

O estado do Jornalismo de Dados no cenário luso-brasileiro

Isabella Cristina Moura

Dissertação de Mestrado em Jornalismo

(Versão corrigida e melhorada após defesa pública)

Março, 2018

Dissertação apresentada para cumprimento dos requisitos necessários à obtenção do grau de Mestre em Jornalismo, realizada sob a orientação científica do Professor Doutor António Granado

*À minha irmã,
Marcela Moura (in memoriam)*

AGRADECIMENTOS

Gratidão é um sentimento tão completo, que simplifica tudo o que queremos dizer. É o sentimento que mais combina com o que sinto pelos meus pais. Se hoje apresento esta dissertação, devo à minha mãe Helena e o meu pai Carlos por tudo o que fazem por mim, o meu amor aos dois.

Aos meus irmãos, Ana Carolina e Enzo por ensinarem valores como partilha e compreensão e por fazerem de mim uma pessoa mais feliz.

Ao professor Ricardo Morais, o meu reconhecimento de grande mestre que é e o meu obrigada por tanto me ensinar e por estar presente em todas as etapas do meu percurso académico, na licenciatura na Covilhã, na candidatura ao mestrado e no apoio nesta dissertação.

Ao meu orientador professor doutor António Granado por ter aceite o desafio, por ter indicado os caminhos a serem seguidos e pelos ensinamentos dados desde o primeiro dia de aulas.

Aos meus amigos e familiares, por estarem comigo quando precisei e por compreenderem a ausência e impaciência.

Aos jornalistas que colaboraram na construção deste trabalho através de entrevistas e de respostas ao questionário.

O ESTADO DO JORNALISMO DE DADOS NO CENÁRIO LUSO-BRASILEIRO

ISABELLA CRISTINA MOURA

RESUMO

A existência de fenómenos estimulados por esferas cívicas e tecnológicas, como o movimento pela disponibilização de dados e a popularização de aparatos técnicos que permitiram o crescimento da informação armazenada e acessível, foi essencial para que as bases de dados se tornassem fontes na construção jornalística. A consolidação do conceito de Jornalismo de Dados revelou-se um processo gradativo, que na presente década adquiriu proporções que possibilitaram a produção ativa nas redações e o fortalecimento como especialização jornalística. Este trabalho tem como objetivo expor sob quais condições e como é feito o Jornalismo de Dados no Brasil e em Portugal, usando como critério o testemunho de jornalistas de ambos os países.

PALAVRAS-CHAVE: Bases de dados, *Big Data*, Jornalismo de Dados, Dados abertos, Brasil, Portugal

THE STATE OF DATA JOURNALISM IN THE LUSO-BRAZILIAN SCENARIO

ISABELLA CRISTINA MOURA

ABSTRACT

The existence of phenomena stimulated by civic and technological spheres, as long as the open data movement and the popularity of the technical appliances that allowed the growth of the stored and available information, was essential for the databases to become sources in journalistic construction. The concept consolidation of Data Journalism reveals a gradual process, that in the present decade acquired proportions that have made possible the active production in the essays and the fortification as journalistic specialization. This study aims to expose under any conditions and how is done the Data Journalism in Brazil and in Portugal, using as criteria the testimony of journalists from both countries.

KEYWORDS: Databases, Big Data, Data journalism, Open data, Brazil, Portugal

ÍNDICE

Introdução.....	1
1. Contextualização histórica.....	3
1.1. O surgimento das bases de dados.....	3
1.2. Computer-assited reporting: o início	5
1.3. <i>Precision journalism</i> : da metodologia à epistemologia	9
1.4. Transparência e <i>open access</i>	12
2. Conceitualização.....	16
2.1. Propostas de definição do conceito de Jornalismo de Dados	16
2.2. Visualização de dados: uma componente indissociável	26
2.3. Desafios profissionais na adaptação ao Jornalismo de Dados	28
3. O estado do Jornalismo de Dados	32
3.1. O Jornalismo de Dados no Brasil.....	32
3.2. O Jornalismo de Dados em Portugal.....	34
4. Metodologia.....	36
5. Análise dos resultados	38
5.1. Vias de acesso ao Jornalismo de Dados	38
5.2. Aquisição de novas habilidades	41
5.3. O Jornalismo de Dados dentro das redações.....	44
5.4. A origem dos dados	46
5.5. A autoridade das fontes baseadas em dados	49
5.6. As etapas de produção de conteúdos baseados em dados.....	52
5.7. De que forma o investimento é sentido.....	57
5.8. Constrangimentos ao crescimento do Jornalismo de Dados	60
5.9. Quem faz Jornalismo de Dados no Brasil.....	63
5.10. Quem faz Jornalismo de Dados em Portugal	67
Conclusão	71
Referências bibliográficas	76
Anexos.....	81

Lista de figuras

- Figura 1 – Mapas utilizados na contextualização da reportagem
- Figura 2 - Pirâmides propostas por Paul Bradshaw (2011) na construção de peças através de Jornalismo de Dados
- Figura 3 - Componentes de classificação dos níveis de Jornalismo de Dados , in Mancini e Vasconcellos (2016)
- Figura 4 - Diagrama proposto por Florence Nightingale para demonstrar as principais causas de morte dos soldados britânicos.
- Figura 5. Etapas jornalísticas seguidas por Cristian Weiss no trabalho com dados
- Figura 6. Etapas jornalísticas seguidas por Juan Torres no trabalho com dados
- Figura 7. Etapas jornalísticas seguidas por Daniel Bramatti no trabalho com dados
- Figura 8. Etapas jornalísticas seguidas por Marlen Couto no trabalho com dados
- Figura 9. Etapas jornalísticas seguidas por Raphael Hernandez no trabalho com dados
- Figura 10. Etapas jornalísticas seguidas por Roberto Maleson no trabalho com dados
- Figura 11. Etapas jornalísticas seguidas por Katia Brembatti no trabalho com dados
- Figura 12. Etapas jornalísticas seguidas por Rita Costa no trabalho com dados
- Figura 13. Etapas jornalísticas seguidas por Rita Costa no trabalho com dados
- Figura 14. Etapas jornalísticas seguidas por Rui Barros no trabalho com dados

Lista de tabelas

- Tabela 1. Fórmula conceitual de Jornalismo de Dados (Crucianelli, 2013)
- Tabela 2. Redações brasileiras sem estrutura fixa para o trabalho de Jornalismo de Dados
- Tabela 3. Redações brasileiras com equipas ou núcleos de dados
- Tabela 4. Como o Jornalismo de Dados está organizado nas redações portuguesas
- Tabela 5. Redações com produção de bases de dados
- Tabela 6. Redações sem produção de bases de dados
- Tabela 7. Regularidade da produção de bases de dados nas redações portuguesas exploradas
- Tabela 8. Domínio de ferramentas pelos inquiridos brasileiro
- Tabela 9. Domínio de ferramentas pelos inquiridos portugueses

Lista de gráficos

- Gráfico 1. Cidade de trabalho dos jornalistas (Brasil)
- Gráfico 2. Regime editorial onde os jornalistas estão inseridos (Brasil)
- Gráfico 3. Principal método de aprendizagem adotado(Brasil)
- Gráfico 4. Distribuição por sexo (Brasil)
- Gráfico 5. Domínio de programação pelos jornalistas (Brasil)
- Gráfico 6. Distribuição da responsabilidade de produção de visualização (Brasil)
- Gráfico 7. Cidade de trabalho dos jornalistas (Portugal)
- Gráfico 8. Regime editorial onde os jornalistas estão inseridos (Portugal)
- Gráfico 9. Principal método de aprendizagem adotado(Portugal)
- Gráfico 10. Distribuição por sexo (Potugal)

INTRODUÇÃO

Desde que o jornalismo online deixou de ser uma réplica dos conteúdos veiculados no papel, o campo da comunicação social assistiu ao surgimento de novos produtos. As potencialidades do meio, a necessidade de reinventar o modelo de negócios e o surgimento de nichos foram algumas das condicionantes que permitiram a emergência de novas práticas.

A criação de significados através do trabalho jornalístico também se tem alterado consoante a evolução tecnológica. Após a imprensa escrita surgiu a oralidade da gravação, que foi seguida pela imagem estática, pela imagem em movimento e agora, pelos números. Cada vez mais, o jornalismo parece se apropriar da grande massa de informação disponível e acessível utilizando aquilo que é definido como Jornalismo de Dados. A internet é o suporte que reúne e permite todas essas alterações.

A produção de dados estimulada pelas Tecnologias da Informação e pela ubiquidade da internet, propiciaram o surgimento do *Big Data*, que tem como parte constituinte as bases de dados que podem ser explicadas como modelo de sistematização de coleta e armazenamento de informação. Como refere Stray (2014), os “dados não são de origem natural”, as bases de dados são construções deliberadas de informação, são registos que são utilizados para conhecer uma determinada realidade.

O Jornalismo de Dados prioriza as informações massificadas e organizadas por números ou outros caracteres para encontrar tendências e padrões que sejam capazes de explicar eventos e fenómenos. “Se a quantificação transforma o mundo em dados, a análise transforma os dados em conhecimento” – a abordagem de Stray (2014) é uma tentativa de explicar que métodos quantitativos e estatística fazem parte do leque de novos conhecimentos que o Jornalismo de Dados exige dos jornalistas na comunicação dos factos.

As principais alterações definidoras do conceito de Jornalismo de Dados estão centradas nos procedimentos técnicos adotados pelos profissionais e também na origem da fonte. Os conceitos de Reportagem Assistida por Computador (RAC) e de Jornalismo de Precisão que, como aponta Anderson (2014), foi o ponto mais evidente de “fusão” entre as ciências sociais, o armazenamento de dados e as práticas jornalísticas, representam o estágio inicial da prática.

A lógica de que o jornalismo deve servir como instrumento de vigilância da máquina do Estado é acentuada com a proposta do Jornalismo de Dados que, como expõe Träsel (2014), recorre à “publicitação de dados gerados pelas instituições públicas e, eventualmente, organizações privadas que transacionem com essas instituições”, para assegurar a transparência da dinâmica governamental.

Entendemos que o crescimento do Jornalismo de Dados representa não só uma alteração processual na rotina jornalística, mas também uma adequação dos conteúdos a uma sociedade em rede, que necessita cada vez mais de compreender a informação pública disponibilizada, mas que não possui habilidades para decodificar os dados e gerar significação.

Para apresentar o estado atual – no Brasil e em Portugal – deste procedimento jornalístico, que utiliza técnicas de extração e análise de bases de dados na construção social da realidade, tentamos, num primeiro momento, evidenciar o crescimento do Jornalismo de Dados, que contou com o estímulo de movimentos da comunidade jornalística iniciados na década de 1950 e suportados pela evolução computacional. Seguidamente, a partir de elementos qualitativos, baseados em entrevistas com jornalistas que atuam no campo do Jornalismo de Dados em redações brasileiras e portuguesas, e quantitativos, suportados pela resposta a dois questionários que contam com a mesma estrutura e diferem apenas na adaptação do título para o país dos inquiridos, procuramos expor a realidade do trabalho jornalístico com dados nos dois países, através da combinação e análise das duas componentes.

Este trabalho assume o propósito de gerar conhecimento numa área que procura afirmação profissional e que ainda se apresenta como pouco investigada no universo académico. Ainda que com limitações – derivadas de um trabalho de dissertação que requer a seleção de uma amostra que, por si só, já é reduzida – consideramos que esta pesquisa agrega valor na compreensão do Jornalismo de Dados.

1. Contextualização histórica

1.1. O surgimento das bases de dados

Desde o século XVII que o recenseamento demográfico é tido como costume na produção de conhecimento sobre a população. A coleta de dados, que nos estágios iniciais representava um processo lento, é transformada em estatística e posteriormente analisada como principal fonte na percepção de como vivem os cidadãos.

O ciclo de coleta das repostas somado aos anos necessários para a leitura e formulação de resultados, retirava a atualidade aos dados. Para suprir a necessidade do governo no levantamento das respostas obtidas, surgiu o primeiro instrumento da computação moderna (COX, 2000). Herman Hollerith foi o responsável por, em 1890, criar uma máquina que permitiu a automatização da contagem das informações recolhidas pelo recenseamento da população dos Estados Unidos. O que demonstra que antes mesmo da ubiquidade da internet e dos computadores, o armazenamento estruturado de informação aliado a aparatos tecnológicos representava uma combinação essencial na compreensão da realidade.

Na década de 1950, o armazenamento de informação assente num sistema manual rudimentar com fitas foi substituído por processos computacionais propiciados pelo desenvolvimento de discos rígidos. As bases de dados permitem adicionar, remover, coletar e disponibilizar informação e foram concebidas, num primeiro momento, para fins comerciais e institucionais. Essa proposta de gestão de informação surgiu como forma uma de facilitar tarefas administrativas (HAIGH, 2009) e é caracterizada como um sistema de programação que possui duas capacidades essenciais: a de gerenciar dados de forma permanente e a de permitir a gestão de um grande volume de informação (ULLMAN, 1988).

A sistematização da internet está assente em bases de dados que acumulam informação e permitem o funcionamento correto da máquina. Esse ciclo tecnológico é o precursor da onda da informação, destacada pelo *Big Data*, que trouxe transformações na rotina industrial e pública. A presença de dados na construção do molde de funcionamento da Internet pode ser retratada através da comparação formulada por Lev Manovich, que sugere que as bases de dados são para a era digital aquilo que a narrativa foi para a era moderna das novelas e do cinema (SCHUDSON, 2010).

As bases de dados, enquanto plataforma de gerenciamento de informação, para serem consideradas como tal têm de possuir, em simultâneo, um local onde os dados são armazenados e indicações de como os dados estão organizados e relacionados entre si (HARRINGTON, 1998) possibilitando assim a sua leitura. Os processos relacionados ao gerenciamento de bases de dados requerem uma linguagem específica no acesso/leitura dos dados (ULLMAN, 1988). O SQL é uma das linguagens de leitura.

