

UNIVERSIDADE CATÓLICA DE PERNAMBUCO
PROGRAMA DE PÓS-GRADUAÇÃO EM INDÚSTRIAS CRIATIVAS
MESTRADO PROFISSIONAL

JORNALISMO DE DADOS
e o redesenho algorítmico do processo de construção da notícia

ARLINE ALESSANA LIRA LINS

Recife

2020

ARLINE ALESSANA LIRA LINS

JORNALISMO DE DADOS

e o redesenho algorítmico do processo de construção da notícia

Dissertação apresentada ao Programa de Pós-Graduação da Universidade Católica de Pernambuco para obtenção do título de mestre em Indústrias Criativas.

Área de concentração: Indústrias Criativas. Linha: tecnologias, linguagens e produtos.

Orientador: Prof. Dr. Cláudio Roberto de Araújo Bezerra

Co-orientador: Prof. Dr. Anthony José da Cunha Carneiro Lins

Recife

2020

L759j

Lins, Arline Alessana Lira

Jornalismo de dados e o redesenho algorítmico do processo de construção da notícia / Arline Alessana Lira Lins, 2020

122 f. : il.

Orientador: Cláudio Roberto de Araújo Bezerra

Co-orientador: Anthony José da Cunha Carneiro Lins

Dissertação (Mestrado) - Universidade Católica de Pernambuco.

Programa de Pós-graduação em Indústrias Criativas. Mestrado

Profissional em Indústrias Criativas, 2020.

1. Jornalismo – Processamento de dados. 2. Tecnologia da informação.
3. Algoritmos. I. Título.

CDU 070

Luciana Vidal – CRB-4/1338

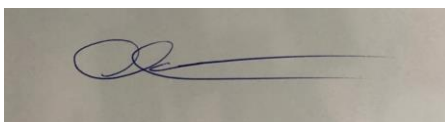
FOLHA DE APROVAÇÃO

ARLINE ALESSANA LIRA LINS

Trabalho de conclusão do mestrado apresentado ao Programa de Pós-Graduação em Indústrias Criativas – Mestrado Profissional da Universidade Católica de Pernambuco, com o título JORNALISMO DE DADOS E O REDESENHO ALGORÍTMICO DO PROCESSO DE CONSTRUÇÃO DA NOTÍCIA, submetida à banca examinadora, como requisito parcial para obtenção do título de mestre.

Data de aprovação – 24/04/2020

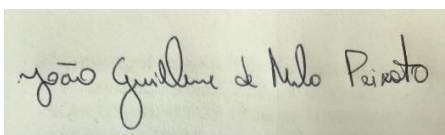
BANCA EXAMINADORA



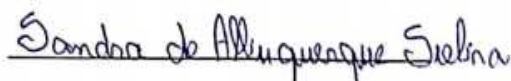
Prof. Dr. Cláudio Roberto de Araújo Bezerra (orientador e presidente da banca)
Universidade Católica de Pernambuco (UNICAP)



Prof. Dr. Anthony José da Cunha Carneiro Lins (co-orientador)
Universidade Católica de Pernambuco (UNICAP)



Prof. Dr. João Guilherme de Melo Peixoto (Titular Interno)
Universidade Católica de Pernambuco (UNICAP)



Profª. Drª. Sandra de Albuquerque Siebra (Titular Externa)
Universidade Federal de Pernambuco (UFPE)

Recife, 2020

DEDICATÓRIA

A Maria José Freire de Lira, vó, meu maior exemplo de ser humano. Você me ensinou sobre valores imutáveis e escassos. Para mim, ninguém conseguiu ser maior espelho de humildade e honestidade. Me provou que a recompensa não é vencer sempre, mas sobreviver. Acreditar que somos capazes de atravessar o impossível é o que nos move. Cada conquista da minha vida sempre vai carregar a herança de suas batalhas, de suas dores e de seu suor. E, mais que tudo, as vitórias da vida me farão lembrar que o amor precisa estar no centro de qualquer sacrifício para encontrarmos sentido na nossa jornada.

A Maria Aparecida de Lira, mainha. Você me prova, há muitos anos, que renunciar a si mesmo é sinal de amor. Cada noite mal dormida por causa da distância me fortaleceu. A dor da separação se tornou minha espada diária. A vida precisava encontrar rumo. E encontrou. Tentei, todos os dias, ser a melhor versão de mim pra justificar todos os sacrifícios que você fez para que eu me tornasse uma pessoa digna. Busquei serenidade diante dos momentos mais cruéis e, por esse motivo, consegui dar os passos necessários e chegar aqui. Espero que o título de mestre seja mais uma razão de orgulho e certeza de que você cumpriu e continua a cumprir sua missão.

A Vilma Lúcia de Lira, Pipa. Você é minha base de suporte, aquela que tem a mão estendida para me levantar mesmo antes que eu caia. Não existe nada neste mundo que seja capaz de expressar a gratidão que levo comigo por cada minuto que você passou ao meu lado e cuidou de mim (e ainda cuida). Deus é o único ser capaz de retribuir na medida em que você me ofereceu. Assim como vó e mainha, você é fundamentalmente responsável por cada conquista que eu já tive e que ainda vou ter enquanto viva estiver. Neste mundo, para mim, você sempre foi sinal de segurança e amparo, doação e luz. E sempre vai ser.

A Sofia, que apesar de não ser humana, guarda valores e sentimentos que me deram estrutura para chegar ao fim deste trabalho. Cada página escrita foi amparada por seu carinho e seu amor. Seu companheirismo me transformou como pessoa. Embora você não consiga ler esta mensagem, eu a escrevo para prometer a mim mesma que todo esse tempo ocupado com os estudos vão agora se tornar minutos, horas e dias que passaremos juntas.

AGRADECIMENTOS

A Suzanna Borba, pelo companheirismo e cumplicidade incansáveis que me fizeram seguir em meio a tantas adversidades e escolhas difíceis. Um dos meus maiores símbolos de força e determinação, soube me inspirar e me fazer reconhecer que eu podia mais do que sabia. Você é responsável, em grande parte, pela força que adquiri e pelo meu crescimento como pessoa.

A Cláudio Bezerra, pelo apoio emocional e orientação indispensável. Conselhos, puxões de orelha e, principalmente, compreensão para os momentos mais duros da rotina. A Anthony Lins, pelas brilhantes contribuições e observações ao longo da construção do trabalho. E, claro, pelas conversas que se transformaram – por vezes – em sessões terapêuticas.

A Angélica Tasso, que tanto me incentiva a ser mais e melhor, que encheu o coração de amor para me impulsionar a dar os primeiros passos no mestrado e que sempre se mostrou um modelo de profissional. A Marjorye Cavalcanti, que descobriu há tantos anos como me acolher, me instruir, me mostrar diferentes perspectivas e que soube celebrar cada pequena conquista a meu lado. Você me ensina todos os dias sobre coragem. A Jô Mazzarolo, que me deu espaço para criar, recriar e sugerir novos caminhos de trabalho com base nos meus estudos sem perder a convicção no meu potencial, me entusiasmando na medida certa.

A Leonardo Burgos e Carol Vasconcelos, que compartilharam das minhas dores e angústias do mestrado, me deram choques de realidade, e nunca se importaram em ser meus ombros de suporte quando precisei.

A Caroline Rangel, minha dupla acadêmica por quem nutro tanto carinho e respeito, por me guiar com conselhos de quem já galgou – brilhantemente – estes mesmos caminhos.

A Gabriela Cavalcanti e a Danielly Castro, por serem portos seguros. Por me darem sustentação nos diferentes terrenos que a vida impõe e serem, simplesmente, amor, lealdade e carinho.

A Mariana Clarissa e a Germana Macambira, pela companhia nos dias de aula, com risadas, jantares na Rua do Lazer, e sessões privadas de desespero e comédia.

"The incredible diversity of information available gives great credit to the creativity and ingenuity of information providers, and points to a very exciting future."

(Tim Berners-Lee)

RESUMO

Dados e jornalismo sempre estiveram conectados. O primeiro auxilia o segundo a se consolidar e entregar informações com credibilidade ao público. Com o surgimento dos computadores e da Internet, o elo entre os dois universos se estreitou. Na década de 1950, quando se utilizavam termos como RAC (Reportagem Assistida por Computador) e Jornalismo de Precisão, os dados eram difíceis de ser acessados. Então o foco estava na obtenção desses elementos que dariam mais credibilidade à notícia e supriria uma fraqueza de apuração nas redações. Com as evoluções tecnológicas, o foco agora está no processamento dos dados e em como eles permeiam o processo da produção jornalística. O Jornalismo de Dados representa uma convergência de campos que são significativos por si próprios - Comunicação e Tecnologia da Informação -, desde a pesquisa investigativa e estatística, até o design, a programação e a formatação dos dados para que eles sejam publicados. Nesta pesquisa, buscamos propor um caminho para otimizar e incorporar o trabalho com dados computacionais nas redações de jornalismo. Os dados são apropriados para definir a composição e a estrutura da apresentação dos conteúdos de natureza jornalística e traduzem os fatos a partir dos números. Partimos de um levantamento bibliográfico de conceitos de jornalismo e mineração de dados para uma proposta de remodelação do processo de construção de notícia. Após a leitura, com a percepção de modelo utilizado em várias empresas de comunicação no mundo, formatamos uma sugestão para formar uma equipe de Jornalismo de Dados dentro da redação. Percebemos que há vários passos a serem seguidos, a exemplo de formar adequadamente os profissionais que tiverem interesse na área, investir em equipamentos e programas computacionais, bem como apostar numa equipe heterogênea. No novo fluxo desenhado para as redações, é possível enxergar um modelo de trabalho híbrido, que transpassa diversas áreas do conhecimento. Acreditamos que esta pesquisa pode contribuir para estudantes e pesquisadores que busquem unir Comunicação e Tecnologia da Informação, bem como servir de base para desenvolvedores de softwares que precisam compreender as necessidades tecnológicas de uma empresa de jornalismo e, assim, solucionar problemas existentes. O trabalho também revela uma importância acadêmica para a formação de jornalistas e cientistas da computação que queiram se aprimorar no universo de uma Ciência de Dados que não está voltada para o Jornalismo, mas que estabelece um diálogo entre as duas áreas de conhecimento, de igual para igual.

Palavras-chave: jornalismo de dados; algoritmo; big data; dados abertos; programação.

ABSTRACT

Data and journalism have always been connected. The first helps the second to consolidate itself and deliver credible information to the public. With the advent of computers and of the Internet, the link between the two universes has narrowed. In the 1950s, when terms such as RAC (Computer Assisted Reporting) and Precision Journalism were used, it was difficult to access data. So the focus was on obtaining those elements that would give more credibility to the news and supply a weakness in the newsrooms. With technological developments, the focus is now on data processing and how it permeates the journalistic production process. Data Journalism represents a convergence of fields that are significant in their own right - Communication and Information Technology -, from investigative and statistical research, to design, programming and formatting of data, so that they are published. In this research, we seek to propose a way to optimize and incorporate the work with computational data in journalism newsrooms. The data are appropriate to define the composition and structure of the presentation of content of a journalistic nature and translate the facts from the numbers. We started from a bibliographic survey of journalism and data mining concepts for a proposal to remodel the news construction process. After this, with the perception of the model used in several communication companies around the world, we formatted a suggestion to form a Data Journalism team within the newsrooms. We realize that there are several steps to be followed, such as adequately training professionals who are interested in this area, investing in computer equipment and programs, as well as betting on a heterogeneous team. In the new flow designed for newsrooms, it is possible to see a hybrid work model, which crosses several areas of knowledge. We believe that this research can contribute to students and researchers who seek to unite Communication and Information Technology, as well as serve as a basis for software developers who need to understand the technological needs of a journalism company and, thus, solve existing problems. The work also reveals an academic importance for the training of journalists and computer scientists who want to improve themselves in a Data Science universe that is not focused on Journalism, but that establishes an equal dialogue between the two areas of knowledge.

Key-words: data journalism; algorithm; big data; open data; programming.

RECADOS AOS LEITORES

Os textos em língua estrangeira receberam livre tradução desta autora e as versões no idioma original foram alocadas em notas de rodapés.

O tema do trabalho nos forçou a utilizar termos específicos da área de Tecnologia da Informação e do cotidiano jornalístico e, por esse motivo, optamos por criar um glossário que facilite a compreensão do texto por todos os públicos. Se encontra nas páginas finais da dissertação.

Ao longo do texto, utilizaremos o termo ‘Jornalismo de Dados’ sempre em caixa alta e baixa para destacar o nosso objeto de estudo.

LISTA DE ILUSTRAÇÕES

| | |
|---|----|
| Figura 1 – Trecho da primeira reportagem que utilizou visualização de dados, publicada pelo The Guardian, 1851. | 21 |
| Figura 2 - Mapa da cólera desenhado pelo dr. John Snow, 1854. | 23 |
| Figura 3 – Aplicação de busca de escolas por código postal ou região. | 33 |
| Figura 4 - Homepage do projeto com aplicação de pesquisa dos dados. | 34 |
| Figura 5 - Visão geral da aplicação disponibilizada na época. | 35 |
| Figura 6 – Estruturas do neurônio biológico e do perceptron mostram similaridades entre os dois modelos de processamento de informações. | 39 |
| Figura 7 – Infográfico detalha diferença de tamanhos entre as unidades de medida de dados utilizadas na computação. | 42 |
| Figura 8 – Ciclo de vida de adoção das tecnologias..... | 43 |
| Figura 9 – Exemplo de consulta (query). | 46 |
| Figura 10 – Modelo de Jornalismo de Dados proposto por Gehrke (2016). | 51 |
| Figura 11 – Homepage do ChicagoCrime.org, 2005. | 58 |
| Figura 12 – Página de verificação das declarações do presidente Donald Trump. | 59 |
| Figura 13 – O contraste populacional da mesma área vista em um intervalo de 3 meses..... | 61 |
| Figura 14 - Exemplo de banheiro ao lado de fonte de água. | 61 |
| Figura 15– Ilustração do mapa de buscas criado pelo La Nación. | 62 |
| Figura 16 – Exemplo de busca realizada para ilustrar o poder da ferramenta. | 63 |
| Figura 17 – Detalhes de servidor público com o salário mais alto do Texas. | 66 |
| Figura 18 - Íntegra do texto publicado por Quakebot, o robô do LA Times. | 68 |
| Figura 19 – Dimensões do Jornalismo de Dados. | 72 |
| Figura 20 – Modelo atual de fluxo de construção da notícia com base em dados. | 78 |
| Figura 21 – Proposta de novo fluxo de construção da notícia com base em dados. | 82 |
| Figura 22 - Passos para mudanças na configuração da redação | 88 |
| Figura 23 – Pirâmide invertida tradicional do jornalismo. | 90 |
| Figura 24 – Pirâmide invertida de Jornalismo de Dados. | 91 |
| Figura 25 – Pirâmide Invertida de Jornalismo de Dados completa. | 92 |

LISTA DE QUADROS

| | |
|---|----|
| Quadro 1 – Dados que estão no trecho destacado da reportagem. | 22 |
| Quadro 2 – Comparação para entender os dados integrais utilizado para falar sobre capacidade de memória e armazenamento dos hardwares. | 41 |
| Quadro 3 – Processo de mineração de dados segundo Han e Kamber (2001). | 44 |
| Quadro 4 - Matriz de habilidades e competências no Jornalismo de Dados. | 70 |

SUMÁRIO

| | |
|---|-----|
| 1 INTRODUÇÃO | 11 |
| 2 JORNALISMO E DADOS: OTIMIZANDO UM VÍNCULO ANTIGO COM O USO DA TECNOLOGIA | 19 |
| 2.1 MÍDIAS CONCATENADAS NUMA PERSPECTIVA DE MUDANÇAS | 25 |
| 2.2 PONDERAÇÕES SOBRE UM FUTURO QUE ALCANÇOU O PRESENTE | 29 |
| 3 A MISSÃO DE APROXIMAR JORNALISMO E TECNOLOGIA DA INFORMAÇÃO | 36 |
| 4 O JORNALISMO DE DADOS EM MEIO A UMA VASTA COLEÇÃO DE TERMOS: PRINCIPAIS CONCEITOS E DIFERENÇAS | 49 |
| 4.1 ABORDAGENS CONCEITUAIS, PRÁTICAS E TECNOLÓGICAS INCORPORADAS AO JORNALISMO DE DADOS | 60 |
| 4.2 UM PONTO DE INFLEXÃO NA CURVA DO JORNALISMO DE DADOS | 67 |
| 5 REDESENHO ALGORÍTMICO NO PROCESSO PARA CONSTRUIR A NOTÍCIA | 72 |
| 5.1 O FLUXO ATUAL DA NOTÍCIA COM BASE EM DADOS | 76 |
| 5.2 REBOOT NO DNA DA EQUIPE | 81 |
| 5.3 PASSO A PASSO x CONDIÇÕES ADVERSAS | 88 |
| 6 CONSIDERAÇÕES FINAIS | 94 |
| REFERÊNCIAS | 100 |
| GLOSSÁRIO | 114 |

1 INTRODUÇÃO

Jornalismo e Tecnologia têm um elo antigo. Não é de hoje que os avanços na área técnica impulsionam a difusão e democratização da comunicação. Ainda no século XV, a produção escrita passou dos poucos milhares para o grupo dos milhões em menos de 50 anos com a reinvenção da imprensa¹ realizada pelo alemão Johann Gutenberg (BRIGGS; BURKE, 2002). Uma novidade que rapidamente se espalhou pelo mundo e encontrou na Grã-Bretanha e nos Estados Unidos espaços de dominação na área jornalística, fato este que acabou provocando mudanças no que o mundo entendia por comunicação. Imprensa passou a ser uma alcunha ligada aos profissionais que publicavam os fatos cotidianos de relevância pública nos jornais: os chamados jornalistas.

A explosão de novas tecnologias de comunicação teria começado entre o fim do século XIX e o início do século XX com o uso de aparatos eletro-eletrônicos como o telégrafo, o rádio, o telefone e o cinema. Ela trouxe, de acordo com Lemos (2002), mudanças impactantes para as práticas sociais. Durante os últimos cem anos as transformações foram, de fato, marcantes. O surgimento dos meios de comunicação de massa deu lugar a uma sociedade que exigia a diminuição urgente de barreiras para acessar informação (ARMAND; MATTERLART, 1995) porque acabara a compatibilidade de vivências sociais com o segredo, ou seja, a ausência dessa informação. As pessoas estavam ávidas por saber o que se passava no mundo.

No jornalismo, esse processo disruptivo ganhou ainda mais força nos idos dos anos 1970 e 1980 com a chegada dos computadores às redações – um processo de mudanças que foi estimulado em todo o planeta pelas Primeira e Segunda Guerras Mundiais e pela Guerra Fria. Os computadores deixaram de ser máquinas de calcular e se tornaram um acessório fundamental nos vários ambientes de trabalho da época (BRIGGS; BURKE, 2002). A criação de uma rede interconectada de computadores levou a humanidade a descobrir rumos que ressignificaram o que antes se entendia por comunicação e vivências sociais, visto que as novidades surgiram de um desejo humano de buscar identidade. “Em um mundo de fluxos globais de riqueza, poder e imagens, a busca da identidade, coletiva ou individual, atribuída ou construída, torna-se a fonte básica de significado social” (CASTELLS, 1999, p. 41). Cria-se,

¹ Os autores utilizam o termo reinvenção porque a tecnologia de impressão já era praticada desde o século VIII na China e no Japão, mas com uma metodologia diferente, que não se aplicava ao alfabeto de letras e, por esse motivo, não teria se sustentado durante muito tempo. Briggs e Burke (2002) comentam que no início do século XV os coreanos estavam trabalhando na evolução da experiência de seus vizinhos. E a notícia dessa possível remodelação no oriente pode ter influenciado Gutenberg a desenvolver sua própria tecnologia.

então, um novo espaço, que não existe no mundo físico, mas que tem um poder de conexão infinitamente maior que qualquer outro já visto: o ciberespaço. Nesta dissertação, utilizaremos o conceito de ciberespaço de Lévy (1999), que o identifica como um novo meio de comunicação surgido a partir da conexão da rede mundial de computadores. O autor diz que também é apropriado chama-lo de rede. “O termo especifica não apenas a infraestrutura material da comunicação digital, mas também o universo oceânico de informações que ela abriga, assim como os seres humanos que navegam e alimentam esse universo” (LÉVY, 1999, p. 7).

Essa nova rede tem uma influência grande nesta sociedade de informação que passa por um processo chamado de cibercultura. Para Lévy (1999), a cibercultura é o “conjunto de técnicas (materiais e intelectuais), de práticas, de atitudes, de modos de pensamento e de valores que se desenvolvem juntamente com o crescimento do ciberespaço” (LÉVY, 1999, p.17). Tal conjunto de inovações encontra no usuário (humano que utiliza o computador) um poder transformador em potencial e impulsiona a sociedade a usar a tecnologia em prol da evolução das relações humanas. O fato é que o ciberespaço proporcionou a convergência dos vários meios de comunicação que haviam surgido até então numa única rede, compartilhada simultaneamente por um grande número de pessoas: a *internet*. Ela representa uma mutação importante no mundo e também para este estudo, que enfatiza os avanços tecnológicos no jornalismo com a utilização intensa do computador nos modos de trabalho.

A tecnologia não está separada do comportamento social. Na verdade, em alguns casos ela pode ser determinante e, em outros, condicionante para tal (LÉVY, 1999), mas nunca estará separada do ser e de como ele escolhe viver em comunidade e, naturalmente, de como esse grupo busca identidade para evoluir. Cultura, sociedade e tecnologia não são elementos apartados. Pelo contrário, se transmutam em um conceito universal, o da cibercultura.

“O universal” significa a presença virtual da humanidade para si mesma. O universal abriga o aqui e agora da espécie, seu ponto de encontro, um aqui e agora paradoxal, sem lugar nem tempo claramente definíveis. (LÉVY, 1999, p. 247)

Essa concepção de universal de Pierre Lévy está mais atual do que nunca. Em todas as redes, nos lugares mais remotos do mundo, de alguma forma, a tecnologia e o ciberespaço se fazem presentes e abarcam, cada vez mais, um número imenso de pessoas em busca de novas formas de se comunicar e de se conectar. O jornalismo, que trilhou e noticiou as transformações ocorridas nessa caminhada, também buscou nutrir-se dos avanços e possibilidades que o

ambiente da *internet* oferece. De tal forma que o modo de produzir e distribuir a notícia vive em estado de mudança constante. A tecnologia da rede de computadores encurtou distâncias, transcendeu os conceitos de tempo (GARCÍA; RODRÍGUEZ; WESTLUND, 2015) e dilatou todo e qualquer conhecimento sobre a difusão da notícia.

As rupturas causadas pela evolução do mundo *high-tech* têm impactado o jornalismo de maneira decisiva. E uma vez que esse universo *online* se estabelece, se alicerça e finca raízes satisfazendo demandas humanas consideradas essenciais, ele força todos os meios antigos a tecerem um acordo de impelida convivência até que se alcance a sonhada paz (JENKINS, 2009). Isso só acontece quando o ser humano compreende que algumas mídias são substituídas e outras são complementadas. Na verdade, “a convergência é a oportunidade de converter empresas jornalísticas do século XX em empresas informáticas do século XXI” (NEGREDO; SALAVERRÍA, 2008, P. 17)². É preciso ir aonde a audiência está.

O conteúdo de um meio pode mudar [...], seu público pode mudar [...] e seu status social pode subir ou cair [...], mas uma vez que um meio se estabelece, ao satisfazer alguma demanda humana essencial, ele continua a funcionar dentro de um sistema maior de opções de comunicação. [...] Cada meio antigo foi forçado a viver com os meios convergentes (JENKINS, 2009, p. 41).

É precisamente o que aconteceu com a *internet*. Alguns especialistas indicam que estamos inseridos em uma rede de quase 50 bilhões de objetos conectados e que, por isso, nosso foco é, cada vez mais, o *design* da experiência do usuário e não do produto (GROOPMAN, 2017). Esse usuário seria o início e o fim de quaisquer produtos que venhamos a produzir. A notícia não está – e nem pode estar – fora desses parâmetros. E como ela já saiu das bancas (imprensa escrita), passou para os nossos ouvidos e olhos (radiodifusão) e chegou aos nossos bolsos e braços (dispositivos móveis e vestíveis), o jornalismo percebeu que o foco é o desejo de quem está do outro lado do eixo de comunicação.

Foi indispensável recorrer à criatividade³ para não perder a audiência, que foi se transfigurando ao longo do tempo, e mudar os processos de produção da notícia. A comunicação precisou ficar atenta a um universo mutante e reconfigurável. Os consumidores

² La convergencia periodística es una oportunidad para reconvertir la organización de las empresas periodísticas del siglo XX en empresas informativas del siglo XXI.

³ Para entender criatividade utilizamos o conceito de Bendassolli et al que a define como a "expressão do potencial humano de realização, que se manifesta mediante atividades geradoras de produtos tangíveis, ou seja, como a capacidade de o indivíduo manipular objetos do mundo externo a partir de um desenvolvimento simultâneo de seus recursos pessoais" ou como "a capacidade, detida por indivíduos ou grupos, de manipular símbolos e significados com o intuito de gerar algo inovador". (BENDASSOLLI ET AL, 2009, p. 13)

de notícia se tornaram uma série de seres que se comunicam – consciente ou inconscientemente – através de números binários (zeros e uns), com mensagens codificadas e armazenadas *online*. O nascimento da *web* enquanto local de armazenamento de dados deu, então, lugar a um novo cenário: o uso dos dados que estavam empilhados nas pastas de computadores para a elaboração de notícias.

Para o jornalismo, o uso de dados não é novidade, mas a massificação e o acesso facilitado a esses números e estatísticas distribuídos e armazenados em uma grande quantidade de bases proporcionaram novos formatos de trabalho e, principalmente, de modelo de produção da notícia. E, por consequência, da relação com a audiência. Mudou a mensuração de quem é o público, a forma de apuração e também de distribuição do modelo de negócio que o jornalismo conheceu outrora. Mello (2016) aponta que vivemos um momento em que a notícia usa suportes lúdicos para estabelecer comunicação com o usuário. São *games* e infográficos interativos que misturam “texto e imagem baseados em notícia ou acontecimentos em tempo real de forma dinâmica” (MELLO, 2016, p. 407). A polivalência da *internet*, incluindo os *games*, entrega nas mãos do usuário o desejo não só de consumir a notícia, mas de se envolver com ela (KAO; HARRELL, 2017). A *web* é o ‘lugar’ onde quase tudo é possível. Em termos de comunicação interativa são incalculáveis as possibilidades de narrativas que podem ser criadas a partir das várias plataformas que utilizem *internet*.

A descoberta do que o público quer absorver e de qual jornada de experiência ele quer percorrer deve ser uma provocação contínua para quem quer trabalhar com o Jornalismo de Dados, seja ele visto como um novo tipo de jornalismo ou como uma ferramenta para as redações. Neste trabalho, nos guiaremos pelo conceito de Bradshaw (2012), que afirma que o Jornalismo de Dados não significa meramente uma junção linguística entre ‘jornalismo’ e ‘dados’ porque tratam-se de conceitos problemáticos e complexos. Para ele, “os dados podem ser a fonte do Jornalismo de Dados ou a ferramenta com a qual a história é contada – ou podem ser os dois” (BRADSHAW, 2012, sem página). Com base nessa afirmação, buscar conhecimento em novas áreas para descobrir pautas nas bases disponibilizadas por fontes públicas, compilar, limpar e minerar os números, transformando-os em informações compreensíveis para a população, é fundamental para promover novas vivências ao contar histórias. Os meios de comunicação sempre atravessam momentos de mudanças profundas. Precisam constantemente encontrar novos modelos de produção que se adaptem às alterações tecnológicas, sociais, políticas e econômicas do público. É também uma maneira que o

jornalismo precisa encontrar, a cada alteração, de se manter vivo nos vários suportes existentes (BACCIN; CANAVILHAS, 2015) e não só mais nos meios de comunicação de massa que dominavam a audiência, a exemplo da televisão (PORCELLO, 2012). Como as práticas sociais e modelos de trabalho têm uma tendência a evoluir com mais lentidão do que as evoluções técnicas (BRIGGS; BURKE, 2002), a pergunta que deu origem a este trabalho foi: como as redações podem se ajustar para incorporar o trabalho com bases de dados computacionais no processo de construção da notícia?

Os tempos são outros. O modelo se atualizou. Por exigir uma formação extracurricular, compreender outros grupos de conhecimento, como o da Tecnologia da Informação, pode simbolizar um empecilho para muitos profissionais de redações. Um outro obstáculo para a aproximação entre jornalistas e dados é a necessidade de investimento financeiro em equipamentos e treinamentos de profissionais para gerar uma mão de obra especializada. Trabalhar com bancos de dados complexos e grandes exige uma rede de computadores e máquinas robustas para que sejam capazes de trabalhar de forma ágil e eficaz. É fundamental também que haja profissionais capazes de trabalhar com Jornalismo de Dados e isso envolve não só o treinamento dos jornalistas, como também a contratação de personagens que antes não faziam parte do quadro comum de funcionários nas redações. Desenvolvedores e *designers*, por exemplo. São adversidades que merecem a atenção dos que quiserem formar uma equipe de Jornalismo de Dados na empresa.

É com base nesses apontamentos que nosso objetivo principal nesta pesquisa é propor um caminho para incorporar o trabalho com bases de dados computacionais no processo de construção da notícia. Além desse objetivo geral, a pesquisa tem também objetivos específicos, sendo eles:

- Propor uma metodologia de formação de equipes de Jornalismo de Dados e de um modelo de construção da notícia nas redações;
- Mapear conceitos de mineração e análise de dados na área de Tecnologia da Informação;
- Fazer ponderações sobre os principais conceitos de Jornalismo de Dados que dão origem a essa expressão;
- Sintetizar a convergência entre o campo do Jornalismo e o da Tecnologia da Informação;

- Ampliar o acesso de profissionais dessas duas áreas a esse conteúdo comum para que encontrem meios de dialogar e criar, juntos, novas possibilidades narrativas para o público;
- Tornar mais plural o acesso a uma bibliografia transdisciplinar sobre o tema.

Desta forma, se faz pertinente o estudo sobre esse tema num programa de pesquisa de pós-graduação que atua na área das Indústrias Criativas e demanda uma procura por produtos que não atendam a padrões específicos, mas que deem um novo significado ao que já existe (HOWKINS, 2013). É possível que esta pesquisa simbolize uma inovação nas redações e repercuta positivamente entre os jornalistas que querem ter acesso ao Jornalismo de Dados, mas ainda não sabem como. Cabe a cada profissional decidir até que ponto imergir nesse universo, mas o estudo apresenta possibilidades em diferentes áreas de conhecimento, com um vasto conteúdo a ser explorado. A pesquisa pode dar origem, não só a novos modelos de trabalho, mas a ferramentas tecnológicas que podem ser desenvolvidas a partir das necessidades que as equipes descobrirem durante o processo de mudança nas redações. Um efeito que alimenta também o mercado de desenvolvimento de softwares e soluções tecnológicas para a Ciência de Dados. Além desses pontos, é importante destacar que estimular jornalistas a dominarem melhor a interpretação dos dados para gerar informação e conhecimento, ganha-se a oportunidade de formar a sociedade em diversos assuntos de interesse público.

A pesquisa tem caráter exploratório porque, de início, se baseia em um levantamento bibliográfico que procura familiaridade com o tema e o problema, de modo a construir hipóteses fundamentadas no embasamento teórico proposto (GIL, 2002) a partir de buscas e análises de grande parte do material já publicado sobre o assunto em livros, revistas e outras publicações (LAKATOS, 2002). É também de natureza aplicada, já que se propõe a gerar conhecimentos para a solução de problemas específicos como a formatação ideal de uma equipe de Jornalismo de Dados. O primeiro passo constituiu-se em fazer um levantamento bibliográfico sobre a evolução tecnológica na comunicação e, em especial, no jornalismo, assim como em buscar conceitos importantes na área de Tecnologia da Informação que versam sobre mineração de dados para entender as concepções aplicadas ao Jornalismo de Dados e elaborar nossa proposta final. Portanto, nos baseamos em métodos qualitativos para obter uma visão mais completa do fenômeno estudado.

No início, a pesquisa envolvia uma experiência interdisciplinar e propunha a criação de um sistema automatizado para coletar, minerar e analisar dados desenvolvidos na TV Globo do Recife (PE) para elaboração de reportagens jornalísticas, mas por questões jurídicas⁴ o processo se tornou demasiado longo para que fosse explorado neste trabalho. Por esse motivo, após a qualificação, optamos por deixar a construção da ferramenta de lado e redirecionamos nosso foco para o objetivo principal desta pesquisa, que é propor um caminho para incorporar o trabalho com bases de dados computacionais no processo de construção da notícia. Configuramos, assim, a proposta de um novo processo.

A dissertação está dividida em cinco capítulos e considerações finais. O primeiro capítulo traz uma introdução ao tema e o segundo, demonstra que o jornalismo e dados sempre estiveram juntos, unidos pela tecnologia vigente em cada época. A começar pela prensa de Gutenberg, que sistematizou o processo de confecção dos jornais e permitiu que eles fossem distribuídos com mais agilidade e a um maior número de pessoas. Demonstraremos que dados numéricos servem de base para as reportagens e são publicados em jornais importantes desde o século XIX. E acompanharemos como as mudanças tecnológicas vêm transformando esse cenário.

No terceiro capítulo iremos conhecer alguns conceitos básicos da área de Tecnologia da Informação que dizem respeito diretamente à Mineração de dados. São definições básicas fundamentais para os profissionais de Comunicação que desejam começar a trabalhar com dados e não têm noção do conhecimento necessário que precisam absorver para dominar esse tipo de conteúdo. Trabalhamos tanto com os conceitos de dados, como também de linguagens de programação acessíveis a profissionais que não tiveram uma formação na área computacional na graduação.

