêmio e os números mais sorteados	11.09.2020
No MPDFT, valor médio do contracheque em agosto foi de R\$ 40 mil	17.09.2020
Friends 25 anos: aquele com vários dados curiosos	22.09.2020
A VOZ DO PRESIDENTE: SAIBA O QUE SE PASSA NA CABEÇA DE JAIR BOLSONARO	27.09.2020
Roda presa: uso de bicicletas compartilhadas do GDF cai pela metade	06.10.2020
Em 7 anos, servidor do Senado embolsou R\$ 406 mil em horas extras	14.10.2020
Interesse em K-pop cresce quase 485% nos últimos 2 anos no Brasil	19.10.2020
Motoristas do DF têm a maior média de multas de trânsito do país	26.10.2020

Senadores gastam 3 vezes mais que deputados em viagens oficiais	02.11.2020
Quase 90 milhões de brasileiros vivem em cidades sem cinema	06.11.2020
R\$ 19 mi: veja quanto cada parlamentar gastou em aluguel de carro	10.11.2020
TCDF tem média salarial mais alta entre servidores: R\$ 25,4 mil	18.11.2020
Brasilienses aderem à campanha pelo fim da violência de gênero	23.11.2020
Marília Mendonça ou Anitta: quem é a mais popular do Brasil?	30.11.2019
Quais as profissões que mais sofrem acidentes de trabalho no DF?	03.12.2020
BPC: 480 mil pessoas só conseguem benefício após ações judiciais	09.12.2020
(M)dados: Bolsonaro nomeou ao menos 660 filiados do Centrão	15.12.2020
No 1º ano, Bolsonaro aprovou menos MPs que Dilma, Lula e Temer	23.12.2020
Presidente tuiteiro: um ano do governo Bolsonaro na rede social	31.12.2020
Sertanejo domina as mais tocadas do Brasil em 2019 no Spotify	10.01.2020
Poder e fé: deputado evangélico fala mais de Deus do que católico	16.01.2020
Bolão do Oscar: veja como aumentar chances de vencer	19.01.2020
Educação no DF: problemas de saúde causaram 56,9 mil afastamentos	28.01.2020
Senado: gastos com despesas médicas explodem e chegam a R\$ 14 mi	02.02.2020
Oscar 2020: os números mostram quais serão os vencedores	09.02.2020
Em mês de recesso, Senado gastou R\$ 567 mil em horas extras	16.02.2020
Neymar? Messi? CR7? Como Haaland se compara com outros craques	24.02.2020
Consumo de álcool causa mais de 16 mil mortes por ano no Brasil	01.03.2020
Li o que você apagou. Os tweets deletados da gestão Bolsonaro	10.03.2020
Senado mantém R\$ 29 milhões em gastos anuais sem comprovação	15.03.2020
Covid-19: Ministério da Saúde divulga menos casos que secretarias	24.03.2020
Bolsa Família tem a maior fila de espera desde a crise de 2015	01.04.2020
Covid-19 se interioriza e chega a 581 cidades. Veja quais	08.04.2020
Coronavírus: 164 cidades do país recorrem ao DF para internações	16.04.2020

3.1. Resultados da análise de conteúdo

Das 34 peças jornalísticas analisadas, 31 delas apresentaram 191 hiperlinks. Observa-se uma média de 6 hipertextos por reportagem. Em tal quesito, as produções do (M)Dados se mostraram de acordo com este aspecto do jornalismo digital. O recurso da

hipertextualidade cumpre quatro funções dentro dos produtos jornalísticos. Primeiro, ele cria vínculos entre a reportagem e a página de instituições que são mencionadas na apuração, muitas vezes sendo as instituições que forneceram os dados para a apuração da matéria. Os vínculos também são feitos com reportagens progressas que abordam o assunto da pauta, a fim de munir o leitor com mais informações a respeito do tema, ou de apresentar em qual contexto a reportagem se interfere. Terceiro, os links de outros conteúdos que foram hospedados em plataformas como *Youtube*, *Spotify* e *Twitter* permite ao leitor ter contato com outros formatos midiáticos. Este terceiro aspecto também é bastante semelhante ao segundo, pois contextualiza a reportagem dentro de um universo midiático mais amplo, mas também amplia a estética do produto jornalístico ao integrar formatos diversos (áudio, texto e vídeo). Por último, os hipertexto da reportagem em si permite que os conteúdos (M)Dados sejam localizáveis pelos mecanismos de busca e compartilháveis por leitores em outras plataformas, como redes sociais e e-mail.