Os dados acumulados e estruturados são depois analisados em busca de padrões capazes de justificar escolhas. A mineração de dados faz parte das técnicas relacionadas ao uso de bases de dados e é tida como ferramenta capaz de prever “tendências e comportamentos” com base no conhecimento que está implícito nas bases de dados (LIMA JUNIOR, 2006).

Na visão de Lev Manovich (2001) a narrativa e as bases de dados estão em polos opostos, mas ambas são representações do mundo. As bases de dados produzem sistemas de significação através de uma coleta ordenada e rigorosa de informação:

“As BDs deslocam linhas de força e passam a exercer papel fundamental, seja registrando, diariamente, bilhões de ações e hábitos de todos nós, produzindo perfis, formatos novos padrões de consumo – ou a personalização em massa -, gerando um novo “sujeito”, forjando produtos culturais, armazenando, estruturando e distribuindo conhecimento. Ou seja, um mundo traduzido em bases de dados.” (BARBOSA, 2008)

Em 2010, quando o Jornalismo de Dados (JBD)¹ dava os seus primeiros passos, Schudson (2010) já antevia que “uma base de dados não é jornalismo, mas, cada vez mais, o jornalismo sofisticado depende de bases de dados de qualidade, acessíveis e pesquisáveis”. Diversas alterações, como na rotina de produção, nos formatos narrativos e na interatividade, são visíveis com a mudança da forma como as informações são armazenadas (PALÁCIOS, 2014):

“Usando bases de dados localizadas em máquinas com crescente capacidade de processamento e armazenamento, com possibilidade de acesso assíncrono da parte do consumidor, o

¹ Como referência ao termo “Jornalismo de Dados”, que é uma simplificação de “Jornalismo de Bases de Dados”, optamos por utilizar a sigla “JBD” para designar o conceito ao longo desta dissertação.

jornalismo online encoraja o usuário (Interatividade) a juntar-se aos produtores e acrescentar informação a essas bases de dados (Participação e Atualização Contínua). Bases de dados jornalísticas e não jornalísticas podem ser conjugadas, interligadas e estabelecer comunicação entre elas (Hiperligação, Hipertextualidade, Multimídia). Para propósitos práticos, as redes digitais disponibilizam espaço virtualmente ilimitado para ao armazenamento de informação que pode ser produzida, recuperada, associada e colocada à disposição dos públicos alvos visados.”

De facto, as bases de dados podem estar presentes no jornalismo de diferentes modos. Como referiu Palácios (2014), essa presença foi explorada num primeiro momento como ferramenta de recuperação de informação, uma memória digital com a funcionalidade de contextualizar. As bases de dados como recurso jornalístico surgiram, em primeiro lugar, como facilitadores de tarefas de documentação da informação e sua distribuição e, como num processo evolutivo, tornaram-se elementos-chave da Reportagem Assistida por Computador (BARBOSA, 2008).

Antes mesmo de ser convencionalizado que o uso de determinadas ferramentas juntamente com um formato específico de informação seria Jornalismo de Base de Dados, surgiram movimentos que propiciaram condições para o seu crescimento. O Jornalismo de Dados é fruto de diversos processos baseados em dimensões tecnológicas, culturais e políticas (BRADSHAW, 2017).

1.2. *Computer-assisted Reporting*: o início

Estudos no campo do jornalismo defendem que o JBD é uma continuidade da Reportagem Assistida por Computador (RAC), assente na investigação e nas questões públicas (BOUNEGRU, 2012; HAMMOND, 2015). A RAC é vista como uma ferramenta capaz de permitir que a história seja fundamentada, mas não deve ser encarada como a própria história (CODDINGTON, 2015; FELLE, 2016), ela requer que os profissionais de jornalismo sejam dotados de capacidades na filtragem e significação dos dados (CODDINGTON, 2015).

Os primeiros registos do *Computer-assisted Reporting* (CAR) apontam para o ano de 1952 e a sua definição engloba um conjunto de técnicas que explora as buscas de dados online, a interpretação e análise de mapas, estatísticas e planilhas provenientes de dados

públicos, e inclui também a possibilidade de trocar e-mails – com o intuito de obter informações – como parte do processo jornalístico (GYNNILD, 2014). As ferramentas sinalizaram esta reportagem como uma alternativa ao jornalismo investigativo concentrado em análises manuais e, portanto, ela esteve sempre associada a cobertura de assuntos ligados a esfera pública.

Durante anos, os dilemas que ocupavam a área eram baseados na necessidade de compreender até que ponto o jornalismo é uma representação precisa da realidade. A fonte humana e a representação da experiência foram o alicerce da metodologia de trabalho no jornalismo, dito tradicional, em busca da objetividade. Apesar de estar intrinsecamente ligada aos princípios do jornalismo “*mainstream*”, a Reportagem Assistida por Computador promoveu uma inversão de prioridade na rotina de trabalho: a presença de descrições baseadas em experiências pessoais e o recurso a fontes oficiais passam a disputar espaço com análises de dados e métodos das ciências sociais (CODDINGTON, 2015). A técnica, baseada em pesquisa independente, confere “autonomia” ao trabalho jornalístico face à dependência de testemunhos institucionais (TRÄSEL, 2014).

Em 1951, na ocasião da cobertura das eleições presidenciais norte-americanas, a rede de televisão CBS destacou-se como a primeira organização a disponibilizar uma reportagem tendo como suporte as técnicas de RAC. A emissora trabalhou com previsões sustentadas por um modelo estatístico e acertou ao apontar que Dwight D. Eisenhower seria o vencedor das eleições. O trabalho foi realizado em conjunto com outros profissionais, como programadores, que criaram um algoritmo capaz de estabelecer padrões recorrendo a estatísticas de eleições anteriores, e assim fazer previsões rigorosas (HOWARD, 2014).

Nas primeiras décadas da implementação da Reportagem Assistida por Computador nos Estados Unidos, várias investigações jornalísticas foram indicadas e acabaram mesmo por conquistar o *Pulitzer Prize*. Coddington (2015) refere, citando Houston (1998), que a relevância do prêmio e a apreciação dada aos vencedores foram uma importante forma de “validação profissional” e um catalisador para a afirmação da RAC como parte do universo das redações.

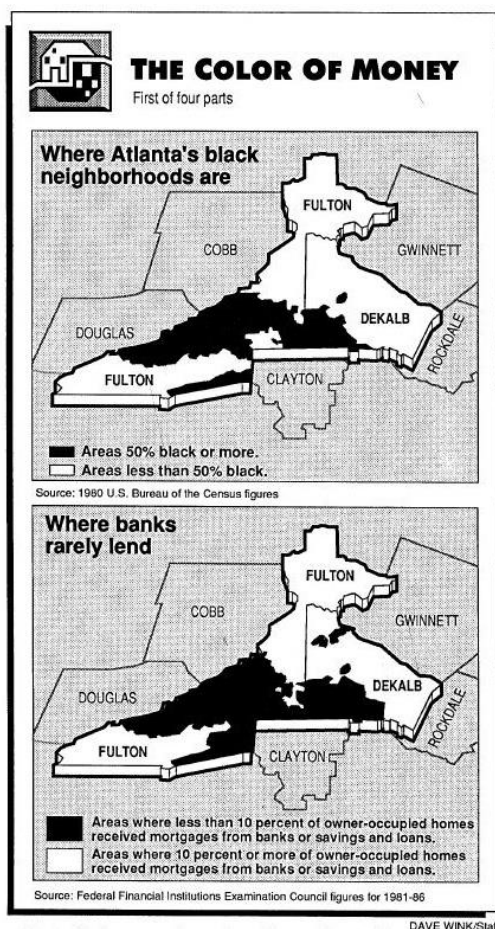


Figura 1 – Mapas utilizados na contextualização da reportagem

Em 1989, “*The Color of Money*”, uma série de reportagens de Bill Dedman, a cargo do *The Atlanta Journal-Constitution*, evidenciaram a discriminação racial dos bancos na aprovação de empréstimos bancários para a compra de casas, sendo este um dos trabalhos agraciado com o prêmio *Pulitzer*. A investigação foi possível através de uma análise computacional guiada pelo software *Framework*, que reunia e integrava diversas ferramentas, como a gestão de bases de dados e planilhas e softwares de produção gráfica. O ponto alto do trabalho foi a apresentação de dois mapas: um era sustentado pelos censos mais recentes e apresentava uma zona da cidade em que mais de 50% das casas eram ocupadas por famílias de raça negra; para comparar, Dedman utilizou outro mapa que indicava as zonas da cidade em que menos de 10% das casas eram fruto de empréstimos de bancos ou instituições financeiras. Com a comparação, foi possível visualizar que a zona da cidade com menos financiamento habitacional coincidia com a área ocupada, maioritariamente, pelas famílias de raça negra (MEYER, 1991).

Em 1993, um ano após o furacão Andrew atingir o estado da Flórida, Steve Doig, jornalista do Miami Herald, realizou um trabalho sobre como o governo teve influência

nos danos causados pela tempestade. A reportagem, sustentada pela análise de dados e mapas, foi a vencedora da categoria “Serviço público”. Na publicação do trabalho, Doig fez uma chamada de capa: “Esta fotografia aérea conta a história. O furacão Andrew não destruiu todos os bairros igualmente”². O jornalista comparou e correlacionou relatórios meteorológicos com relatórios de inspeções das construções e edifícios danificados para apontar que a destruição foi influenciada por erro humano.

A RAC emergiu muito antes do início da digitalização dos meios (GYNNILD, 2014) e foi afetada por constrangimentos de ordem tecnológica e profissional. Nos primeiros anos após a introdução da técnica, eram raras as organizações que possuíam e recorriam a computador na busca por materiais de caráter noticioso. Na década de 1980, a popularização dos computadores atingiu os repórteres e redações como ferramenta de edição de texto e a investigação com base na RAC ficou concentrada num pequeno grupo de profissionais (COX, 2000):

“Após as eleições de 1952, a introdução de computadores nas salas de redação prosseguiu no que Reavy (1996) classifica como três fases sobrepostas. Essas três fases foram a de negócios, a de produção e a de informação. Durante a fase de negócios, o uso de computadores foi principalmente para fins de contabilidade e circulação (Reavy, 1996). Em meados da década de 1960, os computadores estavam disponíveis na sala de redação para tarefas tão diversas quanto o controle de estoque e no contacto com informações adquiridas a agências. Na década de 1970, os jornais também começaram a usar computadores para a produção, o que levou à fase de informação.”

Além dos esforços de profissionais vistos como pioneiros no uso de softwares para a obtenção e o processamento da informação, a criação do *National Institute for Computer-Assisted Reporting* (NICAR) foi de suma importância para a afirmação da técnica como parte integrante do jornalismo, ao valorizar e promover formações na área. Com a expansão dos computadores e o natural desenvolvimento das redações em torno das tecnologias de informação, o termo *Computer-assisted reporting*, enquanto técnica diferenciada da produção de produtos jornalísticos entrou em desuso (CODDINGTON, 2015).

² Capa do *The Miami Herald* com o trabalho de Steve Doig sobre o furacão Andrew: <http://www.derivativeworks.com/2008/09/complete-scans.html>

Enquanto os Estados Unidos caminhavam para a consagração da tecnologia como instrumento natural da rotina jornalística, diversos países davam o primeiro passo rumo à aprendizagem das capacidades associadas às tecnologias de informação. A discrepância existente entre o surgimento e a afirmação da RAC nos EUA, comparado com a adoção da técnica por organizações de outros países, como o Reino Unido, era evidente (HEWETT, 2017). Hewett citou o NICAR ao apontar a falta de entidades que promovessem a partilha de “ideias e habilidades” como ponto fundamental do “contraste” entre a potência americana e os outros países na prática da RAC.

Em 2003, quando os primeiros cursos e workshops de investigação chegaram ao Reino Unido, o Centro de Jornalismo de Investigação (CIJ) apresentou a RAC como uma oportunidade para aprender uma técnica nova (HEWETT, 2017), muito embora do outro lado do atlântico o trabalho jornalístico com computadores já estivesse a caminhar para complexas inovações.

1.3. *Precision journalism*: da metodologia à epistemologia

O carácter quantitativo da informação, mesmo sob a justificativa de estar alinhado e correlacionado, não deve ser assumido, instantaneamente, como conhecimento, a etapa da interpretação deve ser parte da produção jornalística (LEWIS E WESTLUND, 2015). Os números, com a crença de serem irrefutáveis, veem a sua autoridade ser questionada no jornalismo, como refere Meyer (1991) “um número não faz história”. Embora os dados sejam o alicerce dessa nova – ou em expansão – área do jornalismo, o seu carácter quantitativo não pode ser critério único na presunção da verdade:

“Os dados não são um reflexo perfeito do mundo. Os conjuntos de dados públicos – sem surpresa – não nos dão informações perfeitas sobre o mundo. São representações do mundo reunidas, geradas, selecionadas, organizadas, filtradas, colhidas, analisadas e corrigidas para fins específicos - propósito tão diverso como contabilidade do setor público, controle de tráfego, previsão do tempo, planejamento urbano e avaliação de políticas. Os dados geralmente são incompletos, inoperantes, imprecisos ou desatualizados”. (ROGERS, 2013)

O jornalismo de precisão, tal como foi proposto por Philip Meyer, foi formulado como uma crítica ao “*New journalism*”, que era sustentado por técnicas literárias e ficcionais que comprometiam a busca pela objetividade (BOUNEGRU, 2012). Essa tentativa de

aproximar o jornalismo da arte não era compatível com os ideais do jornalismo. Na visão de Meyer (1991), o ideal seria a busca por uma “verdade verificável”, através do rigor das análises de dados das ciências sociais.