O quarto capítulo aborda os principais conceitos discutidos no mundo acadêmico que envolvem o Jornalismo de Dados. Fazemos um resgate de terminologias do século XX, quando as redações começavam a lidar com os computadores, e chegamos aos registros mais recentes.

⁴ A TV Globo – assim como outras empresas consolidadas no mercado de comunicação – tem uma série de parâmetros e métodos para os trabalhos de desenvolvimento de soluções tecnológicas para as várias áreas que a compõe. É preciso seguir um protocolo de aprovação de projeto com a sede da corporação, que fica no Rio de Janeiro, assim como submeter o trabalho e a necessidade de publicação acadêmica a uma revisão jurídica feita pelos especialistas da Globo. Esse processo não tinha prazo para ser concluído e por isso optamos por seguir um caminho alternativo, ainda que com base empírica na experiência de trabalho na redação do Recife.

Demonstraremos também exemplos de reportagens que ilustram os conceitos abordados que consideramos fundamentais na pesquisa. Também abordaremos brevemente o conceito de Jornalismo Algorítmico que nasce com o Jornalismo de Dados, mas utiliza como base a máquina e não o humano. É neste capítulo que mostraremos o estado da arte do nosso objeto de estudo para que consigamos elaborar uma proposta com conceito criativo ao fim do trabalho.

O quinto capítulo é composto pela sugestão de mudança no processo de construção da notícia que passa, necessariamente, por transformações profundas no quadro profissional da empresa com capacitações dos profissionais e readequação de atividades. Falaremos também dos principais desafios desse processo e detalharemos um plano de ação para que ele seja colocado em prática. Tudo com base no modelo atual de trabalho com dados nas redações, que foi construído com base no levantamento bibliográfico, e com a formatação proposta no resultado deste trabalho.

Por fim, apresentamos as considerações finais sobre o trabalho e as possibilidades vislumbradas de pesquisas futuras com base nesta dissertação. Logo após as referências oferecemos um glossário com expressões que utilizamos ao longo do texto para consultas durante a leitura.

2 JORNALISMO E DADOS: OTIMIZANDO UM VÍNCULO ANTIGO COM O USO DA TECNOLOGIA

Há muito tempo o jornalismo utiliza os números advindos de fontes confiáveis para produzir notícias bem fundamentadas e apresentar credibilidade. Os dados numéricos conferem ao texto informativo um caráter essencial para a missão do jornalismo: a busca pela verdade a ser apresentada de uma maneira compreensível para o público de maneira objetiva. O Manual de Redação (2001) do jornal Folha de São Paulo recomenda trocar o tom mais dramático da narrativa por informações baseadas nos conjuntos de algarismos. “Tudo pode ser provado pelos números” (PENA, 2005, p. 53). Como a notícia assinala acontecimentos para tornar públicos fatos através de informações (SODRÉ, 1986), os dados numéricos ajudam a produzir inferências sobre o tema apresentado na reportagem, colaboram com a formação de opinião e até de conteúdo da audiência, além de permitir uma leitura mais analítica do que é publicado.

Esse é um dos motivos pelos quais os dados e o jornalismo já apresentam uma união estável há centenas de anos. Números e estatísticas aprofundam a matéria-prima do jornalismo, a notícia, que representa tanto o que o público necessita saber, como o que ele deseja falar (TRAQUINA, 1993; VIZEU, 2000). Significam mais um mecanismo de busca pela verdade, utilizando a observação de métodos científicos (LIPPMANN, 1922; MEYER, 1973). O advento de novas tecnologias aperfeiçoou essa relação e acelerou o processo de conexão com novos formatos de trabalho. A incorporação dos computadores às redações começou na segunda metade do século XX no Brasil (BALDESSAR, 2001) e impôs a necessidade, aos jornalistas, de adquirir habilidades comum a outras profissões.

Associar as técnicas de coleta, mineração e tratamento de dados a uma investigação social dos fatos faz com que o jornalista exerça atividades similares ou até iguais às de um *data analyst* ou *data scientist* (profissionais da área de Tecnologia da Informação que se especializam no trabalho com os dados) e consiga ir além do papel de intérprete e transmissor de notícias. É aí que começa o desafio.

A tradição de que o mundo do conhecimento é povoado por dois tipos diferentes de pessoas – aquelas que leem e aquelas que contam - está decaindo rapidamente. Cada vez mais, são necessárias duas habilidades para se dar bem, e não apenas no mundo acadêmico. Os formuladores de políticas públicas estão aprendendo a contar e medir, ou pelo menos a trabalhar com pessoas que o fazem. Essa tendência significa que aqueles de nós que optaram por habilidades verbais adotando o jornalismo têm que se atualizar. O primeiro passo para recuperar o atraso é livrar-se do nosso medo natural dos números.

E a maneira mais fácil de fazer é obter uma régua de cálculo de dez centavos ou uma calculadora de bolso (MEYER, 1973, p. 64)⁵.

Utilizar planilhas de cálculos para contextualizações, checar informações, ilustrar reportagens com tabelas e gráficos converteu-se, pois, em realidade. Números e estatísticas amparam a notícia e lhe tornam mais crível. Atualizar-se sempre foi uma premissa básica do jornalismo, mas quando falamos de dados esse conceito se complexifica porque não se trata apenas de tomar conhecimento das últimas inovações na área da comunicação. É, no entanto, abrir a mente a outro campo do conhecimento antes visto quase como antagônico ao jornalismo: o universo da matemática dos computadores. O fato de utilizar as máquinas não significa saber programa-las e configurá-las. O contato com os dados existia no meio jornalístico muito antes da chegada dos computadores. Eles sempre foram utilizados como uma das principais fontes de informação das reportagens (ROGERS, 2011) para relatar fatos de interesse público à sociedade (TRAQUINA, 2001).

O jornal *The Guardian* (Reino Unido) publicou a primeira reportagem com dados em 5 de maio de 1821 para fazer uma relação entre a frequência de alunos nas escolas das cidades de *Manchester* e *Salford* com os gastos anuais médios do governo na área de educação (Figura 1). As tabelas, construídas com dados cedidos por uma fonte anônima ao repórter, foram capazes de mostrar quantas crianças pobres havia nas cidades e quantos estudantes recebiam educação gratuita. Nessa época, de acordo com o *post* do jornal (*THE GUARDIAN*, 2011), as únicas estatísticas oficiais no país eram disponibilizadas por quatro clérigos. Como os braços eram poucos para dar conta da grande quantidade de trabalho, os números conhecidos até então eram imprecisos.

⁵ The tradition that the world of knowledge is populated by two different kinds of people – those who read and those who count – is swiftly decaying. Increasingly, it takes both skills to get along, and not just in the academic world. Public policy makers are learning to count and measure, or, at least, to work with people who do. This trend means that those of us who opted for verbal skills by taking up journalism have some catching up to do. The first step in catching up is to get rid of our natural fear of numbers. And the easiest way to do is to get yourself a dime-store slide rule or a pocket calculator.

Figura 1 – Trecho da primeira reportagem que utilizou visualização de dados, publicada pelo The Guardian, 1851.

| DAY SCHOOLS.— <i>Establishments</i> | Boys | Girls | Total | Ann. Exp. | Remarks. | |
|-------------------------------------|------|-------|-------|-----------|--|------------------------------|
| Grammar School | 155 | | 155 | 1800 | } Endowments. | |
| Blue Coat ditto..... | 80 | | 80 | 2000 | | Taught, clothed and boarded. |
| Green Coat ditto..... | 50 | | 50 | 200 | | Taught and clothed. |
| Collegiate Church ditto..... | | 50 | 50 | 40 | | And offertory money: do. do. |
| Strangeways ditto..... | 10 | | 10 | 100 | | |
| St. Mary's ditto..... | 12 | 12 | 24 | 40 | { (Suppose)—Taught and clothed. Funds arising from Sacramental Offerings. | |
| St. John's ditto..... | 9 | | 9 | 40 | { (Suppose)—Expences raised by voluntary Subscription. | |
| St. Paul's ditto | 20 | | 20 | | | |
| Ladies' Jubilee | | 30 | 30 | 250 | { Taught, clothed and boarded. by voluntary Subscription. | |
| Back King-street..... | 21 | | 21 | 40 | { (Suppose)—Taught and partly clothed. This School is supported by the benevolence of a single individual. | |
| NATIONAL SCHOOLS, Granby-row..... | 194 | 119 | 313 | 600 | { Voluntary Subscription, and Collections at Churches. | |
| Bolton-street, Salford | 300 | 170 | 470 | | | |
| | 851 | 381 | 1232 | £5410 | | |

Fonte: *The Guardian, Datablog*, 2011.

Na figura 1, pode-se ver que os dados foram tabulados, com linhas e colunas, formando uma espécie de tabela para que os leitores entendessem as categorias que existiam. Neste trecho que escolhemos destacar os dados estavam organizados desta forma: na primeira coluna estavam o tipo de escola ou igreja (diária ou dominical, estas eram as escolas em que estudavam crianças que trabalhavam durante a semana. No país, a educação não era obrigatória até 1880). Nesta coluna, o tipo da escola vem sempre acompanhado da palavra “estabelecimento”, que indicava se a escola era da Igreja Anglicana (também chamada de Igreja da Inglaterra) ou não. As outras religiões não eram oficiais e por isso os tipos das escolas se diferenciam. Depois vinham três colunas com a quantidade de meninos, a de meninas e o total de alunos. Em seguida estava a coluna de gastos anuais calculados em libras, a moeda do país. E por último a coluna de observações que diziam respeito, entre outras informações, às origens do dinheiro e também se se havia gastos com fardamento dos alunos, por exemplo. Cada linha representa uma escola. Isso significa que as linhas eram os registros de instituições que seriam analisadas na tabela. Atualmente ainda utilizamos esse modelo para estruturar dados em uma tabela, com categorias nas colunas e registros nas linhas. No quadro 1 detalhamos os dados que estão dispostos na figura 1.

Quadro 1 – Dados que estão no trecho destacado da reportagem.

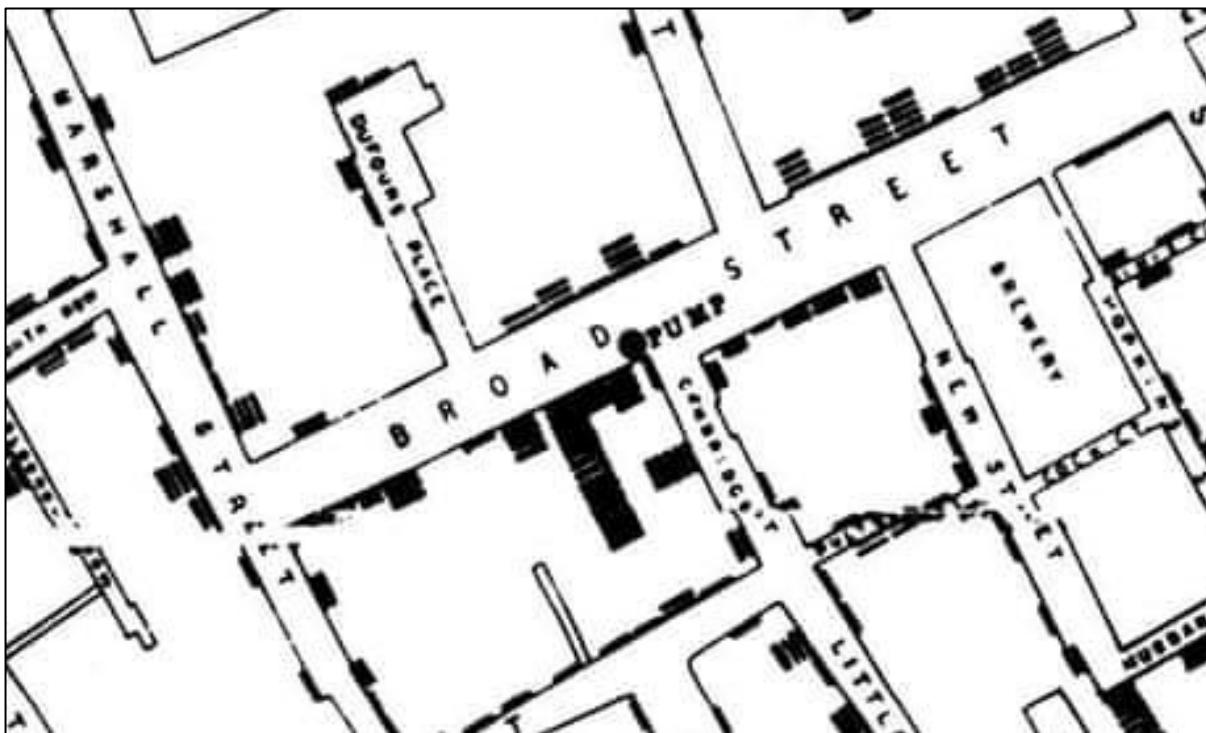
| Escola/ igreja | Tipo | Meninos | Meninas | Total | Gastos anuais £ | Observações |
|-------------------------------------|---------------------------------------|----------------|----------------|--------------|------------------------|--|
| <i>Grammar school</i> | Escola diária, estabelecimento | 155 | | 155 | 1,800 | Doação. |
| <i>Blue coat</i> | Escola diária, estabelecimento | 80 | | 80 | 2,000 | Ensino, fardamento, doação. |
| <i>Green coat</i> | Escola diária, estabelecimento | 50 | | 50 | 200 | Fardamento, doação. |
| <i>Collegiate church</i> | Escola diária, estabelecimento | | 50 | 50 | 40 | Dinheiro de ofertório, fardamentos, doação. |
| <i>Strangeways</i> | Escola diária, estabelecimento | 10 | | 10 | 100 | Doação. |
| <i>St Mary's</i> | Escola diária, estabelecimento | 12 | 12 | 24 | 40 | Ensino, fardamento, fundos provenientes de ofertas sacramentais. |
| <i>St John's</i> | Escola diária, estabelecimento | 9 | | 9 | 40 | Gastos arrecadados por inscrição voluntária. |
| <i>St Paul's</i> | Escola diária, estabelecimento | 20 | | 20 | 40 | Gastos arrecadados por inscrição voluntária. |
| <i>Ladies' jubilee</i> | Escola diária, estabelecimento | | 30 | 30 | 250 | Ensino, fardamento e deslocamento por inscrição voluntária. |
| <i>Back King-street</i> | Escola diária, estabelecimento | 21 | | 21 | 40 | Ensino e parte do fardamento. Esta escola é mantida pela benevolência de uma única pessoa. |
| <i>National schools, Granby row</i> | Escola diária, estabelecimento | 194 | 119 | 313 | 600 | Inscrição voluntária e coletas nas igrejas |
| <i>Bolton street, Salford</i> | Escola diária, estabelecimento | 300 | 170 | 470 | 600 | Inscrição voluntária e coletas nas igrejas |
| TOTAL | Escola diária, estabelecimento | 851 | 381 | 1,232 | 5,110 | |

Fonte: *The Guardian, Datablog*, 2011 / Traduzido pela autora

A reportagem do *The Guardian* surpreendeu os leitores ao mostrar que cerca de 25 mil crianças recebiam educação gratuita, um número muito maior do que os de 8 mil alunos previamente apresentados como dados oficiais pela apuração dos clérigos. Três anos mais tarde o médico John Snow também revolucionou uma publicação jornalística no mesmo *The Guardian*, mas desta vez utilizando - pela primeira vez - visualização de dados (ROGERS, 2013). Na época, a cólera era uma doença se espalhava com facilidade e o bairro de *Soho*, em Londres, no Reino Unido, passava por um surto misterioso. Doutor Snow, então, teve a ideia

de estruturar em papel os casos de contaminação do bairro. Com um mapa local em mãos, o médico desenhou pequenos traços ao longo das ruas, como visto na figura 2. Cada um deles simbolizava uma pessoa que havia ido a óbito por causa da doença. Ao fim do trabalho, ele percebeu que as mortes estavam concentradas em torno de uma bomba de água situada na Rua *Broad* (hoje Rua *Broadwick*). A descoberta levou o médico a perceber que a água dessa bomba estava extremamente contaminada e era a fonte de contágio. Tomadas as medidas cabíveis para afastar a população do contato com a água que saía da referida bomba, o número de pessoas doentes começou a cair.

Figura 2 - Mapa da cólera desenhado pelo Dr. John Snow, 1854.



Fonte: The Guardian, Datablog, 2013.

Nessa época, os dados eram de difícil acesso. Então o foco estava na obtenção desses elementos que dariam mais credibilidade à notícia e supririam uma fraqueza de apuração nas redações (MEYER, 1973). Hoje, com a rede de computadores, o foco está no processamento dos dados e em como eles permeiam o processo de produção jornalística. Aos poucos o movimento das novas máquinas nas redações a partir da década de 1980 foi gerando um processo de metamorfose no ato de produzir notícia. Uma das mudanças mais importantes foi

a possibilidade de conectar o jornalista com dados estruturados em tabelas e bancos que tornaram a fonte das informações um ambiente mais confiável e de fácil acesso.

Deu-se início a um novo modelo de trabalho nas empresas: a RAC – Reportagem Assistida por Computador, que “consiste na utilização de instrumentos tecnológicos com o objetivo de aproximar ao máximo o jornalista da informação primária, proporcionando-lhe condições mais adequadas para interpretar a realidade” (PENA, 2005, p. 174). O papel fundamental do jornalismo, reforçando, atravessa mais uma mudança, mas permanece o mesmo: ter compromisso com a verdade, ser fiel com o cidadão, ter disciplina com a checagem dos fatos, funcionar como um monitor independente do poder, transformar fatos significantes em notícias relevantes e interessantes, assim como preocupar-se com a necessidade de produzi-las (as notícias) de maneira compreensível para todo o público, sem distinções (TRAQUINA, 2004).

2.1 MÍDIAS CONCATENADAS NUMA PERSPECTIVA DE MUDANÇAS

Passado o primeiro momento de transformações, em que os meios de comunicação tradicionais acreditavam na competição entre as plataformas, e agora, em vez de disputar lugares, as mídias perceberam que no contexto da convergência se retroalimentam e não podem ser ignoradas entre si (ROST, 2014), ficamos diante de uma quebra de paradigma sem precedentes. O 'fazer reportagem' tem se tornado um caminho complexo, múltiplo e interdisciplinar. Essas oscilações de padrão e rupturas de tradições não são uma prerrogativa do jornalismo. Vivemos em um período de revolução no modo de consumir. O que determina o progresso de uma sociedade é a extensão com que ela adota a inovação tecnológica.

É assim desde a Primeira Revolução Industrial. Já se passaram centenas de anos desde que o tear mecanizado e a máquina a vapor nasceram na Inglaterra. A Segunda Revolução Industrial ainda não foi vivida por 100% da população mundial e a chegada dos computadores no século 20, marcando a terceira fase de mudanças, precisa alcançar quase metade do planeta (SCHWAB, 2016). Nos últimos anos a chamada Indústria 4.0 trouxe um modelo de produção 'inteligente', em que os mundos físico e virtual cooperam entre si de maneira integrada.

A quarta revolução, no entanto, não diz respeito apenas a sistemas e máquinas inteligentes e conectadas. Seu escopo é muito mais amplo. Ondas de novas descobertas ocorrem simultaneamente em áreas que vão desde o sequenciamento genético até a nanotecnologia, das energias renováveis à computação quântica. O que torna a quarta revolução industrial fundamentalmente diferente das anteriores é a fusão dessas tecnologias e a interação entre os domínios físicos, digitais e biológicos. (SCHWAB, 2016, p. 16)

Cada vez mais, a tecnologia tem se apresentado além do aparato e se tornado extensão do nosso corpo. Estamos rodeados de dispositivos que são capazes de nos ouvir, responder e tomar decisões que antecipam nossas escolhas e, por vezes, até o nosso querer. É claro que “nem todas as indústrias estão no mesmo ponto de ruptura, mas todas estão sendo levadas a uma curva de transformações pelas forças da quarta revolução industrial” (SCHWAB, 2016, p. 68). Enquanto vemos empresas de tecnologia como a *Google* e a *Apple* anunciarem a produção de automóveis, nos deparamos também com algumas corporações que ainda não conseguiram se justapor à curva dos avanços. Essas, se não correrem contra o tempo, estarão fadadas a um desvanecimento dentro em pouco. E definitivo. Afinal de contas, o modelo muda porque as pessoas buscam novos rumos e tal demanda movimenta igualmente o formato de consumo.

No jornalismo não é diferente. Como explicitamos, do jornal impresso para o rádio e a televisão, algumas alterações e atualizações foram significativas, mas nenhuma abalou tanto a comunicação quanto a *internet*. É como se as pessoas agora tivessem uma persona virtual que precisa estar atenta a esse universo mutante e reconfigurável. Não existe mais tempo, nem lugar. Somos uma série de seres que se comunicam – consciente ou inconscientemente – através de números binários (zeros e uns), com mensagens codificadas e armazenadas online. E o jornalismo é um dos segmentos da comunicação que mais se beneficia com essas novidades tecnológicas e quebras de padrão, que acabam transformando os modos de produção. As redações de jornalismo ao redor do mundo precisaram se adaptar às novidades para não serem devoradas por elas no caminho.

Um dos fatores que possibilitou essa ruptura é o poder que a *internet* tem de agregar diferentes pilares de comunicação em um único 'lugar'. A chamada multimídia, que Salaverría (2014) diz ser uma característica intrínseca ao ser humano, já que sempre usamos mais de um sentido corporal para emitir e receber mensagens uns dos outros. A *web* seria, então, o espaço onde as mídias convencionais convergem com um único propósito: colocar à disposição do público a informação usando mais de um elemento ao mesmo tempo – texto, foto, vídeo, gráficos, iconografia ou ilustrações, vídeo, animação, discursos orais, músicas e efeitos sonoros, entre outros.

Ainda que o universo multimídia não seja novidade para os leitores, ouvintes, telespectadores ou, simplesmente, usuários, de acordo com a definição de Salaverría (2014), os dispositivos usados para que essa multimídia exista são frutos das descobertas tecnológicas diárias. E elas acontecem dentro de um processo de mutação constante dentro dele mesmo.

Nos primeiros anos do século XXI vivemos um primeiro momento de agitação em torno das narrativas multimídia devido, sobretudo, ao êxito da infografia digital. As principais publicações de internet utilizaram este gênero como trampolim para dar um salto qualitativo na sua oferta de conteúdo. Poucos anos depois, o interesse pela infografia decaiu, em grande medida, com a aparição fulgurante do vídeo enquanto novo formato estrela. Uma vez assimilada a novidade oferecida pelo vídeo, os meios digitais parecem apostar na atualidade na exploração de narrativas multimídia integradas. Já não se trata, apenas, de apostar neste ou naquele formato em particular, mas antes de investigar as possibilidades oferecidas pela articulação de vários formatos (SALAVERRÍA, 2014, p. 48).

Significa dizer que a tecnologia não só permitiu o diálogo entre meios de comunicação diferentes, como também possibilitou que todos se tornassem um único e novo meio sem perder

as particularidades de cada. A *web* tem mesmo um poder consolidador e agregador em várias áreas do conhecimento, mas na comunicação e, em especial, no jornalismo, ela se mostra um tanto além de um caráter evolucionista e alcança um patamar que nega afirmações de que os jornalistas (ou pelo menos parte deles) desapareceriam com o surgimento da *internet* (PALACIOS, 2003). Ela agora é símbolo de uma primeira ruptura: a dos espaços físicos.

Se antes os profissionais sempre foram limitados a uma quantidade de linhas ou tempo de transmissão nos meios de radiodifusão, na *web* o tempo perde a finitude do mundo real. Ao processo em que vários tipos de mídia se encontram num mesmo ambiente Jenkins (2009) dá o nome de convergência. Segundo ele, esta é uma palavra que consegue definir vários tipos de transformações, sejam elas culturais, mercadológicas, sociais e até tecnológicas, dependendo de quem é o público-alvo. É uma mudança na maneira de encarar o mundo e a comunicação porque a *web* se apresenta como um lugar de experimentação, onde as pessoas e os meios podem se encontrar. Além de romper com o tempo, ela quebra os paradigmas de espaço e todas as restrições geográficas com que os comunicadores dos meios de comunicação de massa sempre foram obrigados a lidar. Isso acontece porque a rede de computadores *online* permite que qualquer pessoa esteja em vários lugares a todo momento que desejar.

Para a comunicação é um feito extraordinário. Em um contexto profissional macro seria possível dizer que talvez esteja no inconsciente coletivo⁶ dos jornalistas o sonho de que suas reportagens alcancem o maior número de pessoas possível, para que sejam lidas, ouvidas ou vistas. Esse fenômeno leva o nome de ubiquidade, uma palavra que significa estar em vários lugares, de preferência, simultaneamente. No jornalismo essa possibilidade tem proporções gigantescas.

No contexto da mídia, ubiquidade implica que qualquer um, em qualquer lugar, tem acesso potencial a uma rede de comunicação interativa em tempo real. Quer dizer que todos podem não apenas acessar notícias e entretenimento, mas participar e fornecer sua própria contribuição com conteúdo para compartilhamento e distribuição global (PAVLIK, 2014, p.160).

A ubiquidade carrega consigo a oportunidade de criar novas formas narrativas e imersivas, já que oferece uma integração ímpar nos avanços espaço-temporais com os quais o

⁶ Utilizamos o conceito de inconsciente coletivo apresentado por Jung (2003). Ele diz que se trata de uma área profunda da psique humana de natureza universal, isto é, não carrega traços de comportamento e conteúdos que pertencem ao espectro pessoal do indivíduo. Pelo contrário: essas informações e códigos são idênticos em todos os seres humanos, um substrato comum de natureza psíquica supra pessoal.

jornalista está em uma constante fase de adaptação. A conectividade ubíqua também abre caminhos para um jornalismo guiado por dados porque ela propicia, de maneira gigantesca, um ambiente favorável a coleta de dados dos usuários da *internet* espalhados pelo mundo. O uso constante das bases de dados armazenadas online traz novas habilidades para os processos criativos nas redações porque proporcionam contextualização para as reportagens que antes pareciam sem fundamento.

A rede de computadores *online* tem ainda, em sua essência, um atributo que imita em tudo os seres humanos: a memória. É que essas máquinas são capazes de armazenar uma enorme quantidade de dados de maneira estruturada ou não e, no jornalismo, isso também faz parte do processo de construção da narrativa, ou seja, da concatenação dos fatos e da notícia a ser publicada. "As bases de dados e a disponibilização da informação na rede, os arquivos disponíveis para o acionamento da memória, no momento da construção do discurso, tornam-se não somente acessíveis e facilmente pesquisáveis, mas também múltiplos" (PALACIOS, 2014, p. 96). A *web*, enquanto lugar de armazenamento, se tornou uma forte aliada das redações por ter otimizado o processo de coleta e estocagem de dados. Em um primeiro momento, era tudo que parecia ser possível fazer com os computadores. Mas já há alguns anos os profissionais de comunicação começaram a esquadrihar novos caminhos a serem percorridos nessa evolução que, como já deixamos claro, não tem nada de simplista.

A limpeza, mineração e análise desses dados guardados na memória dos computadores e da *internet* amplificou o universo do jornalismo e dos modelos de produção e distribuição de notícia. Se achávamos que o Jornalismo *Online* havia sido uma personificação mais recente do futuro da comunicação na sociedade de informação, a união de todas as características da *web* somada a um contato inédito entre jornalistas e cientistas de dados criou mais uma mutação nos moldes das redações no Brasil e no mundo. É claro que se trata de mais uma metamorfose na produção da notícia e no formato de consumo que o usuário – já livre para escolher suas próprias experiências com a notícia – terá num lugar onde as mídias se cruzam, com um poder de interatividade cada vez mais imprevisível e veloz. Até porque o uso do processamento de dados na construção das reportagens tem uma novidade em relação ao Jornalismo *Online*: apresentar a oportunidade de produzir notícias para os meios tradicionais de um jeito novo e atualizar os processos das redações dos grandes conglomerados de mídia que conservam os meios de comunicação de massa como principal elo com o público.

2.2 PONDERAÇÕES SOBRE UM FUTURO QUE ALCANÇOU O PRESENTE

Tirar os dados do papel de coadjuvante tem sido, talvez, o principal desafio desta empreitada. Apesar do número grande de definições para esta prática, continua o chamado Jornalismo de Dados sendo jornalismo fundamentalmente, mesmo provocando agitação e medo em algumas redações do mundo (SUNNE, 2016) e exigindo um número maior de profissionais que sejam tecnicamente qualificados para programar e fazer as análises dos números (HOWARD, 2014). Como toda mudança gera preocupações, os jornalistas estão apreensivos porque precisam se adaptar às novidades e à transdisciplinaridade particular desse molde para produzir as reportagens.

Ele (Jornalismo de Dados) representa a convergência de um número de campos que são significativos por si próprios - desde a pesquisa investigativa e estatística até o design e a programação. A ideia de combinar essas habilidades para contar histórias importantes é poderosa, mas também intimidadora (BRADSHAW, 2010, sem página)⁷.

Intimida porque exige que o jornalista domine – ainda que não completamente em alguns casos – outros campos do conhecimento ou que, ao menos, reconheça a igual importância de um programador e de um designer, por exemplo, para a publicação da notícia. É essa hibridização de saberes que proporciona novas experiências um fator intrigante e instigante para esta pesquisa, já que os “dados são essenciais para fazer o jornalismo de hoje mais forte do que antes” (SUNNE, 2016, sem página)⁸ e um de seus maiores benefícios é entregar ao usuário informações de interesse público – que antes estavam restritas a grupos específicos e hoje estão disponíveis graças às leis de acesso à informação – de uma maneira compreensível e inteligível para a mente humana (GRANDIN, 2014).

Os dados não falam por si só ou eximem o jornalista da responsabilidade de checar os fatos e construir uma notícia com base nos princípios éticos da profissão, mas eles oferecem caminhos para construir a informação de uma maneira inovadora, inclusive para descobrir a que tipo de audiência estamos nos dirigindo. Se antes nos baseávamos em pesquisas de institutos, por exemplo, hoje a reportagem disponível *online* dá um retorno à empresa de que

⁷It represents the convergence of a number of fields which are significant in their own right - from investigative research and statistics to design and programming. The idea of combining those skills to tell important stories is powerful - but also intimidating. Disponível em <https://www.theguardian.com/news/datablog/2010/oct/01/data-journalism-how-to-guide>. Acesso em 11 de outubro de 2018, às 20h31.

⁸ Data is essential to making the journalism of today stronger than what came before. Disponível em <https://www.americanpressinstitute.org/publications/reports/strategy-studies/data-journalism/single-page/>. Acesso em 09 de fevereiro de 2020, às 10h57.

tipos de dados o usuário quer consumir. Manovich (2001) afirma que as bases de dados traduzem o mundo a partir de um conjunto de dados cada vez mais concatenados e recombinantes. São elas que definem a composição e a estrutura da apresentação dos conteúdos de natureza jornalística (ALMEIDA; BARBOSA; TORRES, 2016) e, por isso, é importante saber que experiência queremos proporcionar ao nosso público.

As previsões para a próxima década apresentadas pelo Instituto Reuters para o Estudo do Jornalismo⁹ apontam que estreitar a conexão e recuperar a credibilidade das empresas junto ao público, assim como uma maior regulamentação da internet são alguns dos objetivos cruciais. O resultado da pesquisa também mostra que os próximos dez anos serão abalados por “uma nova onda de rupturas tecnológicas envolvendo a automação baseada em Inteligência Artificial, *Big Data*, e novas interfaces visuais e baseadas em voz”¹⁰. Para 2020, os entrevistados da pesquisa afirmaram que reconhecem a importância da Inteligência Artificial, mas destacam o medo de empresas de porte pequeno ficarem para trás, bem como de reter profissionais de tecnologia capazes de trabalhar com o dilúvio de dados no jornalismo uma vez que as empresas dessa área oferecem salários muito maiores do que a média dos conglomerados de mídia (NEWMAN, 2020). Com vistas a recuperar relevância no mercado, como buscam as empresas, o jornalismo precisa manter-se atento às atualizações impulsionadas pelo universo do *Big Data* porque deve ser mais uma época marcada pela tensão intergeracional nas redações (KUENG, 2020) e se faz necessária uma mudança de perspectiva.

As bases de dados podem ser vistas também como uma forma cultural simbólica que dão um novo sentido a várias áreas do conhecimento, incluindo o Jornalismo. Elas representam ruptura, remediação e potencialidade para esse campo da comunicação (BARBOSA, 2007). Antes elas tinham a função de organizar e estruturar uma grande quantidade de dados, uma espécie de sistema de registros computadorizados, principalmente nas primeiras máquinas de computação, que de tão grandes chegavam a ser do tamanho de uma sala. À medida que a tecnologia foi evoluindo e os computadores foram alcançando o uso individual, o acesso a esse

⁹ Relatório disponível em <http://www.digitalnewsreport.org/publications/2020/journalism-media-and-technology-trends-and-predictions-2020/>. Acesso em 01 de fevereiro de 2020, às 11h47. De acordo com a descrição de metodologia utilizada no relatório, “233 pessoas responderam a uma pesquisa fechada em dezembro de 2019. Os participantes, selecionados de 32 países, foram selecionados porque ocupavam cargos seniores (editorial, comercial ou produto) em editoras tradicionais ou nascidas em formato digital e eram responsáveis por aspectos da mídia digital ou mais ampla estratégia. Os resultados refletem essa amostra estratégica de líderes selecionados do setor, não uma amostra representativa.”

¹⁰ It will also be rocked by the next wave of technological disruption from AI-driven automation, big data, and new visual and voice-based interfaces.

tipo de conteúdo também foi se democratizando. Por isso, Manovich (2001) aponta as bases de dados como uma forma cultural, uma vez que elas passam a forjar o conteúdo da nova mídia, que agora está no ambiente virtual. Elas tornam-se importantes para todo mundo – e não só para os jornalistas – porque traduzem novos perfis de consumo dos cidadãos que buscam dados abertos.