A partir da análise do emprego do hipertexto nas peças jornalísticas do (M)Dados, observa-se a recorrência dos quatro aspectos do jornalismo digital (MIELNICZUK, 2004): interatividade, memória, hipermedialidade e conectividade. A editoria analisada surgiu nativamente no meio digital e, portanto, apresenta todas as características necessárias para ser enquadrada como um produto do jornalismo digital.

Uma das características que a hipermedialidade confere ao conteúdo web é a possibilidade de explorar diferentes estilos e formatos dentro de uma mesma página (CANAVILHAS, 2015). Neste quesito, o (M)Dados também apresentou as características de um produto com variedade midiática e estilística. Das 34 matérias, oito possuíam apenas fotos como suporte visual. Vinte e quatro matérias apresentavam mais de um tipo de recurso visual. Vinte matérias apresentaram infográficos, sendo nove deles infográficos interativos. Ao todo foram dez recursos visuais encontrados, entre slideshows, cards, prints, tabelas, vídeos, ilustrações, mapas interativos, infográficos e infográficos interativos.

A incidência desses formatos indica que a maioria dos produtos do (M)Dados investem em uma das características centrais do Jornalismo de Dados, segundo Gray et al (2014): a visualização. Conforme discutido no capítulo 1.2, as “formas diversas” de visualização são um dos aspectos mais distintos do JD. A hipermedialidade se mostra aqui tanto como uma marca do jornalismo digital mas também do JD, por tornar a visualização

da informação, com diferentes recursos imagéticos, um dos objetivos principais da reportagem.

De acordo com Holovaty (2006), a visualização dos dados serve para dar camadas de atribuição à história, ou seja, a variedade de fotos, misturadas com infográficos, vídeos e outros, permite que o leitor observe o conteúdo a partir de perspectivas distintas e atribua significações diferentes aos dados. Percebe-se também que o uso múltiplo de recursos visuais atribui às peças um certo didatismo, ou seja, um intuito de transmitir e explicar a informação por vários caminhos.

Neste ponto, a editoria analisada também evidenciou estar em consonância com as características do JD expressas na fundamentação teórica. Para Bulla (2015), a priorização de números na narrativa jornalística serve para atribuir ao produto jornalístico um aspecto de objetividade e credibilidade encontrado em trabalhos científicos, que frequentemente usam de recursos estatísticos e matemáticos para embasar argumentos. Ao empregar números com frequência nas reportagens, o (M)Dados visa dar respaldo às denúncias e atribuir maior credibilidade aos argumentos apresentados nas matérias.

Mas enquanto reportagens buscam ganhar respaldo e credibilidade com entrevistas e afirmações de especialistas, autoridades e personagens colocadas no corpo do texto, os conteúdos do (M)Dados parecem se embasar prioritariamente nos dados coletados nas apurações em bancos de dados. Apenas 16 das 34 matérias apresentam aspas de fontes, com o total de 30 fontes entrevistadas. É como se a base de dados falasse em nome das instituições que as organizaram e o jornalista de dados agisse como um entrevistador. A forma como os dados assumem a centralidade no produto jornalístico, ocupando o local que poderia ser assumido por um especialista no assunto, mas sem excluir a possibilidade de ambos coexistirem, corrobora com a visão de Gray et al (2014) da oficialidade que a informação extraída de um banco de dados carrega. Em um primeiro momento, pode parecer que o jornalista de dados gera a própria informação que reporta, quando na verdade ele age como vetor da informação por meio dos processos de limpeza, estruturação e análise dos dados.