A aproximação entre o jornalismo e a ciência pode ser explicada pelo aumento do número de computadores – e conseqüente crescimento da informação disponível para análise – e também pela mudança no consumo de produtos noticiosos, que abandonou a produção massiva, em escala, para acompanhar as tendências relacionadas com o consumo mais direcionado (MEYER, 1991). Os mesmos fatores que foram causadores do modelo proposto por Meyer podem ser considerados agentes promotores de alterações na epistemologia do jornalismo:

“À medida que a epistemologia do jornalismo é permeada pelas culturas de dados, ela se move de um lado para o outro, entre relatórios e programação, narração e codificação, pesquisa orientada por hipóteses e pesquisa direcionada por dados, pensamento jornalístico e pensamento computacional, interatividade e estática, visualidade e textualidade, centralidade e localidade, individualidade e colaboração ou paradigmas humanos e paradigmas mecanizados, para citar alguns.”
(BORGES-REY, 2017)

A construção do jornalismo enquanto área de conhecimento tem ancorada uma etapa epistemológica que foi sedimentada ao longo de décadas, mas que se viu na obrigação de ser repensada após o surgimento de novos meios de trabalho e da ascensão da quantidade de dados disponíveis (LEWIS E WESTLUND, 2015). Parasie e Dagiral (2013) assumem o termo epistemologia, no contexto jornalístico, como um manifesto dos jornalistas em busca do reconhecimento das suas práticas como conhecimento dos factos. Por outro lado, Lewis e Westlund (2015) colocam o jornalismo em outro patamar, ao afirmar que a área é uma das mais influentes na produção de conhecimento.

O modelo investigativo do jornalismo de precisão é fundamentado por Meyer através da “aplicação de métodos de pesquisa das ciências sociais e comportamentais” e apropriação da “objetividade científica e ideais científicos” ao fazer jornalístico; nessa proposta o papel do jornalista enquanto mediador da realidade dos factos não é afetado, mas alargado pela adição de novas práticas.

Como conta Meyer (1991), a primeira aplicação do conceito ocorreu na cobertura das manifestações de Detroit, onde durante cinco dias a luta racial tornou-se num dos episódios mais violentos da cidade. Meyer, através da *Detroit Free Press*, procurou contestar duas teorias populares: a de que cidadãos com níveis de escolaridade mais baixos eram mais propensos a participar nos tumultos; e a de que a dificuldade de adaptação dos migrantes negros, provenientes do sul, era uma das causas dos atos. A investigação contou com a colaboração de pesquisadores da Universidade do Michigan que auxiliaram na realização de uma pesquisa em que a amostra era constituída por residentes das zonas onde a manifestação teve lugar e com uma programadora que ficou responsável pela organização e cruzamento dos resultados. Apesar de as hipóteses serem defendidas pela população, “a teoria não foi suportada pelos dados”: a comparação, baseada em modelos científicos, provou que estudantes do ensino superior têm a mesma predisposição para participarem nos atos que os indivíduos que não concluíram o ensino secundário. Em relação à incidência de manifestantes provenientes do sul ou do norte, ficou comprovado que a probabilidade do militante ser do Norte era três vezes maior do que de ter sido criado no sul.

Meyer (2011) refere que o diferencial do trabalho consistiu na definição prévia de hipóteses assentes em teorias criadas pela população, o que permitiu que a abordagem do jornal possuísse um ponto de foco – o teste das hipóteses – e não fosse apenas um conjunto de artigos com temáticas aleatórias, com possíveis ligações entre eles, sobre o mesmo evento.

A prática jornalística tradicional baseada em relatos de testemunhas, autoridades e especialistas, quando comparada ao modelo científico, apresenta uma falha metodológica por não conseguir contornar a parcialidade dos pontos de vistas (MEYER, 1991). Por outro lado, a utilização de técnicas de pesquisa das ciências sociais, e análises de dados provenientes de registos públicos, fornece à reportagem autoridade suficiente para não depender do testemunho dos atores envolvidos nos acontecimentos (TRÄSEL, 2014).

A proposta do jornalismo de precisão vai além da componente tecnológica, ela está formulada sob um modelo de pesquisa que promove novas práticas de investigação e apesar da RAC alterar a dinâmica do trabalho jornalístico ao introduzir computadores num ambiente focado em resoluções manuais, o modelo de Meyer foi alavancado por um conjunto de proposições sobre o exercício da atividade (ANDERSON, 2014).

A ausência de interesse acadêmico na investigação sobre o Jornalismo de Precisão pode ser justificada pelo baixo número de pares a desempenhar esse tipo de jornalismo (TRÄSEL, 2014). No entanto, o trabalho de Meyer promoveu diversas discussões sobre o fazer jornalístico.

O Jornalismo de Dados traz a afirmação de duas componentes: a tecnológica, com a afirmação do computador como ambiente natural de trabalho, e a alteração do agir jornalístico com a formulação de um novo método de seleção, apuração, análise e publicação.

1.4. Transparência e open access

Antes da popularização do *Open Government* e da afirmação do Jornalismo de Dados como tendência de prática do jornalismo contemporâneo, outras comunidades atuaram na busca pela liberação da informação. Iniciativas como a criação da *Fair Play Alliance*, uma plataforma de vigilância eslovaca desenvolvida em 2000 e focada nos movimentos do fluxo de dinheiro administrativo, tornaram-se comuns, juntamente com a colaboração de organizações cuja bandeira estava na defesa do acesso aberto e na crença de “que as habilidades de desenvolvimento deveriam ser usadas para reunir e comunicar informações socialmente relevantes” (KAYSER-BRIL, VALEEVA & RADCHENKO, 2015).

Apesar de em 1966, um ano antes da distinção do *Pultizer Prize* ser entregue a Philip Meyer pelo trabalho na cobertura das manifestações de Detroit, os Estados Unidos da América terem avançado com a promulgação da *Freedom Of Information* (FOI), a lei que permite que os cidadãos tenham acesso aos dados do governo, o ponto crucial do movimento de abertura foi, segundo Rogers (2013), em 2009, altura da posse de Barack Obama e do anúncio do novo portal do governo americano, o *data.gov*. A plataforma foi apresentada sob a promessa de disponibilizar informações referentes aos departamentos governamentais dos EUA. O portal, associado a lutas pela transparência política, promoveu a adoção de medidas de abertura ao redor do mundo, com diversos países tomando como prática a disponibilização de informação.

Como fonte e suporte do processamento da informação na construção de trabalhos jornalísticos ancorados na análise de dados, a criação de uma cultura política de transparência alimentou as construções conceptuais em torno do Jornalismo de Dados. A transformação da internet num espaço democrático, onde há liberdade de acesso à

informação, mas também de publicá-la é parte dos movimentos cívicos e do ativismo público no estabelecimento de direitos (SCHUDSON, 2010):

“O crescimento dos observatórios políticos, o avanço da vigilância do governo (e vigilância de outros centros de poder chave nos Estados Unidos e em todo o mundo) tanto fora ("sociedade civil") quanto dentro do próprio governo e a nova disponibilidade de bancos de dados para o interesse público, representam apenas uma característica do futuro das notícias. Uma característica vital que até o momento recebeu pouco destaque. Os observatórios políticos não substituem os jornalistas, nem os bancos de dados ignoram as narrativas. Mas os observatórios são parceiros cada vez mais valiosos para jornalistas, e as bases de dados estabelecem novas bases para a narrativa. Ambos oferecem promessa de desenvolver o tipo de informação pública que torna possível a democracia.”

Os dados administrados pelo governo reúnem informações sobre os cidadãos e sobre a atuação da máquina governamental. A adoção de práticas de transparência, com a consequente disponibilização dos *datasets*, representa a democratização de ferramentas que antes estavam em posse de pequenos grupos (LIMA JUNIOR, 2012):

“A fim de manter o governo responsável por suas ações, cidadãos devem conhecer sobre essas ações. Para isso acontecer, eles devem insistir que o governo atue de forma aberta e transparente possível. No século XXI, isso é colocar os dados disponíveis de forma online e com fácil acesso. Se o dado do governo é colocado disponível em formatos úteis e flexíveis, cidadãos podem ser capazes de utilizar modernas ferramentas em software livre para lançar luz sobre as atividades do governo.”

O movimento de abertura de dados não pode ser um ato iniciado e encerrado em si mesmo, como parte representante do estado de democracia, deve ser tratado como um processo focado na melhoria da qualidade e da relevância dos dados na vida dos cidadãos (STONEMAN, 2015). Apesar do evidente aumento dos dados disponíveis, o trabalho do jornalista encontra barreiras na obtenção e tratamento da informação. Mesmo em democracias com políticas de *Open Data*, os dados não possuem um padrão de publicação e necessitam da atuação do jornalista na correção de imperfeições com “dados sujos”,

metodologia de coleta inconsistente e disponibilização em configurações menos eficientes, como casos em que grandes conjuntos de dados são entregues em formato PDF (SUMAN-PEÑA, 2014).

Em concordância com a necessidade de padronizar a interpretação do conceito e a aplicação de medidas de dados abertos, ativistas e organizações formularam em 2007, oito princípios para o *Open Government Data*³:

1. *Completos*: Todos os dados públicos estão disponíveis. Os dados públicos são dados que não estão sujeitos a limitações de privacidade, segurança ou privilégios.
2. *Primários*: os dados são coletados na fonte, com o maior nível possível de granularidade, não em formas agregadas ou modificadas.
3. *Atuais*: os dados são disponibilizados o mais rapidamente possível para preservar o seu valor.
4. *Acessíveis*: os dados estão disponíveis para a mais ampla gama de usuários, para a mais ampla gama de propósitos.
5. *Processáveis por máquina*: os dados são razoavelmente estruturados para permitir o processamento automatizado.
6. *Não discriminatório*: os dados estão disponíveis para qualquer pessoa, sem exigência de registro.
7. *Não proprietário*: os dados estão disponíveis num formato sobre o qual nenhuma entidade possui controle exclusivo.
8. *Livre de licença*: os dados não estão sujeitos a nenhuma regulamentação de direitos autorais, patentes, marcas registradas. Podem ser permitidas restrições razoáveis de privacidade, segurança e privilégio.

Contra a ideia de “monopólio da informação”, organizações como a *Open Knowledge Foundation* se posicionam em sentido oposto a procedimentos governamentais de publicação de relatórios com interpretações dos dados, na visão dos ativistas, o conhecimento aberto está necessariamente dependente de dados abertos – em bruto – para

³ Representantes de diversas organizações propuseram um conjunto de princípios para facilitar a abertura dos dados e a sua usabilidade: <https://opengovdata.org/>

que haja liberdade na interpretação dos dados e na criação do conhecimento (BAACK, 2015).

Apesar de a RAC ter introduzido mudanças significativas na relação entre o jornalismo e o uso de dados, o jornalismo no cumprimento do papel de “quarto poder” já fazia uso de informações estatais na contextualização de assuntos públicos, recorrendo a dados estatísticos e gráficos (FELLE, 2016). No entanto, foi o movimento universal de abertura de dados, que se tornou prática comum dos governos, que foi visto pela comunidade jornalística como uma das saídas para a crise enfrentada pela indústria, que necessariamente forneceria mão-de-obra para a análise e explicação dos dados disponibilizados (STONEMAN, 2015). A liberdade de informação, vista como direito constitucional de acesso a dados públicos, é apontada por pesquisadores como um dos fatores responsáveis pela emergência do Jornalismo de Dados. A vigilância dos poderes públicos foi facilitada pela transição do armazenamento analógico para informações em formato eletrônico, que associadas à criação de direitos de acesso, favoreceu o desempenho dos jornalistas (FELLE, 2016).

2. Conceitualização

2.1. Propostas de definição do conceito de Jornalismo de Dados

O Jornalismo de Dados é um dos campos do Jornalismo de Investigação (FLORES E SALINAS, 2013) que surgiu como uma das variantes possíveis do jornalismo *online* – como meio de disseminação – ou do jornalismo com recurso a técnicas computacionais – como ferramenta que auxilia o processo de produção.

Na tentativa de defini-lo, torna-se necessário mencionar a Reportagem assistida por computador. A RAC, no entanto, não deve ser utilizada em alternativa ao termo JBD, uma vez que a primeira ganha sentido ao trabalhar com pequenos volumes de dados, ao passo que o segundo é um importante meio na análise de grandes conjuntos de informação (CRUCIANELLI, 2013).

No Jornalismo de Dados e na RAC, os pressupostos que guiam a ação do jornalista estão assentes na necessidade de acompanhar o desempenho de questões que afetam a vida pública; no entanto, enquanto a Reportagem Assistida por Computador é encarada como uma facilitadora da tarefa jornalística, o JBD permite um maior envolvimento do público e transforma o jornalista em mediador dos dados (CODDINGTON, 2015) ao possibilitar, através da apresentação e contextualização, que o leitor interprete e produza sentido individualmente.

O Jornalismo de Dados pode ser visto como o resultado da adaptação da profissão jornalística à cibercultura e à economia do conhecimento (TRÄSEL, 2014), uma vez que permite que habilidades informáticas aproximem um vasto conjunto de informação que, espalhada e desconectada, passa a poder ser trabalhada e alinhada em prol de questões públicas. De um conjunto de ações voltadas ao manuseio de dados que, por vezes brutos, necessitam ser lapidados para tomar forma jornalística, é possível gerar novos produtos e conhecimento:

“O JGD tem por objetivo 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 distribuição 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 trabalho de comparar diferentes bases de dados, para a produção de novo conhecimento sobre a sociedade, criando *mash-ups* ou relatando os resultados em texto, audiovisual e imagem.” (TRÄSEL, 2017)

Crucianelli opta por denominar essa vertente do jornalismo como «Jornalismo de Base de Dados», isso porque o “Jornalismo de Dados” é encarado pela pesquisadora como uma redundância – o que pode induzir ao erro – pois assim como outros autores referem (BOUNEGRU, 2012), os dados desde sempre fizeram parte do trabalho jornalístico.