Com esse fim, a apresentação visual das reportagens se apresenta como outra peça fundamental de aproximação entre conteúdo e usuário, já que os dados são trabalhados em algoritmos e tabelas para apresentá-los de maneira legível ao ser humano, representando uma nova era na visualização da informação. Um tempo que mostra novos caminhos visuais de conhecimento, representando, pois, mais um potencial singular e inovador desse jornalismo (BARBOSA, 2007). Isso acontece, ainda de acordo com Barbosa (2007), com base na Teoria da Difusão das Inovações (Curva do S), um conceito de Everett Rogers que define a difusão como um processo de inovação que é divulgado por meio de canais de comunicação durante determinado tempo em uma sociedade.

É dizer que o jornalismo se reconfigurou enquanto modelo de negócio e interação com o público a partir do contato direto com as bases de dados. A partir desse processo, o que Barbosa chama de Jornalismo Digital em Base de Dados (JDBD) se estabelece como um novo formato de produção da notícia. Acreditamos ser de fundamental importância elaborar um estudo que combine as duas áreas do conhecimento e possa agregar valor de leitura para ambos os grupos que lidam ou precisarão lidar, daqui em diante, com essa ferramenta híbrida e transversal que é o Jornalismo de Dados.

A necessidade de aproximação com as Ciências Sociais e com os números vem sendo proposta por Philip Meyer desde 1973 e ainda há uma escassez no número de reportagens publicadas com base em dados se compararmos ao universo do *Big Data* e ao que ele disponibiliza. Talvez um dos fatores responsáveis por essa lentidão na mudança para um celeiro de notícias com bases em dados seja um obstáculo real e comum: ainda são poucos os cursos de graduação que preparam o estudante de jornalismo para um mercado de trabalho que os utilizem com frequência como elementos fundamentais da construção da notícia (MALULY, 2016). Não obstante o perfil do jornalista esteja em constante mutação. E ela não é sutil na maioria das vezes. Pelo contrário, acontece de forma frenética.

As formas de produção jornalística se alteraram radicalmente. Hoje, o profissional participa de todo o processo jornalístico. Ele apura, fotografa, escreve, edita e publica e, para isso, deve conhecer todas as linguagens e dominar a utilização de softwares. [...] Acredito, também, que precisamos de mais trabalhos de grandes reportagens, análises, jornalismo interpretativo e a utilização de bancos de dados (MALULY, 2016, p. 38).

Essa análise feita em uma conversa com Maluly em 2016 pela professora Maria Elisabete Antonioli, que estuda a formação no Jornalismo, retrata a necessidade de orientar melhor os profissionais para essa perspectiva de mercado. No Brasil atual, como afirmado, são raras as matrizes curriculares que incluem conhecimentos relevantes para o chamado Jornalismo de Dados. De acordo com dados do Ministério da Educação (MEC)¹¹, há 420 Instituições do Ensino do Superior com o curso de Jornalismo em atividade no Brasil. Destas, 60 alcançaram nota 5 (máxima) em pelo menos uma das três principais avaliações do MEC. E outras nove obtiveram nota 5 em dois ou mais conceitos.

Analisando as matrizes curriculares das nove IES que obtiveram as maiores pontuações na avaliação do MEC, percebemos que apenas uma delas indica uma disciplina obrigatória com foco em Jornalismo de Dados. É a Escola Superior de Propaganda e Marketing (ESPM), com 36 horas-aula. Já a Universidade Federal da Bahia (UFBA), também uma das mais bem pontuadas no MEC, menciona conceitos como *Big Data* e Cultura do Algoritmo no conteúdo programático da disciplina obrigatória de Comunicação e Tecnologia. Fora dessa lista, a Universidade Federal de Minas Gerais incluiu uma disciplina de Jornalismo de Dados na grade do curso de bacharelado em jornalismo. Aos futuros profissionais que pretendem trabalhar com dados abertos usando os softwares disponíveis no mercado, resta, por enquanto, fazer cursos complementares para se aprimorarem e redescobrirem o mundo dos números e cálculos como principal opção.

Em termos de mercado, a Corporação Britânica de Radiodifusão, mais conhecida como BBC (*British Broadcasting Corporation*), tem uma equipe de dados formada por cerca de 20 jornalistas, desenvolvedores e designers. Eles são responsáveis por produzir todas os projetos com base em dados (desenvolvimento, reportagem e visualização) e têm como pré-requisito do

¹¹ Relatório processado em 25/11/2018, às 15h37. Disponível em <https://emec.mec.gov.br/emec/nova>. Instituições com matrizes curriculares analisadas: Universidade Universus Veritas Guarulhos, Universidade Federal da Bahia, Centro Universitário Ibmecc, Universidade de Brasília, Faculdade Casper Líbero, Universidade Federal de Minas Gerais, Centro Universitário da Grande Dourados, Centro Universitário Metodista e Escola Superior de Propaganda e Marketing.

trabalho saber manipular aplicativos de planilhas para analisar dados, como *Excel*, *Fusion Tables* e *Google Sheets* (LEIMDORFER, 2012). O time também utiliza softwares para análise de dados como *MySQL*, *Solr* e *Access* (LEIMDORFER, 2012). Eles acreditam que a personalização de conteúdo é o que mais chama a atenção do público (LEIMDORFER, 2012; LORENZ, 2014) e vêm criando projetos com base nessa premissa desde 1999. O *English primary school tables* usa dados publicados anualmente pelo governo. É possível comparar os indicadores das instituições de ensino primário a partir de vários filtros.

Além de ser interessante para o cidadão britânico, a página serve de fonte para os profissionais da BBC que são setoristas de educação que produzem reportagens a partir da atualização da base de dados. Antes o trabalho era feito na própria página da rede de comunicação, mas como o próprio governo passou a disponibilizar as BDs de maneira compreensível aos leitores britânicos, a BBC hoje oferece um link direto para os dados. O leitor precisa digitar apenas o código-postal ou área onde fica a escola, conforme visualizado na Figura 3.

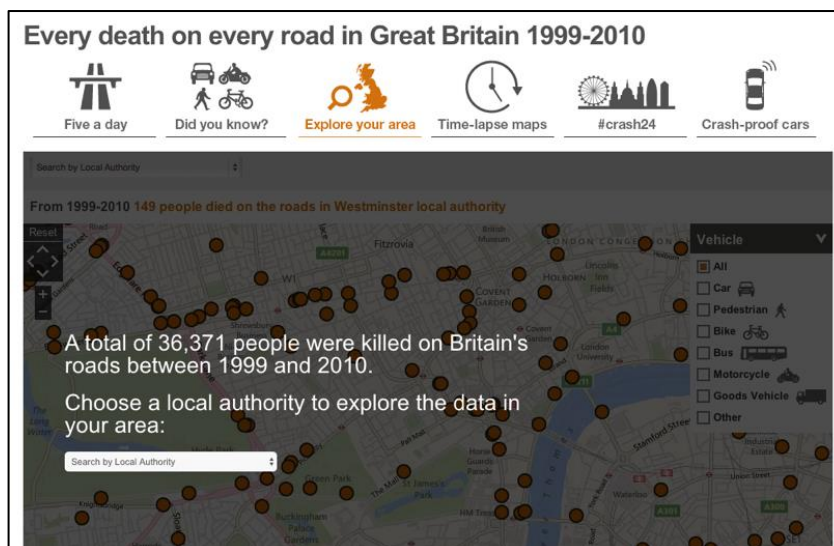
Figura 3 – Aplicação de busca de escolas por código postal ou região.

Fonte: BBC, 1999.

Leimdorfer (2012) defende que a disponibilização de dados precisa obedecer a um critério comum ao jornalismo: eles precisam ser de interesse público. Em 2011 eles criaram um relatório especial com um grande conjunto de dados sobre as pessoas que morreram nas estradas do Reino Unido entre 1999 e 2010. A pesquisa pode ser feita por códigos postais e mostram ao usuário a localização de todas as fatalidades que ocorreram no período. Para dar mais

visibilidade ao projeto, a BBC criou uma experiência transmídia¹². Durante 24 horas a rádio e a TV da empresa acompanharam ao vivo todos os acidentes que aconteceram no país durante esse dia. Os relatos também chegaram à redação via *Twitter* com o uso da *hashtag* #Crash24 e ampliaram a exploração dos dados, conforme visto na Figura 4.

Figura 4 - Homepage do projeto com aplicação de pesquisa dos dados.

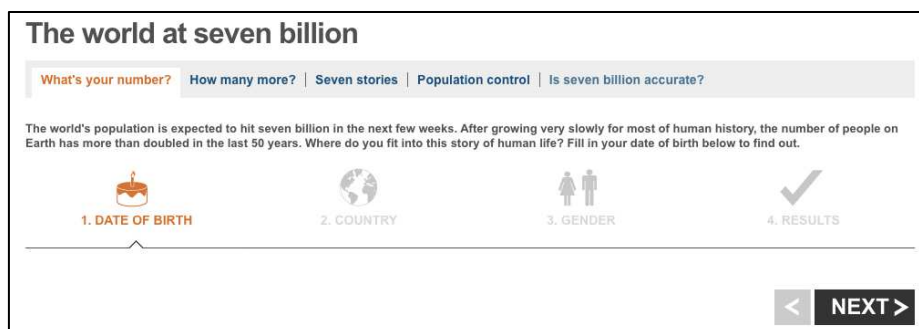


Fonte: BBC, 2011.

O uso de ferramentas simples com base em dados também atrai o leitor e faz parte dos projetos que tiveram sucesso na BBC. Elaborar uma notícia a partir de um fato pessoal aparece aqui como uma estratégia de aproximação do público com o site. Para informar que a população mundial havia chegado a 7 bilhões de pessoas, a equipe utilizou bases de dados disponibilizadas pela ONU e permitiu que o leitor descobrisse qual era o número de pessoas que habitavam o planeta na data em que ele nascera. Ao fim da pesquisa, era possível publicar o resultado no *Facebook* ou no *Twitter*. Em 2011, o link da página, demonstrado na Figura 5, se tornou o link mais compartilhado no *Facebook* no Reino Unido.

¹² Utilizamos o conceito proposto por Carlos A. Scolari que define as narrativas transmídia como "um tipo de relato onde a história se desenrola através de múltiplos meios e plataformas de comunicação, e no qual uma parte dos consumidores assume um papel ativo nesse processo de expansão" (SCOLARI, 2013, p. 46). Texto original: Un tipo de relato donde la historia se despliega a través de múltiples medios y plataformas de comunicación, y en el cual una parte de los consumidores asume un rol activo en ese proceso de expansión.

Figura 5 - Visão geral da aplicação disponibilizada na época.



Fonte: BBC, 2011.

Outras empresas importantes de comunicação também desenvolveram projetos ao longo da última década. A gigante *Australian Broadcasting Corporation* (ABC), por exemplo, é considerada - até hoje - a organização de notícias mais confiável da Austrália (CARLISLE, 2012). Em 2011, eles lançaram um projeto multiplataforma com cinco páginas de texto, mapas interativos e visualização de dados que mostrava poços de gás e arrendamentos de carvão no país. Os dados foram extraídos do governo e, para eles, são mais uma ferramenta de possibilidades no mundo digital. O alemão *Zeit Online* também se esforçou para adentrar esse universo e publicou um projeto de comparação de riqueza baseado em dados do Relatório Mundial de Educação. A análise permitiu que eles comparassem - de maneira única até então - os padrões de vida em vários países (VENOHR, 2012). O resultado deu uma grande quantidade de acesso ao site, o que justificou aos donos da empresa um investimento financeiro maior no Jornalismo de Dados. Bem feito, é uma boa fonte de geração de recursos para a empresa além de desempenhar um papel social necessário. O importante para fazer esse trabalho é ser uma fonte confiável para o público (LORENZ, 2012) e, quanto mais isso se torna uma realidade, mais esse lucro pode se converter em novas vagas de emprego nas redações.

3 A MISSÃO DE APROXIMAR JORNALISMO E TECNOLOGIA DA INFORMAÇÃO

Apesar da existência de uma proximidade entre jornalismo e dados, como demonstrado, esse encontro ainda não é uma tarefa tão simples porque muitos profissionais têm dificuldade em entender como funcionam as engrenagens por trás dos aparatos que utilizam. Alguns, sequer, sentem apreço pelo conteúdo numérico que envolve a Tecnologia da Informação. Nas redações, quando há a necessidade de manipular ajustes computacionais de várias ordens, o mais comum de se ouvir é “isso é com a equipe de TI”. Fundamentalmente é, mas talvez essa realidade encontre novas perspectivas.

A era do *Big Data* enquanto novo modelo de produção de conhecimento amplia o horizonte de prismas para buscar respostas sobre o que o uso intensivo e o atual dilúvio de dados representam para a sociedade (ARONOVA; OERTZEN; SEPKOSKI, 2013) e, certamente, para a comunicação e o jornalismo. Escrevendo o prefácio do livro *Data Mining: concepts and techniques* (HAN; KAMBER, 2001), Jim Gray, que faz parte do time de pesquisa da Microsoft, afirma que estamos – todos – inundados por dados de quaisquer áreas do conhecimento e, por esse motivo, se tornou impossível, em caráter humano, absorvê-los e aproveitá-los na sua completude.

A atenção humana se tornou um recurso precioso. Então nós precisamos encontrar caminhos para automaticamente: analisar os dados, classificá-los, resumi-los, descobrir e caracterizar suas tendências e sinalizar anomalias. Essa é uma das áreas mais ativas e empolgantes da comunidade de pesquisa de banco de dados (HAN; KAMBER, 2001, p. 7)¹³.

Essa comunidade é formada por profissionais de áreas que trabalham com visualização, estatística, inteligência artificial, aprendizado de máquina, entre outras. O fato é que os dados espelham um universo tão complexo e amplo, que para compreender como eles vêm evoluindo é preciso elaborar um caminho intrincado, por vezes. São muitas vertentes. Mas como estamos tentando traçar um percurso para compreender o que desse campo compete aos jornalistas para redesenhar nossos processos de construção de notícia, partamos do princípio: “dado é um valor atribuído a alguma coisa” (ESCOLA DE DADOS, 2013, sem página). Ele pode ser classificado,

¹³ Human attention has become a precious resource. So, we must find ways to automatically analyze the data, to automatically classify it, to automatically summarize it, to automatically discover and characterize trends in it, and to automatically flag anomalies. This is one of the most active and exciting areas of the database research community.

de acordo com a Escola de Dados (2013), desta forma: dado qualitativo (aquilo que se refere à qualidade de algo, como uma descrição de cores); dado quantitativo (naturalmente tem a ver com números que delineiam quantidade); dado categórico (aquele que coloca o item que está sendo descrito em uma categoria); dado discreto (números com brechas na sequência entre eles, ou seja, apenas números inteiros – 0, 1, 2, 3, etc.); e dado contínuo (aquele em que todos os valores são possíveis).

Para lidar com essa gama enorme de informações numéricas, a ciência precisou se atualizar e criou, entre outras invenções, os computadores. Por esse motivo as habilidades para gerar e coletar dados cresceram ao longo dos anos, principalmente depois do surgimento da *World Wide Web* como um sistema de informações global ou como uma grande base de dados (BARBOSA, 2007). Essa rede foi desenvolvida para ser “uma piscina de conhecimento humano, que permitiria a colaboradores em locais remotos compartilharem suas ideias” (BERNERS-LEE et al., 1994, p. 76)¹⁴ e, nos últimos anos, passou por muitas atualizações até ficar com a aparência e usabilidade que conhecemos hoje.

A essa altura da história da computação, o mundo já conhecia uma ciência surgida no período posterior à Segunda Guerra Mundial. A Inteligência Artificial, nome cunhado em 1956, se tornou relevante para várias esferas da atividade intelectual humana por automatizar e sistematizar tarefas (NORVIG; RUSSEL, 2004) e se tornar uma forma de aprendizado de máquina (*machine learning*). Ela começou a surgir como indústria nos anos 1980, em especial com o primeiro sistema especialista e comercial que ajudou a configurar pedidos de novos sistemas de computadores na *Digital Equipment Corporation* em 1982. A empresa conseguia, já em 1986, uma economia de 40 milhões de dólares por ano com a nova tecnologia. Turing (1950), um dos pais da Ciência da Computação e da Inteligência Artificial, começa um de seus mais famosos testes tentando responder à pergunta “máquinas podem pensar?” (TURING, 1950, p. 433)¹⁵. Alguns teóricos defendem que, se submetidas ao teste criado por Turing¹⁶, as máquinas não conseguiriam o resultado de seres pensantes, mas de simuladoras de pensamento.

¹⁴ The World-Wide Web (W3) was developed to be a pool of human knowledge, which would allow collaborators in remote sites do share their ideas and all aspects of a common project.

¹⁵ I propose to consider the question, "Can machines think?"

¹⁶ O objetivo do Teste de Turing é descobrir se uma inteligência artificial consegue se passar por um ser humano. Funciona como um jogo de perguntas e respostas. O computador teria que enganar pessoas respondendo às perguntas feitas pelos humanos. Se 30% dos humanos consultados acreditarem que as respostas foram dadas por outra pessoa, a máquina passa no teste.

O próprio Turing rejeitava a hipótese de que máquinas poderiam pensar. Em 2014, quase 70 anos depois da criação do teste, uma máquina conseguiu enganar 33% da banca de jurados da *Royal Society* na Universidade de *Reading*, em Londres (Reino Unido). *Eugene Goostman*, um *chatbot* criado por uma equipe russa, se passou por um menino de 13 anos que morava na Ucrânia e foi considerada a primeira inteligência artificial a ser aprovada no teste. Foi um marco histórico para a Ciência da Computação porque demonstrou o quanto avançamos em menos de um século com as novas aplicações que a IA permitiu criar (NORVIG; RUSSEL, 2004). É um bom começo, mas ainda há um longo caminho para que o homem crie uma tecnologia que seja capaz, de fato, de imitar o cérebro humano.

O cérebro é o computador – sistema de processamento de informação - altamente complexo, não-linear e paralelo. Ele tem a capacidade de organizar seus constituintes estruturais, conhecidos por neurônios, de forma a realizar certos processamentos muito mais rapidamente que o mais rápido computador digital hoje existente. Uma rede neural é um processador maciçamente paralelamente distribuído constituído de unidades de processamento simples, que têm a propensão natural para armazenar conhecimento experimental e torná-lo disponível para uso. Ela se assemelha ao cérebro em dois aspectos:

1. O conhecimento é adquirido pela rede a partir de seu ambiente através de um processo de aprendizagem.
2. Forças de conexão entre neurônios, conhecidas como pesos sinápticos, são utilizadas para armazenar o conhecimento adquirido (HAYKIN, 2001, p. 28).

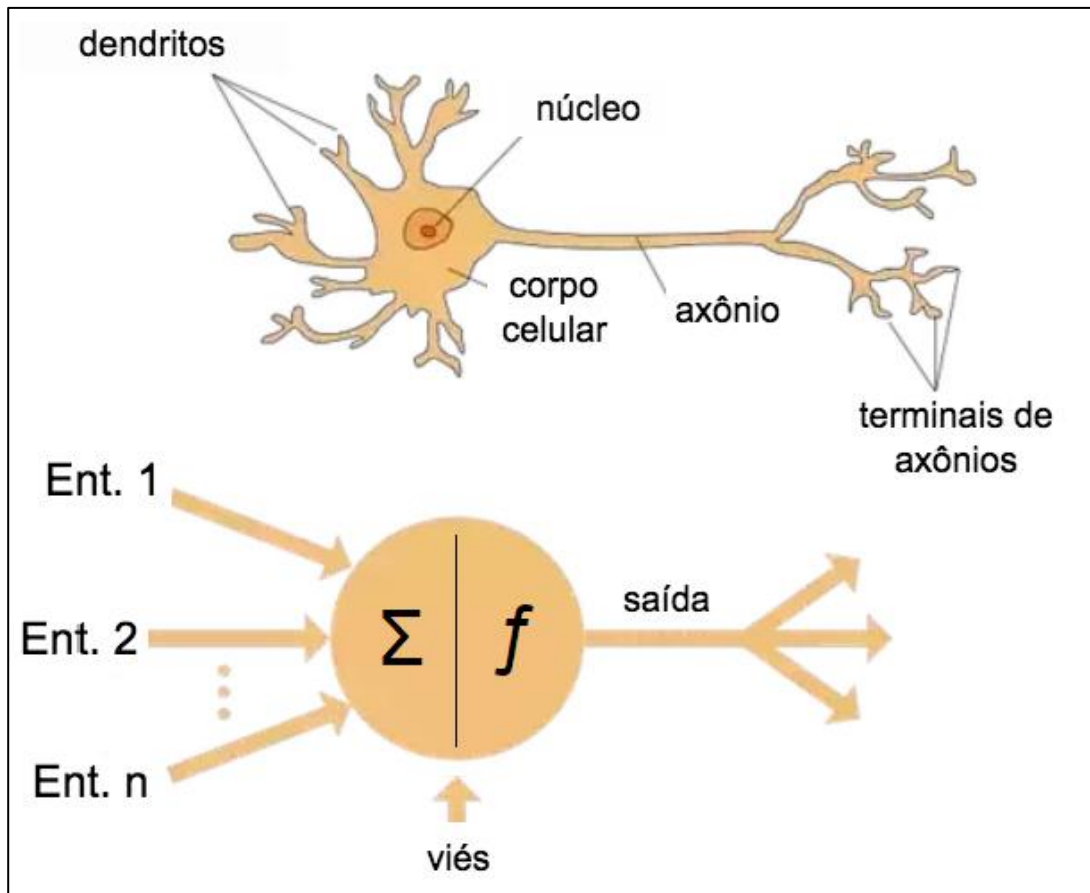
Essas redes neurais representam a forma que a Ciência da Computação encontrou para tentar mapear os conceitos do cérebro humano e desenvolver metodologias aperfeiçoadas que as aproximassem de técnicas como reconhecimento de padrões, aprendizado de máquina (*machine learning*) e estatística. O aprendizado de máquina também se dá por analogia, por indução, por evolução, por reforço e por probabilidades. No caso das redes neurais, esse aprendizado tem algumas classificações. Pode ser supervisionado, semi-supervisionado, não supervisionado, instantâneo, por pacotes, contínuo, ativo, por classificação ou por aproximação de funções (OSÓRIO, 1999). Cabe ao especialista que irá desenvolver a aplicação escolher o tipo mais adequado para a análise que se quer fazer ou problema a ser solucionado.

Para trabalhar com o método das redes neurais, os cientistas precisaram se inspirar no funcionamento dos neurônios biológicos porque acreditavam ser possível imitar algumas partes do processo cerebral humano (NAGYFI, 2018). O neurônio é formado por: corpo celular; dendritos, que são unidades estimuladoras; e axônios, que são unidades transmissoras (LUNDY-EKMAN, 2007). Os dendritos recebem sinais, como se fossem comandos. Esses

sinais são transportados pelos axônios através do corpo celular até sair de um neurônio para outro, como pode ser visto na figura 6.

A reprodução dessa atividade teve como ferramenta principal a matemática, através de modelos simplificados de funções de entrada e saída (*input e output*). Assim, entre as décadas de 1950 e 1960, Frank Rosenblatt criou o Mark 1, um *perceptron* (espécie de predecessor dos neurônios artificiais da atualidade). À época, o cientista teve a intenção de criar uma máquina que realizasse tarefas de reconhecimento de imagens para a marinha dos Estados Unidos¹⁷. Era promissor por parecer muito semelhante a um neurônio biológico, mas apresentou problemas.

Figura 6 – Estruturas do neurônio biológico e do *perceptron* mostram similaridades entre os dois modelos de processamento de informações.



Fonte: NAGYFI, 2018 / Traduzida pela autora

¹⁷ O jornal The New York Times publicou uma notícia, na época, falando sobre a nova descoberta da marinha e disponibilizando o manual para operar o *perceptron* Mark 1. Resumo da matéria do jornal disponível em <https://www.nytimes.com/1958/07/08/archives/new-navy-device-learns-by-doing-psychologist-shows-embryo-of.html>. Acesso em março de 2020. Manual disponível em <https://apps.dtic.mil/dtic/tr/fulltext/u2/236965.pdf>. Acesso em março de 2020.

O Mark 1 não foi capaz de resolver problemas complexos porque um dado tem muitas camadas e a máquina de Rosenblatt não era capaz de processar todas elas, principalmente as que ficavam ocultas. Anos mais tarde surgiu um novo tipo de neurônio artificial, com pequenas alterações no modelo criado na década de 1950. Apesar de poucas mudanças, a nova descoberta resolveu o problema inicial e foi capaz de processar um número variado de camadas sem comprometimentos. Um dos resultados dessa evolução é a tecnologia denominada Mineração de Dados, que acabou gerando uma promissora indústria nos últimos anos (CAMPOS; SAITO, 2004; HAYKIN, 2001; NORVIG; RUSSEL, 2004). Mineração de dados é um campo multidisciplinar. Envolve conhecimentos em base de dados, inteligência artificial, *machine learning*, estatística, reconhecimento de padrão e visualização de dados.

O conceito pode ser visto como uma evolução natural da tecnologia da informação (HAN; KAMBER, 2001). Para os autores, o caminho de evolução se deu a partir da descoberta de metodologias para estas etapas: coleta de dados e criação dos bancos de dados (1960); gerenciamento dos dados (1970 e 1980) [sistemas avançados (1980 até hoje); sistemas de bases de dados na web (1990 até hoje); *Data warehousing* e *data mining* (1980 até hoje)] e compreensão e análise dos dados (nova geração de sistemas de informação integrados).

Ao pé da letra, mineração de dados é extrair ou minerar conhecimento de grandes quantidades de dados. Do ponto de vista da morfologia da palavra, mineração deriva do latim *mineralis* – relativo a ‘mina’ e a ‘minerais’ (AMARAL; FILHO, 2007) - e quer dizer "ato ou efeito de minerar; trabalho de extração do minério; depuração do minério extraído das minas" (HOUAISS; VILLAR; FRANCO, 2009, p. 1293). Levando em consideração que a mineração, em essência, caracteriza o processo de extrair pedras preciosas de uma grande quantidade de matéria-prima, entendemos que a utilização desse termo para a tecnologia emergente que envolve os dados não só demonstra semelhanças nos processos de trabalho entre os campos geológico e computacional, como também posiciona os dados num patamar de riqueza a ser explorada ou, como alguns preferem afirmar, diz que os “dados são o novo petróleo” (HUMBY, 2006, p. 2)¹⁸. Para termos uma noção da força que os dados representam hoje vamos apresentar, de forma breve, alguns elementos básicos para quem quer começar a se aproximar dos conceitos de Tecnologia da Informação.

¹⁸ Data is the new oil.

O menor dígito possível gerado em um computador foi batizado *bit* (*Binary digiT* – dígito binário). Ele é chamado de binário porque pode corresponder a valores de dois números: 0 e 1. Assim, separados em dígitos, apenas os profissionais de TI conseguem entender os significados, mas a junção deles em grupos de oito *bits* forma 1 *byte*, um agrupado de informações compreensível para pessoas leigas em computação. O conjunto de 1.024 *bytes* forma 1 *kilobyte*, um termo mais conhecido do público geral por ter sido muito utilizado nas décadas de 1980 e 1990 quando da disseminação de computadores (geralmente quando precisávamos baixar uma música da internet, a velocidade era dada em *kilobytes* por segundo – kB/s). Multiplicando 1 *kilobyte* por 1.000 chegamos aos *megabytes*. Repetindo a multiplicação, chegamos ao *gigabyte*. Mais uma atualização: 1.000 *gigabytes* formam 1 *terabyte*, que talvez seja um nome já mais conhecido entre os leigos em tecnologia, já que atualmente a maioria dos computadores para uso pessoal vêm com esse poder de armazenamento.

Seguindo os cálculos, o conjunto de 1.000 *terabytes* forma 1 *petabyte*. Multiplicando os *petabytes* por 1.000 chegamos ao *exabyte*, que ainda é um número distante dos computadores comuns, mas já é realidade na internet mundial. Depois vem o *zettabyte*, com uma capacidade de armazenamento de dados que já é utilizada por empresas como o *Google*. E, por último, o *yottabyte*, que equivale a 1 trilhão de *terabytes* ou 1 quadrilhão de *gigabytes*. Ainda é uma quantia de dados inimaginável, mas essa afirmação pode estar completamente obsoleta se alguém estiver lendo-a em 2030. Para facilitar o entendimento de todos esses números, utilizamos a analogia desenhada por Bechtold (2014) no quadro 2 com algumas adaptações nossas e utilizamos o infográfico criado pela equipe do site Tecmundo (2011), conforme a figura 7.

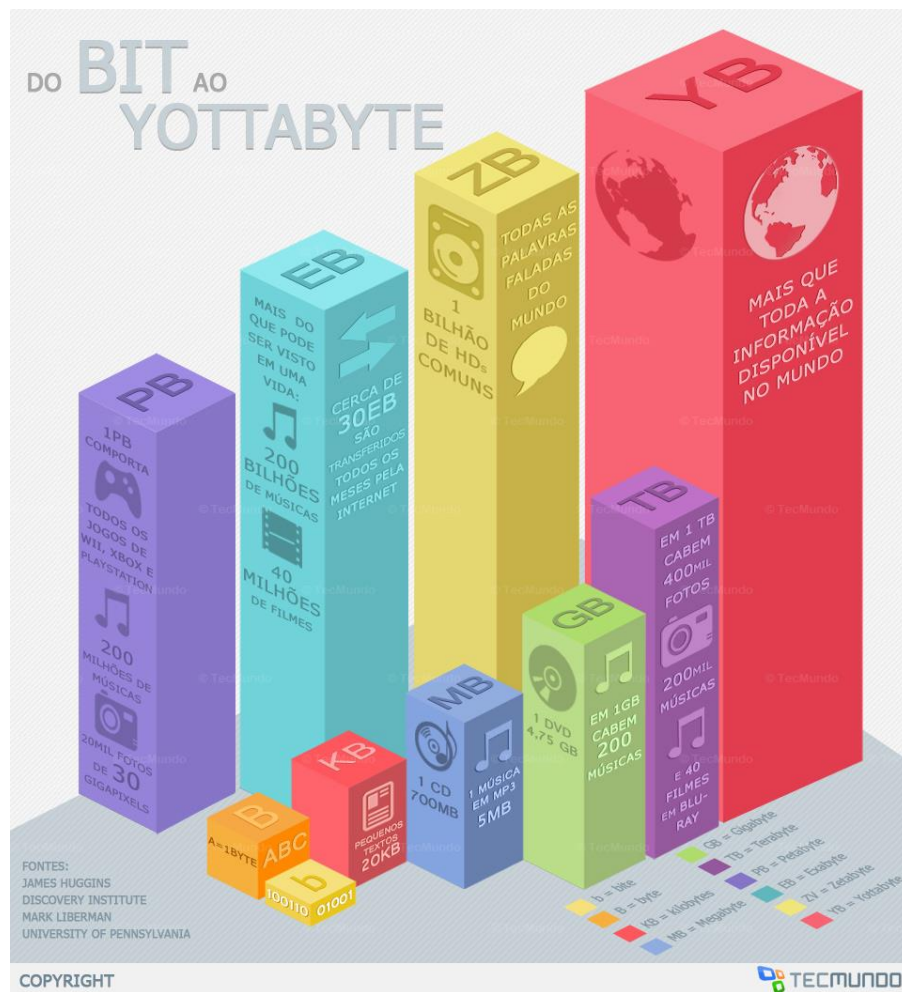
Quadro 2 – Comparação para entender os dados integrais utilizado para falar sobre capacidade de memória e armazenamento dos *hardwares*.

| Nome | Símbolo | Analogia de Bechtold (2014) |
|-----------------|---------|---|
| <i>Byte</i> | B | Um grão de arroz |
| <i>Kilobyte</i> | KB | Um copo de arroz |
| <i>Megabyte</i> | MB | Oito sacas de arroz |
| <i>Gigabyte</i> | GB | Três caminhões de tamanho médio cheios de arroz |
| <i>Terabyte</i> | TB | Dois navios de carga cheios de arroz |

| | | |
|------------------|----|--|
| <i>Petabyte</i> | PB | Cobre o povoado de Manhattan (NY, EUA) com arroz ¹⁹ |
| <i>Exabyte</i> | EB | Cobre os estados da Costa Oeste dos Estados Unidos com arroz ²⁰ |
| <i>Zettabyte</i> | ZB | Preenche a área do Oceano Pacífico com arroz ²¹ |
| <i>Yottabyte</i> | YB | Uma bola de arroz do tamanho do Planeta Terra |

Fonte: Autora / Bechtold, 2014

Figura 7 – Infográfico detalha diferença de tamanhos entre as unidades de medida de dados utilizadas na computação.



Fonte: HAMANN (2011)

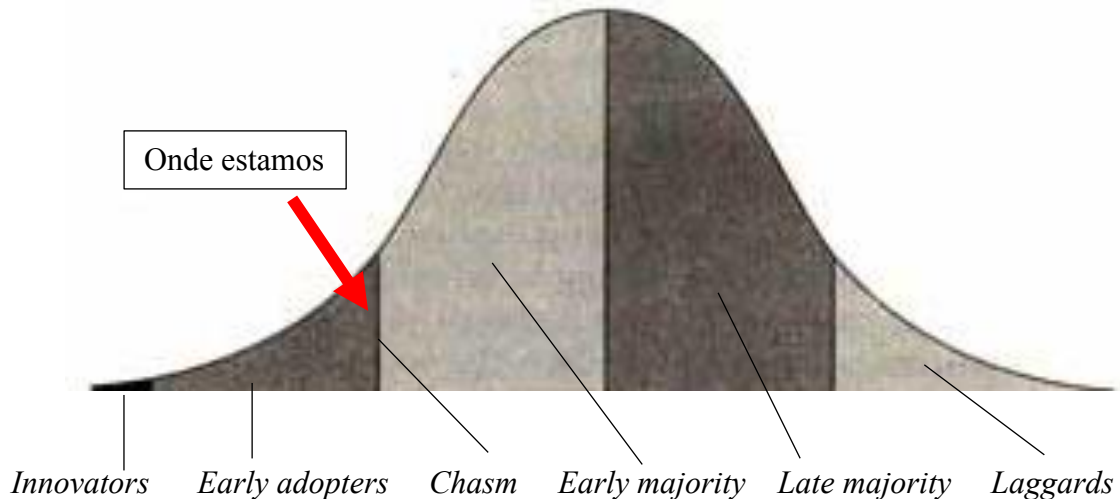
¹⁹ Correspondente a uma área de 59,1 km². Equivalente, em números aproximados, a 5.920 campos de futebol. Para fazer este cálculo e os que seguem, consideramos que 1km² equivale a 100 hectares e 1 campo de futebol tem uma área aproximada de 1 hectare.