Quanto à forma que o jornalista de dados do projeto encontra para creditar a informação encontrada, nota-se que 32 das 34 matérias especificam de onde os dados foram retirados, muitas vezes por meio do nome da instituição que disponibilizou os dados acompanhado de um hyperlink.

Nos casos em que mais de uma base de dados foi usada, o corpo da matéria explica que as informações possuem fontes distintas e que tais bases foram cruzadas, sempre mencionando o nome das bases de dados, ou do portal das instituições.

Ao analisar o crédito da fonte das informações, nota-se que 17 casos, ou seja, por volta da metade dos bancos de dados utilizados são mantidos por instituições públicas. O uso recorrente de bases de dados públicas reitera a importância dos princípios de transparência, *open government* e *accountability* que foram apresentados no capítulo 4.3 e 4.6. A disponibilização de informações de entidades públicas se faz essencial para que o (M)Dados encontre os insumos para as apurações. E o uso analítico e recorrente dessas bases de dados também marca uma característica em comum entre o (M)Dados e outras iniciativas de jornalismo de dados apresentadas na fundamentação teórica. Também foram utilizadas como fonte para as apurações plataformas e mídias sociais (6), iniciativas de dados abertos da sociedade civil e outros (3).

O último aspecto revelado pela análise de conteúdo foi a variedade temática que os conteúdos do (M)Dados cobrem. Ao todo foram 8 temas identificados, sendo eles: economia, política, entretenimento, transporte, saúde, esporte, educação e sociedade. A temática mais explorada foi política (15) sendo que 8 delas focam em assuntos relativos ao orçamento público. Nesse ponto, o projeto não parece se especializar em um tipo de cobertura, mas aproveita da variedade de dados e plataformas disponíveis para tornar a cobertura mais diversa.

3.2. Resultado da entrevista em profundidade

De acordo com Marchesini (2020), a apuração de uma reportagem de jornalismo de dados pode ser compreendida em 5 etapas: busca do banco de dados, limpeza, cruzamento de informação, checagem e redação. A primeira etapa, que é encontrar o banco de dados, costuma ser a mais simples. O jornalista procura em mecanismos de busca, ou em sites de instituições públicas que já são conhecidas por possuírem bases de dados públicas, ou empresas de tecnologia como Youtube, Spotify e serviços de streaming que costumam disponibilizar informações sobre as preferências do público.

A alta proporção de reportagens que utilizam base de dados governamentais, como foi evidenciado na análise de conteúdo, revela a importância da cultura *open government* para a produção do jornalismo de dados. A transparência na web se mostra

como elemento fundamental para o dia a dia na redação e o acesso das bases de dados serve de matéria prima para o jornalista de dados.

A segunda etapa é o que Marchesini (2020) define como a “limpeza” dos dados. Os bancos de dados veem abarrotados de informações, muitas das quais não serão úteis para a apuração e que, portanto, farão ruído no momento de postular a pergunta central da pauta para o banco de dados. Marchesini (2020) também define limpar uma base de dados como “estruturar”, ou seja, uma base de dados mal estruturada é aquela com excesso de ruído. Existe o caso em que a base de dados sequer existe em um conjunto, são apenas informações que podem ser encontradas sob o guardachuva de um endereço na web. Nesse caso, deve-se voltar uma etapa e realizar o *scrapping*, que consiste em elaborar um código capaz de arrancar desse endereço web informações que irão alimentar a estruturação de um banco de dados.

A terceira etapa é mais pontual e envolve de fato responder a pergunta, ou perguntas, que geraram a pauta. Essa pergunta é postulada ao banco de dados na forma de um cruzamento de informações. Esse cruzamento é feito em planilhas, ou se usa a linguagem de programação Python, dependendo da complexidade da informação que se busca.

A quarta etapa é um misto de checagem da informação encontrada no cruzamento e continuação da apuração, sendo basicamente uma consulta, ou entrevista com um especialista. “Antes de encaixar a informação no texto eu procuro alguém que entenda do assunto, para poder me falar se aquela informação parece plausível. Porque o número muitas vezes é só o começo de uma história. Às vezes ele é um fim em si mesmo, ainda mais quando você tem domínio do assunto”, explica.