O jornalismo de bases de dados é representado pela soma de diversas disciplinas do jornalismo. O que o JBD acrescenta ao conjunto dessas disciplinas é o manuseio de um grande volume de dados, a possibilidade de produzir visualizações de dados interativas e também a integração do trabalho do programador entre as práticas jornalísticas (CRUCIANELLI, 2013).

Tabela 1. Fórmula conceitual de Jornalismo de Dados (Crucianelli, 2013)

JORNALISMO DE BASE DE DADOS = JP + JI+JP+RAC+JA+ VOLUME DE DADOS + VISUALIZAÇÕES + PROGRAMADOR	
Jornalismo em profundidade	Pode verificar-se consoante o tema
Jornalismo investigativo	Utilizam-se técnicas que são próprias do jornalismo investigativo
Jornalismo de precisão	Em alguns casos os métodos de investigação propostos por Philip Meyer são utilizados
Reportagem Assistida por Computador	Quase sempre segue os procedimentos associados ao RAC
Jornalismo analítico	Há apropriação de métodos analíticos na análise dos dados

Tabela de autoria própria

A conciliação entre ferramentas de análise e possíveis aplicações de algoritmos podem ter como resultado um conjunto de padrões, capazes de assumir a condição de tendência

e a produção de inferências; se antes o relato humano era fundamental na compreensão dos fenómenos, agora esse poder também é transferido aos dados, que assumem o papel de fonte e devem ser analisados (HOWARD, 2014).

O crescimento da quantidade de dados produzidos e existentes no mundo está longe de possuir uma fórmula de previsão, o crescimento é inconstante em valores, mas tendencialmente está sempre a aumentar. A quantidade de dados existentes em 2015 representa uma parcela de 10% de todos os dados disponíveis em 2017, o que significa que 90% dos dados disponíveis atualmente foram criados nos últimos dois anos⁴.

Os dados podem ser vistos como recursos ubíquos – por estarem em todo o lado e pela facilidade de obtenção– que podem ser tratados de forma a produzir novas perceções (MCCANDLESS, 2010), visão que se alinha com a de Lorenz (2012), autor que defende que nas redes de conexões propiciadas pela internet há conjuntos de informações que devem ser vistos do ângulo certo para que seja possível a produção de sentido. Lorenz afirma ainda que com o Jornalismo de Dados há uma mudança no comportamento face ao jornalismo tradicional: a prioridade deixa de ser a de dar a notícia em primeira mão, para passar a ser a de transmitir informações com significado a partir dos dados analisados.

Os dados, que são a espinha dorsal do JBD, podem ser classificados em estruturados, semi-estruturados e não-estruturados, acompanhando a lógica do *Big Data*. Por estruturados entende-se os dados que estão padronizados em bases de dados, como por exemplo em linguagem SQL e que seguem lógicas de organização em blocos semânticos. Os semi-estruturados são conjuntos de dados que apresentam variações na organização. Por fim, os não-estruturados são dados independentes, sem padrões evidentes e que podem estar disponíveis em formatos variáveis como documentos Word ou arquivos HTML (BERTOCCHI, 2013).

Os principais exemplos de dados são públicos, como registros de impostos, orçamentos, dados do censo e dados privados, como informações disponíveis em sites de redes sociais e também informações de interesse público que são disponibilizadas através de fugas de informação – são esses conjuntos de dados que são utilizados nas investigações e nas explicações dos factos sob a orientação de jornalistas (STAVELIN, 2013).

⁴ Em 2017, 2.5 milhões de *terabytes* de dados foram gerados *online*: <https://medium.com/@Bahareh/tedx-talk-how-data-journalism-is-changing-the-newsroom-18353e63da88>

Na rotina de produção de trabalhos jornalísticos guiados por dados, os dados podem ocupar diferentes funções: serem a fonte que sustenta o trabalho ou a ferramenta capaz de dar significação aos factos – ou ambos (BRADSHAW, 2012). Rogers, Schwabish e Bowers (2017) propõem uma categorização mais específica do uso dos dados no jornalismo, tendo em conta os papéis que estes podem desempenhar. A proposta dos autores é subdividida em três tópicos: histórias que são enriquecidas por dados (os dados são adicionadas para dar suporte a narrativa); histórias que utilizam dados para investigar (em muitos conjuntos de dados, a relevância jornalística não é evidente, nesses casos os processos podem ser mais demorados e exigir uma combinação mais complexa de ferramentas, para que dos dados sejam retirados elementos de interesse para uma história); e histórias que explicam dados (com o evidente aumento da informação disponível, cabe ao jornalista o papel de decodificar a informação escondida nos dados). Já Crucianelli (2013) defende que do JBD podem surgir quatro produtos jornalísticos que se complementam, esses produtos são classificados consoante a aplicação dos dados e o consumo por parte do público. Dos grandes volumes de dados podem ser retirados padrões específicos que servem como contextualização de peças pouco desenvolvidas, que também podem resultar de cruzamentos de informação entre diferentes bases de dados, dando origem aos **artigos baseados em dados**. Muitas vezes, o trabalho desenvolvido no Jornalismo de Dados é remetido ao uso de informação visual que funciona como um elemento adicional na interpretação da notícia, no entanto, em alguns casos, o elemento infográfico, individualmente, ocupa o papel central do trabalho e dispensa a contextualização através de linguagem de texto, esses produtos são denominados de **visualização interativa**. Tradicionalmente, o uso das técnicas de jornalismo guiado por dados tem como resultado final peças em formato jornalístico. No entanto, Crucianelli defende que o **acesso a conjuntos de dados** é também um produto jornalístico. O ato de disponibilizar bases de dados que, por motivos técnicos ou opção governamental, não se encontram acessíveis ao público em geral é também um importante contributo. Esses produtos são habituais em países sem políticas de dados abertos ou com tendências para disponibilizar informação em formatos que dificultam a legibilidade. O desenvolvimento do JBD permite gerenciar a abundância de informação característica da era do *Big Data*, apesar disso, em alguns casos, o grande volume de informação torna-se um obstáculo na seleção dos pontos de interesse, o chamado «cabide» da história. Uma das soluções encontradas passa pelas chamadas «*news apps*» ou **aplicações de notícias**, em que os

dados são apresentados em função do leitor. Neste caso, é necessária a convergência entre o jornalismo e a programação.

Estes quatro produtos jornalísticos defendidos por Crucianelli (2013) só são possíveis graças à digitalização do processo jornalístico. Ao contrário do que acontecia com a RAC, onde o diferencial jornalístico estava nas ferramentas utilizadas, com o Jornalismo de Dados o *storytelling* tem também um papel central (FELLE, 2016) e pode ser apenas a distância que separa uma interpretação incorreta de uma real compreensão dos dados.

As *news apps* facilitam a compreensão ao transformar grandes quantidades de dados em informação (HOWARD, 2014). Essa possibilidade e sobretudo a capacidade de personalização da informação é hoje uma das referências do Jornalismo de Dados, na medida em que «eu» do leitor é introduzido em aplicações suportadas por bases de dados e tem como resultado a contextualização da história na primeira pessoa:

“Hoje, muitos jornalistas dedicam atenção não só a encontrar dados para investigações, mas a publicá-lo ao lado de histórias vivas, ou aplicativos de notícias. As aplicações de notícias são uma das mais importantes novas formas de contar histórias deste jovem milênio, nativas do media digital e, muitas vezes, acessíveis em todos os navegadores, dispositivos e sistemas operacionais na Web aberta.”
(HOWARD, 2014)

A importância do desenvolvimento do JBD não está reduzida ao público, ela surge como esperança no mercado mediático, corrompido pela instabilidade de modelos de negócio. Além da possibilidade da automatização de tarefas, que podem gerar produtos para serem utilizadas por outros profissionais que não dominam as técnicas, o JBD tem a capacidade de atrair novos públicos (HOWARD, 2014).

A estatística e a regra da proporcionalidade são chaves importantes na aplicação do Jornalismo de Dados e uma falha pode induzir ao erro. Sempre que os dados estatísticos forem requisito de uma peça, é indispensável o enquadramento da amostra ao universo representado, de forma a que o cenário divulgado aponte para a perspectiva real (MCCANDLESS, 2010). A presença da matemática, não significa, porém, que o resultado do trabalho no Jornalismo de Dados seja infalível, apesar do estreitamento entre as áreas, o jornalismo continua a pertencer aos limites da experiência humana (STRAY, 2014).

Uma vez que as falhas de comunicação eram frequentes nos primórdios do jornalismo, a pirâmide invertida foi inserida como uma prática de segurança. Era necessário enviar para a redação primeiramente os *takes* de maior importância, sob o risco de se perder informação ao longo da comunicação. Após a estabilização das linhas de correspondência, a pirâmide invertida continuou a ter protagonismo nas redações, agora exclusivamente como estratégia de “venda” da notícia. A quantidade de leitores que leem os primeiros parágrafos é maior do que o número de pessoas que chegam aos últimos parágrafos, portanto os pontos de interesse do evento devem estar nas primeiras linhas da peça.

De forma a sintetizar os processos que permitem que de uma grande quantidade de informação seja possível encontrar indicadores para comunicar dados de relevância pública, a regra de afunilamento da informação foi adaptada ao Jornalismo de Dados por Bradshaw (2011), na sua proposta “A pirâmide invertida do Jornalismo de Dados”:

Compilação: É o processo inicial e implica a reunião de informação para posterior análise. Essa compilação de dados pode ser feita de diferentes formas: Recorrer a dados fornecidos diretamente por organizações governamentais e privadas (1); Através de técnicas de pesquisa avançada para navegar e aceder a dados que não estão visíveis nos sites do governo (2); Extração de bancos de dados escondidos on-line usando ferramentas como o OutWit Hub, o Scaperwiki e o Web Scraper (3); Se os ficheiros forem obsoletos, ainda assim podem ser utilizados através da conversão dos documentos com recurso a ferramentas próprias (4); Tirando informações de *APIs*, que são padrões de informação que permitem o funcionamento de aplicações e *websites* (5); Criação própria de dados por meio de observação, pesquisas, formulários online ou *crowdsourcing* (6).

Limpeza: A estruturação dos dados nas bases por vezes apresenta inconsistências que podem ser propagadas nos trabalhos, gerando erros de interpretação. Como tal, é necessária uma análise do estado dos dados e a correção das imperfeições encontradas. Segundo Bradshaw, esse estágio do processo deve ser feito de duas formas: através da remoção do erro humano e com a adequação dos dados usados no trabalho a um formato comum. O erro pode estar em espaçamentos, letras maiúsculas ou minúsculas, diferentes nomes para o mesmo objeto e a limpeza pode ser feita a partir do *Excel*, do *Google Docs* e do *Google Refine*.

Contexto: Consoante a fonte de origem dos dados, a credibilidade é inerente, mas deve ser sempre assumida como objeto falível. Como tal, é importante compreender em quais condições o conjunto de dados selecionado como fonte do trabalho jornalístico foi formulado. É necessário responder a algumas perguntas sobre os dados como: Quem coletou? Quando? Por que razões? Assim como ter conhecimento da metodologia adotada na coleta dos dados.

Combinar: O cruzamento de um conjunto de dados com outros dados permite uma maior fiabilidade da informação. As inferências produzidas a partir de cruzamentos de dados que com outras indicações têm mais autonomia e autoridade na cobertura de assuntos e na criação de sentido.

Comunicação: É nesta etapa que os dados e os resultados provenientes das análises são preparados para serem publicados. As visualizações, muitas vezes representadas por mapas, gráficos e infografias, são as escolhas mais frequentes.

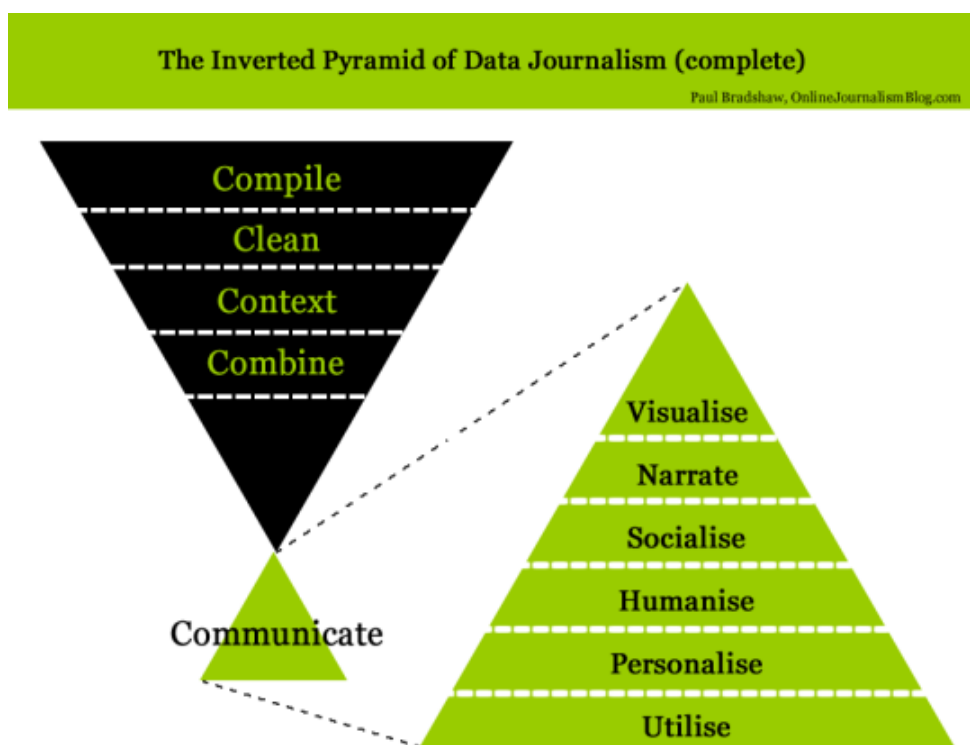


Figura 2 - Pirâmides propostas por Paul Bradshaw (2011) na construção de peças através de Jornalismo de Dados

Como complemento à pirâmide invertida, Bradshaw propõe a pirâmide da comunicação e justifica a criação como uma forma de mostrar que, apesar da evidente relação entre o JD e as visualizações, existem outros caminhos para comunicar os dados. Visualizar é a

opção mais prática na comunicação dos resultados, a facilidade de leitura e interpretação de aspetos gráficos é uma das justificativas para a popularidade dos conteúdos visuais. Narrar os resultados pode ser uma das alternativas na comunicação dos dados, para tal, faz-se necessário delinear estratégias na organização da informação. Socializar os dados no jornalismo pode significar a transferência de autonomia para o público, ao fornecer os dados e permitir que parta deles a indicação dos pontos de interesse. Humanizar a comunicação nada mais é do que encontrar pessoas que representem os números, ir atrás do testemunho de alguém que faz parte daquele universo quantificado. Personalizar os dados é uma das vantagens da interatividade, os números apresentados podem ser moldados à realidade do leitor. A utilização engloba a criação de ferramentas e aplicações baseadas nos dados.