²⁰ A Costa Oeste dos Estados Unidos é formada, basicamente, por três estados: Califórnia, Oregon e Washington. Juntos, eles somam 863.823 km², o que equivale, aproximadamente, a 86 milhões e 382 mil campos de futebol.

²¹ O Oceano Pacífico tem uma área aproximada de 161.800.000 km². Seriam 161 bilhões e 800 milhões de campos de futebol.

É natural que essa imensa quantidade de dados que começaram a ser gerados no mundo tenha dado origem a uma necessidade urgente de criar novas técnicas e ferramentas automatizadas que assistam o trabalho humano e sejam capazes de transformá-los em informações úteis. Han; Kamber (2001) descrevem um ciclo pelo qual todo tipo de tecnologia passa, incluindo a mineração de dados, conforme visto na Figura 8.

Figura 8 – Ciclo de vida de adoção das tecnologias



Fonte: HAN; KAMBER (2001)

Um primeiro momento é quando a tecnologia começa a tomar forma e a ser estudada (*Innovators*). Um segundo, em que o interesse por essa tecnologia começa a crescer e passam a ser desenvolvidos diferentes métodos para manipular a tecnologia (*Early adopters*). Um terceiro, que representa a fase de obstáculos e desafios que são encontrados antes que a tecnologia se espalhe se torne mais acessível (*Chasm*). Um quarto, quando a maturidade é alcançada e a tecnologia é bem aceita e utilizada por muita gente (*Early majority*). Depois uma maturidade já avançada em que o interesse pela tecnologia começa a cair ou porque o problema inicial não é mais de interesse ou outras necessidades surgiram (*Late majority*). E um último em que o uso da tecnologia se torna obsoleto (*Laggards*). Acreditamos que, dentro desta curva do tempo, nos encontramos em algum lugar entre o segundo e o terceiro momentos, já que lidamos com alguns obstáculos presentes na fase do abismo (*Chasm*), mas ainda há uma grande quantidade de pessoas que estão conhecendo esse universo (*Early adopters*).

Os autores entendem que, originalmente, *data mining* é uma parte do processo, mas como esse termo se popularizou rapidamente nos meios de comunicação, eles escolhem utilizar “*data mining*” como uma espécie de abreviação para descrever o conceito de que "mineração

de dados é o processo para descobrir conhecimento relevante a partir de grandes quantidades de dados armazenados, seja em bases de dados, em *data warehouses*, ou em outros repositórios (HAN; KAMBER, 2001, p. 7)²².

Mineração de dados envolve uma integração de técnicas multidisciplinares como tecnologia de bases de dados, estatística, aprendizado de máquina, computação de alta performance, reconhecimento de padrões, redes neurais, visualização de dados, recuperação de informações, processamento de imagem e sinal e análise espacial de dados (HAN; KAMBER, 2001, p. 9)²³.

Como eles afirmam que *data mining* é um processo de busca pelo conhecimento que vai muito além da mineração em si, os autores descrevem os passos necessários para que uma tarefa se cumpra através das técnicas da *data mining*.

Quadro 3 – Processo de mineração de dados

| | | |
|---|------------------------|---|
| 1 | Limpeza de dados | Remover ruídos e dados inconsistentes. |
| 2 | Integração de dados | Combinação de diferentes fontes de dados. |
| 3 | Seleção de dados | Escolher os dados relevantes para o objetivo do analista. |
| 4 | Transformação de dados | Consolidação dos dados em formatos apropriados para a mineração. |
| 5 | Mineração de dados | Momento essencial em que se aplicam métodos inteligentes para extrair padrões de dados. |
| 6 | Avaliação de padrões | Identifica os padrões que são, de fato, relevantes para o objetivo de análise. |
| 7 | Resultados | Apresentação do conhecimento extraído da análise para o usuário. |

Fonte: HAN; KAMBER, 2001, p. 7.²⁴

Para seguir esse modelo de trabalho é preciso, ainda segundo os autores, se fazer alguns questionamentos antes de começar a análise dos dados. Afinal de contas, tudo começa com perguntas (ESCOLA DE DADOS, 2013). Escolhemos chamar esse questionário de ‘Os cinco

²² Data mining is the process of discovering interesting knowledge from large amounts of data stored either in databases, data warehouses, or other information repositories.

²³ Data mining involves an integration of techniques from multiple disciplines such as database technology, statistics, machine learning, high-performance computing, pattern recognition, neural networks, data visualization, information retrieval, image and signal processing, and spatial data analysis.

²⁴ Data cleaning (to remove noise and inconsistent data); Data integration (where multiple data sources may be combined); Data selection (where data relevant to the analysis task are retrieved from the database); Data transformation (where data are transformed or consolidated into forms appropriate for mining by performing summary or aggregation operations, for instance); Data mining (an essential process where intelligent methods are applied in order to extract data patterns); Pattern evaluation (to identify the truly interesting patterns representing knowledge based on some interestingness measures); Knowledge presentation (where visualization and knowledge representation techniques are used to present the mined knowledge to the user).

quês’. Que grupo de dados quero minerar? Que tipo de conhecimento quero extrair desses dados? Que tipo de hierarquia vou utilizar na mineração? Que tipo de parâmetros serão usados para analisar os padrões encontrados? De que forma quero apresentar os resultados? (HAN; KAMBER, 2001). É importante destacar que o Jornalismo de Dados, atualmente, trabalha massivamente com bases de dados relacionais, que são apresentadas de forma estruturada, com linhas e colunas organizando os dados em uma sequência lógica para os softwares utilizados na análise. E esse processo se parece muito a uma entrevista jornalística porque se baseia em consultar os dados com perguntas elaboradas em uma linguagem diferente: SQL - *Structured Query Language* ou Linguagem de Consulta Estruturada.

Para os jornalistas que querem trabalhar com dados, a SQL se tornou uma espécie de idioma obrigatório no currículo. Dominando essa linguagem é possível elaborar consultas – chamadas de *queries* ou *query*, no singular – que vão dos níveis mais simples aos mais complexos. Reconhecendo a mineração de dados como uma tecnologia, nos damos conta de que ela “exige tempo e esforço para pesquisa, desenvolvimento e maturação” (HAN; KAMBER, 2001, p. 473). Não só para seu desenvolvimento, mas também para que os profissionais aprendam a lidar com ela. Essa afirmação se aplica perfeitamente à batalha diária travada por jornalistas que tentam aprender a minerar e analisar dados.

A linguagem SQL pode até ser simples para quem está dentro do ecossistema de Tecnologia da Informação, mas causa tensão em alguns profissionais do ramo da Comunicação. Apresentaremos um exemplo de *query* básica feita em SQL para ilustrar as diferenças entre a linguagem idiomática e a computacional. Vamos supor que queremos examinar uma tabela do Ministério da Educação com dados de todas as Instituições de Ensino Superior (IES) do país e a nossa pergunta seja: quais IES na cidade do Recife têm mais de 10 funcionários técnicos com doutorado? Essa pergunta traduzida para a SQL e seus resultados seriam apresentados conforme a Figura 9.

Figura 9 – Exemplo de consulta (*query*).

```

1 SELECT*
2 FROM DM_IES
3 WHERE NO_MUNICIPIO_IES = "Recife"
4 AND QT_TEC_DOUTORADO_FEM >= 10
5 AND QT_TEC_DOUTORADO_MASC >= 10
6
7

```

| | CO_IES | NO_IES | SGL_IES |
|---|--------|------------------------------------|---------|
| 1 | 580 | UNIVERSIDADE FEDERAL DE PERNAMBUCO | UFPE |
| 2 | 409 | UNIVERSIDADE DE PERNAMBUCO | UPE |

Fonte: elaborada pela autora na aplicação DB Browser for SQLite.

Base utilizada: Microdados do Censo da Educação Superior 2014. Disponível em:

http://download.inep.gov.br/microdados/microdados_censo_superior_2014.zip. Acesso em mar 2020.

Apesar de assustadora, à primeira vista, essa tem sido a rotina de muitos jornalistas em várias redações do mundo nos últimos anos. Principalmente depois do surgimento do que conhecemos por Dados Abertos²⁵. Eles começaram a entrar no contexto do Jornalismo de Dados no fim da década de 2000 com o lançamento do data.gov nos Estados Unidos, o primeiro portal governamental a disponibilizar dados abertos. Um mês depois, em junho de 2009, o Reino Unido lançou o 'data.gov.uk', um site que tinha a mesma lógica do estadunidense, apadrinhado por Tim Berners-Lee, o fundador da *World-Wide-Web*.

No Brasil, esse movimento chegou um pouco mais tarde. Só em 2011 é que foi promulgada a Lei 12.527 (BRASIL, 2011), também chamada de Lei de Acesso à Informação (LAI). Apesar do atraso de divulgação dos dados em relação a outras nações, o Brasil ocupa o 9º lugar no ranking de transparência da *Open Knowledge Foundation*, ficando à frente,

²⁵ "Dados abertos são dados que podem ser livremente usados, reutilizados e redistribuídos por qualquer pessoa - sujeitos, no máximo, à exigência de atribuição da fonte e compartilhamento pelas mesmas regras" (DIETRICH *et al.*, 2011, sem página).

inclusive, dos Estados Unidos²⁶. A LAI regula o acesso a informações da esfera pública e obriga das instituições estatais a divulgarem todos os dados de interesse público nos denominados Portais da Transparência.

Nos últimos anos ficou clara uma relação direta entre dados abertos e a evolução do Jornalismo de Dados, o grande volume do primeiro sendo essencial para elevar o potencial de desenvolvimento do segundo (MARTINHO, 2014). A justificativa está na definição de dados abertos: “são dados que podem ser livremente usados, reutilizados e redistribuídos por qualquer pessoa - sujeitos, no máximo, à exigência de atribuição da fonte e compartilhamento pelas mesmas regras” (DIETRICH *et al.*, 2009, sem página). O acesso a esse tipo de informação ajuda os cidadãos a compreenderem melhor a sociedade em que vivem e os modelos políticos e econômicos em que estão inseridos para que consigam fiscalizar os poderes governamentais. Para isso é ideal que ele siga alguns critérios de disponibilização, tais como completude, primazia por serem coletados diretamente da fonte maior, pontualidade na divulgação, acessibilidade e ausência de discriminação para que todos tenham acesso, ausência de decoro para que não haja interferências de entidades no processo e também ausência de requisitos de licença (SCHELLONG; STEPANETS, 2011; SIRKKUNEN, 2011). Tais características permitem que o Jornalismo de Dados não seja um mero negócio (LEIMDORFER; THEREAUX, 2012), mas simbolize uma distinção na qualidade informativa das empresas de notícia (JUNIOR; OLIVEIRA, 2016).

Manter os cidadãos em busca do consumo de notícias não é uma tarefa simples, requer sensibilidade do profissional para perceber que tipo de conteúdo produzir e como apresentá-lo ao público. Entramos em uma era que coloca na base da notícia, além do jornalista, um pedestal para outros profissionais de áreas distintas. Faz parte do processo criativo, que apresenta novidades e novos significados a uma profissão consolidada no mercado (HOWKINS, 2013) que enfrenta desafios diários para fidelizar o usuário, leitor, ouvinte ou telespectador. É um processo constante e interdisciplinar. Buscar conhecimento em novas áreas para formular pautas nas bases de dados disponibilizadas por fontes públicas, além de compilar, limpar e minerar os dados, transformando-os em informações compreensíveis para a população, é fundamental para promover novas experiências nas construções narrativas do jornalismo.

²⁶ Os últimos dados foram coletados em 2016. Disponível em <https://index.okfn.org/place/>. Acesso em 18 de fevereiro de 2020, às 22h53.

O processo criativo tem início quando nos deparamos com dados brutos que não são de fácil leitura para o público e começamos a planejar uma decodificação que vai dar forma à notícia. E até que a humanidade exija o contrário, o papel do jornalista é disponibilizar o maior número de informações possível, de modos tão distintos, que possam ser contados, sistematizados e interpretados. Os filtros dependem do repertório de cada um. A provocação para o profissional de jornalismo impulsiona-o a mergulhar no tal dilúvio de dados que tem ‘engolido’ o mundo.

Cabe, por fim, ressaltar que os esclarecimentos feitos nesse capítulo não se propõem a um aprofundamento nos conceitos de mineração de dados do ponto de vista da Tecnologia da Informação, mas a servir como uma base de conhecimento para profissionais de comunicação que buscam uma aproximação da linguagem em questão. Muitas vezes o jornalista desperta a curiosidade para aprender sobre dados, mas tem dificuldade em traduzir tantos termos técnicos comuns aos especialistas da área.

4 O JORNALISMO DE DADOS EM MEIO A UMA VASTA COLEÇÃO DE TERMOS: PRINCIPAIS CONCEITOS E DIFERENÇAS

É grande o número de expressões utilizadas pelos jornalistas para se referir ao jornalismo feito com dados. Uma das primeiras a surgir foi Jornalismo Digital em Base de Dados (JDBD), cunhada por Barbosa (2007). Segundo a autora, o JDBD inaugura uma nova geração para os profissionais porque os processos se tornam mais complexos e o público também se torna mais envolvido e exigente no ciberespaço, atribuindo a ele uma ideia de remediação por oferecer aos usuários uma informação nunca antes oferecida nesse novo estilo – interativo e interpretado de maneira distinta. Analisar dados vai além de calcular números e cruzar tabelas. Eles precisam fazer sentido para que se demonstrem de interesse público.

Para Barbosa (2007), tal novidade simboliza uma espécie de subcampo do Jornalismo Digital – daí esse termo estar presente na nomenclatura que a autora utiliza. Ela afirma que os processos de apuração da notícia demonstram pequenas mudanças, já que se baseiam em coletar dados através de mineração e raspagem nas bases; tratamento dos dados, que corresponderia ao cruzamento de tabelas e ao entendimento do contexto e informação que os dados representam; e, por fim, a apresentação dessa informação usando características puras da web, tais como hipertextualidade, interatividade, multimídia, personalização, instantaneidade, memória e ubiquidade (CANAVILHAS, 2014). Esse cenário representa um desafio porque envolve trabalhar também com sistemas de gestão das bases de dados que, geralmente, envolvem linguagens de programação das mais variadas pelas quais os dados precisam passar para serem traduzidos. Uma tarefa para jornalistas que não foram formados para lidar com essa nova realidade.

Reforçando a necessidade de enxergar as bases de dados como uma forma cultural que estrutura os novos produtos de mídia, Barbosa (2007) delinea o conceito de Jornalismo Digital em Base de Dados (JDBD) da seguinte forma:

“é o modelo que tem as bases de dados como definidoras da estrutura e da organização, bem como da apresentação dos conteúdos de natureza jornalística, de acordo com funcionalidades e categorias específicas, que vão permitir a criação, a manutenção, a atualização, a disponibilização e a circulação de produtos jornalísticos digitais dinâmicos” (BARBOSA, 2007, p. 218).

Esse novo formato, então, possibilitaria a criação de produtos digitais metainformativos, inteligentes e ativos, uma vez que com os sistemas de produção, edição e distribuição são

automatizados e com a utilização das tecnologias das BDs a notícia muda de configuração, ou seja, o processo jornalístico de atualização da notícia além de ser automático envolve também a mudança na forma de informar. Por esse motivo, o modo de apresentação das informações evolui com novos conceitos para a elaboração do conteúdo e propõe mudanças também na figura do profissional, que precisa dominar outras áreas do conhecimento – ainda que de maneira inicialmente superficial – para lidar com os novos processos. Bem como o novo tipo de leitor que vai surgindo, cada vez mais participativo e ocupando também os papéis de fonte e coautor da notícia.

O trabalho com dados posiciona o jornalista em um papel de cientista em busca de uma verdade precisa e verificável, como sugeriu Meyer (1991). Os dados brutos, conforme explicitamos anteriormente, não dizem nada quando sozinhos, crus. Cabe ao profissional de comunicação compreendê-los, exercer um papel intermediário para análise e disponibilizá-los de maneira inteligível para o público (GEHRKE, 2016). O jornalismo ocuparia, neste caso, um espaço híbrido entre meio de comunicação e meio de conhecimento. Para Gehrke (2016), o Jornalismo de Dados segue um processo específico de produção com base na metodologia científica proposta por Meyer (1973) e no conceito do jornalismo como canal de conhecimento proposto por Park (1940). Ela utiliza o modelo apresentado na Figura 10.

Figura 10 – Modelo de Jornalismo de Dados proposto por Gehrke (2016).



Fonte: GEHRKE, 2016, p. 11

Para Gehrke (2016), o conceito de Jornalismo de Dados nasce da junção entre o jornalismo como forma de conhecimento e o jornalismo de precisão, ambos propostos no século 20. Eles formariam, juntos, um aprimoramento do modelo de produção da notícia e quebram o processo metabólico dos moldes tradicionais adotados pelas redações. O que, conforme já afirmado, representa um estímulo à inovação e à evolução dos métodos no jornalismo, que, ao fim, é a palavra-chave de todo e qualquer conceito de Jornalismo de Dados (SUNNE, 2016).

O desenvolvimento de produções jornalísticas baseadas em dados implica um novo desafio para os meios e profissionais da comunicação que começam a analisar aspectos relacionados com a gestão de novas habilidades, a capacitação, o trabalho em equipe e os modelos de negócios que tornem essa narrativa rentável (BRIZUELA, 2015, p. 151)²⁷.

²⁷ El desarrollo de producciones periodísticas basadas en datos implica un nuevo desafío para los medios y profesionales de la comunicación que comienzan a analizar aspectos relacionados con la gestión de nuevas habilidades, la capacitación, el trabajo en equipo y los modelos de negocios que permitan hacer rentable esta narrativa.

Analisando as empresas de comunicação da América Latina, a autora afirma que é importante refletir sobre a importância do Jornalismo de Dados no futuro da profissão e diz que há cinco razões que justificam essa ponderação. A primeira, segundo Brizuela (2015), é que as empresas de comunicação estão demonstrando ser possível fazer reportagens sólidas com base em dados e passam por uma fase de reestruturação e planejamento para traçar as estratégias de trabalho nesse novo padrão. A segunda razão é o fato do Jornalismo de Dados se mostrar desafiador porque envolve um esforço árduo dos profissionais para entrarem no universo dos números e estarem confiantes o suficiente para não cometerem erros na análise de dados e na construção da reportagem porque isso seria fatal para a credibilidade do profissional e do veículo para o qual ele trabalhar.

A terceira envolve a multidisciplinaridade e colaboração como características fundamentais desse novo processo, já que os jornalistas precisam focar em descobrir as histórias de interesse público por trás dos números enquanto confiam no apoio de uma *squad* formada por cientistas de dados e designers capazes de finalizar a reportagem, uma relação que enriquece o resultado por propiciar diferentes olhares sobre uma mesma realidade. Brizuela (2015) vai além e diz que a quarta razão para se refletir sobre o *Database Journalism* é que ele revaloriza a audiência porque permite que os usuários interajam com as plataformas criadas por algumas reportagens e até personalizem o conteúdo, estreitando laços com o jornalismo enquanto informação com poder de formação. A quinta e última razão para reflexão é porque o JD proporciona uma mudança social positiva em favor da democracia, agindo como um fiscalizador do poder público uma vez que irá buscar portais de transparência atualizados e devidamente preenchidos.

O conceito de Jornalismo Digital em Base de Dados exposto por Barbosa (2007) ganhou outra nomenclatura com o passar dos anos. É que com tantas evoluções desde o Jornalismo de Precisão, passando pela RAC (Reportagem Assistida por Computador) e pelo Jornalismo de Dados como um braço do Jornalismo Digital. Barbosa; Torres (2013) afirmam que o JDBD é um paradigma em expansão contínua, justificando a terminologia que começa a surgir na segunda década do século 21: *data-driven journalism* ou Jornalismo Guiado por Dados (BRADSHAW, ROHUMAA, 2011; GRAY, BOUNEGRU, CHAMBERS, 2012). Para Lewis; Usher (2013) ele seria a perfeita fusão de jornalismo e ciência da computação num cenário que reconfigura a rotina jornalística porque põe os programadores no centro da redação.

Apesar de estarem trabalhando em um mesmo ambiente, eles não ocupam o cerne do processo de construção da notícia porque essa nova terminologia devolve o protagonismo do jornalista. Seria o Jornalismo Guiado por Dados “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, 2014a, P. 90). Essa definição recorda um alerta que Meyer (1991) dava para o perigo de colocar um foco excessivo na tecnologia dentro do campo de abrangência jornalística. Segundo ele, o aspecto mais importante da RAC é o conhecimento gerado a partir dela e não as ferramentas utilizadas durante o processo de construção da reportagem nas redações.

Por esse motivo, a expressão Jornalismo Guiado por Dados teria uma vantagem semântica ao direcionar a atenção para os resultados jornalísticos do processo que chegam ao público através da reportagem.

O JGD tem por objetivo, justamente, a 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 (computadores pessoais, *smartphones*, *tablets*), na geração de visualizações e infografias. Principalmente, as técnicas de JGD permitem ao jornalista encontrar informação com valor noticioso em bases de dados com milhares ou milhões de registros, dificilmente manejáveis sem a ajuda de computadores. Facilitam, ainda, o cruzamento de diferentes bases de dados, para a produção de novo conhecimento sobre a sociedade, a ser apresentado em narrativas que se estendem dos jogos eletrônicos e *mash-ups*²⁸ às matérias tradicionais em texto, audiovisual e imagem (TRÄSEL, 2013, p. 2).

Esses objetivos colocam o JGD não só como extensão do JDBD (BARBOSA; TORRES, 2013), mas também como descendente direto do Jornalismo de Precisão (MEYER, 1973) porque guiam o jornalismo para um caminho contrário ao do senso comum, utilizando metodologias de computação em busca da informação para o público. Nesse caso, a tecnologia não substitui o repórter e, sim, expande o cardápio de pautas que podem se tornar reportagens menos ou mais complexas para revelar histórias que antes não eram facilmente encontradas. Não seria um novo tipo de jornalismo, mas um reforço na missão que esta profissão assumiu com a defesa daquilo que se mostra de interesse público, bem como “uma prática capaz de criar novos caminhos para a narrativa noticiosa e um possível atrativo para uma audiência composta, cada vez mais, por leitores nascidos e criados na cibercultura” (TRÄSEL, 2013, p. 15). O

²⁸ [Grifo de Träsel (2013)] Um *mash-up* é compreendido, no contexto da informática, como uma aplicação para a Web produzida pelo amálgama de duas ou mais fontes diferentes, gerando um terceiro produto. Em geral, as fontes colocam os dados voluntariamente à disposição.

Jornalismo Guiado por Dados seria, então, uma ferramenta para trazer informação relevante e de credibilidade, apresentando-as de inúmeras formas narrativas ao público (AITAMURTO, LEHTONEN, SIRKKUNEN, 2013). Esse método é importante porque democratiza o acesso aos recursos e às fontes de dados quando permite que as pessoas não só leiam as reportagens, mas também verifiquem a lógica utilizada pelo jornalista e possam replicá-la para fazer suas próprias análises além de comprovar o que foi feito na reportagem, assim como faz a ciência (BOUNEGRU, 2012; ANDERTON-YANG et. al., 2012). É mais uma alteração no padrão de produção e de consumo da notícia, que exige uma mudança de ou no público.

Se antes as pessoas esperavam comprar o jornal ou ver a TV para saber o que aconteceu na noite anterior, hoje as notícias acontecem a todo tempo e o uso de dados permite análises mais profundas do que entendemos por notícia e abre a capacidade do jornalista pensar a produção dessa notícia de uma maneira completamente diferente, pegando números amplos e abstratos e transformando em uma informação de valor público que seja legível por qualquer pessoa (LORENZ, 2012; BRADSHAW, 2012). Na verdade, é papel do jornalista dar sentido a esses números que, à primeira vista, parecem não dizer muita coisa, se mostram cansativos e entediantes, mas quando observados sob uma perspectiva analítica contêm *insights* poderosos, que acabam tornando as informações menos impenetráveis.

Estabelecer uma relação direta entre a RAC – Reportagem Assistida por Computador – e o que conhecemos por Jornalismo de Dados pode se mostrar uma tarefa complexa. Alguns pesquisadores argumentam não haver diferença significativa entre os dois, mas reconhecem que há diferenças no processo de tratamento e análise das bases em entre essas abordagens nas práticas mais recentes dos meios de comunicação (BOUNEGRU, 2012).

Alguns argumentam que há uma diferença entre a RAC e o jornalismo de dados. Eles dizem que a RAC é uma técnica para coletar e analisar dados como uma maneira de aprimorar a reportagem (geralmente investigativa), enquanto o jornalismo de dados presta atenção à maneira como os dados ficam em todo o fluxo de trabalho jornalístico. (BOUNEGRU, 2012, sem página)²⁹.

De acordo com o Instituto Nacional de Reportagem Assistida por Computador dos Estados Unidos (NICAR), o termo 'reportagem assistida por computador (RAC)' foi muito utilizado nos últimos anos para descrever o que muitas pessoas chamam de 'Jornalismo de

²⁹ Some argue that there is a difference between CAR and data journalism. They say that CAR is a technique for gathering and analyzing data as a way of enhancing (usually investigative) reportage, whereas data journalism pays attention to the way that data sits within the whole journalistic workflow.

Dados³⁰. Hammond (2017), por outro lado, coloca o Jornalismo de Dados como a antítese da RAC porque acredita que esta está mais centrada no desenvolvimento científico da notícia apontado por Meyer (1973). Já o Jornalismo de Dados atual, para Hammond (2017), ecoa essas afirmações de cientificidade e democracia de acesso a informações escondidas nos números, mas não por um mérito puro do jornalista e sim pela facilidade de acesso a eles com a chegada da era do *Big Data*.

Opinião semelhante é compartilhada por Anderson (2008), que defende as diferenças entre causalidade e correlação nesse cenário. Para ele, a ciência que ainda busca os porquês está obsoleta já que a quantidade de dados existentes hoje, uma espécie de dilúvio, comprova que as teorias de causa e consequência – ou hipótese, modelo e teste – podem ser, na verdade, mera coincidência. Seria como dizer que a Física se transformou em uma área de especulações teóricas porque ainda não fomos capazes – nós, humanos – de executar os experimentos que falsificariam as hipóteses de Isaac Newton e de Albert Einstein, por exemplo. É um outro momento: a era dos *Petabytes*.

Kilobytes foram armazenados em disquetes. *Megabytes* foram armazenados em discos rígidos. *Terabytes* foram armazenados em matrizes de disco. *Petabytes* são armazenados na nuvem. À medida que avançamos nessa progressão, passamos da analogia da pasta para a analogia do gabinete de arquivo para a analogia da biblioteca para - bem, em *petabytes* ficamos sem analogias organizacionais. Na escala de *petabytes*, a informação não é uma questão simples de taxonomia e ordem tridimensionais ou quadridimensionais, mas de estatísticas agnósticas dimensionalmente. Ela exige uma abordagem totalmente diferente, que requer a perda das amarras dos dados como algo que pode ser visualizado em sua totalidade. Nos obriga a visualizá-los matematicamente primeiro e a estabelecer um contexto para eles posteriormente (ANDERSON, 2008, sem página)³¹.

Significa dizer que, segundo Anderson, números falam por si só diante da quantidade massiva de dados que há atualmente. Eles não dependem mais de um modelo que os contextualizem e os ofereçam produção de sentido. Para Anderson (2008) esse é o modelo antigo da ciência. É preciso recordar, no entanto, que mesmo tendo essa possibilidade no mundo jornalístico, como já estabelecido, o papel humano e analítico do jornalista permite

³⁰ Disponível em <https://www.ire.org/nicar/about>. Acesso em 24 de janeiro de 2020, às 18h40.

³¹ Kilobytes were stored on floppy disks. Megabytes were stored on hard disks. Terabytes were stored in disk arrays. Petabytes are stored in the cloud. As we moved along that progression, we went from the folder analogy to the file cabinet analogy to the library analogy to — well, at petabytes we ran out of organizational analogies. At the petabyte scale, information is not a matter of simple three- and four-dimensional taxonomy and order but of dimensionally agnostic statistics. It calls for an entirely different approach, one that requires us to lose the tether of data as something that can be visualized in its totality. It forces us to view data mathematically first and establish a context for it later.

interpretações e construções de narrativas importantes para que uma informação de qualidade e útil chegue ao público.

O fato de existirem máquinas com sistemas capazes de escrever uma reportagem baseada em códigos não justifica a substituição do jornalista porque este está habilitado a compreender outras variáveis que envolvem o mundo da notícia: os contextos econômico e social do público e as formas culturais nas quais ele está inserido, por exemplo. É claro que a programação de computadores simboliza uma ruptura de padrão significativa – e reconhecemos esse fato, – mas seria um tanto imprudente defender a completa reposição do profissional de jornalismo pelo raciocínio puramente lógico dos códigos binários. É preferível sustentar a ideia de que a tecnologia ampara e legitima o trabalho jornalístico, não o suprime.

Com os dados não seria diferente. Eles não dizem muito sozinhos, ao contrário do que afirma Anderson (2008), porque ainda não conseguem ser apresentados de uma maneira inteligível para a mente sem que haja interferência humana. Geralmente um dado isolado ganha significado quando o comparamos com outra coisa. Vamos tomar como exemplo a distância quilômetro (KM). O dado '500 km', solitário como se mostra, não parece ter muita importância. Mas se digo que a cidade de Xangai, na China, estabeleceu um novo recorde ao ter o maior sistema de metrô do mundo com um comprimento de 500 km estou produzindo e dando sentido a esse dado. Temos uma informação.

O jornalismo vai além. Ao elaborar uma reportagem comparando o tamanho da linha do metrô de Xangai com as linhas de Paris, Nova Iorque e de São Paulo, essa informação passou por três processos: foi aprendida, aplicada e entendida. Por fim, produziu conhecimento (ESCOLA DE DADOS, 2013). O Jornalismo Guiado por Dados seria alterar um pouco a centralidade da atividade jornalística. Holovaty (2006) defende que ela saia do campo da história e passe para a esfera dos dados.

Não estou dizendo que os jornais devam se concentrar completamente em vastas coleções de dados, abandonando completamente o formato de uma notícia. Os artigos de notícias são ótimos para contar histórias, analisar questões complexas e todo tipo de outras coisas. Um artigo - uma “grande bolha de texto” - geralmente é a melhor maneira de explicar conceitos. As nuances do idioma inglês não são mapeadas corretamente para fontes de dados manipuláveis por máquina. (Essa entrada, que você está lendo no momento, é um excelente exemplo de algo que não pode ser substituído por um banco de dados.) Quando digo que "os jornais precisam interromper a visão de mundo centrada na história", não quero dizer "os jornais precisam abolir as histórias." As duas formas de divulgação de informações podem coexistir e se

complementam. [...] Digamos que um jornal tenha escrito uma história sobre um incêndio local. Ser capaz de ler essa história em um telefone celular é bom e elegante. Viva, tecnologia! Mas o que eu realmente quero fazer é explorar os fatos brutos dessa história, um por um, com camadas de atribuição e uma infraestrutura para comparar os detalhes do incêndio - data, hora, local, vítimas, número do quartel dos bombeiros, distância do corpo de bombeiros, nomes e anos de experiência de bombeiros em cena, tempo que levou para os bombeiros chegarem - com os detalhes de incêndios anteriores. E incêndios subsequentes, sempre que acontecem (HOLOVATY, 2006, sem página)³².

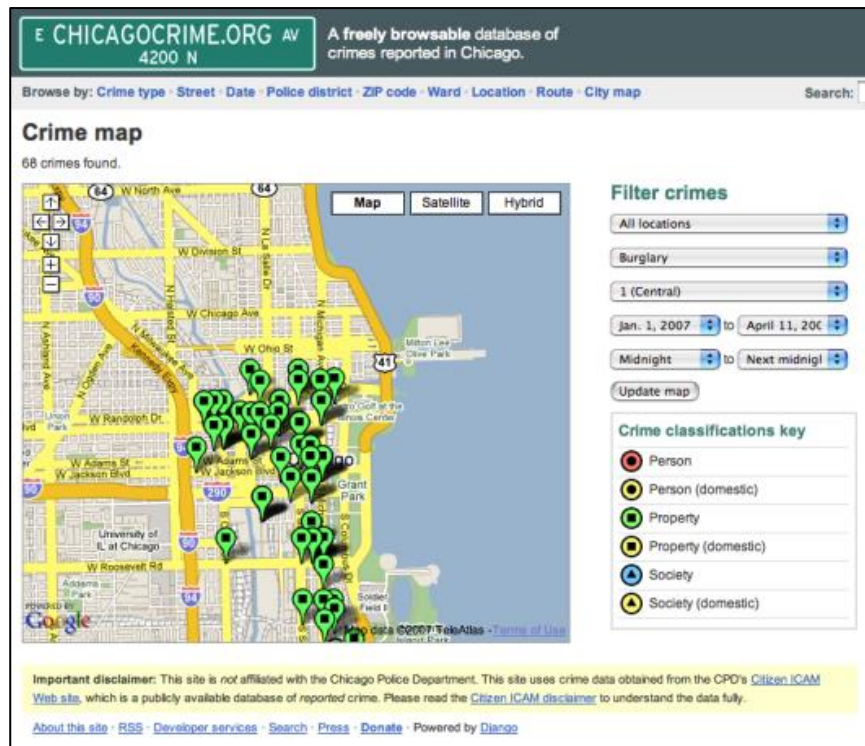
Adrian Holovaty é um desenvolvedor web formado em jornalismo que foi responsável por um dos primeiros registros da terminologia simplificada (não simplista) de Jornalismo de Dados (HOLOVATY, 2006). Ele foi um dos primeiros a trabalhar o jornalismo através da programação de computadores entre o fim da década de 1990 e os anos 2000. Em 2005 criou o famoso projeto *Chicago Crime.Org*, um projeto que ficou disponível de 2005 a 2008 e consistia em um *mash-up* de mapas com um banco de dados de crimes relatados na cidade de Chicago, nos Estados Unidos.