É possível fazer um paralelo entre a quarta etapa e a baixa ocorrência de entrevistados nos textos, mostrada na análise de conteúdo (apenas 16 das 34 matérias apresentam entrevistas). Vemos aqui que a consulta com um especialista é usada a título de checagem. A informação da base de dados ocupa uma posição privilegiada no lide e segue sendo o recurso central que dá embasamento ao texto, aspecto que se corrobora na rotina produtiva de Marchesini (2020).

A quinta e última etapa é a redação da reportagem, ou a representação dos dados com imagens. Para escrever a reportagem permanece a técnica de pirâmide invertida e as normativas de manuais de redação. Marchesini (2020) revelou que costuma passar os dados direto ao departamento de arte do jornal, então não executa diretamente o processo

de elaborar o visual da reportagem. Em alguns casos, quando o departamento de arte se encontra muito atribulado, Marchesini prepara gráficos para as reportagens. E ele enfatiza a importância de manter contato com o departamento responsável para que as reportagens ganhem bons recursos estéticos. O uso de fotos, vídeos e infográficos evidencia a diversidade de formatos encontrados nas reportagens como um ponto central na rotina produtiva do (M) Dados.

Quanto às ferramentas usadas por Marchesini (2020) no cotidiano de seu trabalho, estão mecanismos de buscas para apuração, planilhas (*excel* e *google sheets*), programação em Python, acrescidas das bibliotecas que disponibilizam códigos prontos, como por exemplo a *Panda* e um programa para extrair informações de páginas web e arquivos PDF chamado *Open Refine*.

Para fazer a visualização dos dados, o jornalista comentou ter domínio da linguagem *Javascript* e do uso da biblioteca G3, que o habilita a criar infográficos do zero. Mas afirmou usar a ferramenta online *Flourish* no dia a dia para fazer conteúdos visuais pela praticidade que ela apresenta.

4. Considerações Finais

O trabalho de análise dos conteúdos da editoria (M)Dados, assim como a fundamentação teórica acerca do jornalismo de dados e das vertentes que o antecederam, permitiram observar a evolução e a confluência de diversas características do webjornalismo, assim como essa vertente evoluiu até o advento do jornalismo de dados. A multimídia, o uso de hiperlinks para contextualizar as histórias, o emprego de fontes documentais nas apurações, a explicação do método de apuração para o leitor e coberturas que mesclam eventos públicos e discussões nas mídias sociais.

Já a entrevista em profundidade permitiu observar praticamente como se estrutura a rotina de um jornalista de dados, quais são as técnicas usadas por esses profissionais e o que difere essa prática do jornalismo convencional. Os dois grandes diferenciais para a produção de reportagens com a perspectiva de dados é o tempo que se gasta procurando, estruturando e analisando um banco de dados, assim como concebendo recursos visuais para acompanhar o texto. A impressão final é que o trabalho de um jornalista tradicional complementa o de um jornalista de dados e vice-versa.

Um dos aspectos investigativos do jornalismo de dados que tornam a sua cobertura distinta é o de tentar dar sentido a grandes audiências de plataformas digitais e

debates em redes sociais. Existe nesse tipo de cobertura a possibilidade de uma apuração feita nas redes, para as redes, com profundidade. Vemos nesse tipo de apuração o estudo da repetição de termos, da audiência de conteúdos e da interação entre usuários. Esse tipo de análise se mostra como uma alternativa mais interessante para o jornalismo do que reportagens de jornalismo declaratório voltadas para personalidades nas redes sociais. Muitos repórteres atualmente transcrevem o que foi dito nas redes sociais por políticos e figuras influentes, sendo que se pode acessar a rede social diretamente para se ter essa informação.