Visto como uma evolução de conceitos como a Reportagem Assistida por Computador e o Jornalismo de Precisão, o uso de dados com a finalidade informativa é anterior à invenção da imprensa, quando em 1600 documentos eram produzidos e vendidos semanalmente para quantificar o número de mortos e de nascimentos em Londres (ROGERS, SCHWABISH E BOWERS, 2017).

No entanto, o primeiro registo em que o uso da análise de dados está relacionado com as práticas jornalísticas remete para uma publicação do *The Guardian* – ainda sob o nome de *Manchester Guardian*. Na ocasião, em maio de 1821, um conjunto de dados provenientes de uma fonte não identificada, mas “confiável”, exibiu os números das despesas gastas na educação em escolas de Manchester e Salford. Os dados divulgados e analisados indicaram que cerca de 25 mil alunos beneficiavam de ensino gratuito, o que derrubou as estimativas oficiais que indicavam um benefício por parte de 8 mil estudantes (ROGERS, 2013). O exemplo demonstra que antes mesmo do termo ser convencionado, o recurso a bases de dados, ainda em papel, já era uma abordagem adotada pelos jornais como parte das suas práticas – ainda que esporadicamente e sem o auxílio das ferramentas que hoje aceleram e dão rigor ao trabalho.

O surgimento de novos termos na área do jornalismo anuncia a dinâmica de inovação da área e a sua conseqüente aceitação com a adoção das novas práticas por parte dos profissionais (GYNNILD, 2014). Desde a introdução do computador ao processamento jornalístico, que essa variante tecnológica forçou a designação de nomes que pudessem ser indicadores das práticas utilizadas. A nomenclatura utilizada para referir as tarefas jornalísticas guiadas por dados nunca foi um consenso, o que torna importante verificar

se as variações existentes representam alterações conceptuais ou apenas um desencontro de convenções.

Apesar de serem muitas vezes utilizados como sinónimos, Martinho (2014) faz uma distinção entre os termos, tendo em conta as diferentes abordagens e os produtos criados através da prática jornalística:

- **Jornalismo de Dados** (*Data journalism*): é um termo abrangente que não reivindica uma prática exclusiva, está ligado ao jornalismo de investigação e a outras disciplinas como o RAC e o Jornalismo de Precisão, sendo inclusive confundido com essas áreas.

- **Jornalismo de base de dados** (*Database journalism*): tem como foco a usabilidade das bases de dados na criação de produtos informativos, normalmente tem como resultado peças que fogem do formato jornalístico.

- **Jornalismo guiado por dados** (*Data-driven journalism*): os dados são os mediadores do processo jornalístico que, neste contexto, dirige o foco para as histórias jornalísticas.

A emancipação do Jornalismo de Dados enquanto variante do jornalismo permitiu verificar que a partir de aparatos já existentes foi possível consolidar práticas novas:

“Definir o JD pela dimensão tecnológica somente, por exemplo, implicaria em aceitar um argumento de natureza determinística: uma vez adotado o aparato técnico, estaríamos diante de uma necessária mudança no fazer jornalístico, no caso, uma mudança inexorável rumo ao JD. Nesse sentido, como a internet e os softwares tornaram-se aparatos acessíveis às redações, logo todo jornalismo com dados seria, necessariamente, Jornalismo de Dados, o que não nos parece correto. Por outro lado, se somente a atitude investigativa do jornalismo (*watchdog*) dimensiona a relação que o jornalismo estabelece com os dados a que ele tem acesso, então o JD seria nada mais do que uma variação semântica do jornalismo investigativo, incorporado às redações desde muito antes do advento da internet?” (MANCINI E VASCONCELLOS, 2016)

Segundo os autores, entre as mudanças essenciais que permitem a sua conceptualização, está o desenvolvimento de competências e da constituição metodológica do trabalho adotado por parte dos profissionais. Apesar de a tecnologia facilitar o processo como um

todo, a definição de JBD encontra-se na própria ação, no trabalho com os dados e na forma como eles são abordados. Para Bounegru (2012), o foco do debate em torno da área não deve ser sobre o Jornalismo de Dados como “uma novidade” ou como um critério de comparação entre outras áreas, mas deve ser visto como oportunidade para trabalhar com uma grande quantidade de dados que anteriormente não estava disponível, e de contar com ferramentas que antes não existiam.

De modo a verificar o uso dos dados na criação de produtos jornalísticos, Mancini e Vasconcellos (2016) propõem a formulação de uma matriz definidora. Com as proposições defendidas pela matriz, é possível verificar que o uso de dados no jornalismo pode ter finalidades distintas representando um Jornalismo COM dados (onde os dados quantitativos ajudam a estruturação da reportagem) e o próprio Jornalismo de Dados (em que os dados sustentam a pauta e são o fio condutor da peça). Se juntam a essas duas componentes, três dimensões que avaliam as escolhas jornalísticas na construção das publicações: a dimensão de investigação representa o trabalho do jornalista na extração e organização dos dados que serão utilizados na reportagem; a dimensão interpretativa procura constatar se há análise por parte do jornalista ou de outras fontes dos dados selecionados; e a dimensão comunicativa, que se refere ao incremento de componentes visuais, como gráficos e infográficos, que auxiliem na compreensão da reportagem.

Categorias X Níveis de JD		Busca e/ou Elaboração própria dos dados (criação da base)	Estrutura da base (séries temporais, categorias <i>rankings</i> , tabelas)	Visualização dos dados (infografia)	Interpretação dos dados (texto)
Jornalismo Com Dados → De Dados	Nível 1	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	Nível 2	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
	Nível 3			<input type="checkbox"/>	<input type="checkbox"/>
	Nível 4			<input type="checkbox"/>	
	Nível 5				

Figura 3 - Componentes de classificação dos níveis de Jornalismo de Dados, in Mancini e Vasconcellos (2016)

De um modo geral, é possível constatar que apesar da ausência de consenso na formulação do conceito de Jornalismo de Dados, é unânime o incremento de novas competências, fruto do aumento do material disponível para análise e do surgimento de ferramentas que auxiliam o processo jornalístico desde a coleta dos dados, até a sua comunicação, com o

próprio jornalista a desempenhar papéis na produção de visualizações. Áreas como o jornalismo computacional, o jornalismo automatizado(*robot*) e também o jornalismo visual reivindicam a sua própria autoridade na produção de sentido, mas podem ser um aliado do jornalismo, como defende Gynnild (2014), na medida em que promovem um reconhecimento dos termos entre os profissionais e podem ajudar no crescimento da área, uma vez que permitem a valorização e potenciam a colaboração interdisciplinar.

2.2. Visualização de dados: uma componente indissociável

O trabalho de Florence Nightingale é citado (ROGERS, 2013) como emblemático na construção histórica do conceito de Jornalismo de Dados. Enfermeira de profissão e responsável pelos cuidados de saúde do exército britânico durante a Guerra da Crimeia, Nightingale fez uso dos seus conhecimentos estatísticos e dos dados reunidos durante os anos ao lado do exército, para denunciar a falta de condições no tratamento aos feridos e propor um conjunto de alterações no sistema de saúde.

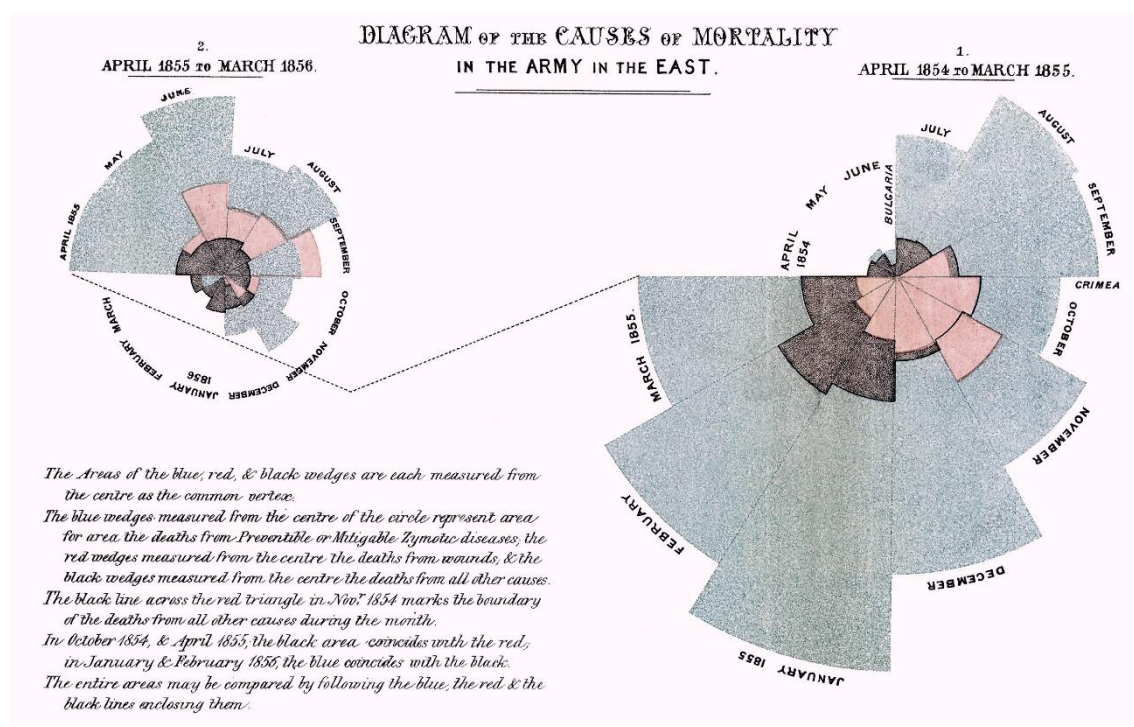


Figura 4 - Diagrama proposto por Florence Nightingale para demonstrar as principais causas de morte dos soldados britânicos.

Através da combinação de um conjunto de gráficos, a enfermeira relacionou a mortalidade dos soldados – apresentada acima em espirais – com um conjunto de fatores de ordem sanitária. As áreas preenchidas a azul representam mortes por doenças infecciosas, a cor de rosa mortes por feridas e a preto por outras causas. No total, durante

a guerra que durou três anos, 2.600 militares morreram em combate e 16.000 morreram por doenças ou falta de condições no tratamento. Para Florence Nightingale, a adoção de alterações nas medidas de higiene, como a limpeza dos solos, seria suficiente para diminuir o número de mortes dos soldados.

Em 1858, altura da publicação dos gráficos, a organização visual da informação sustentada por dados estruturados tornou-se realidade na defesa do interesse público. No Jornalismo de Dados no estágio contemporâneo, a adoção de componentes visuais como complemento ao trabalho é considerada como determinante.

A visualização de dados pode ser encarada com uma das disciplinas do Jornalismo de Dados, ligada ao design de informação, que procura a descodificação dos valores e informações quantitativas, numa mistura semiótica que tenta facilitar a interpretação. A componente visual que acompanha muitas histórias construídas a partir da análise e interpretação de grandes quantidades de dados é confundida e em alguns casos utilizada em alternativa ao conceito de Jornalismo de Dados. As visualizações de dados podem ser inseridas em forma de mapas, gráficos, infografias e construções visuais dotadas de interatividade.

Transformar a informação quantificável – no seu estado bruto – em visualizações, retira aos dados o seu carácter de abstração (BARLOW, 2015), uma vez que estes passam a ser o resultado da combinação de um série de escolhas e adquirem um significação que antes estava latente.

Trabalhar com visualizações tornou-se simples com a quantidade de softwares disponíveis gratuitamente em rede. No entanto, em alguns casos, incluir esse elemento como parte da peça pode exigir a presença colaborativa de outros profissionais. A introdução da disciplina visual, pode indicar a inclusão de novos métodos de trabalho, como a adoção do *storyboard*, que para além de ser uma ferramenta que ajuda a estruturar a narração, permite que não haja perda do propósito jornalístico durante a troca colaborativa (WEBER & RALL, 2012).

Se em exemplos anteriores de trabalhos com dados, onde o papel era o principal meio de publicação massiva a presença de componentes visuais apresentava-se como um aliado na desconstrução de temas, com a internet e a sofisticação de ferramentas, os gráficos ganham outra força. No Jornalismo de Dados, fruto da emergência da Web 2.0, os gráficos surgem aliados à interatividade. Definidos como uma forma híbrida, os gráficos

interativos são derivados da combinação entre elementos verbais – texto e áudio – e visuais – fotografias, ilustrações, mapas e símbolos (WEBER e RALL, 2012).