Os *scripts* desenvolvidos em *Python* coletavam os dados do Departamento de Polícia da cidade uma vez por semana e utilizavam os mapas do Google para possibilitar a visualização de crimes de maneira dinâmica, além de identificar a área policial a qual o morador pertencia. Um outro recurso permitia visualizar detalhes dos crimes agrupados por bairros da cidade. O projeto venceu o *Batten Awards* de Inovações em Jornalismo no ano de 2005³³ e vários moradores entraram em contato com Holovaty para agradecer a criação do projeto e contar que estavam marcando reuniões policiais com os vereadores da cidade para discutirem os dados mostrados no site. De acordo com o autor, o *Chicago Crime.Org*, demonstrado na Figura 11, também ajudou o Google a abrir a API de mapeamento geográfico – hoje o tão comumente utilizado ‘*Google Maps*’.

³² I’m not saying newspapers should turn completely to vast collections of data, completely abandoning the format of a news article. News articles are great for telling stories, analyzing complex issues and all sorts of other things. An article — a “big blob of text” — is often the best way to explain concepts. The nuances of the English language do not map neatly to machine-manipulatable data sources. (This very entry, which you’re reading right now, is a prime example of something that could not be replaced with a database.) When I say “newspapers need to stop the story-centric worldview,” I don’t mean “newspapers need to abolish stories.” The two forms of information dissemination can coexist and complement each other. Say a newspaper has written a story about a local fire. Being able to read that story on a cell phone is fine and dandy. Hooray, technology! But what I really want to be able to do is explore the raw facts of that story, one by one, with layers of attribution, and an infrastructure for comparing the details of the fire — date, time, place, victims, fire station number, distance from fire department, names and years of experience of firemen on the scene, time it took for firemen to arrive — with the details of previous fires. And subsequent fires, whenever they happen.

³³ Prêmio do Laboratório de Jornalismo do Instituto para o Jornalismo Interativo (J-Lab: The Institute for Interactive Journalism). Disponível em <https://www.j-lab.org/2005/09/12/chicagocrime-org-wins-10000-knight-batten-award/>. Acesso em 11 de janeiro de 2020, às 8h08.

Figura 11 – Homepage do Chicago Crime.Org, 2005.



Fonte: J-Lab - The Institute for Interactive Journalism, 2005.

Ao escrever o ensaio ‘Uma maneira fundamental de mudança para os sites de Jornalismo’³⁴, Holovaty (2006) inspirou a criação do PolitiFact³⁵, apresentado na Figura 12, que ganhou o prêmio Pulitzer de Relatórios Nacionais pela cobertura das eleições presidenciais em 2008. O projeto foi lançado em agosto de 2007 e examinou mais de 750 declarações políticas utilizando bancos de dados online para verificar a veracidade das declarações que envolviam os candidatos, incluindo as que eles próprios faziam durante os atos de campanha. O trabalho foi feito com uma equipe de cinco repórteres e editores do St. Petersburg Times.

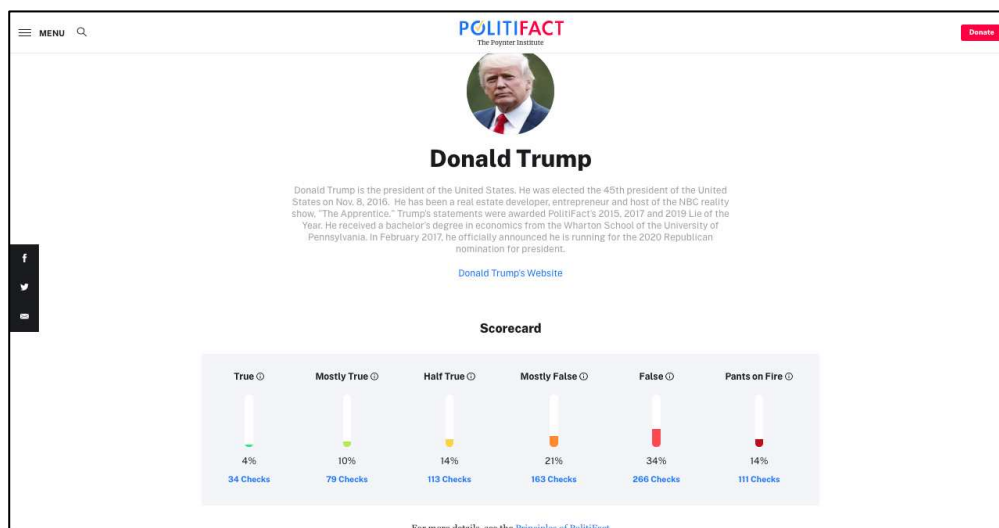
A equipe criou uma espécie de medidor da verdade, o *Truth-O-Meter*, com seis níveis Verdadeiro; Principalmente verdadeiro; Meio verdadeiro; Quase verdadeiro; Falso; e *Pants on fire* (expressão em inglês utilizada para dizer que alguém foi pego na mentira). O projeto também ganhou um prêmio da *National Press Foundation* para jornalismo online, um prêmio *Knight Batten* de Inovação em Jornalismo, um *Digital Edge Award* de melhor site de notícias da Associação de Jornais dos Estados Unidos e primeiro lugar para reportagem política online

³⁴ A fundamental way newspaper sites need to change. Disponível em <http://www.holovaty.com/writing/fundamental-change/>. Acesso em 18 de novembro, às 18h20.

³⁵ Disponível em <https://www.politifact.com/>.

no *Green Eyeshade Awards*. O PolitiFact continua ativo e com projetos de verificação de fatos utilizando bancos de dados online. A equipe já começou a trabalhar para a campanha presidencial de 2020, checando as declarações feitas pelos pré-candidatos. Entre eles, o atual presidente do país, Donald Trump.

Figura 12 – Página de verificação das declarações do presidente Donald Trump.



Fonte: PolitiFact, 2020.

É importante frisar que o Jornalismo de Dados não significa meramente uma junção entre ‘jornalismo’ e ‘dados’. Esses dois conceitos são problemáticos e complexos. Poderíamos resumir dizendo que “os dados podem ser a fonte do Jornalismo de Dados ou a ferramenta com a qual a história é contada – ou podem ser os dois” (BRADSHAW, 2012, sem página). Ele se diferencia da RAC porque não envolve somente o uso de computadores, mas exige que o profissional tenha um conhecimento básico em sistemas computacionais, métodos de as Ciências Sociais e estatísticas (BOUNEGRU; CHAMBERS; GRAY, 2012; HOWARD, 2014). Também vai além dos conceitos de JDBD e do JCD porque apenas a utilização de dados nas reportagens não implica uma mudança na estrutura do processo de produção da notícia. O Jornalismo de Dados, então, "estaria mais próximo de uma fusão com os princípios da cultura dos dados abertos e com o trabalho computacional (MANCINI; VASCONCELLOS, 2016, p. 73). O processo começa na quantificação dos números e é complementado pela análise dos dados que levam, por fim, ao conhecimento moldado na informação.

4.1 ABORDAGENS CONCEITUAIS, PRÁTICAS E TECNOLÓGICAS INCORPORADAS AO JORNALISMO DE DADOS

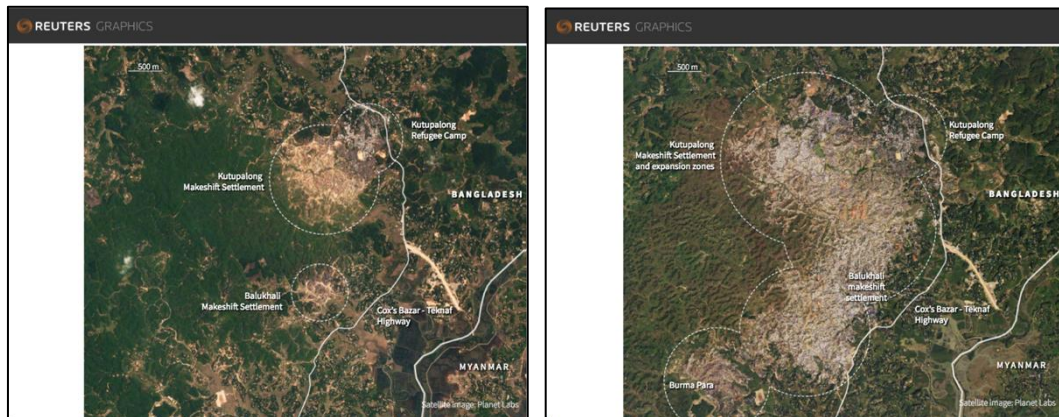
Os dados acrescentam novos elementos ao fazer jornalístico e geram novos produtos: artigos baseados em dados com uso de programação para extração, mineração e análise; visualizações interativas através de tratamento de um grande volume de dados e trabalho de designers e desenvolvedores para a criação dos *dashboards*; além de conjuntos de dados abertos e aplicações de notícias, que também utilizam a programação para tratar os *datasets* e compartilhar com o público os dados tratados ou até mesmo brutos, com uma limpeza superficial de estruturação para análise direta do leitor (MARTINHO, 2013; CRUCIANELLI, 2013). Esses produtos são encontrados na mídia com uma maior facilidade nos últimos anos.

Em 2017, a reportagem ‘Vida nos acampamentos’ (*Life in the camps, Reuters, UK*³⁶) apresentou um gráfico produzido pelos designers Simon Scarr e Weivi Cai combinando vídeo, fotografia, imagem de satélite, mapas e textos para mostrar as condições de vida nos acampamentos de refugiados dos povos *Rohingya*, situados em Bangladesh. São mais de 800 mil muçulmanos que vivem na área que tem um crescimento populacional veloz dado o êxodo em larga escala do grupo de origem no Myanmar, tendo um salto de quase 114 mil moradores para 598 mil em apenas três meses daquele ano, conforme demonstrado na Figura 13.

A reportagem denuncia as más condições em que essas pessoas vivem e dá ao uso de mapas um novo sentido na narrativa jornalística. Foi através deles que a equipe conseguiu identificar o modelo de acesso a água e esgoto, constatando que em algumas áreas com maiores índices endêmicos de doenças, os canais da água a ser consumida estavam muito próximos a banheiros improvisados construídos pelos próprios refugiados. Esses dados foram comparados às diretrizes básicas de saúde pública das Nações Unidas e os fotógrafos da Reuters documentaram e reproduziram o que diziam os mapas, como apresenta a Figura 14.

³⁶ Disponível em <http://fingfx.thomsonreuters.com/gfx/rngs/MYANMAR-ROHINGYA/010051VB46G/index.html>. Acesso em junho de 2018.

Figura 13 – O contraste populacional da mesma área vista em um intervalo de três meses.



Fonte: Reuters Graphics, 2017.

Figura 14 - Exemplo de banheiro ao lado de fonte de água.



Fonte: Reuters Graphics, 2017.

A reportagem concluiu que a sujeira na água consumida pelos refugiados era responsável por grande parte das enfermidades: Infecção respiratória aguda (29%); Diarreia aguda causada por água (21%); doenças de pele (9%) e febres sem explicação (28%). A mesma história poderia ser contada numa narrativa audiovisual ou mesmo num texto de jornal e revista, mas o impacto causado no usuário ao abrir o *website* e lidar com tantas informações visuais que fazem o cérebro colocar dezenas de neurônios para trabalhar a imagem torna a experiência completamente diferente e requalifica o modo de estruturar uma reportagem jornalística.

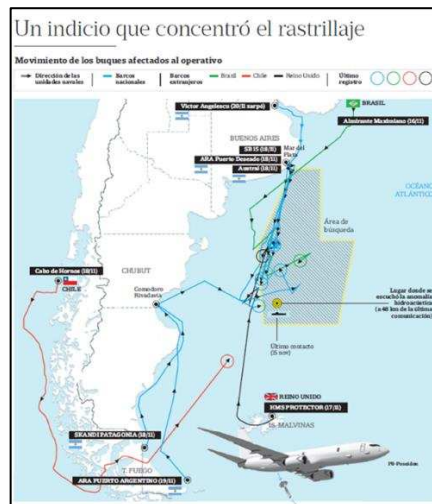
O Jornalismo de Dados acaba por representar um conjunto de mudanças na sociedade, que busca um consumir reportagens menos baseadas em crenças pessoais e mais focadas em

métricas quantitativas que possam oferecer análises e reflexões sobre o contexto em que se vive (ARAÚJO, 2016; MEYER, 1991).

Vários estudos de práticas jornalísticas verificam que os jornalistas geralmente trabalham com dados de outras fontes – governos, instituições públicas, empresas de pesquisa, etc. Apenas em alguns casos de jornalismo investigativo os profissionais foram capazes de encontrar recursos para coletar e analisar os próprios dados e publicar os resultados encontrados. Mas a situação está mudando rapidamente. Cada vez mais, há dados dos mais diversos disponíveis na internet e ferramentas eficientes com as quais qualquer pessoa pode analisar, visualizar e publicar grandes quantidades de dados (SIRKKUNEN, 2011, p. 6)³⁷.

A operação de busca pelo submarino *Ara San Juan* foi alvo de uma cobertura em tempo real de 36 horas do *La Nación* (Argentina)³⁸. A reportagem demonstra como dados, infografia e instantaneidade podem prender a atenção do usuário. Como um artefato de jornalismo do futuro, o jornal utilizou dados de tráfego de embarcações extraídos de imagens de satélites, sensores e outros tipos de observação direta (Figura 15). A equipe foi a única a acessar as rotas de cada embarcação que buscava o submarino na costa argentina atualizando em tempo real o status de cada uma. Consequentemente, foram os primeiros a noticiar qualquer tipo de novidade relacionada às buscas.

Figura 15 – Ilustração do mapa de buscas criado pelo *La Nación*.



Fonte: La Nación, 2017.

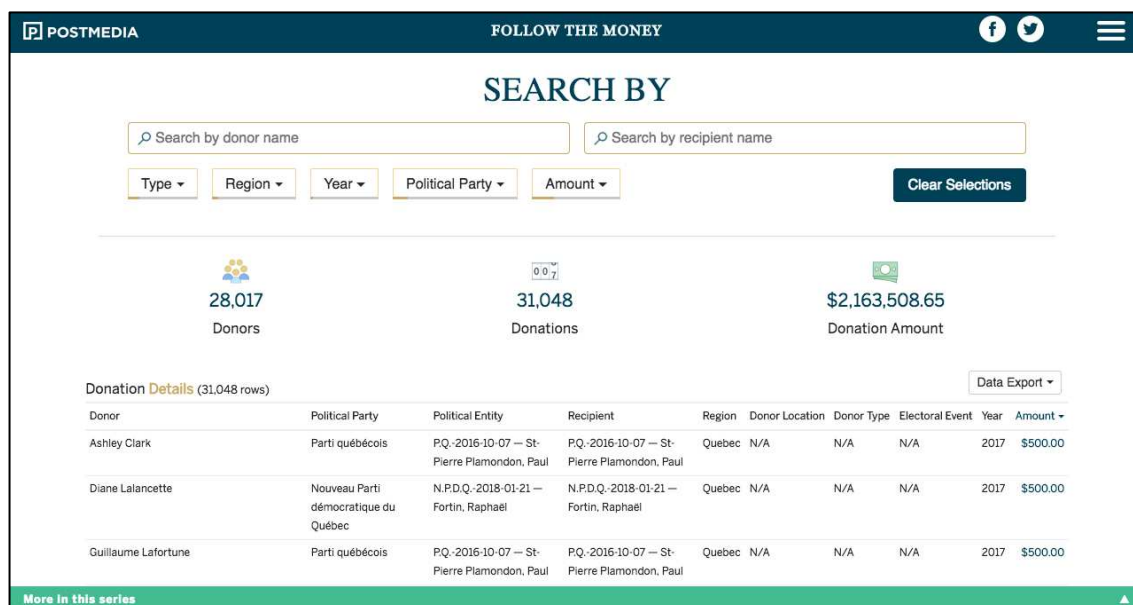
³⁷ Various studies on journalistic practices verify that journalists most often rely on data provided by other sources—governments, city officials, research companies, etc. Only in some cases of investigative journalism have journalists been able to find resources to gather and analyze their own data and publish their results. But the situation is changing rapidly. More and more free data are available online and there are efficient tools with which anyone can analyze, visualize and publish large amounts of data.

³⁸ Disponível em <https://www.lanacion.com.ar/2084935-submarino-ara-san-juan-un-video-del-recorrido-de-los-barcos-que-buscan-en-la-zona>. Acesso em junho de 2018.

Um outro exemplo de reportagem reforça a afirmação de Sirkkunen (2011) que propõe um acesso maior do público a dados abertos para que façam suas próprias análises. O projeto ‘Siga o dinheiro’ (*Follow the Money*³⁹) do jornal canadense *Postmedia* disponibilizou a primeira base de dados de fácil compreensão para doações políticas no país completamente aberta para cidadãos em geral, acadêmicos e jornalistas de outros grupos de comunicação que estivessem buscando entender o caminho que o dinheiro traça: quem está doando, quanto, para quem? Essas perguntas eram quase impossíveis de serem respondidas no passado do Canadá porque os dados estavam em um lugar de difícil acesso ou estruturados em uma linguagem complexa para o público.

No projeto, o usuário pode filtrar as buscas por tempo, região e maiores doadores. Vale destacar que todas as pesquisas que o usuário fizer podem ser baixadas nos computadores individuais para análises mais detalhadas. Uma busca rápida, ilustrada abaixo (Figura 16), mostra que em 2017 houve 31.048 doações na região de Quebec, advindas de 28.017 doadores, resultando numa quantia de \$ 2.163.508,65. Ainda é possível refinar melhor se escolhermos o partido, o tipo de doação e a quantia doada por cada um.

Figura 16 – Exemplo de busca realizada para ilustrar o poder da ferramenta.



Fonte: Follow the Money, 2017.

³⁹ Disponível em < <http://special.nationalpost.com/follow-the-money/database> >. Acesso em junho de 2018.

Nos primeiros dias depois de lançado, o projeto já havia atraído mais de 40 mil visualizações. O fato de utilizar tecnologias de buscadores e filtros deixa a experiência do usuário ainda mais próxima do dia a dia porque estabelece um vínculo com o consumo de outros tipos de conteúdo na internet e personaliza o que cada um gostaria de absorver a partir de interesses próprios (LORENZ, 2014), além de reconquistar a credibilidade do público ao oferecer informações mais precisas resolvendo um velho problema do jornalismo de ter um melhor contato e empatia com a audiência (BALDESSAR; FEIJÓ; TOASSI, 2013). Se levarmos em consideração que os meios de comunicação de massa já sofrem com o afastamento do público das notícias, nesse caso, a *Internet* se mostra mais agregadora do que concorrente.

Apesar da membrana de complexidade que envolve o trabalho de análise dos dados e de proximidade dos profissionais que estão no mercado com diferentes níveis de tecnologia, é perceptível um movimento em direção ao caminho guiado pela massificação de mídias digitais que permite a existência de um novo modelo de apresentação jornalística, que promove uma construção social da realidade (GRANDIN, 2014; ARAÚJO, 2016). Não se trata de uma substituição de modelos, mas de um acréscimo importante aos padrões tradicionais adaptado às condições atuais em que o mundo vive.

É comum ouvirmos comentários que expressam incerteza em relação aos benefícios da tecnologia porque parece que o mundo percorreu um caminho inverso: o número de dados gerados aumentou, mas a humanidade padece de falta de informação de qualidade. Na verdade, é possível que o problema real seja uma assimetria de informação uma quantidade tão grande que as pessoas não conseguem absorvê-las e assimilá-las na velocidade em que elas chegam até nós (BOUNEGRU; CHAMBERS; GRAY, 2012). Enquanto o público é bombardeado com um grande número de dados e tenta compreendê-los, os jornalistas tentam vencer as barreiras que têm o trabalho com pensamento lógico e numérico e correm contra o tempo para encontrar um lugar que os deixe transitar entre o lado escritor e o lado cientista e os permita fazer o jornalismo da era do *Big Data* (LEWIS, 2015).

À medida que o jornalismo quantitativo se torna mais central ao núcleo profissional do jornalismo e à medida que as técnicas computacionais e algorítmicas também se entrelaçam com os modelos de negócios nos quais o jornalismo é apoiado, questões críticas emergem continuamente sobre a relação sócio material do *Big Data*, do jornalismo e do trabalho amplo da mídia. Até que ponto a autoridade cultural e as práticas tecnológicas do jornalismo estão mudando no contexto (embora não necessariamente por causa) do *Big Data*? E como essas mudanças podem ser conectadas ao público das notícias, formas de histórias, arranjos organizacionais, canais de

distribuição e valores e ética das notícias, entre muitas outras coisas? (LEWIS, 2015, p. 327)⁴⁰.

O conceito de jornalismo quantitativo se baseia nas afirmações de que a informação, com o advento da Internet, se tornou um combinado de zeros e uns e obrigou o jornalismo a computar, reunir e analisar essas informações de maneira quantitativa (CODDINGTON, 2015; PETRE, 2013). É uma mudança que coloca desenvolvedores e jornalistas trabalhando lado a lado nas redações e incentiva o compartilhamento de ideias advindas de diferentes perspectivas e visões de mundo, agregando valor ao que será noticiado e enriquecendo o processo de construção da notícia (CRUCIANELLI; ZANCHELLI, 2011). Essas novas conexões entre jornalistas e programadores promove muito mais benefícios ao público do que se esperava porque não só revela informações escondidas em dados disponíveis publicamente, como aproxima a audiência de ferramentas que permitem um acesso completo a esses dados (DAGIRAL; PARASIE, 2013). É que além de escrever reportagens guiadas por dados, alguns veículos desenvolvem recursos que dão ao usuário o poder de analisar e interpretar os dados por conta própria.

Isso acaba instigando um sentimento de fiscalização cidadã e contribui para uma reflexão social desse público. Até porque o Jornalismo de Dados “pode incluir todas as possibilidades desde a tradicional RAC (usando o dado como uma ‘fonte’) até visualizações de dados e aplicações de notícia com tecnologia de ponta” (BOUNEGRU; CHAMBERS; GRAY, 2012, sem página)⁴¹. Em 2017 o *The Texas Tribune* executou um projeto que explicita como os usuários podem se beneficiar do acesso fácil a dados que, até então, ficavam escondidos do público. O *Government Salaries Explorer*⁴² é um banco de dados com mais 331 mil registros que estão disponíveis e atualizados para o público. É possível fazer buscas por nome, valores recebidos ou alocação dos funcionários.

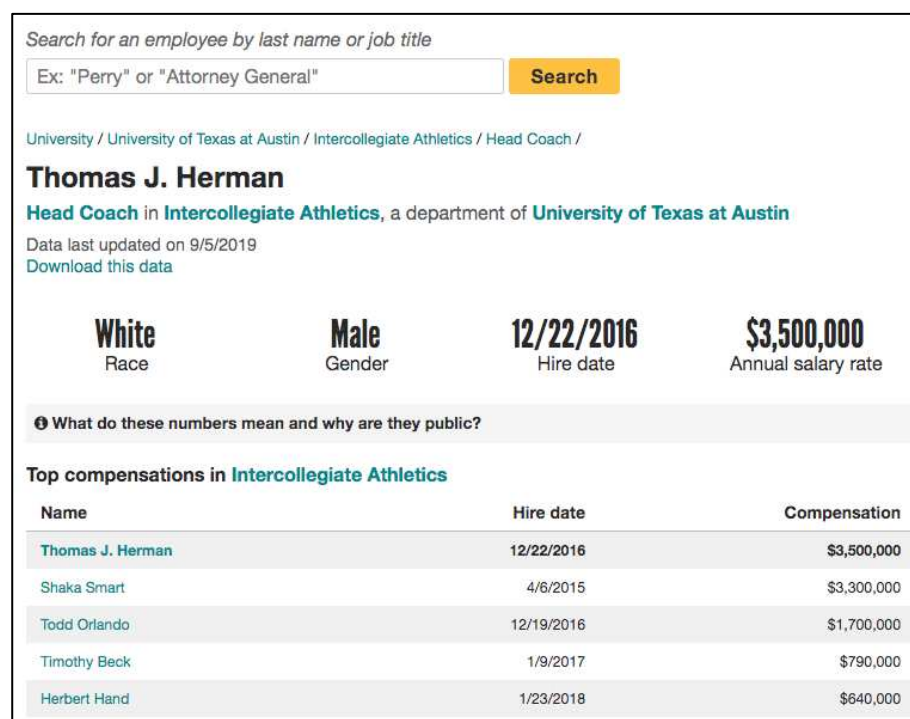
⁴⁰ As quantitative journalism becomes more central to journalism's professional core, and as computational and algorithmic techniques likewise become intertwined with the business models on which journalism is supported, critical questions will continually emerge about the socio-material relationship of big data, journalism, and media work broadly. To what extent are journalism's cultural authority and technological practices changing in the context of (though not necessarily because of) big data? And how might such changes be connected with news audiences, story forms, organizational arrangements, distribution channels, and news values and ethics, among many other things?

⁴¹ It can include everything from traditional computer-assisted reporting (using data as a ‘source’) to the most cutting edge data visualization and news applications.

⁴² Disponível em <https://salaries.texastribune.org/>. Acesso em 17 de fevereiro de 2020, às 19h01.

O projeto já disponibiliza um *ranking* dos maiores salários pagos aos servidores do governo com informações detalhadas do cargo ocupado, a qual entidade e departamento pertence o funcionário, em que data ele foi contratado e qual valor ele recebe por ano. Ao detalhar a pesquisa, o leitor também tem acesso a informações baseadas em cruzamentos de dados que mostram os outros servidores com os salários mais altos na mesma instituição, como demonstra a Figura 17.

Figura 17 – Busca mostra detalhes de servidor público com o salário mais alto do Texas.



Fonte: The Texas Tribune, 2017

Essas experiências reforçam o pensamento de que as narrativas passaram a ser resultado de uma harmonização entre dados e conteúdo informativo desenhadas em formatos com propósitos claros e pré-determinados que perpassam diferentes camadas e granularidades durante o processo de mineração até chegar ao resultado final (BERTOCCHI *et. al*, 2015; ESTEVANIM, 2016). Esse conceito nas redações faz parte de um processo de transmutação perene ao qual o jornalismo está submetido para que sobreviva do longo do tempo. Os tempos de reconfiguração dos modelos de produção e distribuição de notícia são sempre caóticos, mas envolvem também oportunidades de crescimento não só para os profissionais, como para o Jornalismo enquanto área de atuação da Comunicação Social que tem o privilégio de ser uma ponte entre o conteúdo conhecido e o escondido da opinião pública.

4.2 UM PONTO DE INFLEXÃO NA CURVA DO JORNALISMO DE DADOS

Historicamente o jornalismo vai se apropriando do ciberespaço (JÚNIOR, 2011) para remodelar seus processos criativos porque já não há como restaurar os mesmos formatos em que foi praticado no século XX (ANDERSON; BELL; SHIRKY, 2013; MIELNICZUK; TRÄSEL, 2017). Eis que surge uma oportunidade: incluir nos desafios a possibilidade de fazer jornalismo pensando como um cientista da computação, que tem como principal habilidade resolver problemas (DOWNEY, 2016). Essa capacidade começa na formulação da pergunta, no design de uma solução criativa - que pode ser através da mineração de dados, por exemplo - e no desenvolvimento de uma solução precisa, clara.

É uma ferramenta que permite ao jornalista trabalhar de forma especializada na análise de dados, que exige uma interferência humana para construir a notícia (JUNIOR, 2006). Mas já conseguimos encontrar experiências em que o processo de notícia acontece, quase que inteiramente, sem a participação humana. Seria uma espécie de Jornalismo Semântico (COSTA, 2014) ou Jornalismo Algorítmico (DIAKOPOULOS, 2015) em que robôs funcionam como repórteres. De acordo com Costa (2014), eles simbolizam uma produção pura de *Hard News* e não ameaça o papel humano do jornalista que se mantém insubstituível nos processos de investigação aprofundada e contextualização social dos dados antes da publicação.

A utilização de *bots*, no entanto, seria uma maneira simples de criar textos automáticos que respondam às perguntas básicas do *lead* jornalístico: ‘Quem? O quê? Quando? Onde? Como? Por quê?’. Tais tecnologias estão inseridas no universo do Jornalismo Pós-Industrial, conceituado no relatório do Tow Center, da Universidade de Columbia, que analisa o papel dos profissionais de comunicação inseridos numa sociedade completamente midiaticizada que deixou de seguir a lógica do maquinário de produção (ANDERSON; BELL; SHIRKY, 2013).

Não é um novo jornalismo, ou um jornalismo superior, mas um novo lugar do qual lançar os olhos sobre o objeto jornalismo – e do qual este, por sua vez, também possa lançar os olhos sobre a sociedade. Um espaço onde as tensões ‘homem x técnica’, ‘repórter x máquina’, ‘emissor x receptor’ se resolvam e deem origem a novas práticas, linguagens e organizações (COSTA, 2014, p. 7).

É difícil pensar, historicamente, em um fim do jornalismo. Ele se mostra mais como uma profissão cíclica (SALAVERRÍA, 2014) porque é extremamente adaptável a inovações de qualquer natureza. Exemplo disto é a notícia sobre um terremoto ocorrido no estado da

Califórnia (EUA) em março de 2014. O *Los Angeles Times* publicou uma matéria informando o que havia acontecido minutos depois do abalo sísmico e utilizou como única fonte a base de dados do *U.S. Geological Survey*, o serviço de notificação de terremotos utilizado na região.

Além do fato recente, a reportagem trazia informações sobre abalos sísmicos ocorridos nos 10 dias anteriores com detalhes das localidades afetadas. O que a princípio parece uma notícia comum revela uma surpresa no fim do texto: o post havia sido criado por um algoritmo, desenvolvido pelo autor, Ken Schwencke (Figura 18). Ele é um jornalista e programador que integra a equipe do *LA Times*. Foi a primeira notícia do jornal escrita por um robô.

Figura 18 - Íntegra do texto publicado por *Quakebot*, o robô do *LA Times*.

Here's Monday morning's initial Quakebot report:

A shallow magnitude 4.7 earthquake was reported Monday morning five miles from Westwood, California, according to the U.S. Geological Survey. The temblor occurred at 6:25 a.m. Pacific time at a depth of 5.0 miles.

According to the USGS, the epicenter was six miles from Beverly Hills, California, seven miles from Universal City, California, seven miles from Santa Monica, California and 348 miles from Sacramento, California. In the past ten days, there have been no earthquakes magnitude 3.0 and greater centered nearby.

This information comes from the USGS Earthquake Notification Service and this post was created by an algorithm written by the author.

Read more about [Southern California earthquakes](#).

Fonte: Los Angeles Times, 2014.

Oremus (2014) reforça que o objetivo do algoritmo não é escrever uma história complexa e perspicaz porque isso é tarefa do quadro humano do jornal. A ideia é que o *Quakebot* publique informações básicas com a maior precisão e velocidade possíveis (OREMUS, 2014). Para ele, enquanto o robô informava o mundo sobre o terremoto, os jornalistas estavam preocupados com outras nuances do fato: alguém estava ferido? Houve danos materiais? O que o *U.S. Geological Survey* já podia dizer sobre o assunto?

A intenção com que foi construída o *Quakebot* está de acordo com a própria definição do algoritmo. “Um algoritmo pode ser definido como uma série de etapas realizadas para resolver um problema específico ou obter um resultado definido” (DIAKOPOULOS, 2015, p.

400). O autor apresenta esse conceito considerando os algoritmos que operam via computadores digitais e que tomam decisões atômicas que ele qualifica como priorização, classificação, associação e filtragem. A priorização enfatiza ou direciona a atenção para determinadas coisas em relação a outras. A classificação acontece quando o algoritmo separa os objetos e os agrupa em categorias. Já as decisões por associação se dão com as relações entre as entidades. E a filtragem envolve a inclusão ou a exclusão de informações a partir de regras e critérios pré-estabelecidos. Diakopoulos (2015) acredita que a transparência durante todo o processo de produção algorítmica reforça a confiança do público nas notícias com base nesse tipo de tecnologia.

Ser protagonista no processo que acaba de ser descrito é, mais uma vez, situar a figura do jornalista num meio entre a comunicação e a cultura Hacker (TRÄSEL, 2014b; ARAÚJO; CARREIRA; ROTH, 2015), que vai revelar um profissional capaz de lidar com a mineração e análise de dados e de manter sua habilidade original de contar histórias. Afinal de contas, nem sempre dá para deixar tudo sob a responsabilidade dos algoritmos. Pelo menos por enquanto.

Por algum tempo, no entanto, haverá um papel e uma necessidade de que os seres humanos verifiquem os algoritmos que geram notícias automatizadas a partir de dados, adicionando contexto, moldando visualmente narrativas convincentes e conduzindo o jornalismo investigativo que algoritmos por si só não podem (HOWARD, 2014, p. 29)⁴³.

Ampliou-se o âmbito das atribuições do jornalista (BALDESSAR, 2001). Não há dúvidas. Angeluci e Oliveira (2019) sugerem uma matriz de habilidades e competências necessárias no Jornalismo de Dados. Ela se baseia em seis categorias: um currículo diferenciado, bases teórico-práticas para o uso de linguagens de programação, narrativas de apuração e visualização, projetos e parcerias, equipes híbridas e aprendizado auto direcionado. São novas habilidades a serem desenvolvidas. Por enquanto elas não são responsáveis por manter o profissional no emprego, mas constituem um diferencial importante.