A segunda conclusão que surgiu após a análise de conteúdo é a capacidade do jornalismo de dados de promover accountability nos gastos do setor público. Boa parte das bases de dados usadas são de instituições públicas e várias reportagens focam em aspectos do orçamento público. Como o jornalista de dados está habituado a apurar em planilhas e boa parte das informações de gastos do governo também estão contidas em planilha de acesso aberto, a apuração se mostrou eficaz em estudar e apontar discrepância nos gastos estatais. Considerando o valor-notícia que reportagens sobre orçamento carregam, e a capacidade do jornalismo de dados de identificar furos antes deles serem vazados por alguém, as técnicas do jornalismo de dados se mostraram formas eficientes de encontrar notícias quentes sem a informação ter que passar por um intermediário (autoridades, assessores e partes interessadas).

Referências bibliográficas

BARDOEL, Jo et al. **'Network journalism': converging competencies of old and new media professionals**. Australian journalism review, v. 23, n. 2, p. 91, 2001.

BARTHES, Roland. **Elementos de semiologia**. Editora Cultrix, 2012.

BRADSHAW, Paul. Data journalism. **Ethics for digital journalists: Emerging best practices**. University of London: 2014.

BULLA, Olívia. **As Relações de Comunicação e Trabalho no Jornalismo Evidenciadas pela Objetividade dos Números na Notícia**, 2015.

CANAVILHAS, João; BACCIN, Alciane Nolibos. **Contextualização de reportagens hipermídia: narrativa hipermídia e imersão**. Brazilian Journalism Research, v. 11, n. 1, p. 10-27, 2015.

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

DUARTE, Jorge. **Entrevista em profundidade. Métodos e técnicas de pesquisa em comunicação**.v. 1. São Paulo: Atlas, 2005..

FLEW, T.; SPURGEON, C.; DANIEL, A.; SWIFT, A. 2012. **The promise of computational journalism. Journalism Practice**, 6(2):57-171. Site: <http://dx.doi.org/10.1080/17512786.2011.616655>

GRAY, Jonathan; CHAMBERS, Lucy; BOUNEGRU, Liliana. **The data journalism handbook: how journalists can use data to improve the news**. O'Reilly Media, Inc, 2012.

HARAWAY, Donna J. **Manifestly**. U of Minnesota Press, 2016.

HOLOVATY, Andrian. **A fundamental way newspaper sites need to change. 6 de setembro de 2016**. Disponível em <<http://www.holovaty.com/writing/fundamental-change/>>. Acesso em: 21 de outubro de 2019.

KIM, Joon Ho. **Cibernética, ciborgues e ciberespaço: notas sobre as origens da cibernética e sua reinvenção cultural**. Horizontes antropológicos, v. 10, n. 21, p. 199-219, 2004.

MACHADO, Elias. **A Base de Dados como formato no Jornalismo Digital**. Actas do VII Lusocom. Covilhã: Universidade Beira Interior, 2005.

MCLUHAN, Marshall. **La galaxia gutenberg**. Círculo de Lectores, 1998.

MEYER, Philip. **Precision journalism: A reporter's introduction to social science methods**. Rowman & Littlefield Publishers, 2002.

MIELNICZUK, Luciana. **Jornalismo na Web: uma contribuição para o estudo do formato da notícia na escrita hipertextual**. 2003.

_____, Luciana. **Webjornalismo de Terceira Geração: continuidades e rupturas no jornalismo desenvolvido para a web**. In: CONGRESSO BRASILEIRO DE CIÊNCIAS DA COMUNICAÇÃO. 2004.

SANTOS, Daniela Filipa Real. **Jornalismo 3.0 e os agregadores de notícias online: Novas formas de produção e consumo de notícias em Portugal**. Universidade Católica Portuguesa, 2016.

SIBILIA, Paula. **O homem pós-orgânico: corpo, subjetividade e tecnologias digitais**. In: O homem pós-orgânico: corpo, subjetividade e tecnologias digitais. 2002.

TRÄSEL, Marcelo. **Entrevistando planilhas: estudo das crenças e do ethos de um grupo de profissionais de jornalismo guiado por dados no Brasil**. 2014. 314 f. 2014. Tese de Doutorado. Tese (Doutorado em Comunicação Social)—Faculdade de Comunicação Social, Pontifícia Universidade Católica do Rio Grande do Sul.

YIN, Robert K. **Estudo de Caso: Planejamento e métodos**. Bookman editora, 2015.