Barlow (2015) relembra que é comum a aproximação, metafórica, do conceito de apresentação gráfica ao de língua universal, algo como uma “língua franca” capaz de criar pontes facilitadoras da comunicação. Já para Cairo (2013), as visualizações podem ser interpretadas como uma tecnologia, que auxilia na criação de novos objetos, essenciais ao desempenho de tarefas. Durante a construção dos trabalhos guiados por dados, a visualização da informação pode ser tida como uma etapa de análise ou como uma das componentes da comunicação da informação. A visualização está inserida nos estágios finais do fluxo de trabalho no Jornalismo de Dados (BRADSHAW, 2011). No entanto, em alguns casos, a complexidade dos dados pode ser diminuída e padrões ocultos podem se tornar evidentes se a visualização dos dados for tida como um dos recursos de análise (MCCANDLESS, 2010), ou seja, além da aplicabilidade das visualizações ser fundamental na comunicação de algumas peças, pode ser adotada também, antes da publicação, como auxílio ao jornalista na análise da informação reunida.

2.3. Desafios profissionais na adaptação ao Jornalismo de Dados

A sofisticação das ferramentas disponíveis online para a raspagem, limpeza, análise e visualização de dados – funcionalidades básicas para o desempenho do Jornalismo de Dados – juntamente com o aumento progressivo da informação disponível, muito relacionada com a transparência governamental, permitiu a popularização da técnica, inclusive em redações pequenas e com recursos limitados (HOWARD, 2014). À medida que há um acréscimo tecnológico nas rotinas jornalísticas, a própria função do jornalista é repensada. Associada a novas formas de trabalhar surgem assim, também, novas responsabilidades.

Hacker-journalism

Julian Assange, hacker e ativista, fundou em 2006 a *Wikileaks*, uma plataforma online sustentada pela disseminação da filosofia hacker de que “a informação quer ser livre”.⁵ A organização, apresentada como sem fins lucrativos, foi responsável por tornar públicos centenas de milhares de documentos referentes a atuação dos EUA na política externa, em casos como a guerra do Afeganistão e do Iraque. Os vazamentos dos dados foram

⁵ Julian Assange é autor do livro “Cypherpunks: Liberdade e o futuro da Internet” onde defende a circulação da informação: https://resistir.info/varios/assange_livro_port.pdf

distribuídos em três atos, a partir de tecnologia hacker e através de um sistema de criptografia encarregue de preservar a identidade de quem submete informação, o que tornou possível que jornalistas e cidadãos tivessem acesso a informações restritas.

O projeto liderado por Assange foi fundamental para compreender o posicionamento do Jornalismo de Dados e levantou questões sobre a preparação da comunidade para trabalhar com dados que reuniam características fora do habitual: eram dados sensíveis, reunidos em grandes quantidades e de fonte sigilosa (MARTINHO, 2014). O *Wikileaks* contribuiu não só para a transformação do Jornalismo de Dados como tendência, mas também permitiu que o jornalismo como um todo repensasse as formas de obtenção de informação e as formas de contar as histórias (BAACK, 2011). Para Baack, a quantidade de dados vazados evidenciou, para a comunidade jornalística, a utilidade das técnicas de Jornalismo de Dados na análise da informação e na criação de visualizações para a apresentação do trabalho.

Segundo o autor, a manifestação conjunta das organizações em torno da cobertura do caso foi fundamental para evidenciar as vantagens do Jornalismo de Dados a nível internacional. E a exemplo disso, citou a eficiência de jornais mais desenvolvidos no trabalho com dados, como o *The Guardian* e o *New York Times*, na análise dos documentos.

Os movimentos que atuaram na luta pelo acesso e circulação da informação foram responsáveis pela criação de uma rede de colaboração multidisciplinar, responsável por imprimir nos jornalistas, novas habilidades. A cultura de defesa por informação aberta, por exemplo, que representa a aproximação entre o *hack* (jornalista) e o *hacker* – enquanto programador (CODDINGTON, 2015), ficou bem patente no caso *Wikileaks*.

A possibilidade de criação de parcerias com a construção de redes de colaboração entre jornalistas e não-jornalistas representa uma valiosa fonte de desenvolvimento dos media (SUSMAN-PEÑA, 2014). Competências de programação podem significar um avanço face as habilidades tradicionais do jornalista, na medida em que podem permitir o contato com informações privadas, mas com potencial interesse público e social (LIMA JUNIOR, 2011). Os *hackathons*, eventos que reúnem interessados na produção ou liberação de conteúdos, têm se tornando ponto de encontro dos jornalistas com outras comunidades, estimulando a troca de conhecimento e o surgimento de novos projetos que por vezes não têm espaço dentro das organizações de *media* (SUSMAN-PEÑA, 2014).

Outras organizações como a *Hacks/Hackers*, uma rede movida pela tentativa de repensar novos rumos para a indústria da informação, promovem o estreitamento da relação entre as duas comunidades que dão origem ao nome. Eles acreditam que reunir *Hackers*, que utilizam a tecnologia para obter informação através da criação de códigos, com *Hacks*, que fazem uso da mesma tecnologia para dar significado a palavras, pode servir como propósito para reinventar o jornalismo.

Acontecimentos como o *Wikileaks* e os *Panama Papers* trouxeram para o jornalismo a necessidade de debater o “*hacking journalism*” e a legitimidade do uso de ferramentas *hacker* na obtenção de informação com potencial para ser transformada em fonte para histórias jornalísticas. Os dados acumulados por governos e instituições ficam armazenados em bases de dados e representam um elevado volume de informação. O jornalista é inserido nesse contexto como autoridade encarregue de sintetizar informações relevantes através de competências da sua prática profissional e de mediar a circulação dos resultados das análises.

Curadoria de dados

Em 2009, o *Wikileaks* enviou cópias dos documentos mais influentes aos jornais de todo o mundo, com um objetivo claro: acelerar a análise dos documentos. O ato fez com que cada jornal reivindicasse enquadramentos e abordagens próprias da análise dos dados.

Com a emergência do Jornalismo de Dados, o processo de curadoria intensificou-se. O aumento de dados disponíveis influencia a sistematização dos factos que representam uma realidade e pode representar uma nova necessidade, a de que o jornalista desempenhe tarefas, demoradas, de busca de informação (ROGERS, 2013). Como refere Bradshaw (2011), o Jornalismo de Dados pode começar de duas formas: ou o jornalista tem uma pergunta e necessita de dados para procurar uma resposta ou o profissional tem um conjunto de dados que requer questionamentos.

No Jornalismo de Dados é prática comum a procura constante por bases de dados, e em bases de dados a busca por assuntos que possam ser encarados como de interesse público. A ausência da literacia para a análise dos dados brutos promoveu o jornalista ao papel de decodificador dessa realidade. A curadoria é uma mescla entre processos automatizados e atuação humana na resolução de tarefas (BAKKER, 2012). No jornalismo ela é representada pela seleção, recolha e reutilização de informação disponível em formato digital para a construção de narrativas (MARTIN, 2015).

As notícias quando baseadas em dados têm, em muitos casos, a necessidade de um filtro num universo marcado pelo *Big Data*. O poder de decisão dos jornalistas ao definir a construção da notícia e as combinações necessárias para a construção da informação demonstram que todo o ato jornalístico é um ato de curadoria (BRADSHAW, 2013).

3. O estado do Jornalismo de Dados

Os avanços nas técnicas baseadas em análise de dados evoluíram consoante o surgimento de ferramentas digitais. Entre os primeiros trabalhos publicados no séc. XIX e o Jornalismo de Dados conhecido no presente século, a principal diferença está na forma como os dados eram armazenados e disponibilizados; a informação compactada em folhas de papel é hoje encontrada em ficheiros computacionais (ROGERS, 2013). O crescimento do Jornalismo de Dados é visto como uma resposta do jornalismo à “datatificação” da sociedade (LOOSEN, REIMER E SILVA-SCHMIDT, 2017) e à necessidade de que alguma instituição assuma a responsabilidade de transformar os dados em material inteligível. O Jornalismo de Dados representa um avanço perante o crescimento das *fake news*, na medida em que permite que a transformação dos eventos em notícia seja ancorada em processos verificáveis e assentes em factos (HERAVI, 2017).

A emancipação do Jornalismo de Dados é apontada por alguns pesquisadores como resultado do aumento do volume de dados e do surgimento de ferramentas gratuitas que permitem o manuseio da informação. De facto, a predominância da cobertura política (LOOSEN, REIMER E SILVA-SCHMIDT, 2017) por parte dos jornalistas de dados é reveladora da própria prática governamental de disponibilizar dados públicos, possibilitando um processo retroativo. No entanto, a disseminação dos conceitos e das práticas do Jornalismo de Dados pode ser vista como parte do pioneirismo dos jornalistas na produção e sugestão de trabalhos que envolvam um alto volume de dados (TRÄSEL, 2014).

3.1. Jornalismo de Dados no Brasil

Uma das primeiras reportagens jornalísticas sustentadas pelo trabalho em bases de dados no Brasil, remete para o ano de 1991. Mário Rosa, jornalista do Jornal do Brasil, teve acesso ao Sistema Integrado de Administração Financeira do Governo Federal (SIAFI) – cuja a permissão de entrada era de exclusividade dos parlamentares. Na base de dados citada, Mário Rosa teve acesso aos gastos do governo e conseguiu identificar manobras de superfaturamento na compra de leite em pó pela Legião Brasileira de Assistência, presidida pela primeira-dama em atividade, Rosane Collor.

O trabalho de Mário Rosa abriu precedência, juntamente com outras reportagens, para que o SIAFI se tornasse a primeira base de dados governamental a poder ser consultada, mesmo que parcialmente, por jornalistas⁶.

No Brasil, a Associação Brasileira de Jornalismo Investigativo (Abraji) foi responsável pela propagação inicial da técnica de Reportagem Assistida por Computador, tendo fornecido treinamento para mais de quatro mil jornalistas (TRÄSEL, 2014). Outras instituições também se tornaram importantes meios de divulgação e formação da área, como a Escola de Dados, um projeto global em defesa da literacia de dados, que atua no Brasil desde 2013 e estima que mais de seis mil pessoas já receberam formação do projeto que “ajuda organizações sociais, jornalistas, agentes públicos e cidadãos a usar dados de forma eficiente para resolver problemas reais em prol de sociedades mais conscientes, sustentáveis e justas”⁷. A Conferência Brasileira de Jornalismo de Dados e Métodos Digitais denominada Coda.Br é uma das iniciativas da Escola de Dados e conta com o apoio da Abraji. A duas edições do encontro aconteceram em São Paulo, em 2017, e reuniram profissionais e pesquisadores brasileiros e internacionais no debate sobre dados.

Em maio de 2012, o Estado de São Paulo anunciou a criação do primeiro núcleo brasileiro de Jornalismo de Dados – o Estadão Dados – que em formato de blog dedicou-se a publicar peças centradas na análise de dados, oriundas do trabalho conjunto entre jornalistas, designers e programadores:

Um núcleo de profissionais dedicados a capturar e tratar informações usando técnicas estatísticas, algoritmos e formas visuais de apresentação. Além de organizar os dados, o produto final permite, na internet e nos tablets, que os usuários façam suas próprias análises. Para o jornal impresso, gera infográficos de última geração. (ESTADÃO, 2012)

O ponto chave da estabilização do núcleo, foi a criação do Basômetro, uma ferramenta criada para medir o nível de governabilidade dos deputados a partir da análise dos votos dados em assembleia. A ferramenta foi descrita por um dos seus criadores como uma “revolução na narrativa jornalística”, que permite que os leitores acessem ao conteúdo e tirem as suas próprias conclusões.

⁶ <http://manual.periodismodedatos.org/marcelo-trasel.php>

⁷ <https://escoladedados.org>

Na mesma semana da criação do Estadão Dados, a Lei de acesso à Informação foi implementada, após ser aprovada em 2011 pelo Senado Federal.

Os prêmios recebidos no jornalismo são um carimbo do reconhecimento, não das técnicas adotadas, mas da utilidade do resultado do trabalho. Em 2015, a reportagem “Farra no Fies”, fruto de investigações dos jornalistas José Roberto Toledo, Rodrigo Burgarelli e Paulo Saldaña foi a vencedora da categoria principal do *Prémio ExxonMobil* de Jornalismo, antigo *Prémio Esso*, o mais importante galardão da área no Brasil. Pela primeira vez na história da premiação, uma reportagem de Jornalismo de Dados recebeu a distinção, uma evidente prova do “amadurecimento” do trabalho com dados por parte dos *media* (KNIGHT CENTER, 2015).

Ao longo das seis edições do *Data Journalism Awards*, a maior premiação da área, diversos projetos brasileiros foram indicados ao prêmio. Em 2012, ano da primeira edição, três trabalhos concorreram a um lugar de destaque, entre eles a série de reportagens 1000 vidas, produzida pelo *Correio*, que faz um balanço dos homicídios registados na capital da Bahia. Na mais recente edição do prêmio, a *news app* produzida pela *Folha de São Paulo*, Ranking de Eficiência dos Municípios, foi finalista do *Data Journalism Award 2017*. Nela, o jornal fez um levantamento de indicadores de mais de 5 mil municípios brasileiros para constatar “quem entrega mais saúde, educação e saneamento gastando menos” recursos. A aplicação concorreu na categoria de “*News Data App of the Year*”.

3.2. Jornalismo de Dados em Portugal

A revolução do 25 de abril aparenta ser um ponto comum nos estudos de Jornalismo de Dados em Portugal (SILVA, 2014; ALEXANDRE, 2014), ao representar o ponto de viragem no fazer jornalístico, que durante décadas esteve condicionado pelas políticas ditatoriais simbolizadas pelo lápis azul da censura.

Para Silva (2014), a prática do Jornalismo de Dados pelas redações portuguesas foi fomentada pela atuação do meio académico, responsável por permitir colaborações entre os *media* e os investigadores, como é o caso do projeto *Retrieval, Extraction and Aggregation Computing Technology for Integrating and Organizing News* (REACTION), que entre 2010 e 2014 permitiu a publicação de trabalhos jornalísticos focados na exploração de dados.