Cada vez mais os jornalistas precisam se agregar diferentes áreas do conhecimento para estabelecer novos parâmetros na profissão e desenvolver modelos adaptados de redações responsivas às alterações provocadas pelo advento do *Big Data*. Não podemos esquecer que a

⁴³ For some time to come, however, there will be a role and need for humans to fact-check the algorithms generating automated news from data, adding context, shaping visually compelling narratives, and conducting investigative journalism that algorithms alone cannot.

tecnologia transforma, principalmente, as relações de consumo do ser humano e isso impacta diretamente no nosso produto fundamental, que é a notícia. Essa compreensão é urgente.

Quadro 4 - Matriz de habilidades e competências no Jornalismo de Dados

| A categoria... | ... aponta para a necessidade de desenvolvimento de competência em... | ... adquirida a partir do desenvolvimento de habilidade para... |
|--|--|---|
| 1. currículo acadêmico diferenciado. | aplicar um raciocínio mais lógico, analítico e voltado ao digital. | trabalhar com planilhas Excel, realizar modelos e simulações em Estatística e Matemática. |
| 2. bases para cultura de programação | contar histórias e revelar dados ocultos no <i>big data</i> . Possuir independência dos programadores. | manusear bancos de dados e linguagens de programação como Python e R. |
| 3. narrativas de apuração e visualização | analisar dados complexos extraídos de grandes bases de dados (<i>big data</i>) e gerar narrativas digitais atraentes para os leitores em rede. | aplicar dados em linguagens gráficas para apresentação de números e resultados, recursos de infografia e técnicas de visualização. |
| 4. projetos e parcerias | estabelecer parcerias interdisciplinares com universidades, mídias e associações independentes para formação de jornalistas de dados. | compreender linguagens e jargões de áreas diferentes, transitar em diferentes setores. Buscar treinamento enquanto aluno. Fazer capacitação enquanto professor. |
| 5. equipes híbridas | ser um jornalista mais completo e sofisticado com visão mais analítica, sem deixar de lado o viés social da notícia ancorado em transparência. | atuar como programador, infografista e jornalista. Interagir de forma dinâmica com equipes interdisciplinares. |
| 6. aprendizado autodirecionado | ser um profissional com perfil mais amplo, interdisciplinar e aderente às demandas de um mercado em evolução constante. | buscar individualmente por treinamentos e conhecimentos, de acordo com seus interesses dentro do ambiente digital. |

Fonte: ANGELUCI; OLIVEIRA, 2019, p. 407.

Essa nova dimensão de competências apresentadas como necessárias por Angeluci e Oliveira (2019) ultrapassa a evolução do jornalismo com dados, passa pelo Jornalismo de Dados e chega ao jornalismo algorítmico (MANCINI; VASCONCELLOS, 2016), que é:

"o processo (semi) automatizado de geração de linguagem natural pela seleção de dados eletrônicos de bancos de dados privados ou públicos (entrada), a atribuição de relevância de características de dados selecionadas ou não selecionadas, o processamento e a estruturação dos conjuntos de dados relevantes em uma estrutura semântica (taxa de transferência) e a publicação do texto final em uma plataforma *online* ou *offline* com um certo alcance (saída) (DÖRR, 2016, p. 4)⁴⁴.

⁴⁴ The (semi)-automated process of natural language generation by the selection of electronic data from private or public databases (input), the assignment of relevance of pre-selected or non-selected data characteristics, the processing and structuring of the relevant data sets to a semantic structure (throughput), and the publishing of the final text on an online or offline platform with a certain reach (output).

Na verdade, ela alcança um futuro do qual ainda não temos perspectiva, não sabemos que desafios ele guarda. O fato é que já podemos perceber que as empresas de comunicação têm entrado no carrossel de evolução do Jornalismo e buscam implantar soluções para o tratamento de dados e o desenvolvimento de produtos noticiosos automatizados (JUNIOR; OLIVEIRA, 2016). Como o objetivo desta pesquisa é propor um modelo de redesenho algorítmico do processo de construção da notícia, vamos agora nos debruçar sobre a nossa proposta para as redações.

Com base em todos os conceitos e exemplos práticos que vimos acima, bem como na experiência como jornalista na redação da TV Globo no Recife, buscaremos um modelo que não pretende ser único, mas relevante para que profissionais e empresários consigam estabelecer e reconhecer esse ponto de inflexão comum e benéfico tanto para a prática jornalística quanto para o Jornalismo enquanto modelo de negócio.

5 REDESENHO ALGORÍTMICO DO PROCESSO PARA CONSTRUIR A NOTÍCIA

Para que as redações desenvolvam um modelo de trabalho com o Jornalismo de Dados, base fundamental de nossa proposta, se faz necessário elucidar novos pontos em discussões que já tivemos anteriormente. Meyer (1991) resume o trabalho de dados a duas fases essenciais: a de *input*, em que os dados são recolhidos e analisados; e a etapa que ele chama de *output*, em que os dados são preparados para serem absorvidos pelo leitor. É importante salientar que o modelo que iremos propor é mais facilmente aplicável para trabalhos com jornalismo investigativo, que oferecem ao jornalista um intervalo de tempo maior para a construção da notícia. O *hard news*, as notícias factuais do dia a dia, encontraria dificuldades nesse modelo de trabalho. Ele se encaixa melhor em atividades mais demoradas.

O primeiro passo, pois, em direção à geração de uma equipe voltada a esse trabalho com dados precisamos definir que habilidades são essenciais ao time do projeto. Se antes o Jornalismo de Dados era um trabalho solitário do jornalista que dispunha de aptidão para executar as tarefas, hoje é uma atividade de equipe (GORDON, 2013). E para que seja um método eficaz para construir a notícia, de acordo com o autor, é preciso conhecer as três dimensões de Jornalismo de Dados que ele sugere. São elas: repórter com auxílio de computador; desenvolvedor de aplicações de notícia e especialista em visualização de dados (Figura 19).

Figura 19 – Dimensões do Jornalismo de Dados.

As 3 dimensões do Jornalismo de Dados



Fonte: GORDON (2013)

Para Gordon (2013), o jornalista com auxílio de computador é aquele que tem a capacidade de encontrar e analisar nos dados numéricos novos pontos de vista sobre o fato para construir a notícia. Eles conseguem entrevistar bases de dados para descobrir rumos alternativos a uma reportagem tradicional. Já o desenvolvedor de aplicações de notícia é o indivíduo que tem uma formação computacional ou um jornalista que conhece níveis avançados de programação. Ele é capaz de criar *softwares* para disponibilizar dados de forma inteligível para o público. Por fim, o especialista em visualização de dados, é o profissional do *design* de interface, que vai pegar o trabalho desenvolvido pelos outros dois integrantes e encaixá-los num molde atrativo aos olhos do consumidor dessa notícia. De acordo com a Figura 20, há também profissionais com mais de uma habilidade. Eles são representados nos pontos de interseção 2, 3 e 4 (habilidades jornalísticas e desenvolvimento de aplicações; desenvolvimento de aplicações e design; habilidades jornalísticas e design, respectivamente). A interseção 1 representa o profissional que tem todas as habilidades do diagrama.

É claro que essas figuras já fazem parte da rotina de redações há bastante tempo, mas geralmente são exploradas de maneira isolada. Ou seja: o jornalista tem uma necessidade de criação com base em dados e faz essa solicitação à área de tecnologia da empresa. Nesse setor, um funcionário com conhecimento em programação vai preparar o *software* que melhor atenda à demanda da equipe de jornalismo. Depois de pronto, esse produto vai para as mãos do time de *designers*, que irá finalizar o trabalho a ser publicado, uma espécie de embalagem para entrega.

Os potenciais especialistas em visualização de dados, comumente, se interessam em trabalhar com matemática e gráficos com base em dados (GORDON, 2013), além de já terem a curiosidade previamente despertada para o trabalho com aplicações de processamento de dados ou com o uso de linguagens de programação para a construção de seus gráficos. Vale salientar que, por não fazerem parte de uma equipe com dedicação exclusiva ao Jornalismo de Dados, esses profissionais lidam com outras demandas que chegam aos respectivos setores e o trabalho com dados acaba se mostrando mais um projeto esporádico do que uma mudança no processo de construção da notícia (GORDON, 2013).

Se esses profissionais se encontram no ponto de interseção sugerido por Gordon (2013) e trabalham cotidianamente com o mesmo propósito, começamos a reconhecer um trabalho de Jornalismo de Dados na essência das empresas de notícia. De acordo com o levantamento

teórico feito para este trabalho - tanto de conceitos, como de experiências de algumas das corporações de jornalismo ao redor do mundo - é possível inferir que as redações estejam num momento de transição entre os profissionais isolados e a formação de um time. Como se fosse um estágio intermediário de percepção da necessidade de mudanças. Uma adversidade importante é a falta ou baixa habilitação dos profissionais para trabalharem com Jornalismo de Dados.

Um problema que não é prerrogativa apenas do jornalismo, mas da indústria 4.0 como um todo – a demanda do mercado é muito maior que a oferta de profissionais para supri-la (SCWAB, 2014). Como tentativa de resolução desse obstáculo, muitas empresas tentam comprar ferramentas de análise e visualização de dados como *Power BI*⁴⁵ e *Tableau*⁴⁶, mas isso não é suficiente a longo prazo. É um recurso imediatista porque não ensina o jornalista a dominar os conceitos de estatística, nem a se familiarizar com a cultura de trabalho com dados abertos, que são o fundamento do Jornalismo de Dados, conforme apresentado no capítulo 4 deste trabalho. Apenas habilita tecnicamente o profissional para mexer com ferramentas específicas. O processo de formação dessa equipe deve ter como base outros aspectos para que seja mais eficaz.

O ensino do JGD deve focar-se na técnica, não em discussões conceituais: aplicativos de coleta, análise e visualização de dados são ferramentas que permitem materializar processos cognitivos, não fins em si mesmos. Além disso, tais ferramentas estão constantemente evoluindo ou sendo substituídas por versões alternativas, e restrições organizacionais podem impedir jornalistas de usar um ou outro software em sua jornada de trabalho. Assim, centrar o ensino nos processos cognitivos da apuração, redação e edição de notícias a partir de bases de dados pode ser mais produtivo e garantir maior flexibilidade a futuros repórteres do que os tornar especialistas no uso de uma pequena gama de aplicativos (MIELNICZUK; TRÄSEL, 2017, p. 623).

Seria um indicativo de que essa nova configuração de processo de elaboração da notícia começa na formação do jornalista, seja na universidade ou no ambiente profissional. Esse jornalismo moderno e sólido é carente, ao nosso ver, de uma sistematização para a estruturação de uma *squad* voltada ao processo de construção da notícia com base nos dados. Na tentativa

⁴⁵ O Power BI, de acordo com a definição da desenvolvedora - Microsoft - é uma solução de análise de negócios que usa inteligência de mercado (Business Intelligence - BI) e permite que a visualização de dados e compartilhamento de insights para tomada de decisões através de dashboards e relatórios que resultam da análise de dados. Disponível em: <https://powerbi.microsoft.com/pt-br/what-is-power-bi/>. Acesso em mar. 2020.

⁴⁶ O Tableau, de acordo com a definição da desenvolvedora - Tableau Software - é uma plataforma de análise de dados. Ela também utiliza inteligência de mercado (Business Intelligence - BI) e permite a visualização de dados através da criação de dashboards e relatórios a partir do resultado da análise. Disponível em: <https://www.tableau.com/pt-br/products/what-is-tableau>. Acesso em mar. 2020.

de sanar esse problema propomos uma metodologia que respeita um dos princípios do processo criativo que afirma não necessitar de elementos novos inéditos, mas elaborar novas soluções dando um novo significado a elementos que já existentes (HOWKINS, 2013) e, assim, sugerir um caminho para essa transformação no jornalismo.

5.1 O FLUXO ATUAL DA NOTÍCIA COM BASE EM DADOS

Há uma tendência de fazer Jornalismo de Dados sem uma grande interferência do jornalista (GORDON, 2013). A falta de mão de obra especializada entre os profissionais de comunicação misturada à urgência em fazer reportagens com base em dados se consagram em uma boa justificativa para que as empresas recorram a soluções rápidas, porém passageiras, como a automatização de alguns processos apenas para facilitar o trabalho jornalístico, sem mudar sua essência (AITAMURTO, 2011; GORDON, 2013). O ideal seria que as pessoas que gerem equipes de jornalismo começassem a identificar quem são os jornalistas que têm perfil para trabalhar com dados. Nas redações, geralmente, eles se apresentam como aqueles que já trabalham com planilhas e gostam de utilizar cálculos matemáticos e análises estatísticas na hora de elaborar as reportagens.

O mesmo movimento deve acontecer com os líderes das áreas de Tecnologia da Informação na empresa: detectar que funcionário tem habilidades de programação, aptidão para trabalhar com grandes bancos de dados e facilidade para analisar os resultados. É importante que essa pessoa já tenha uma certa familiarização com a rotina da redação e/ou demonstre afinidade com o fazer jornalístico. A dinâmica precisa se repetir na equipe de Design, que em muitas empresas leva é conhecida por ser o Departamento / Editoria de Arte ou Ilustrações Gráficas, entre tantas outras denominações. Esse profissional, que geralmente conhece os fluxos de construção da notícia, precisa apresentar um entendimento do trabalho com dados numéricos para construir gráficos e modelos de apresentação atrativos e de fácil interpretação para o público.

Uma vez que a gestão da empresa identifique essas pessoas com o grupo de habilidades indispensáveis ao trabalho com dados começa a capacitação da equipe, que “costuma ser pequena, tendo entre duas e cinco funcionários, mesmo em grandes empresas” (AITAMURTO, 2011, p. 12). Algumas corporações traçam um caminho paralelo e tentam buscar profissionais de jornalismo no mercado que já atendam à demanda de formação desse novo time.

A corroborar a necessidade de novas competências estão os anúncios de emprego para jornalistas onde são pedidas competências específicas na área da análise de dados, normalmente a par de competências elevadas a nível das tecnologias de informação e comunicação (MARTINHO, 2013, p. 22).

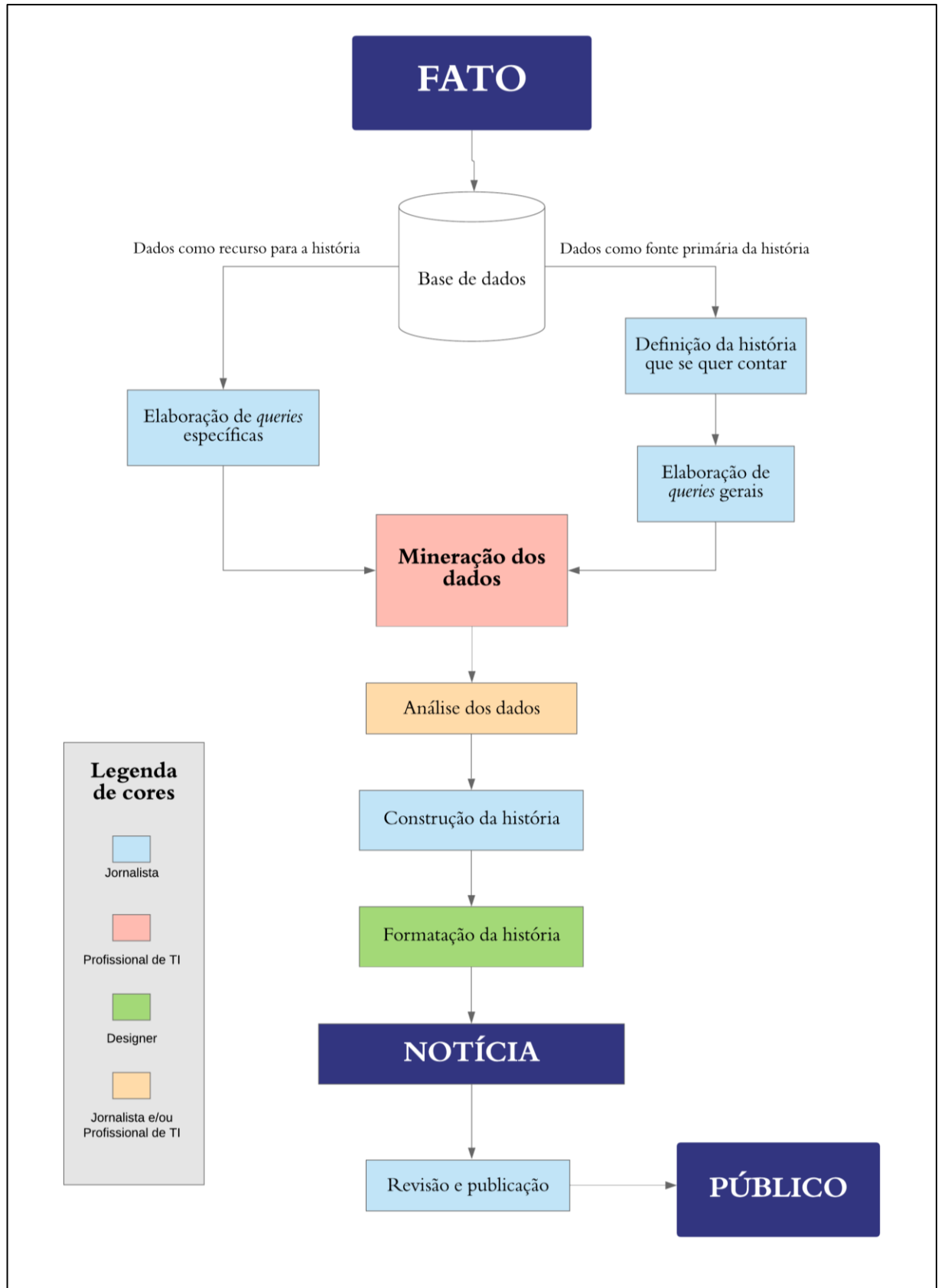
Não é uma tarefa fácil já que, como dito anteriormente, os jornalistas não encontram o trabalho com dados numéricos na formação universitária com frequência. Por isso escolhemos propor uma alternativa com os profissionais que já estão nas redações com equipes consolidadas, investindo em capacitação para que eles adquiram ou aprimorem suas habilidades. Para identificar esses profissionais e dar início ao processo formativo, é preciso compreender o fluxo primário de construção de uma reportagem com dados.

Vamos utilizar o modelo de construção da notícia que norteia a maior parte das redações de jornalismo (MANUAL DA REDAÇÃO, 2001). O processo se dá desta forma: os acontecimentos ou fatos que chegam ao conhecimento dos jornalistas são geralmente debatidos em equipe para que se decida, de acordo com as linhas editoriais da empresa, quais deles se tornarão notícia. A segunda fase é de apuração, momento em que o profissional busca se informar sobre as nuances que circundam o acontecimento que se tornará notícia em breve. Essa atividade, atualmente, pode acontecer de várias formas: pessoalmente, por telefone, por *e-mail* ou por uma conversa em um aplicativo de *chats*. Depois de apurada, a informação passa por uma seleção hierárquica de critérios em que é escolhido qual o encaminhamento que se dará à pauta.

O momento seguinte configura a produção da notícia. O jornalista faz as entrevistas necessárias, o levantamento de dados sobre o assunto, as análises que a pauta exige e, só depois, formata o seu texto. Depois da escrita, geralmente, esse texto passa por diversas revisões – tanto do profissional que o escreveu, como de companheiros de equipe, de editores e editores-chefes da empresa. O trabalho de infografia, quando necessário, pode começar durante o processo da escrita ou da revisão. Ele acontece em conjunto com a editoria de arte da organização. Só depois que o conteúdo é revisado por inteiro novamente é que ele pode ser publicado.

No Jornalismo de Dados há dois caminhos básicos para que os números virem notícia. Ou eles funcionam com um recurso para aprofundar uma história que já está sendo acompanhada pelo jornalista, ou eles são a base para toda a história que será noticiada (AITAMURTO, 2011; BRADSHAW, 2014). Elaboramos uma ilustração a seguir para facilitar o entendimento do caminho que a reportagem com base em dados geralmente percorre, com base no levantamento bibliográfico que fizemos e no contato empírico da autora com reportagens baseadas em dados (Figura 20).

Figura 20 – Modelo atual de fluxo de construção da notícia com base em dados.



Fonte: elaborado pela autora

No caminho em que os dados são um recurso da reportagem, o jornalista já tem conhecimento da história que quer passar ao público e necessita de dados numéricos e estatísticos para tornar a notícia mais sólida. Trata-se de um processo amplamente comum nas redações, como já explanamos anteriormente. Nesse caso, o repórter precisa definir que tipo de informação quer extrair das bases de dados escolhidas durante a fase de apuração e elaborar perguntas (*queries*) que possam elucidar as dúvidas do jornalista. A partir daí o profissional de Tecnologia da Informação entra em cena para minerar os dados em busca dos resultados que lhe foram pedidos.

Em outras situações, quando os dados são a fonte principal da história que será transformada em notícia, é preciso que o jornalista, em primeiro lugar, delineie o caminho a ser seguido – ou seja, escolha a história que quer contar vislumbrando as possibilidades que a base de dados escolhida oferece. Em seguida, ele precisa elaborar uma lista de *queries* mais vasta que a supracitada, uma vez que a história se baseia completamente na análise dos dados. Eles não configuram mais um dos elementos da reportagem e, sim, representam seu cerne. A partir daí, começa a mineração dos dados.

No geral, cabe ao profissional de TI coletar as bases de dados e iniciar o processo de mineração que já conhecemos no item 2: limpeza e correção de erros nos dados brutos, cruzamento com outras bases, seleção dos dados relevantes, formatação única dos dados, mineração com metodologias para extrair os padrões, avaliação dos padrões e entrega dos resultados. Essa tarefa é solitária em grande parte das vezes e tem uma duração longa porque a maioria dos dados, apesar de disponíveis conforme a lei, se apresentam em um formato não-legível para máquinas, tais como o PDF (AITAMURTO, 2011). Nesse momento, o cientista de dados precisa criar mecanismos para converter os arquivos que se encaixem no padrão para, só então, começar a mineração.

Este último item termina quando os dados vão para a análise, que pode ser feita pelo jornalista, pelo próprio profissional de TI ou pelos dois juntos. Em alguns casos é preciso também que essa fase envolva outros profissionais como estatísticos, matemáticos, cientistas políticos, entre tantos outros, a depender do conteúdo da reportagem. O processo de análise passa por inferências a partir dos dados, checagem das informações obtidas – que pode acontecer com uma volta à entrevista com os dados ou numa apuração tradicional com uma fonte externa – e, finalmente, a construção da história. É nesse momento que o jornalista decide

qual modelo irá utilizar, a depender, claro, do veículo em que trabalha. Para essa definição, habitualmente, ele recorre ao método que fazem parte de seu cotidiano: eleger critérios de noticiabilidade, valores-notícia (características do fato que determinam se o assunto deve ser notícia ou não), construção da audiência e rotinas de produção (TRAQUINA, 2001; WOLF, 2002). Uma vez escolhido o escopo da história, os dados vão passar por um processo de formatação visual que depende de outro profissional: o designer.

Às mãos do designer chegam os dados estruturados e a história que eles contam. Isto é: os dados já se transformaram em informação de utilidade pública, mas ainda precisam de um tratamento visual para que alcancem o público com objetividade e cumpram seu papel fundamental, que é informar. Uma boa formatação é essencial para a construção da notícia, uma vez que metade do cérebro humano se dedica a processar informações visuais (COHEN, 2012) e, por isso, elas são mais facilmente compreendidas.

Os gráficos surgiram, justamente, a partir da necessidade de compreender tabelas de dados densas (SUDA, 2012) e mantêm essa missão até hoje. Ao designer, então, compete a tarefa de auxiliar o jornalista na produção da notícia com a criação dos elementos visuais que contam parte da história. Do contrário, explicar o significado de tantos dados numéricos ao público seria uma tarefa árdua que acabaria segregando a audiência, o que configura uma fuga do jornalismo ao coração de sua missão como agente de cidadania (BAZZO; BRAGA; MARTINS, 2017; MARCONDES FILHO, 2009).

Uma vez que os elementos de visualização dos dados estão formatados para se unir à notícia, a reportagem costuma voltar para a redação para um processo de finalização e revisão antes da veiculação para o público. Trata-se de uma etapa que pode ser cumprida pelo repórter que elaborou a notícia e/ou por um jornalista com cargo de chefia dentro da hierarquia da empresa. O fato é que esse vai e vem de dados configura um processo demasiado lento e, por vezes, truncado. Por requerer pessoas de diferentes áreas do conhecimento que nem sempre trabalham juntas por longos períodos de tempo, a reportagem corre o risco de apresentar lacunas de informação ou de clareza no produto final.

É por isso que buscamos propor uma hibridização de saberes e o estabelecimento de uma equipe multidisciplinar dentro das redações que trabalhem exclusivamente com o Jornalismo de Dados, ainda que ela não seja volumosa.

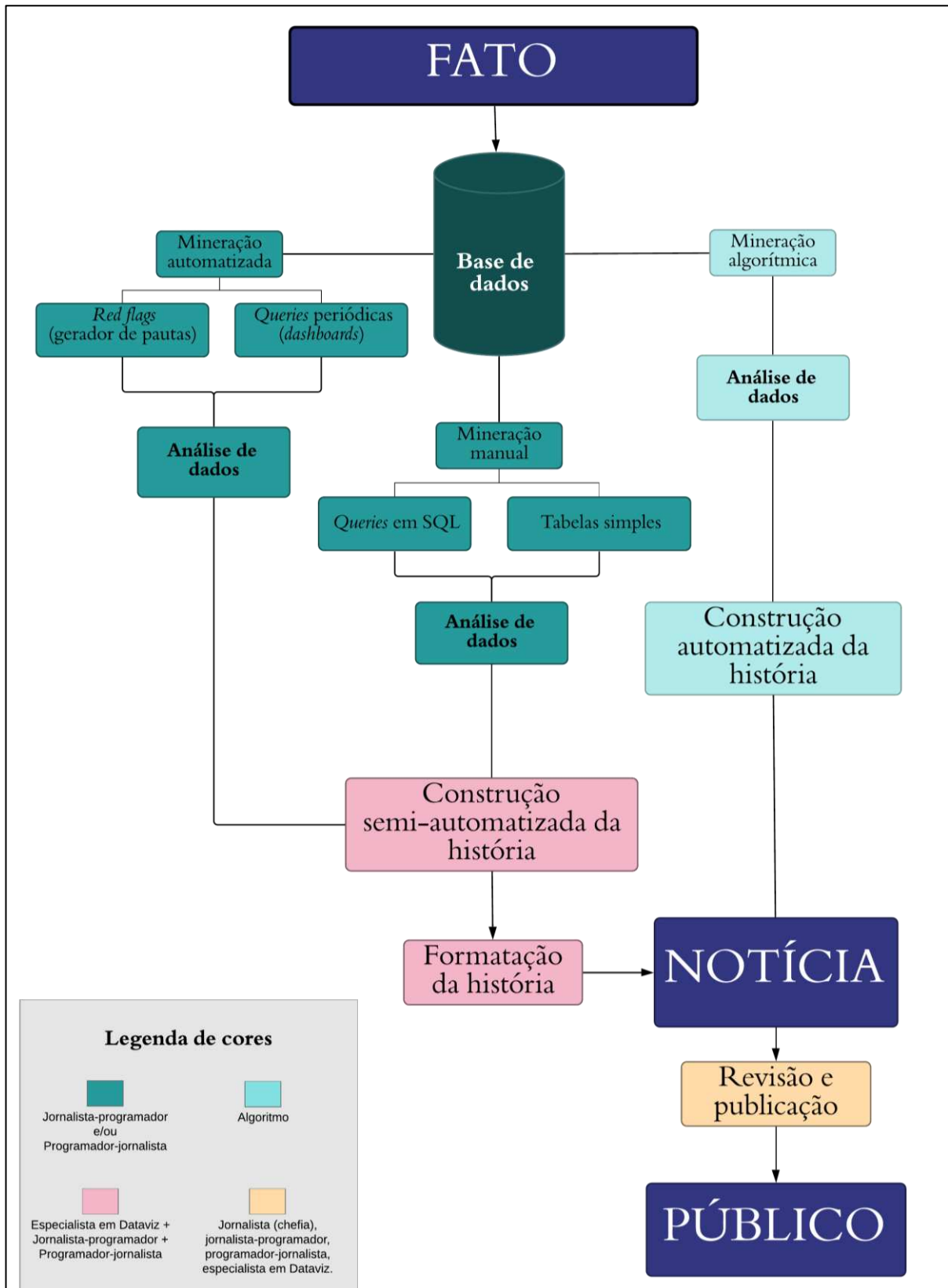
5.2 REBOOT NO DNA DA EQUIPE

“O Jornalismo de Dados acrescenta novos requisitos à profissão do jornalista” (AITAMURTO, 2011, p. 2) e a equipe de dados necessita, primordialmente, combinar as diferentes habilidades que apresentamos no item anterior. Não há atualmente um modelo padrão de negócio para o Jornalismo de Dados, mas a formatação heterogênea que as redações apresentam pode ser aperfeiçoada para deixar o processo de construção da notícia mais robusto e merecedor de crédito da audiência.

A remodelação, no entanto, encontra mais um entrave. A metodologia do Jornalismo de Dados para chegar à notícia leva mais tempo que o jornalismo tradicional e várias empresas ainda acreditam que o Jornalismo de Dados gera impacto no público, mas não produz receita para a corporação (AITAMURTO; LEHTONEN; SIRKKUNEN, 2011). Talvez este seja o motivo para tanta lentidão nas decisões sobre mudanças, principalmente em grandes empresas de comunicação. Mas as possibilidades de elaboração da notícia a partir da imensa multiplicidade de fontes que os dados oferecem (JUNIOR, 2012) é encorajadora. Aos poucos as organizações estão encontrando valor no Jornalismo de Dados, ainda que ele não seja tão significativo do ponto de vista financeiro.

Para descobrir o que esse novo padrão pode oferecer é necessário investir. Não somente na capacitação dos profissionais, mas na inovação de formar uma equipe singular e um tanto estranha, à primeira vista, para o que se conhece de uma redação. É preciso coragem. Nossa proposta para o redesenho algorítmico do processo de construção da notícia não combina com conservadorismo, mas com riscos calculados. Não estamos somente atualizando o modelo. Encontramo-nos num revés do fazer jornalístico tradicional, como demonstraremos adiante, na Figura 21.

Figura 21 – Proposta de novo fluxo de construção da notícia com base em dados.



Fonte: elaborado pela autora

O alfa do processo permanece o mesmo: os fatos precisam estar armazenados em bases de dados que serão a fonte principal do processo de apuração jornalística. No entanto, na nova configuração, os processos de mineração não vão depender apenas da demanda esporádica de projetos jornalísticos. Ou seja: não é necessário que o jornalista solicite ao setor de TI dados pontuais para aprofundar uma notícia ou para descobrir informações relevantes nos *datasets*. Na verdade, o trabalho de mineração e análise de dados ocupa agora a fase inicial do processo de construção da notícia. Ele deixa de ser apenas um meio e se torna o protagonista que dá o pontapé inicial da história a ser formatada.

Dividimos as atividades desse novo processo em três direções distintas. A primeira se dá pela mineração manual, que se inspira no fluxo atual e envolve um trabalho de coleta, limpeza e garimpo dos dados fazendo consultas a bases de dados utilizando softwares para consultas em SQL, por exemplo. A escolha por trilhar esse caminho pode se dar porque o profissional está lidando com um *dataset* ainda desconhecido ou por se tratar de consultas simples, que não envolvam a necessidade de criar um sistema que faça esse trabalho de maneira automática. Algumas tabelas mais triviais, com poucos registros em linhas e colunas, podem ser trabalhadas em programas de amplo acesso como o *Microsoft Excel*, que permite análises estatísticas e inferências de ordem menos complexa. Por exemplo: se o fato é que juízes na esfera estadual estão ganhando salários excessivos e a pauta é fazer uma análise das folhas de pagamento dos últimos 12 meses, as tabelas com os valores divididos por mês - obtidas nos portais da transparência - são suficientes. O profissional faz o *download* dos dados, organiza em uma pasta de planilhas no *software* de sua escolha, e começa a criar *rankings* dos valores recebidos por mês. A finalidade da busca não era complexa, portanto a metodologia para chegar a ela também pode ser simples.

Para diferenciar que tipo de mineração escolher, é muito importante ter discernimento para direcionar os esforços do time e evitar desgastes desnecessários. No caso de análises mais complexas que envolvem, geralmente, bases de dados muito grandes (com milhares e até milhões de registros de linhas), bem como a necessidade de cruzamento de tabelas para obter resultados mais robustos, o caminho seria a mineração automatizada. Essa trilha envolve a necessidade do uso de linguagens de programação para coleta e mineração dos dados. Isto porque o volume é tão grande que fazer uma análise manual, humana, caracteriza uma tarefa penosa e, certamente, muito demorada. Então propomos a criação de códigos que sejam capazes de fazer consultas aos bancos de dados escolhidos de maneira automática e avisem à equipe de

dados sobre os resultados por meio de relatórios gerados em formato de *dashboards* ou pela criação de alertas (*red flags*).

Os *dashboards*, geralmente, são utilizados em análises mais complexas, que precisam ser organizadas em fluxos de resultados para que produzam sentido. Por exemplo, se a pauta é descobrir como a última gestão municipal fez investimentos que beneficiem a população os resultados obtidos devem envolver informações como: detalhamento de receitas e despesas do município para analisar o quanto foi deslocado para a criação de novas unidades básicas de saúde; cruzamento dos dados de receitas e despesas na área da educação com o censo escolar para descobrir se mais investimento pode ser indicativo de melhora no sistema de ensino; quadro de servidores agrupados por setor e despesas com folhas de pagamento para observar se há alguém recebendo mais de um salário por mês e assim por diante. É uma gama de possibilidades gigante.

Já as *red flags* serviriam como um gerador de pautas para a equipe. Funciona como um tipo de alerta enviado toda vez que uma consulta automática for executada e o resultado estiver fora dos parâmetros estabelecidos pelo analista de dados. Uma ótima utilização dessa ferramenta seria para monitorar os gastos de deputados da Câmara Federal com a Cota para o exercício da atividade parlamentar (CEAP)⁴⁷. Eles têm o direito, garantido por lei, de pedir um reembolso de gastos com alimentação, combustível, materiais de escritório e vários outros itens que envolvam a atividade parlamentar.