Em Portugal, a produção de artigos académicos que permitam conhecer o estado de desenvolvimento do Jornalismo de Dados ainda é reduzida. Através de estudos sobre o panorama do ciberjornalismo, é possível perceber que o ritmo de desenvolvimento de produtos com origem no online é lento e, portanto, não é um mal apenas do Jornalismo de Dados:

“Acresce que, nestes primeiros quinze anos, há um aspecto que foi obnubilado pelos debates à volta do aproveitamento das potencialidades da Web, do formato de ciberjornal ideal, das audiências, do web design, dos modelos de negócio, das necessidades de formação, ou das estratégias empresariais e de marketing: o próprio jornalismo. Com nuances consoante as diferentes fases, os media na Web foram sendo encarados menos como plataformas para o desenvolvimento de informação de qualidade própria, isto é, produzida pelos ciberjornalistas, do que como suportes para a reprodução de conteúdos originalmente produzidos para media tradicionais. Raras vezes foi dado aos cibermedia e aos ciberjornalistas espaço para o desenvolvimento de uma identidade própria, que viabilizasse a assunção de alguns papéis tradicionais do jornalismo, a saber, investigar, cultivar fontes próprias, vigiar poderes instituídos, influir na opinião pública, proporcionar análise e interpretação sobre questões complexas.” (BASTOS, 2011)

A visão de Bastos (2011) vai ao encontro do panorama do Jornalismo de Dados exposto por Martinho (2013) que apresenta o “estado embrionário” dessa especialização do jornalismo em Portugal, onde as organizações noticiosas relutam em assumir as potencialidades do online ao priorizar o papel para as publicações de trabalhos que envolvam dados. Essas escolhas editoriais evidenciam o desconhecimento do Jornalismo de Dados como uma tendência do online.

Em 2014, um estudo de Ilo Alexandre, apresentou possíveis indicativos para o número reduzido de trabalhos publicados seguindo as técnicas de Jornalismo de Dados. Na ocasião, de um conjunto de cinco publicações analisadas, apenas um jornal possuía mão-de-obra dedicada ao trabalho com dados. Os outros justificaram a ausência de

investimento como fruto da crise enfrentada pelo jornalismo impresso – ainda o principal subsidiador dos produtos jornalísticos em Portugal.

4. Metodologia

A pesquisa que deu origem a este trabalho tem como proposta estabelecer um ponto, não comparativo, do balanço atual da forma de se fazer Jornalismo de Dados no Brasil e em Portugal e foi suportada por duas componentes, uma qualitativa e outra quantitativa.

Como parte da tentativa de compreender «como é feito» o Jornalismo de Dados em cada um dos países, decidimos pela entrevista estruturada.

As entrevistas foram realizadas por chamada de voz na maioria dos casos, salvo exceção de duas profissionais, a quem foram aplicadas entrevistas presenciais. Das redações brasileiras, sete profissionais responderam ao conjunto de perguntas: Juan Torres, do jornal Correio, de Salvador(Bahia); Cristian Weiss, do Diário Catarinense, de Florianópolis (Santa Catarina); Katia Brembatti da Gazeta do Povo, Curitiba(Paraná); Daniel Bramatti do O Estado de São Paulo (São Paulo); Marlen Couto do O Globo (Rio de Janeiro); Raphael Hernandez da Folha de São Paulo (São Paulo) e Roberto Maleson do GloboEsporte (Rio de Janeiro).

Em Portugal foram entrevistados três jornalistas: Raquel Albuquerque, do Expresso(Lisboa); Rita Costa, do Público(Lisboa); e o jornalista Rui Barros, da Rádio Renascença(Porto).

Os jornalistas brasileiros foram selecionados sob a condição de fazerem, necessariamente, produtos para redações tradicionais, terem uma rotina de produção ativa na área e/ou reconhecimento pelo trabalho produzido. Os exemplos portugueses também produzem para redações tradicionais e representam quase a totalidade da amostra de jornalista dedicados ao trabalho com dados no país.

Com base na análise das respostas, fizemos um levantamento de variantes que caracterizam a forma como o Jornalismo de Dados é feito: «As vias de acesso ao Jornalismo de Dados », «As habilidades essenciais à prática», «Como os jornalistas veem investimento por parte das organizações onde estão inseridos», «Como o Jornalismo de Dados é organizado dentro da redação», «Os processos jornalísticos dominantes» e «Os constrangimentos ao crescimento do Jornalismo de Dados ».

No Brasil, onde o cenário se apresenta mais aberto ao Jornalismo de Dados com mais organizações e profissionais a desempenharem o papel de decodificador dos dados lançados em rede, consideramos relevante a integração de um componente quantitativa – explorada a partir de questionários – para expor o perfil de «quem faz» esse trabalho. Esse elemento da pesquisa foi sustentado com a análise de 50 respostas. Em Portugal, apesar de o número de profissionais ser mais restrito – quatro profissionais responderam ao questionário – entendemos que possuir dados indicativos, apesar de não servir como comparação em relação ao Brasil, complementa a componente suportada pelas entrevistas.

5. Análise dos resultados

As entrevistas exploratórias serviram como mote na percepção de como o JBD está presente nas redações e de que modo fatores externos e internos condicionam o desenvolvimento da área.

Na segunda componente, com base nas respostas obtidas a partir da análise dos questionários enviados aos jornalistas que adotam na sua prática quotidiana técnicas de Jornalismo de Dados, conseguimos apreender alguns tópicos que permitem expor o posicionamento desta especialização no Brasil e em Portugal, nas redações e na comunidade jornalística.

5.1. Vias de acesso ao Jornalismo de Dados

Produzir conteúdos baseados em técnicas de Jornalismo de Dados exige o incremento de novas rotinas produtivas, que implicam um «saber fazer» distinto do jornalismo dito tradicional. Ao longo das entrevistas, dois pontos mostraram-se relevantes na formação dos profissionais para aquisição de competências: os cursos, maioritariamente citados como *online*; e a existência de entidades que dão suporte à capacitação e ao desenvolvimento da área.

No caso dos cursos, eles são citados com objetivos diferentes, para uns foi um acréscimo de habilidades, podendo ser um diferencial em futuros trabalhos, para outros, a realização dos cursos surge com a intenção de suprir a necessidade de trabalhar com dados e acompanhar a tendência.

Em Portugal, os três profissionais que hoje representam o Jornalismo de Dados no país, foram contratados, inicialmente pelas competências na análise de dados, e para desempenharem funções na área ainda que não em regime de exclusividade.

- Brasil

Apesar da consolidação do conceito de Jornalismo de Dados ser mais recente, Cristian Weiss, jornalista do Diário Catarinense, relembra o seu início de carreira, em 2009, no Jornal de Santa Catarina, e refere que faziam Jornalismo de Dados, sem saber que de facto toda aquela metodologia era JBD: “Lá fazíamos muitos trabalhos no dia a dia, que muitas vezes tínhamos que coletar dados porque o poder público não tinha dados para aquela situação. A gente nem sabia que era Jornalismo de Dados, como conceito isso ainda não era tão difundido”. Além do trabalho e do contato direto com a área na redação,

o Jornalista do Diário Catarinense investiu em formações, “muitas online”, algumas através da Abraji e do *Knight Center* da Universidade do Texas, referenciadas como “públicas e acessíveis”.

Para Juan Torres, jornalista do Correio, o termo Jornalismo de Dados era desconhecido mesmo quando já se seguia a lógica do trabalho com bases de dados: “Quando fizemos essa série, eu não sabia que eu estava fazendo Jornalismo de Dados, eu particularmente sempre gostei de trabalhar com planilhas, me formei em administração, tinha essa coisa que normalmente jornalista não gosta que é trabalhar com números e numa lógica mais cartesiana.” Ter sido finalista da primeira edição do *Data Journalism Awards 2012* com a reportagem “1000 vidas”⁸ despertou o interesse do jornalista da publicação sediada em Salvador, que passou a fazer cursos, “a maioria foram cursos online”, a participar em eventos e *hackathons*: “Comecei a frequentar esses fóruns, o próprio Knight Center de Jornalismo das Américas, que é uma organização da Universidade do Texas, também oferece cursos de Jornalismo de Dados, online.(...) Eu fiz muitos cursos na área e hoje damos treinamento pela Escola de Dados no Brasil todo e desenvolvemos projetos em Jornalismo de Dados”.

Kátia Brembatti, da Gazeta do Povo, conta que tomou conhecimento com o conceito de Jornalismo de Dados em 2008, num congresso da Abraji, onde participou numa oficina de Reportagem Assistida por Computador. Apesar de não relacionar o conceito à prática, a jornalista já havia publicado antes disso matérias com “perfil de Jornalismo de Dados”. “Eu já trabalhava com planilhas de Excel, mas de um jeito autodidata, eu não recebi nenhum tipo de treinamento, ou na faculdade, ou no jornal que trabalho, para trabalhar com isso, mas eu comecei a sentir necessidade de mexer nisso”, explica.

Para Daniel Bramatti, o autodidatismo e o interesse na área – que o fizeram publicar trabalhos com dados mesmo antes do projeto de dados do Estado de São Paulo ser criado em 2012 – foram as chaves para o desempenho atual como jornalista de dados: “Eu comecei a fazer matéria de política, não entrevistando pessoas, mas entrevistando bases de dados, foi um filão que eu fui explorando e isso acabou, naturalmente, me colocando

⁸ O “1000 vidas” é uma narrativa jornalística suportada pelo acompanhamento e análise de dados sobre homicídios fornecidos pela Secretaria de Segurança Pública: <http://www.correio24horas.com.br/noticia/nid/homicidios-em-salvador-e-rms-aumentam-27-com-relacao-a-2011>

no Estadão Dados quando a ideia foi apresentada para a direção”. Segundo Bramatti, as habilidades que domina “aprendeu fazendo” ou através de “formações específicas”.

Na visão de Raphael Hernandez, da Folha de São Paulo, “coisas de internet, aprende-se na internet” e considerando que iniciou a carreira num jornal local, a RBC, do interior do Paraná, essa via foi a mais natural, com cursos na *Canvas Network* e na Abraji. Antes de ser contratado pela Folha, fez três *trainees*, dois na atual entidade empregadora, sendo que o último foi determinante: “Durante esse *trainee* eu tentei focar no trabalho com dados, usei inclusive os dados para me vender no processo seletivo, eu era a única pessoa do grupo de que tinha essas habilidades, como a gente estava trabalhando com saúde pública, o uso de dados veio muito bem a calhar”.

Tal como Raphael Hernandez, o jornalista Marlen Couto, do O Globo, começou a trabalhar efetivamente com Jornalismo de Dados a partir de um *trainee*: “Nessa experiência como repórter *trainee* eu tive a oportunidade de fazer algumas reportagens que utilizavam levantamento de dados, acabou por se perceber que eu tinha um perfil para trabalhar com Jornalismo de Dados por essa minha capacidade de fazer análises estatísticas, cruzamentos de informações e até uma habilidade no Excel”. O jornalista do O Globo referiu que não possuía antecedentes em Jornalismo de Dados além de um curso técnico em Excel, que não foi voltado ao JBD.

Roberto Maleson trabalha com desporto, mais especificamente, com a cobertura estatística dos considerados clubes de maior expressão do futebol brasileiro. O jornalista do GloboEsporte (GE) fez mais de quatro cursos na área, o primeiro em 2014, um *Massive Open Online Course* (MOOC) pelo *Knight Center*: “A partir disso que entrei no GE, eu entrei especificamente por conta disso, pelo meu conhecimento em dados”. O profissional considera que o facto de estudar a área ajuda na melhoria da “performance” da equipa de dados do GE, que ainda possui metodologias um pouco rústicas.

- Portugal

Apesar de não possuir experiência anterior na área, Rita Costa, do Público, direcionou a sua formação para essa componente do jornalismo ao frequentar o curso de Comunicação Visual do Instituto Universitário de Lisboa (ISCTE) e participar em cursos do *Knight Center*: “Eu gosto dessa parte de estatística e da análise dos dados e comecei a investir, a fazer cursos online, até que surgiu essa oportunidade no Público”.

No caso de Raquel Albuquerque, do Expresso, a porta de entrada para o Jornalismo de Dados foi um projeto de investigação: “Eu comecei a perceber o que era, estava nessa equipa que tinha vários investigadores na área das engenharias e eu estava na redação do Público, que era um dos parceiros do projeto. À medida que ia descobrindo o que era o Jornalismo de Dados e como é que se fazia, íamos aplicando e os trabalhos iam sendo publicados”. Como complemento ao conhecimento adquirido no trabalho de campo, realizou formações extensivas e cursos online. Apesar de mencionar a ausência de cursos em Portugal, citou o Centro Protocolar de Formação Profissional para Jornalistas (Cenfor), que promove “alguns cursos pequenos” na área.

Rui Barros conta que o facto de acompanhar os trabalhos feitos lá fora, alimentou o seu interesse na área e fez parte do estágio realizado na Rádio Renascença, onde trabalha atualmente como Jornalista de Dados e de outros tópicos, na publicação online do meio. Nesse período, o jornalista refere que apresentou alguns trabalhos e isso “mostrou que havia algo de diferenciador”. No que se refere à capacitação, os conteúdos presentes na web acabaram por ser os meios de aprendizagem: “Eu aprendi online, fui seguindo o que os outros andavam a fazer, o meu primeiro contacto com ferramentas foi um dos cursos do *Knight Center*, depois fui seguindo, vendo tutoriais”.

5.2. Aquisição de novas habilidades

O surgimento do Webjornalismo trouxe a necessidade de compreender que o jornalista passou a desempenhar funções que outrora eram destinadas a outros profissionais. Ser multitarefa passou a ser característica fundamental na produção de conteúdos para o universo online. O Jornalismo de Dados, ainda que pertencente ao Webjornalismo – que por si só já acarreta um novo perfil jornalístico – trouxe consigo novas funções aos profissionais da área.