No caso de refeições funciona assim: o deputado paga a quantia devida ao restaurante no momento em que consome. Com a nota fiscal em mãos, ele dá entrada num pedido de restituição do valor gasto com a refeição e entrega a nota fiscal à Câmara dos Deputados. O

⁴⁷ "O Ato da Mesa nº 43 de 2009, que detalha as regras para o uso da CEAP, determina que só podem ser indenizadas despesas com passagens aéreas; telefonia; serviços postais; manutenção de escritórios de apoio à atividade parlamentar; assinatura de publicações; fornecimento de alimentação ao parlamentar; hospedagem; outras despesas com locomoção, contemplando locação ou fretamento de aeronaves, veículos automotores e embarcações, serviços de táxi, pedágio e estacionamento e passagens terrestres, marítimas ou fluviais; combustíveis e lubrificantes; serviços de segurança; contratação de consultorias e trabalhos técnicos; divulgação da atividade parlamentar, exceto nos 120 dias anteriores às eleições; participação do parlamentar em cursos, palestras, seminários, simpósios, congressos ou eventos congêneres; e a complementação do auxílio-moradia. O valor máximo mensal da cota depende da unidade da federação que o deputado representa. Essa variação ocorre por causa das passagens aéreas e está relacionada ao valor do trecho entre Brasília e o Estado que o deputado representa." (Site da Câmara dos Deputados, disponível em https://www2.camara.leg.br/transparencia/aceso-a-informacao/copy_of_perguntas-frequentes/cota-para-o-exercicio-da-atividade-parlamentar, acesso em fevereiro de 2020).

documento passa por uma análise e, ao fim, o valor é depositado de volta na conta do deputado. Para fiscalizar, as *red flags*, aqui, funcionariam da seguinte maneira: a equipe de dados pode estabelecer que o limite máximo de valor para uma refeição seja de 150 reais. Se for entregue uma nota fiscal com um número acima dessa quantia, o sistema envia um alerta para a equipe. Entra em campo o trabalho jornalístico de entender que refeição foi essa, por quê teve um preço caro e repercutir o assunto. Mas, novamente, é apenas uma das incontáveis possibilidades que esse recurso oferece.

O passo seguinte é construir a notícia. Com base nas inferências que nascem a partir das análises dos dados, a reportagem será construída pela equipe. Escolhemos dizer que esse processo acontece de forma semi-automatizada porque a confecção da matéria pode – e deve – contar com a ajuda de *softwares* que auxiliem o profissional a interpretar as informações extraídas dos dados durante as análises. Não deixa de ser um processo feito em parceria entre humano e máquina. Até porque é nessa fase que a forma de apresentação da história é desenhada. O time de dados decide que tipo de informação vai apresentar como texto e que dados são importantes de serem compartilhados através de gráficos que facilitem a absorção do conteúdo por parte do público.

O fato é que os caminhos de mineração automatizada e manual passam por uma análise de dados que envolve o profissional durante grande parte do processo. Já a terceira direção de trabalho no modelo proposto, a mineração algorítmica, apresenta um esboço de participação humana diferente. Isso acontece porque a interação do homem aparece apenas na construção de um código capaz de apurar os dados, escrever e publicar um texto. A equipe de dados, então, trabalharia manualmente apenas nas fases iniciais do processo, construindo esse algoritmo, que vai ser responsável por todas as atividades de mineração e análise de dados, assim como, pela construção a notícia.

Esse modelo de produção pode se aplicar a casos que envolvem grandes coberturas ou atualizações em tempo real. Seria possível, por exemplo, criar um algoritmo que trabalhasse num *dataset* de defesas civis municipais atualizado em tempo real pelo poder público. Ele iria buscar atualizações sobre pedidos de socorro da população e poderia ser programado para escrever textos automáticos informando quando houvesse um novo incidente que envolvesse o trabalho dos agentes de defesa civil. Seria uma forma de alertar o público e a própria redação sobre um fato novo e relevante para a audiência.

Propomos ainda que todas essas notícias, depois de construídas, passem por um processo de revisão antes de ser publicada. Apesar da confiabilidade dos números e do zelo na hora de elaborar os processos de mineração dos dados, é papel do jornalismo checar as informações que passarão para o público até o último momento. É uma forma de evitar erros – de variados graus de relevância – e de ser fiel à verdade que se busca com a emissão e transmissão dos fatos reais em forma de notícia (TRAQUINA, 2001; SODRÉ, 2014). A equipe de dados não está isenta dos processos de checagem de uma redação jornalística tradicional.

Depois de entender a mudança proposta nos caminhos que o fato percorre até virar notícia, nos voltamos para o olhar sobre o profissional. É um outro cordão de mutações. Se antes demonstramos que a atuação com o Jornalismo de Dados (quando existente) acontece ainda de forma segregada dentro da maioria das empresas de comunicação, sugerimos uma renovação de perspectiva e de pontos de vista. No lugar de profissionais de áreas diferentes da empresa tentando trabalhar juntos num projeto esporádico com o uso de dados, indicamos a alternativa de um time de dados que atue fisicamente dentro da redação. Além do foco exclusivo nas atividades da equipe, sem a distração com demandas de outra natureza dentro da empresa, há um sentimento de pertencimento ao time ao jornalismo. É importante se sentir parte do processo. Mas essas mudanças não acontecem de repente.

Uma diferença fundamental entre o fluxo apresentado no item 4.1 e a nossa proposta de novo modelo apresentada no item 4.2 é a função exercida pelo profissional. Enquanto o modelo vigente em grande parte das redações, apresentado no item 4.1, apresenta uma configuração mais próxima do que conhecemos, com funcionários de áreas separadas trabalhando em projetos pontuais de Jornalismo de Dados, a nova formatação, detalhada no item 4.2, propõe uma hibridização de saberes que enriquecem o processo noticioso (VENTURA, 2018). Quanto mais diversa for a equipe, maior será a multiplicidade de olhares sobre o fato e a notícia. Na nossa proposta aparecem novos personagens na redação. A figura do jornalista-programador indica formação acadêmica em jornalismo e complementar em áreas de Tecnologia da Informação. Configuraria um jornalista tradicional que domina linguagens de programação e as utiliza na análise de dados. O programador-jornalista seria o contrário: um profissional formado em Ciência da Computação que conhece o universo jornalístico e os fluxos de notícia nas redações.

A nossa ideia não é que um substitua o outro completamente, nem que os profissionais passem a acumular funções para diminuir os custos da empresa. Sugerimos, por outro lado, que ter profissionais multidisciplinares numa equipe que lida diretamente com mineração e análise de dados tem muito a contribuir para o engrandecimento do trabalho. “Ter acesso aos bancos de dados não estabelece vínculo direto com a precisão e a pertinência do material jornalístico. O profissional de comunicação precisa saber como manipulá-los e interpretá-los” (FOTIOS, 2014, sem página). Um jornalista que tenha noções básicas de programação, ao menos, se tornará menos dependente do cientista de dados para tarefas simples. Por conseguinte, este profissional teria mais tempo para se dedicar a análises mais complexas que dependam exclusivamente de seu *know-how* na área de TI. Ao mesmo tempo, enquanto coleta e garimpa grandes bases de dados, o cientista pode fazer inferências sobre possíveis pautas que surjam a partir dos números, como uma pré-avaliação jornalística do assunto. Naturalmente não é simples montar uma equipe formada por pessoas com tantas habilidades distintas que se complementem. Mas é possível. Algumas redações começaram a trilhar esses caminhos no mundo e, por isso, já conseguem elaborar reportagens de dados com extrema eficiência.

5.3 PASSO A PASSO x CONDIÇÕES ADVERSAS

Apesar de um dos principais obstáculos para o crescimento de equipes de Jornalismo de Dados ser a falta de geração de receita e, conseqüentemente, de investimento da empresa nesse tipo de atividade, alguns autores acreditam que, à medida que o processo de trabalho com o Jornalismo de Dados evolui e decodifica um número maior de informações para o público, aumenta também o interesse da audiência (MARTINHO, 2014). Por isso sugerimos que algumas etapas a serem seguidas para solucionar esses problemas. São elas: a formação dos profissionais; a aproximação de áreas do conhecimento diferentes; o investimento em equipamentos e programas computacionais; o exercício livre de atuação dos profissionais aliado à confiança das lideranças nos resultados a serem apresentados. Esse é o caminho proposto para se chegar a uma reconfiguração dos modos de construção da notícia, conforme a Figura 22.

Figura 22 – Passos para mudanças na configuração da redação



Fonte: elaborado pela autora

Para dar início a esse processo, o primeiro passo é a escolha e formação dos profissionais. Mas não com os cursos existentes na atualidade. Eles são importantes como inspiração, claro, mas a nossa proposta diferente. É que a maioria do conteúdo disponível hoje ainda é muito focada no ensino para manipular ferramentas. Fotios (2014) analisou o método desenvolvido para o ensino de Jornalismo de Dados aplicado na Escola Superior de Propaganda

e Marketing (ESPM-SP) e na Universidade Metodista de São Paulo (Umesp). Foi criada uma sistemática para a formação dos estudantes que cursavam o sétimo período do curso de jornalismo nas instituições. Os professores perceberam que “o ensino de Jornalismo de Dados carecia de um processo matemático e ao mesmo tempo cognitivo” (FOTIOS, 2014, sem página) e que ensinar os alunos a entender o processo de construção desse tipo de notícia mais do que apenas manipular ferramentas é o caminho.

É um método em que eles escolheram utilizar a analítica criativa dividida em quatro etapas: preparação (reconhecer o problema e observar experiências anteriores), imersão (coleta e análise de dados), incubação (análise mais profunda dos dados e extração das informações relevantes) e ideia (comunicação dos resultados). Os professores concluíram que "alunos de Jornalismo ganham nova competência com a analítica criativa, voltada para a investigação de fatos noticiosos em que a fonte primária deixa de ser um entrevistado ou autoridade, mas o conjunto de dados que estes personagens geram no *big data*" (FOTIOS, 2014, sem página). O processo de formação precisa ser diferente.

Como a nossa proposta de capacitação é para os profissionais que já estão no mercado de trabalho com o intuito de formar uma *squad* de Jornalismo de Dados dentro das redações, o caminho que propomos é um tanto diferente. Ele tem uma característica mais prática e utiliza um método para resolução de problemas. Elaboramos estas etapas:

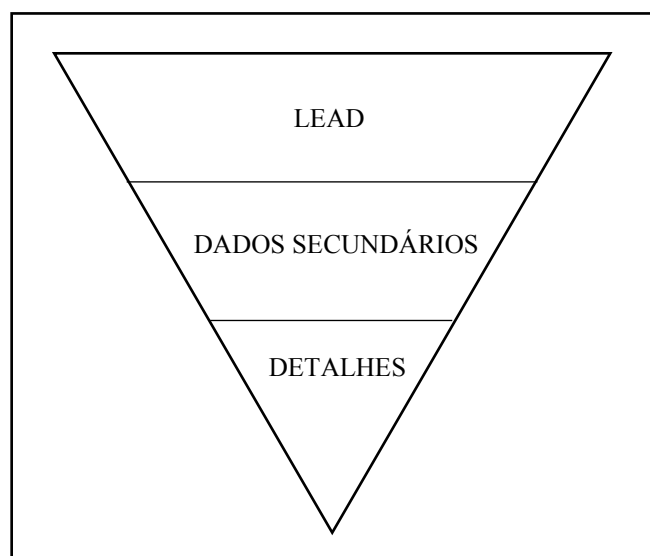
1. Aproximação de linguagens
2. Conceitos básicos de estatística e lógica de programação
3. Fluxos de notícia e conceito de pirâmide invertida
4. Jornalismo de Dados - Prova de conceito e fase de testes
5. Validação do projeto

O primeiro momento caracteriza uma imersão para reconhecimento do momento tecnológico atual para as duas áreas de conhecimento – comunicação e tecnologia da informação. A ideia é que os profissionais possam conhecer expressões linguísticas e processos básicos de trabalho de cada campo para se familiarizarem com o universo particular de cada um. Esse momento inicial serve, inclusive, para que a equipe perceba as habilidades de cada um e se reconheçam como peças importantes no processo de mudança e na formação dessa nova equipe.

O segundo momento serve para aproximar os integrantes da equipe com conceitos básicos de matemática, estatística e lógica de programação. Como já explicitamos, não é comum que jornalistas dominem esse conteúdo, mas é necessário. Nesta fase podem existir aulas específicas de linguagens de programação mais acessíveis aos jornalistas e comumente utilizadas na análise de dados, como *Python* e *R*. É provável também que muitos profissionais de TI, apesar de amplo conhecimento nessa área, se beneficiem desse momento já que, além de absorver novos argumentos, vão poder ajudar os profissionais de comunicação a entenderem os processos matemáticos e computacionais com que trabalham. O inverso também é presumível no terceiro momento da formação. Profissionais de TI entendendo como um fato se torna notícia e que tipo de ferramenta é necessário para que isso aconteça.

Neste momento, os profissionais entram em contato com o modelo básico de construção de uma notícia tradicional: a pirâmide invertida. A regra geral de construção da notícia é começar com o *lead* (ou lide), “um relato sintético do acontecimento logo no começo do texto, respondendo às perguntas básicas do leitor: o quê, quem, como, onde, quando e por quê” (PENA, 2005, p. 42). O *lead* corresponde, geralmente, ao primeiro parágrafo da notícia. A ele se segue o restante do texto, respeitando uma hierarquia que dá forma à chamada pirâmide invertida (Figura 23).

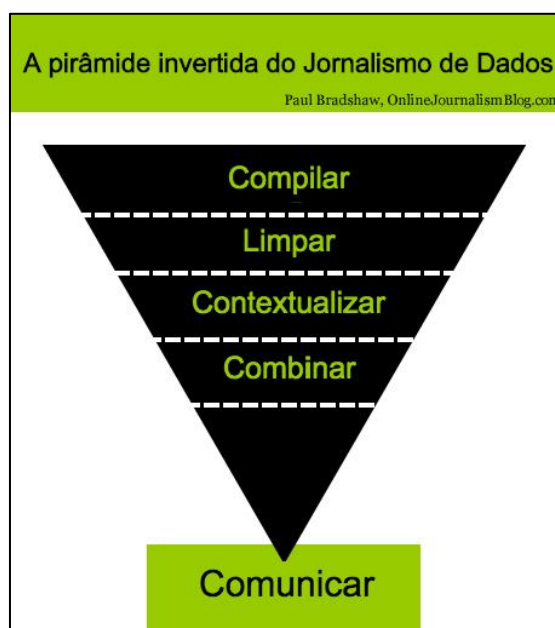
Figura 23 – Pirâmide invertida tradicional do jornalismo.



Fonte: PENA (2005)

Seria hierarquizar as informações por ordem de relevância dos fatos e não por uma sequência cronológica (FILHO, 1987). Tanto que os detalhes podem ser retirados do texto sem que isso prejudique o entendimento da notícia. Conhecendo sua versão clássica, o time de Jornalismo de Dados em formação é apresentado, no terceiro momento da capacitação, à pirâmide invertida do Jornalismo de Dados proposta por Paul Bradshaw (2011). Ele a configura mais como uma forma de trabalho do que uma metodologia de construção do texto (Figura 24). Bradshaw (2011) afirma que para fazer o Jornalismo de Dados é preciso, primeiro, seguir esta hierarquização: compilar, limpar, contextualizar e combinar. É preciso destacar aqui a importância do processo de limpeza porque, na maioria dos casos, os dados disponibilizados pelas instituições públicas apresentam muitos erros na estruturação. É comum, por exemplo, encontrar as mesmas categorias com nomes e registros distintos. Sem fazer uma limpeza na planilha bruta, a análise fica comprometida e, se for finalizada, é provável que apresente erros de informações porque os dados não estavam estruturados da maneira correta.

Figura 24 – Pirâmide Invertida de Jornalismo de Dados.



Fonte: BRADSHAW (2011) / Traduzido pela autora

Bradshaw (2011) acrescenta que esse processo culmina em uma nova pirâmide chamada de “comunicar”. Ela implica em decodificar o que foi analisado nos números para o público e percorre os seguintes passos: visualização, narração, socialização, humanização, personalização e utilização, como ilustrado na Figura 25. O autor criou esses processos para que eles fossem utilizados como uma espécie de cartilha e nenhuma etapa fosse deixada de

lado, já que para ele o “Jornalismo de Dados moderno cresceu ao lado de um enorme crescimento na visualização desses dados” (BRADSHAW, 2011, sem página)⁴⁸. O modo como o público acessa esse tipo de conteúdo e o formato em que ele é apresentado ao espectador ou leitor pode aproximar ou afastar a audiência. Por isso é um processo que precisa de muita atenção enquanto está sendo formatado.

Figura 25 – Pirâmide Invertida de Jornalismo de Dados completa.



Fonte: BRADSHAW (2011)

Compreender esses processos é importante para encerrar a etapa mais teórica da formação. O quarto momento da capacitação proposta nesta pesquisa envolve fazer projetos com Jornalismo de Dados dentro da redação para que os profissionais pratiquem os conhecimentos adquiridos durante as fases anteriores. Mesmo porque grande parte dos profissionais afirmam que programar, assim como outras atividades, é um tipo de habilidade

⁴⁸ Modern data journalism has grown up alongside an enormous growth in visualisation.

que só evolui com bastante prática. Para que os jornalistas saibam cada vez mais usar a programação para analisar dados e que os programadores consigam reconhecer as possibilidades de notícias diante dos fatos, ambos precisam exercitar os conhecimentos.

Nesta fase os profissionais devem formatar notícias respeitando o fluxo proposto de mineração de dados, bem como a elaboração da notícia a partir dos resultados gerados pelas análises. O recomendável é que façam um variado número de reportagens com abordagens distintas para estabelecer uma fase de testes de processos. É o momento de observar a fluidez da equipe, identificar gargalos que atrapalham o rendimento, descobrir quais pontos merecem atenção para efetuar melhoras técnicas e de conteúdo e delinear caminhos que são mais eficazes para a execução completa do processo de construção da notícia.

Em seguida, é importante validar os processos e o resultado com outras pessoas da redação. Sugerimos que seja avaliado quanto tempo a equipe levou para elaborar cada notícia, que nuances do trabalho jornalístico foram contempladas e quais ficaram de fora – se restaram lacunas nas informações, por exemplo –, se o resultado ficou amplamente inteligível para os vários tipos de audiência que podem ser alcançados com a reportagem e, finalmente, se o grupo atingiu o objetivo de compor uma equipe de Jornalismo de Dados que apresente, de fato, um trabalho mais efetivo do que o modelo atual. Acreditamos que o cronograma de ação para formar a equipe pode ser da seguinte forma:

1. Aproximação de linguagens – 2 dias
2. Conceitos básicos de estatística e lógica de programação – 3 dias
3. Fluxos de notícia e conceito de pirâmide invertida – 2 dias
4. Jornalismo de Dados: prova de conceito e fase de testes – 3 a 4 semanas
5. Validação do projeto – 3 a 4 semanas

Ratificando o que foi já explicitado, trata-se de um processo demorado e que precisa de investimento. Os esforços dos profissionais precisam ficar exclusivamente voltados para as atividades da equipe e, certamente, isso envolve uma demanda financeira para a melhoria de máquinas e substituição da mão-de-obra dessas pessoas enquanto elas estiverem testando o modelo da nova equipe. Essas são as maiores adversidades: ter resultados a longo prazo e a necessidade de investimento pessoal e financeiro. No entanto, trata-se de um desafio válido. E pode levar não só a equipe em formação a um novo patamar, como toda a redação de jornalismo.

6 CONSIDERAÇÕES FINAIS

Ao longo desta dissertação, percebemos que o método criativo que redesenha o processo de construção da notícia tem início quando o jornalista se depara com dados brutos e começa a planejar uma decodificação, que vai dar forma ao produto final. Nosso intuito com esta pesquisa foi, num primeiro momento, apresentar aos jornalistas - e aos outros profissionais que se interessam por Jornalismo de Dados - um panorama do trabalho nas redações e propor formas de integração que facilitem a execução de tarefas num ambiente historicamente povoado por comunicadores apenas.

Com o auxílio das reflexões teóricas sobre o tema, foi possível alcançar nosso objetivo principal de elaborar uma sugestão de caminho para incorporar o trabalho computacional com bases de dados no cerne do processo de construção da notícia. Refletindo sobre os principais conceitos do jornalismo inserido num ciclo de inovações tecnológicas, nos demos conta que eles configuram uma profissão em mutação. Eram alguns dos nossos objetivos específicos, mapear conteúdos relacionados a mineração e análise de dados, bem como de Jornalismo de Dados. Nessa fase, tivemos a oportunidade de observar que a notícia tem um caráter sempre novo e revolucionário.

A comunicação não se adaptou aos novos modelos e aparatos técnicos que foram criados. Essas invenções, na verdade, precisaram existir para que a comunicação avançasse. A demanda tem início de forma inversa. Com o tempo, ela se torna uma via de mão dupla: tanto as pessoas precisam se adaptar às novidades, como a tecnologia tem de avançar para impulsionar o progresso. Essa demanda se torna, então, compartilhada. Isso ficou explicitado com clareza quando sintetizamos a convergência existente entre Jornalismo e Tecnologia da Informação. A princípio, duas áreas completamente distintas. Mas, hoje, indiscutivelmente correlatas.

Se a criatividade é o ato de produzir coisas novas e valiosas ou de dar novo significado a coisas antigas (BENDASSOLLI *et. al.*, 2009; HOWKINS, 2013), ela é, também, elemento fundamental para o jornalismo porque o desafio de fidelizar ainda mais o público é constante. O mundo que coaduna meios de comunicação e tecnologias, converge profissões e habilidades que ocupam novos lugares e adaptam antigas formas de trabalho. As reportagens citadas neste trabalho foram construídas não por jornalistas, mas por um pedestal formado por eles,

programadores, designers, analistas estatísticos e profissionais de outras áreas. É um processo contínuo, esse de cruzar atividades interdisciplinares para gerar e gerenciar o conteúdo jornalístico de distintas formas.

Depois de entender como os processos jornalísticos se encaixaram no novo mundo durante os séculos XX e XXI, nos esforçamos para explorar - ainda que superficialmente - uma área de estudos nova, complexa e múltipla. Aprender sobre conceitos básicos de Tecnologia da Informação voltados ao *big data* reforça a ideia de que temos o compromisso coletivo de direcionar toda e qualquer inovação para o interesse público e humano, evoluindo de maneira sustentável (SCHWAB, 2016). O próprio termo Jornalismo de Dados demonstra a força que uma área do conhecimento impõe à outra para que, complementando-se entre si, possam servir de alicerce para a geração de informação de qualidade e útil para a população.

Por esse motivo, entendemos que o desenvolvimento do trabalho nos possibilitou concluir outros dois objetivos traçado no início da pesquisa. Os de ampliar o acesso de profissionais de variadas áreas do conhecimento ao tipo de leitura proposta nesta dissertação e de pluralizar o alcance palpável a uma bibliografia interdisciplinar sobre o tema. Como visto ao longo do trabalho, a união de diferentes campos do saber já é realidade em grande parte das redações pelo mundo, mas pudemos observar que isso se dá, ainda, de maneira improvisada em alguns casos. Por esse motivo nossa proposta apresenta pertinência. Os profissionais que querem fazer jornalismo investigativo utilizando dados podem fazê-lo de uma forma mais automatizada e precisa, para construir a notícia com elementos mais sólidos.

Buscamos propor uma formatação de equipe que consolida as técnicas que envolvem a mineração de dados e enraíza a mudança no processo noticioso para que ela seja mais consistente. Essa metodologia nova era, também, um de nossos objetivos específicos. Com a prática de trabalho como profissional híbrido e a capacitação para entender os conceitos de outras áreas de conhecimento, o profissional não dominaria apenas ferramentas novas, mas mudaria completamente a maneira como pensa a notícia. É uma transformação mais profunda, que vai além das habilidades técnicas. Mesmo porque entender o funcionamento de diferentes softwares não garante que o profissional esteja preparado para lidar com dados. Se a aplicação ficar obsoleta, a sua substituição significa uma volta ao ponto de partida. Mas se esse profissional compreende a lógica do trabalho com dados, só precisará adaptar as habilidades ao novo programa desenvolvido.

Diante do exposto, chegamos à conclusão de que é preciso insistir em uma capacitação mais profunda para fazer alcançar nosso principal objetivo com este trabalho: incorporar o trabalho com bases de dados computacionais no processo de construção da notícia. É importante tanto para os profissionais, como para o resultado que chega ao público. As reportagens têm mais chances de serem bem executadas e, assim, serem apresentadas de forma clara para a maior parte do público. Afinal de contas, números agrupados em tabelas, para muitas pessoas, não passam de uma mensagem codificada e, por conseguinte, inacessível. O jornalismo tem a responsabilidade de tradução e interpretação desses dados, fazendo com que eles saiam do nível de entendimento, geralmente, ineficaz - quando parecem um arranjo aleatório de números. Será possível, então, gerar informação e conhecimento a partir deles.

Acreditamos que a nossa sugestão de modelo contempla os requisitos necessários para dar início a uma mudança complexa e urgente. Enquanto o mundo vai se movendo de uma maneira extremamente veloz, as redações vão descobrindo o valor que os dados agregam às reportagens e como eles contribuem para a sociedade de maneira geral. A pandemia do novo coronavírus, que transformou o comportamento das pessoas globalmente em 2020, demonstrou o quão poderosa é a formatação de uma notícia com um bastidor de programação, mineração, análise de dados, e utilização da infografia. Inúmeros conglomerados de mídia apostaram no processamento computacional de bases de dados abertos para buscar novas abordagens que versaram sobre a pandemia. Mesmo que, às vezes, elas se apresentem de maneira inusitada.

O site brasileiro de notícias Exame publicou uma reportagem com aconselhamentos para combater o novo coronavírus. O único entrevistado presente na narrativa foi um cientista de dados. Ricardo Cappra, que já havia trabalhado na contenção do vírus Ebola em 2015. Cinco anos mais tarde, em 2020, ele reuniu sua equipe para estruturar o que chamou de ‘caixa de ferramentas de dados qualificados sobre a nova doença’, a Covid-19. O intuito era que outros cientistas pudessem analisar os cenários da pandemia e tomar decisões de contenção do vírus. De acordo com o cientista, as fontes de dados ainda são apresentadas de maneira nebulosa. E chegam a confundir mais do que informar. Cappra reforça a necessidade de estruturar bem as bases de dados para que elas guiem o poder público e a tomada de decisões com uma precisão maior e gerem conhecimento público.

A reportagem, que trata sobre um tema de saúde, tomou como base um especialista em dados. Uma cena possivelmente rara no século passado. É sinal de que os tempos mudaram e

de que estamos todos, aos poucos, nos adaptando ao *big data*. Se não o fizermos, ele (o *big data*) nos atropelará. Muitas empresas consolidadas no mercado de notícias têm apostado nos dados para lidar com o grande volume de informação que vem sendo divulgado. A reportagem que escolhemos destacar sequer foi feita com as técnicas discutidas aqui sobre Jornalismo de Dados, mas apresenta argumentos e provas de que os dados são uma riqueza preciosa na atualidade.

Esta pesquisa, portanto, não está fadada a um fim próximo. Pelo contrário. Há muitos desdobramentos factíveis. Um deles é a consequência principal da nova formatação do time de dados dentro das redações. Acreditamos que seria possível transformar essa equipe em uma agência de notícias. Uma espécie de startup dentro da própria empresa. Já que o Jornalismo de Dados ainda enfrenta o desafio de não ser um modelo de negócio rentável, uma alternativa para esse problema seria produzir notícias também para outros veículos de comunicação que não tivessem estrutura para formar uma equipe de dados. O modelo de notícias escritas por agências já existe no mercado de jornalismo há anos. Por que não criar uma nova vertente? As empresas brasileiras importam um vasto conteúdo do exterior através de agências. Quem sabe não se pode conseguir rentabilidade financeira exportando notícias brasileiras com base em análise de dados.

A agência também seria capaz de encontrar meios legais para negociar com anunciantes de publicidade usando não somente a audiência, como se faz, mas também com o compartilhamento de dados, desde que respeitando a Lei 13.709 (BRASIL, 2018). Conhecida como Lei Geral de Proteção de Dados, ela do direito do cidadão à não exposição pública de dados pessoais. Algumas empresas já utilizam esse tipo de informação como modelo de negócio. Precisáramos, porém, analisar essa possibilidade com muita cautela para evitar invasões de privacidade e desrespeito para com os cidadãos. Tal cuidado não impede uma oportunidade de inovação desta pesquisa.

Uma outra perspectiva de futuro é voltar a pesquisa para a área de educação e desenvolver disciplinas eletivas ou mesmo obrigatórias de Jornalismo de Dados nas universidades. Não é nossa intenção ter a experiência de ensinamento sobre o conteúdo em apenas um semestre. E para ir além, poderemos buscar desenvolver ementas com aulas de programação, estatística e design, oferecendo ao aluno não só uma formação mais plural, como

também embasamento para buscar oportunidades no mercado de trabalho que não estejam acorrentadas às poucas vagas existentes nas redações de jornalismo.

É possível também criar, a partir desta dissertação, um manual de formatação de equipes para o trabalho com Jornalismo de Dados. Naturalmente, o texto passaria por um processo de mudanças, para se tornar mais pedagógico. Com as devidas adequações, o material poderia servir de amparo a estudantes de jornalismo, a gestores de comunicação nas empresas e a pesquisadores que busquem novas formas de compreender esse universo.

Há, ainda, uma possibilidade no campo do desenvolvimento de softwares que otimizem ainda mais o trabalho jornalístico. A pesquisa pode tomar um caráter mais tecnológico no futuro, implicando na construção de uma plataforma de análise de dados, que pode ser um produto para diferentes empresas, sejam elas dos setores público ou privado. Essa atividade envolveria a venda de um plano de negócio para investidores e a montagem de uma *startup* voltada ao desenvolvimento da aplicação.

Descobrimos também que as possibilidades de incrementos e aprofundamentos do trabalho que fizemos durante o mestrado, além de serem muitos, são ainda desconhecidas. Como estamos em constante estado de mudança, é natural que haja oportunidades que ainda não conseguimos vislumbrar. No momento, importa mais saber que concluímos um longo período de pesquisas que podem ajudar jornalistas e cientistas de dados a buscarem interesses em comum, que os façam trabalhar em favor do público. É uma responsabilidade conjunta fazer crescer essa nova atividade e desenvolver módulos de trabalho que envolvam, cada vez mais, uma equipe heterogênea. A notícia só tem a ganhar com essa transfiguração.

Além do mais, o jornalismo sempre vai encontrar maneiras de permanecer no centro da sociedade como um serviço essencial. Em uma época de pessoas e objetos mergulhados numa quantidade gigante de dados, as empresas de comunicação demonstram uma importância singular. Não somente para veicular o que é notícia para diferentes públicos, mas cumprindo um papel de curadoria atenta de conteúdo. Em meio a tantas notícias falsas que são facilmente compartilhadas por meio da internet, o Jornalismo de Dados pode ser uma alternativa sensata para entregar os fatos verídicos à população. Os números e análises estatísticas são grandes aliados da notícia.

Com base nessas afirmações, encerramos propondo uma reflexão sobre a missão de informar a sociedade com vistas ao que há de mais essencial no jornalismo: a verdade. O que, de fato, acontece. E suas diferentes interpretações e perspectivas. Há muitos lados em cada história. Os dados permitem que partamos do princípio de que cada um deles é importante e merece atenção. Manualmente seria inconcebível coletar, interpretar e publicar todos. Mas se estiverem organizados e bem estruturados, há solução. Enquanto os profissionais de comunicação se mantiverem preocupados com a evolução de seu trabalho, o Jornalismo de Dados tem lugar no quadro de inovação do mercado.

REFERÊNCIAS

ALMEIDA, Yuri; BARBOSA, Suzana; TORRES, Vitor. **Bases de Dados como Agentes Estruturantes do Jornalismo no Contexto Multiplataforma**. In: MARTINS, E.; PALACIOS, M. (Orgs.). Ferramentas para Análise de Qualidade em Ciberjornalismo - Aplicações. Covilhã, PT: Livros LabCom, 2016.

AITARMURTO, Tanja. *Data journalism strategies in the US and Elsewhere*. Disponível em: http://virtual.vtt.fi/virtual/nextmedia/Deliverables-2011/D3.2.1.2.B_Hyperlocal_Trends_In%20Data_Journalism.pdf. Acesso em dez. 2019.

AITAMURTO, Tanja; LEHTONEN, Pauliina; SIRKUNNEN, Esa. *Trends In Data Journalism*, 2011. Disponível em: http://virtual.vtt.fi/virtual/nextmedia/Deliverables-2011/D3.2.1.2.B_Hyperlocal_Trends_In%20Data_Journalism.pdf. Acesso em dez. 2019.

AMARAL, Antônio José Rodrigues do; FILHO, Clóvis Ático Lima. **Mineração. Agência Nacional de Mineração**, 2007. Disponível em: <https://www.dnpm-pe.gov.br/Geologia/Mineracao.php>. Acesso em fev. 2020.

ANDERSON, Christopher. *The End of Theory: The Data Deluge Makes the Scientific Method Obsolete*. Disponível em: <https://www.wired.com/2008/06/pb-theory/>. Acesso em fev. 2020.

ANDERSON, Christopher W.; BELL, Emily; SHIRKY, Clay. **Jornalismo pós-industrial: adaptação aos novos tempos**. Revista de Jornalismo ESPM, n. 5, abr./jun. 2013.

ANDERTON-YANG et. al. Why Is Data Journalism Important?. In: **Manual de Jornalismo de Dados**, 2012. Disponível em: <https://datajournalism.com/read/handbook/one/introduction/why-is-data-journalism-important>. Acesso em fev. 2019.

ANGELUCI, Alan César Belo; OLIVEIRA, Ana Paula. **Competências e habilidades no jornalismo de dados: percepções sobre o perfil do profissional brasileiro**. Braz. Journal. Res. Brasília, v. 15, n. 1, p. 398-417, ago. 2019.

ARAÚJO, Lucas Vieira de. **A web e o jornalismo de dados: mapeamento de conceitos chave**. Revista Dispositiva, Minas Gerais, v. 5, n. 1, p. 144-163, jan./jul. 2016.

ARAÚJO, Lucas Vieira de; CARREIRA, Krishma; ROTH, Pedro Henrique. **Deep web e jornalismo: como transformar dados desta rede anônima em informação de relevância**

social. In: Pensacom Brasil, São Bernardo do Campo, SP, 2015. Anais eletrônicos. Disponível em: <http://portalintercom.org.br/anais/pensacom2015/resumos/032.pdf>. Acesso em fev. 2020.

ARONOVA, Elena; VON OERTZEN, Christine Von; SEPKOSKI, David. *Historicizing Big Data*. Max Planck Institute for the History of Science, 2013. Disponível em: https://www.mpiwg-berlin.mpg.de/research/projects/DeptII_Aronova_Oertzen_Sepkoski_Historicizing. Acesso em mar. 2020.

ARMAND; MATTELART, Michèle. História das Teorias da Comunicação. Título original: *Histoire des théories de la communication*. 16. ed. São Paulo: Edições Loyola, 1999.

BACCIN, Alciane; CANAVILHAS, João. *Contextualization in Hypermedia news report: narrative and immersion*. In: Brazilian Journalism Research. v. 1, n. 1, pp. 10-27, 2015. Disponível em: <https://bjr.sbpjor.org.br/bjr/article/view/803>. Acesso em jan. 2020.

BALDESSAR, Maria José. **Apontamentos sobre o uso do computador e o cotidiano dos jornalistas**. In: XXIV Congresso Brasileiro de Ciências da Comunicação, 2001, Campo Grande. Anais eletrônicos. Disponível em: <http://www.portcom.intercom.org.br/pdfs/146570658598096418172568944627582558771.pdf>. Acesso em 13 de fev. 2020.

BALDESSAR, Maria José; FEIJÓ, Valéria; TOASSI, Flavio. **Dados e contexto: o uso de dispositivos moveis para o acesso a informação: paralelos entre Brasil e Argentina/Lugares e não-lugares**. In: Congresso Brasileiro de Ciências da Comunicação, nº 36, 2013, Manaus, AM. p. 1-14. Anais eletrônicos. Disponível em: <http://www.intercom.org.br/papers/nacionais/2013/resumos/R8-1645-1.pdf>. Acesso em dez. 2019.

BARBOSA, Suzana. **Jornalismo digital em base de dados (JDBD): um paradigma para produtos jornalísticos digitais dinâmicos**. 2007. 331 f. Tese (Doutorado em Comunicação e Cultura Contemporâneas) – Faculdade de Comunicação, Universidade Federal da Bahia, Bahia.

BARBOSA, S. A.; TORRES, V. **O paradigma ‘Jornalismo Digital em Base de Dados’: modos de narrar, formatos e visualização para conteúdos**. Galaxia (São Paulo, Online), n. 25, p. 152-164, jun. 2013.

BAZZO, Jéssica; BRAGA, Claudomilson Fernandes; MARTINS, Dalton Lopes. **A incorporação da formação em Jornalismo de Dados nas escolas de jornalismo: o caso da Universidade Federal de Goiás**. In: X Simpósio Nacional da ABCiber, São Paulo, dez.

2017. Disponível em: <http://www.abciber.org.br/anais-abciber-2017.pdf>. Acesso em fev. 2020.

BECHTOLD, David. *What is big data?* 2014. Disponível em: <https://oldcolony.us/wp-content/uploads/2014/11/whatisbigdata-DKB-v2.pdf>. Acesso em nov. 2019.

BENDASSOLLI, Pedro F. et al. **Indústrias criativas: definição, limites e possibilidades**. Revista de Administração de Empresas (RAE), São Paulo, v. 49, n.1, 10-18, jan./mar. 2009.

BERNERS-LEE et al. *The World-Wide Web*. Communications of the ACM, v. 37, n. 8, p. 76-82, 1994. Disponível em <https://dl.acm.org/doi/10.1145/179606.179671>. Acesso em janeiro de 2019, às 11h05.

BERTOCCHI, Daniela et. al. **Possibilidades narrativas em dispositivos móveis**. In: CANAVILHAS, João; SATUF, Ivan. *Jornalismo para dispositivos móveis: produção, distribuição e consumo*. Livros Labcom books, 2015. Disponível em: <http://www.livroslabcom.ubi.pt/book/137>. Acesso em dez. 2019.

BOUNEGRU, Liliana. *Data Journalism in Perspective*. In: Manual de Jornalismo de Dados, 2012. Disponível em: <https://datajournalism.com/read/handbook/one/introduction/data-journalism-in-perspective>. Acesso em fev. 2019.

BOUNEGRU, Liliana. CHAMBERS, Lucy. GRAY, Jonathan. **Manual de Jornalismo de Dados**, 2012. Disponível em: <http://datajournalismhandbook.org/pt/index.html>. Acesso em fev. 2019.

BRADSHAW, Paul. *How to be a data journalist*. The Guardian (Datablog), 1 out. 2010. Disponível em: <https://www.theguardian.com/news/datablog/2010/oct/01/data-journalism-how-to-guide>. Acesso em 13 set. 2019.

_____. *The inverted pyramid of data journalism*. Online Journalism Blog, 7 jul. 2011. Disponível em: <http://onlinejournalismblog.com/2011/07/07/the-inverted-pyramid-of-data-journalism>. Acesso em: 25 fev. 2020.

_____. What Is Data Journalism? In: BOUNEGRU, Liliana. CHAMBERS, Lucy. GRAY, Jonathan. **Manual de Jornalismo de Dados**, 2012. Disponível em: <https://datajournalism.com/read/handbook/one/introduction/what-is-data-journalism>. Acesso em mar. 2019.

_____. Data Journalism with Impact. In: BOUNEGRU, Liliana. CHAMBERS, Lucy. GRAY, Jonathan. **Manual de Jornalismo de Dados 2**, 2014. Disponível em: <https://datajournalism.com/read/handbook/two/situating-data-journalism/data-journalism-with-impact>. Acesso em fev. 2020.

BRADSHAW, Paul; ROHUMAA, Liisa. *The online journalism handbook: skills to survive and thrive in the digital age*. Harlow: Longman, 2011.

BRASIL. Ministério do Planejamento, Orçamento e Gestão. **Lei nº 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. Diário Oficial da República Federativa do Brasil. Brasília, nov. 2011b, Seção I.

BRASIL. Ministério do Planejamento, Orçamento e Gestão. **Lei nº 13.709, de 14 de agosto de 2018**. Dispõe sobre a proteção de dados pessoais e altera a Lei nº 12.965, de 23 de abril de 2014 (Marco Civil da Internet). Diário Oficial da República Federativa do Brasil. Brasília, ago. 2018.

BRIGGS, Asa; BURKE, Peter. Uma história social da mídia: de Gutenberg à internet. Título original: *A social history of the media (From Gutenberg to the internet)*. Rio de Janeiro: Jorge Zahar Editora, 2002.

BRIZUELA, Elena. Periodismo de datos: una apuesta necesaria. In: PICCATO, Franco; ROITBERG, Gastón (Org.). *Periodismo disruptivo: dilemas y estrategias para la innovación*. Buenos Aires: La Crujía, 2015. p. 151-158.

CAMPOS, M. M.; SAITO, K. **Sistemas inteligentes em controle e automação de processos**. Rio de Janeiro: Editora Ciência Moderna Ltda., 2004)

CANAVILHAS, João. **Webjornalismo: 7 características que marcam a diferença**. Tradução de Hélder Prior. Covilhã (Portugal): Labcom books, 2014.

CASTELLS, Manuel. **A sociedade em rede**. Tradução de Roneide Venâncio Majer. 4. ed. São Paulo: Paz e Terra, 1999. A era da informação: economia, sociedade e cultura; v. 1.

CARLISLE, Wendy. *The ABC's Data Journalism Play*. In: BOUNEGRU, Liliana. CHAMBERS, Lucy. GRAY, Jonathan. **Manual de Jornalismo de Dados**, 2012. Disponível em: <http://datajournalismhandbook.org/pt/index.html>.

CODDINGTON, Mark. Clarifying Journalism's Quantitative Turn. In: *Digital Journalism*, v. 3, n.3, 2015, p. 398-415. Disponível em: <<https://doi.org/10.1080/21670811.2014.976400>>. Acesso em nov. 2019.

COHEN, Sara. Using visualizations to Tell Stories. In: **Manual de Jornalismo de Dados**, 2012. Disponível em: <https://datajournalism.com/read/handbook/one/delivering-data/using-visualizations-to-tell-stories>. Acesso em abr. 2019.

COSTA, Andriolli. **Os caminhos para um pós-jornalismo**. Observatório da Imprensa, 2014. Disponível em: http://www.observatoriodaimprensa.com.br/imprensa-em-questao/_ed806_os_caminhos_para_um_pos_jornalismo/. Acesso em jan. 2020.

CRUCIANELLI, Sandra. **Guia de ferramentas digitais para jornalistas 2.0**. Austin: Universidade do Texas, 2013.

CRUCIANELLI, Sandra; ZANCHELLI, Michael. *Integrating data journalism into newsrooms*, 2011. Disponível em: <https://ayomenulisfisip.files.wordpress.com/2011/02/integrating-data-journalism.pdf>. Acesso em jan. 2020.

DAGIRAL, Eric; PARASIE, Sylvain. *Data-driven journalism and the public good: "computer-assisted-reporters" and "programmer-journalists" in Chicago*. Revista New media & society, v. 15, Estados Unidos, p. 853-871, set. 2013.

DIAKOPOULOS, Nicholas. *Algorithmic Accountability: journalistic investigation of computational power structures*. In: Digital Journalism, v. 3, n.3, p. 398-415. Disponível em: <https://www.tandfonline.com/doi/abs/10.1080/21670811.2014.976411>. Acesso em out. 2018.

DIETRICH, Daniel et al. O que são Dados Abertos?, 2009. Disponível em: https://opendatahandbook.org/guide/pt_BR/what-is-open-data/. Acesso em jan. 2020.

_____. **Open Data Handbook**, 2011. Disponível em: https://opendatahandbook.org/guide/pt_BR/what-is-open-data/. Acesso em abr. 2020.

DÖRR, Konstantin. *Mapping the field of Algorithmic Journalism*. University of Zurich, 2016. Disponível em: <http://dx.doi.org/10.1080/21670811.2015.1096748>. Acesso em fev. 2020.

DOWNEY, Allen B. **Pense em Python: pense como um cientista da computação**. Califórnia: O'reilly media, 2016.

ESCOLA DE DADOS. **Raspagem e Jornalismo de Dados**. Disponível em <<https://escoladedados.org/tutoriais/raspagem-e-jornalismo-de-dados/>>. Acesso em 04 dez. 2018, às 08h19.

ESTEVANIM, Mayanna. **Processos no jornalismo digital: do Big Data à visualização de dados**. Dissertação (mestrado). Escola de Comunicações e Artes, Universidade de São Paulo, São Paulo, 2016. Disponível em: <https://www.teses.usp.br/teses/disponiveis/27/27152/tde-06032017-094238/publico/MAYANNAESTEVANIMVC.pdf>. Acesso em fev. 2020.

FILHO, Adelmo. **O segredo da pirâmide: para uma teoria marxista do jornalismo**. Porto Alegre: Tchê, 1987.

FOTIOS, Ricardo. **Analítica quantitativa criativa aplicada ao ensino e à prática do Jornalismo de Dados para alunos da graduação em comunicação social**. Academia.edu, 2018. Disponível em: https://www.academia.edu/26373467/Anal%C3%ADtica_no_ensino_do_jornalismo_de_dados. Acesso em mar. 2020.

GARCÍA, José López; RODRÍGUEZ, Alba Silva; WESTLUND, Oscar. **Prensa, smartphones y tablets: la industria de medios impresos se sube al periodismo móvil**. Telos (Cadernos de Comunicação e Inovação), Madri, p. 1-14, Fev-Mai 2015.

GEHRKE, Marília. **Jornalismo de dados e conhecimento científico: uma aproximação possível**. In: SBPJor – Associação Brasileira de Pesquisadores em Jornalismo, nº 14, Palhoça, SC, 2016. Anais eletrônicos. Disponível em: <http://sbpjour.org.br/congresso/index.php/sbpjour/sbpjour2016/paper/viewFile/60/173>. Acesso em nov. 2019.

GIL, Antonio Carlos. **Como elaborar projetos de pesquisa/ Antonio Carlos Gil. - 4. ed. - São Paulo: Atlas, 2002.**

GORDON, Rich. **Want to build a data journalism team? You'll need these three people**. Knight Lab, 2013. Disponível em: <https://knightlab.northwestern.edu/2013/06/28/want-to-build-a-data-journalism-team-youll-need-these-three-people/>. Acesso em fev. 2020.

GRANDIN, Felipe Rodrigues. **Jornalismo guiado por dados como forma contemporânea de produção de sentido**. In: Congresso Brasileiro de Ciências da Comunicação, nº 37, 2014, Foz do Iguaçu, PR. p. 1-15.

GROOPMAN, Jessica. *Design an IoT user experience, not an IoT product*. Disponível em <https://internetofthingsagenda.techtarget.com/feature/Design-an-IoT-user-experience-not-an-IoT-product>. Acesso em junho de 2018.

HAMANN, Renan. **Do bit ao Yottabyte: conheça os tamanhos dos arquivos digitais [infográfico]**. Techmundo, 2011. Disponível em: <https://www.tecmundo.com.br/infografico/10187-do-bit-ao-yottabyte-conheca-os-tamanhos-dos-arquivos-digitais-infografico-.htm>. Acesso em abr. 2020.

HAMMOND, P. (2017). *From computer-assisted to data-driven: Journalism and Big Data*. Journalism, 18(4), 408–424. Londres: Sage Publications.

HAN, J.; KAMBER, M. *Data mining: concepts and techniques*. São Francisco (CA-EUA): Morgan Kaufmann Publishers, 2001.

HAYASHI, C. (1998). **O que é Ciência de Dados? Conceitos Fundamentais e um Exemplo Heurístico**. Em: Hayashi C., Yajima K., Bock HH., Ohsumi N., Tanaka Y., Baba Y. (eds) *Ciência de Dados, Classificação e Métodos Relacionados. Estudos em Classificação, Análise de Dados e Organização do Conhecimento*. Springer, Tóquio.

HAYKIN, Simon. *Redes Neurais: princípios e prática*. Título original: **Neural networks: a comprehensive foundation**. 2ª ed. Porto Alegre: Bookman, 2001.

HOLOVATY, Adrian. **A fundamental way newspaper sites need to change**. Adrian Holovaty, 6 set. 2006. Disponível em: <http://www.holovaty.com/writing/fundamental-change>. Acesso em: fev. 2020.

HOUAISS, Antônio; VILLAR, Mauro de Salles; FRANCO, Francisco Manoel de Mello. **Dicionário Houaiss da língua portuguesa**. Rio de Janeiro: Objetiva, 2009. 1986 págs.

HOWARD, Alexander Benjamin. *The Art and Science of Data-driven Journalism: when journalists combine new technology with narrative skills, they can deliver context, clarity, and a better understanding of the world around us*. Columbia: Columbia Journalism School, 2014.

HOWKINS, John. *Economia criativa : como ganhar dinheiro com ideias criativas*. Título original: *The creative economy : how people make money from ideas*. São Paulo : M. Books, 2013.

HUMBY, Clive. In: *Think Big. Britain's data opportunity*. Wandisco, 2006. Disponível em: https://www.wandisco.com/assets/blt018be6e4943f732d/ThinkBig_BritainsDataOpportunity.pdf. Acesso em fev. 2020.

JENKINS, Henry. **Cultura da Convergência**. Tradução: Susana Alexandria. – 2 ed. – São Paulo: Aleph, 2009.

JUNG C. G. **Os Arquétipos e o Inconsciente Coletivo**. Petrópolis: Vozes, 2003.

JUNIOR, Walter Teixeira Lima. **Big Data, Jornalismo Computacional e Data Journalism: estrutura, pensamento e prática profissional na Web de dados**. Revista Estudos em Comunicação, n. 12. São Paulo: Universidade Metodista de São Paulo, dez. 2012. P. 207-222. Disponível em: <http://www.ec.ubi.pt/ec/12/pdf/EC12-2012Dez-11.pdf>. Acesso em fev. 2020.

_____. **Jornalismo computacional em função da “Era do Big Data”**. Revista Líbero, v. 14, n. 28, p. 45-52, dez. 2011. Disponível em: <http://seer.casperlibero.edu.br/index.php/libero/article/view/329/303>. Acesso em set. 2019.

_____. **Jornalismo inteligente na era do data mining**. Revista Líbero, ano 9, nº 18, p. 119-126, dez. 2006. Disponível em: <https://casperlibero.edu.br/wp-content/uploads/2014/05/Jornalismo-inteligente-na-era-do-data-mining.pdf>. Acesso em set. 2019.

JUNIOR, Walter Teixeira Lima; OLIVEIRA, Andre Rosa de. **Jornalismo estruturado: uso de metadados para enriquecimento de bases noticiosas na web**. Revista Famecos | Mídia, cultura e tecnologia, Porto Alegre, v. 24, n. 3, set-dez. 2017.

KAO, D; HARRELL, D. F. *Toward Understanding the Impact of Visual Themes and Embellishment on Performance, Engagement, and Self-Efficacy in Educational Games*. Disponível em: <http://people.csail.mit.edu/dkao/pdf/kao2017aera-r.pdf>. Acesso em fev. 2020.

KUENG, Kucy. *Handling inter-generational tensions in news media*. Disponível em: <http://www.digitalnewsreport.org/publications/2020/handling-inter-generational-tensions-news-media/>. Acesso em jan. 2020.

LAKATOS, Eva Maria. **Metodologia do trabalho científico: procedimentos básicos, pesquisa bibliográfica, projeto e relatório, publicações e trabalhos científicos** / Eva Maria Lakatos, Marina de Andrade Marconi. 4. ed. São Paulo: Atlas, 1992.

LEIMDORFER, Andrew. Data Journalism at the BBC. In: BOUNEGRU, Liliana. CHAMBERS, Lucy. GRAY, Jonathan. **Manual de Jornalismo de Dados**, 2012. Disponível em: <http://datajournalismhandbook.org/pt/index.html>.

LEIMDORFER, Andrew; THEREAUX, Olivier. *How open data is redefining the roles of the journalist, audience and publisher*, 2012. Disponível em: https://www.w3.org/2012/06/pmod/pmod2012_submission_9.pdf. Acesso em mar. 2020.

LEMOS, André. **Cibercultura: tecnologia e vida social na cultura contemporânea**. 7ª ed - Porto Alegre: Sulina, 2002. 295 págs.

LEVY, Pierre. Cibercultura. Título original: *Ciberculture*. Tradução de Carlos Irineu da Costa. - São Paulo: Editora 34, 1999 (3ª edição). 272 págs.

LEWIS, Seth C. *Journalism In An Era Of Big Data. In: Digital Journalism*, v. 3, n.3, p. 321-330. Disponível em: <https://www.tandfonline.com/doi/pdf/10.1080/21670811.2014.976399>. Acesso em 29 out. 2018.

LEWIS, Seth; USHER, Nikki. *Open source and journalism: toward new frameworks for imagining news innovation*. Media, culture & society, v. 35, n. 5, p. 602-619, 2013.

LIPPMANN, Walter. Opinião pública. Título original: *Public opinion*. Rio de Janeiro: Vozes, 1922.

LORENZ, Mirko. Personalização: Análise aos 6 graus. In: CANAVILHAS, João (org.). **Webjornalismo: 7 características que marcam a diferença**. Tradução de Hélder Prior. Covilhã (Portugal): Labcom books, 2014.

_____. Why Journalists Should Use Data. In: BOUNEGRU, Liliana. CHAMBERS, Lucy. GRAY, Jonathan. **Manual de Jornalismo de Dados**, 2012. Disponível em: <http://datajournalismhandbook.org/pt/index.html>.

LUNDY-EKMAN, Laurie. *Neuroscience: fundamentals for rehabilitation*. Elsevier: Amsterdã, 2007.

MALULY, L. Maria Elisabete Antonioli: **Novos desafios à formação superior em jornalismo**. Revista Alterjor, v. 13, n. 1, p. 34-39, 2 maio 2016.

MANCINI, Leonardo; VASCONCELLOS, Fabio. **Jornalismo de Dados: conceito e categorias**. Revista Fronteiras, v. 18, n. 1. jan-abr 2016. p. 69-82. Disponível em <http://revistas.unisinos.br/index.php/fronteiras/article/view/fem.2016.181.07/5300>. Acesso em jan. 2020.

MANOVICH, Lev. *The Language of New Media*. Massachussets: MIT Press, 2001.

Manual da Redação: Folha de São Paulo. 16. ed. - São Paulo: Publifolha, 2001.

MARCONDES FILHO, Ciro. **Ser jornalista: o desafio das tecnologias e o fim das ilusões**. São Paulo: Paulus, 2009.

MARTINHO, Ana Isabel Pinto. **Jornalismo de dados: caraterização e fluxos de trabalho**. Revista Comunicação e Ciências Empresariais. n. 9., 2014. Disponível em: <https://repositorio.iscte-iul.pt/bitstream/10071/17493/1/n9-B3.pdf>. Acesso em jan. 2020.

_____. **Jornalismo de dados: contributo para uma caracterização do estado da arte em Portugal**. Lisboa: ISCTE-IUL, 2013. Dissertação de mestrado. Disponível <http://hdl.handle.net/10071/8329>. Acesso em 16 de fev. 2020.

MELLO, Fábio Bandeira de. Webjornalismo: a era da personalização em massa. In: **Transmutações no jornalismo** [Livro eletrônico]/ Fernando Firmino da Silva (Org.). – Campina Grande: EDUEPB, 2016.

MEYER, Philip. *Precision Journalism*. Bloomington: Indiana University Press, 1973.

_____. *The new precision journalism*. Bloomington: Indiana University Press, 1991.

MIELNICZUK, Luciana; TRÄSEL, Marcelo. **Jornalismo guiado por dados como inovação profissional e seus desafios para a educação**. Revista Contemporânea | comunicação e cultura, v.15, n.02. mai-ago 2017. p. 609-629. Disponível em: <https://portalseer.ufba.br/index.php/contemporaneaposcom/article/view/21510>. Acesso em out. 2019.

NAGIFY, Richard. *The differences between Artificial and Biological Neural Networks*. Disponível em: <https://towardsdatascience.com/the-differences-between-artificial-and-biological-neural-networks-a8b46db828b7>. Acesso em fev. 2020.

NEGREDO, S.; SALAVERRÍA, R. *Periodismo integrado: convergência de médios y reorganizacion de redacciones*. Barcelona: Editorial Sol90, 2008.

NETO, José Cavalcanti Sobrinho. Novos formatos de distribuição de notícias: The New York Times no relógio inteligente. In: **Transmutações no jornalismo** [Livro eletrônico]./ Fernando Firmino da Silva (Org.). – Campina Grande: EDUEPB, 2016.

NEWMAN, Nic. *Journalism, Media, and Technology Trends and Predictions 2020*. Disponível em: <http://www.digitalnewsreport.org/publications/2020/journalism-media-and-technology-trends-and-predictions-2020/>. Acesso em fev. 2020.

NORVIG, P.; RUSSEL, S. Inteligência Artificial. Título original: *Artificial Intelligence*. Rio de Janeiro: Elsevier, 2004.

OREMUS, Will. *The First News Report on the L.A. Earthquake Was Written by a Robot*. Slate, 2014. Disponível em: <https://slate.com/technology/2014/03/quakebot-los-angeles-times-robot-journalist-writes-article-on-la-earthquake.html>. Acesso em set. 2019.

OSÓRIO, Fernando. **Redes neurais artificiais: do aprendizado natural ao aprendizado artificial**. I Fórum de Inteligência Artificial: Canoas, 1999. Disponível em: <http://osorio.wait4.org/oldsite/IForumIA/fia99-ppt.pdf>. Acesso em mar. 2020.

PALACIOS, Marcos. Memória: Jornalismo, memória e história na era digital. In: CANAVILHAS, João (org.). **Webjornalismo: 7 características que marcam a diferença**. Tradução de Hélder Prior. Covilhã (Portugal): Labcom books, 2014.

_____. Ruptura, Continuidade e Potencialização no Jornalismo Online: o Lugar da Memória. In: **Modelos do Jornalismo Digital**. Salvador: Editora Calandra, 2003.

PARK, Robert. **News as a form of knowledge: a chapter in the sociology of knowledge**. American Journal of Sociology, v. 45, n. 5. Chicago: The University of Chicago Press, 1940. Disponível em: <http://www.jstor.org/stable/2770043>. Acesso em fev. 2020.

PAVLIK, John V. Ubiquidade: O 7º princípio do jornalismo na era digital. In: CANAVILHAS, João (org.). **Webjornalismo: 7 características que marcam a diferença**. Tradução de Hélder Prior. Covilhã (Portugal): Labcom books, 2014.

PENA, Felipe. **Teoria do Jornalismo**. 2 ed. São Paulo: Contexto, 2005.

PETRE, Caitlin. **A Quantitative Turn in Journalism?** Chartbeat, 2013. Disponível em: <https://blog.chartbeat.com/2013/10/31/quantitative-turn-journalism/>. Acesso em mar. 2020.

PORCELLO, Flávio. **De Boni a Boninho: a mudança estética e conceitual da TV brasileira para atrair a nova classe média.** In: XXXV Congresso Brasileiro de Ciências da Comunicação. Anais eletrônicos. Fortaleza: setembro, 2012. Disponível em: <http://www.intercom.org.br/sis/2012/resumos/R7-0450-1.pdf>. Acesso em nov. 2019.

ROGERS, Simon. **John Snow's data journalism: the cholera map that changed the world.** Disponível em: <https://www.theguardian.com/news/datablog/2013/mar/15/john-snow-cholera-map>. Acesso mar. 2020.

_____. 2013. **Data journalism at the Guardian: what is it and how do we do it?** Disponível em: <https://www.theguardian.com/news/datablog/2011/jul/28/data-journalism>. Acesso em jan. 2020.

ROST, Alejandro. Interatividade: Definições, estudos e tendências. In: CANAVILHAS, João (org.). **Webjornalismo: 7 características que marcam a diferença.** Covilhã (Portugal): Labcom books, 2014.

RUIC, Gabriela. **Novo coronavírus é incomparável, diz cientista que ajudou a conter o ebola.** São Paulo: Exame, 2020. Disponível em: <https://exame.abril.com.br/mundo/novo-coronavirus-e-incomparavel-diz-cientista-que-ajudou-a-conter-o-ebola/>. Acesso em mar. 2020.

SALAVERRÍA, Ramón. Multimedialidade: Informar para cinco sentidos. In: CANAVILHAS, João (org.). **Webjornalismo: 7 características que marcam a diferença.** Covilhã (Portugal): Labcom books, 2014.

SCHELLONG, Alexander; STEPANETS, Ekaterina. **Unchartered Waters - The State of Open Data in Europe.** CSC - Public Sector Study Series, 2011. Disponível em: <https://joinup.ec.europa.eu/sites/default/files/document/2014-12/Unchartered%20Waters%20-%20The%20State%20of%20Open%20Data%20in%20Europe.pdf>. Acesso em mar. 2020.

SCHWAB, Klaus. A quarta revolução industrial. Título original: **The fourth industrial revolution.** 1 ed. São Paulo: Editor, 2016.

SIRKKUNEN, Esa. **Overview: Public data and data journalism in some countries.** Trends In Data Journalism, 2011. Disponível em: http://virtual.vtt.fi/virtual/nextmedia/Deliverables-2011/D3.2.1.2.B_Hyperlocal_Trends_In%20Data_Journalism.pdf. Acesso em dez. 2019.

SODRÉ, Muniz. **Técnica de reportagem: notas sobre a narrativa jornalística** / Muniz Sodré, Maria Helena Ferrari. - São Paulo: Summus, 1986. 5 ed. - 141 pags.

SUDA, Brian. Designing With Data. In: **Manual de Jornalismo de Dados**, 2012. Disponível em: <https://datajournalism.com/read/handbook/one/delivering-data/designing-with-data>. Acesso em mar. 2019.

SUNNE, Samantha. **Diving into Data Journalism: Strategies for getting started or going deeper**. Disponível em: <https://www.americanpressinstitute.org/publications/reports/strategy-studies/data-journalism/single-page/>. Acesso em jan. 2020.

TRAQUINA, Nelson. **Jornalismo: questões, teorias e "estórias"**. Lisboa: Vega, 1993.

_____. **O poder do jornalismo: análise e textos da teoria do agendamento**. Coimbra: Minerva, 2001.

_____. **Teorias do jornalismo, por que as notícias são como são**. 2. ed. Florianópolis: Insular, 2004.

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. Tese (Doutorado em Comunicação Social) — Faculdade de Comunicação Social, Pontifícia Universidade Católica do Rio Grande do Sul, 2014a.

_____. **Jornalismo guiado por dados: aproximações entre a identidade jornalística e a cultura hacker**. Estudos em Jornalismo e Mídia, v. 11, n. 1, 2014b.

_____. **Jornalismo guiado por dados: relações da cultura hacker com a cultura jornalística**. In: XXII Encontro Anual da Compós. Anais eletrônicos. Salvador: 2013. Disponível em: http://www.compos.org.br/data/biblioteca_2065.pdf. Acesso em dez. 2019.

_____. **Aprendendo a se deixar guiar por dados: a formação dos jornalistas da equipe Estadão Dados**. Revista Brasileira de Ensino de Jornalismo, Brasília, v. 4, n. 14, p. 85-99, jan./jun. 2014c.

TURING, A. M. **Computing machinery and intelligence**. Revista Mind, v. 59, n. 236, out. 1950, p. 433-460, Reino Unido. Disponível em: <https://academic.oup.com/mind/article/LIX/236/433/986238>. Acesso em jan. 2020.

VENOHR, Sascha. Data Journalism at the Zeit Online. In: BOUNEGRU, Liliana. CHAMBERS, Lucy. GRAY, Jonathan. **Manual de Jornalismo de Dados**, 2012.. Disponível em: <http://datajournalismhandbook.org/pt/index.html>.

VENTURA, Mariana Pires. **Jornalismo de Dados como diferencial: o caso do Nexo**. Pauta Geral - Estudos em Jornalismo, v. 5, n. 2, p. 240-254, dez. 2018. Disponível em: <https://www.revistas2.uepg.br/index.php/pauta/article/view/12290>. Acesso em mar. 2020.

VIZEU, Alfredo. **Decidindo o que é notícia: os bastidores do telejornalismo**. Porto Alegre: Edipucrs, 2000.

WOLF, Mauro. **Teorias da Comunicação**. Lisboa: Presença, 2002.

GLOSSÁRIO

ALGORITMO: é composto por instruções a serem executadas por um programa.

API: termo em inglês que significa Interface de Programação de Aplicação. Se refere a regras características que permitem interações entre programas.

BIG DATA: termo em inglês referente à grande quantidade de dados produzidos no mundo atual.

BOTS: termo em inglês para se referir a robôs. Vem da palavra robot. Agente computacional autônomo.

CHATBOT: termo utilizado para se referir a programas de computador que interagem com humanos, geralmente, por meio de mensagens. Também é um agente computacional autônomo.

DASHBOARD: painéis de visualização de dados para tomada de decisões.

DATA ANALYST: termo utilizado para definir o cargo do profissional que trabalhar com análise de dados.

DATA SCIENTIST: termo em inglês que significa Cientista de Dados. É o profissional capaz de extrair conhecimento dos dados para auxiliar na tomada de decisões em variadas áreas do conhecimento.

DATA WAREHOUSING: termo em inglês utilizado para se referir ao armazenamento de dados num repositório, também chamado de warehouse.

DATASET: termo em inglês para se referir a conjunto de dados.

HACKER: termo em inglês para se referir a uma pessoa com uma ampla habilidade técnica em tecnologia.

HARDWARE: termo em inglês utilizado para se referir às partes físicas do computador.

MACHINE LEARNING: termo em inglês que significa aprendizagem de máquina. Geralmente é associado ao trabalho automático de máquinas que analisam dados, executam tarefas e tomam decisões utilizando algoritmos e sem intervenção humana.

INPUT: termo utilizado para se referir à entrada de dados em um programa.

INSIGHT: termo utilizado para se referir a ideias repentinas.

KNOW-HOW: conhecimento em determinado assunto que é adquirido através de experiência.

OUTPUT: termo utilizado para se referir à saída de dados em um programa.

PYTHON: linguagem de programação muito utilizada para analisar dados e uma das mais conhecidas entre os jornalistas.

QUERY: termo em inglês que significa consulta. Geralmente utilizado como sinônimo de tarefas a serem executadas por um programa de computador para responder as perguntas do usuário.

R: linguagem de programação muito utilizada para desenvolver programas de análise de dados e estatísticas. Depois da Python, é a segunda linguagem mais comum nas redações de jornalismo.

REBOOT: termo em inglês que significa reiniciar. Comumente utilizado para caracterizar um recomeço.

RED FLAG: termo em inglês que significa bandeira vermelha. Geralmente utilizado para se referir a alertas dados por um programa de computador.

SOFTWARE: termo em inglês que se refere ao conjunto de códigos desenvolvidos para executar tarefas, geralmente associado a aplicações/programas de computador.

SQUAD: termo em inglês que significa equipe de trabalho, geralmente voltada a uma atividade específica.

STARTUP: termo em inglês utilizado para se referir a empreendimentos em estágio inicial, com ampla possibilidade de crescimento e geralmente vinculados à área de tecnologia.