Com base nas entrevistas, as características sugeridas indicam um balanço entre a atuação jornalística, o domínio de novas técnicas e de uma nova metodologia, e a necessidade de coordenação mental durante os trabalhos. Se por um lado o conhecimento “básico” em matemática é citado como um diferencial, elementos do foro psicológico como “vontade”, “paciência”, “curiosidade” e “disponibilidade mental” foram lembrados como importantes.

De um modo geral, a programação não é citada, pelos entrevistados, como essencial para o profissional do JBD. Para Juan Torres a “programação hoje é uma incógnita” e o

jornalista defende que é possível “participar de uma equipa de Jornalismo de Dados sem saber programar”. Raphael Hernandez corrobora a opinião do profissional do *Correio*, defendendo que o jornalista de dados “não precisa saber programar”, mas admite que “ajuda”.

Apesar de a programação não ser vista como característica fundamental para se trabalhar com dados, a linguagem de *Python* é frequentemente citada como parte da rotina de trabalho ou de aprendizagem. Há uma preocupação por parte dos jornalistas de incluir essa habilidade para facilitar o processo jornalístico. Não há um consenso quanto à finalidade da aplicação da programação, ela pode ser utilizada para extração, análise, limpeza e visualização dos dados.

- *Brasil*

Katia Brembatti, que já investiu em formações em *Python*, considera que sentiu necessidade de colmatar algumas lacunas ao nível dos conhecimentos, nomeadamente na área da programação: “De uns tempos para cá, cada vez mais eu tenho sentido a dificuldade de não programar e tenho tentado procurar vencer essa demanda. Sinto que eu poderia otimizar e automatizar coisas simples, coisas básicas, que muitas vezes eu dependo de outra pessoa.”

Marlen Couto justifica a decisão de aprender *Python* como uma forma de compreender como a programação pode ser utilizada: “Acho que mais do que eu mesmo fazer extração de dados utilizando a programação, o objetivo para mim, nesse momento, é saber mais o que pedir e o que pode ser feito com programação”.

Para Roberto Maleson a motivação de estar a aprender a programar é ser “independente”, mas admite que apesar de a programação ser útil na raspagem, análise e compilação dos dados, na rotina atual ela não é utilizada para esse fim: “A programação só entra quando temos algum projeto especial grande, é mais para produtos com interatividade, conteúdos maiores”.

Raphael Hernandez sugere que a natureza das bases de dados no Brasil faz com que sejam necessárias “manobras” para trabalhar a informação: “Eu uso muito programação para automatizar algumas tarefas mais simples. Eu não tenho o costume, por exemplo, de analisar dados em *Python*, eu estou estudando para isso agora”. O jornalista refere que essa linguagem é uma “ajuda no dia a dia” para fazer a “limpeza” e “estruturar” os dados.

Cristian Weiss relata que o uso de programação ocorre apenas depois de identificar uma possível história a partir dos dados, porque a criação de ferramentas “toma tempo”, fator que pode levar à utilização de materiais que já estão disponíveis: “Se formos usar o *Python*, tem o *GitHub*, que encontramos muitas ferramentas prontas, muitas vezes para a nossa necessidade só precisamos adaptar alguns scripts”. O jornalista conta que após a extração dos dados, outras linguagens, como o R, podem ser utilizadas “para poder organizar os dados e prosseguir no restante do método”.

Para Daniel Bramatti, a presença de programação na sua rotina de produção de conteúdos está concentrada em “alguns comandos de *SQL*” para permitir a leitura de bases de dados. “Tenho interesse em aprender *R*, em aprender *Python*, acho que isso é um buraco na minha formação”, declara.

Por iniciativa própria e apesar de não fazer parte do cotidiano profissional, Juan Torres, do *Correio*, tem participado de formações: “Estou fazendo cursos de programação online, *Python* básico já fiz, já estou seguindo para um *Python* mais avançado agora”.

- *Portugal*

Raquel Albuquerque assume que não possui conhecimentos de programação, mas justifica o ingresso na pós-graduação em Visualização de Informação como uma forma de adquirir “autonomia” e não ficar assim dependente de outros profissionais: “Há bibliotecas de visualização de dados como D3 e coisas do género, que eu estou a tentar aprender, principalmente para eu conseguir ter essa liberdade de já não ter que depender de um programador para fazer coisas que para eles é básico e eles têm muitas coisas para fazer”.

Para Rita Costa a programação como recurso jornalístico está em fase inicial: “Eu nunca fiz um trabalho em que eu só utilizei o *Python*, mas o que eu tento fazer é, quando tenho um trabalho que eu sei que dá para usar, vou tentando ganhar prática”

Rui Barros ressalta que não é programador, mas admite ter conhecimentos na área e cita algumas aplicações: “Recentemente tive um caso em que eu precisava desesperadamente de uma base de dados que não havia. Eu tive que escrever um *Scraper* em *Python*, fui eu que programei e o *Scraper* fez o trabalho e mesmo as minhas infografias são programadas por mim”. Para o profissional, aplicar códigos abertos encontrados no *GitHub* foi uma das formas de aprendizagem.

5.3. O Jornalismo de Dados dentro das redações

A existência de uma estrutura dentro da redação, traduzida em sectores próprios para a produção de trabalhos baseados em técnicas de extração, análise e publicação de dados, pode ser um indicador da frequência de produção dos conteúdos, da importância do Jornalismo de Dados para a publicação ou do nível de investimento aplicado.

Nos três jornais fora do eixo Rio de Janeiro-São Paulo, casos de “Gazeta do Povo”, “Correio” e “Diário Catarinense”, não há uma estrutura, deliberadamente organizada, para se fazer Jornalismo de Dados.

Em Portugal, nos três jornais mencionados a partir da entrevista aos profissionais, não existe uma secção, um núcleo físico, destinado exclusivamente ao Jornalismo de Dados. Nos três exemplos, apenas um profissional – em cada órgão – tem como função a produção de conteúdos a partir da análise de bases de dados.

A existência de núcleos, onde há uma estruturação definida para trabalhar com dados, não significa, necessariamente, que exista um isolamento em relação a outros profissionais da redação. Essa inclusão junto de outras editorias demonstra que existe uma integração não só entre os profissionais que possuem o mesmo perfil, mas com outros elementos da redação para troca mútua de conhecimentos.

- *Brasil*

Tabela 2. Redações brasileiras sem estrutura fixa para o trabalho de Jornalismo de Dados

Gazeta do Povo	Correio	Diário Catarinense
Não há uma estrutura onde estejam concentrados os profissionais que trabalham com dados, na redação há outros profissionais com “perfil” para esse tipo de trabalho.	Não há uma estrutura e os trabalhos dependem de “iniciativas individuais”.	Não há uma estrutura e existem outros profissionais com “habilidade para trabalhar com Jornalismo de Dados” na redação.

Para Cristian Weiss, a “dificuldade” de fazer essa “integração” das equipas, até mesmo em relação às outras redações do grupo de comunicação de que o Diário faz parte, impede um “fortalecimento dos resultados” obtidos. “O pessoal chama de ‘lobo solitário’, é um jornalista de um lado da redação trabalhando com dados, outro de outro lado da redação trabalhando com dados, cada um fica trabalhando em seu projeto”, explica.

Tabela 3. Redações brasileiras com equipes ou núcleos de dados

Folha de São Paulo	O Globo	O Estado de São Paulo	GloboEsporte
Possui o Núcleo de Inteligência, focado no Jornalismo de Dados e o Núcleo de Audiência e Dados, que tem outro foco, mas também realiza reportagens baseadas em dados. Na redação há outros profissionais, fora dos Núcleos, com “aptidão” para a área.	Possui um núcleo dedicado exclusivamente ao Jornalismo de Dados e ao <i>Fact-Checking</i> .	Possui um Núcleo próprio, o Estadão Dados. Foi a primeira redação do Brasil a criar um grupo para esse fim.	Possui uma secção destinada especificamente ao trabalho com dados, a Editoria Futdados.

Criado em 2012, o Estadão Dados – núcleo do O Estado de São Paulo – começou com três pessoas e “chegou a ter cinco pessoas, das quais dois eram programadores”, algo que para Daniel Bramatti “é característica importante do Jornalismo de Dados”. Atualmente o núcleo conta com dois jornalistas, devido a saída de profissionais que foram “contratados” por outras empresas. “Estamos tentando recompor a equipa nesse momento, mas é um momento de crise no Brasil e isso tudo aliado com a tradicional crise dos media, a gente tem poucos recursos, esses profissionais que a gente procura, geralmente são mais caros”, conta.

Daniel Bramatti salienta ainda o colaborativismo que existe entre as outras secções, com “parcerias praticamente com o jornal inteiro”. Essa prática também se revelou comum noutros núcleos. Raphael Hernandes, da Folha de São Paulo, refere que outros profissionais de outras editorias, “que têm uma relação um pouco mais próxima com os dados”, se “associam”, por vezes, ao Núcleo de Inteligência e também ao Núcleo de Audiência e Dados. No O Globo, Marlen Couto afirma que a existência do Núcleo “não impede” a interação “com as outras editorias da redação”, o que pode levar mesmo ao surgimento de reportagens com participação de diferentes profissionais.

- Portugal

Tabela 4. Como o Jornalismo de Dados está organizado nas redações portuguesas

Público	Expresso	Rádio Renascença
Há um profissional destacado para o trabalho de Jornalismo de Dados .	Há um profissional destacado para o trabalho de Jornalismo de Dados . Está inserida na secção Sociedade, mas consoante os dados, enverga para outros temas.	Há um profissional destacado para o trabalho de Jornalismo de Dados , está integrado na Redação Web.

O estágio inicial de projetos em Jornalismo de Dados é apontado como o principal motivo para a ausência de uma estrutura integrada. Rita Costa justifica a produção de Jornalismo de Dados , pelo facto de a aposta do Público ser “embrionária”. Rui Barros aponta na mesma direção ao referir que “a aposta no Jornalismo de Dados na Renascença é um projeto relativamente recente”.

Rui Barros afirma ainda que apesar de ser o único a trabalhar “exclusivamente com dados”, não significa que “trabalhe sozinho” esses dados, uma vez que muitas vezes conta com a ajuda de outros jornalistas: “Saber limpar uma base de dados é relativamente fácil, agora conseguir interpretar o que os valores dizem, já é preciso ter um conhecimento na área e conto sempre com esse conhecimento da redação”.

5.4. A origem dos dados

A primeira etapa do processo jornalístico de análise de dados é a busca e extração da informação que servirá como mote para a fundamentação da peça. As instituições governamentais, que muitas vezes promovem a adoção de medidas de transparência ou cumprem com leis de acesso ao responder pedidos por informação, são as fontes mais utilizadas pelos entrevistados para acesso às bases de dados. Institutos de pesquisa públicos – responsáveis por estatísticas segmentadas ou recenseamento – Organizações não governamentais e fundações também são citados locais de procedência da informação.

Em alguns casos, inconsistências nos dados, formatos ociosos ou mesmo ausência de informações que cubram determinado tema podem levar à produção própria de bases de dados a partir da construção de uma metodologia de coleta.

Com base nas entrevistas, a produção de bases de dados por parte da redação é uma prática que existe numa minoria dos jornais representados, todos brasileiros. Ainda que haja um processo próprio de coleta em alguns conteúdos produzidos, essa característica não é um elemento definidor do trabalho de Jornalismo de Dados realizado, a frequência de produção é pequena.

- *Brasil*

Tabela 5. Redações com produção de bases de dados

	Entrevistados
Gazeta do Povo <i>Há produção</i>	“Eu gostaria de trabalhar com mais dados próprios, mas isso tem um custo de tempo que normalmente acaba diminuindo o uso desse recurso. Por ano eu faço no mínimo 2 ou 3 matérias com bases próprias”.
GloboEsporte <i>Produção constante</i>	“A nossa compilação é própria, não buscamos em nenhum local na rede para conseguir os dados que precisamos”.
Folha de São Paulo <i>Há produção, pouco frequente</i>	“Faço isso raramente, mas vejo isso acontecer na redação. Eu acabo fazendo pouco por uma questão de tempo”.
O Globo <i>Há produção, pouco frequente</i>	“Dependendo do tipo de informação nem sempre conseguimos achar a informação do jeito que queremos, precisamos montar nossas bases, nossos levantamentos. (...) ainda não é muito comum no O Globo, a gente tem dificuldade e acho que principalmente pela escassez de mão de obra, porque são levantamentos que levam um pouco mais de tempo”.

O GloboEsporte é a única organização que produz a totalidade dos seus conteúdos com recurso ao levantamento próprio de informação. Tal como foi referido durante o processo de construção de peças guiadas por dados, os produtos criados pelo jornal dependem de informações que não são coletadas por outras entidades, são baseadas em elementos estatísticos de jogos de futebol que foram definidos previamente como relevantes para compreensão da realidade dos jogos.

As outras três publicações justificam a coleta irregular como consequência do tempo necessário para a construção das bases. Katia Brebbatti, da Gazeta do Povo, aponta a vantagem de a informação ser exclusiva, mas faz a ressalva de que quem produz e trabalha a partir de bases originais “assume 100% da responsabilidade da informação”.

Tabela 6. Redações sem produção de bases de dados

	Entrevistados
Correio <i>Não produz</i>	“Na redação do correio especificamente, o trabalho de Jornalismo de Dados , inclusive já tem algum tempo que a gente não faz alguma coisa.”
Diário Catarinense <i>Não produz</i>	“Isso é algo que eu estou tentando estimular os colegas a fazerem. (...) vemos a dificuldade muitas vezes do próprio repórter entender que se um dado público não está disponível, porque não tentarmos pelo menos recolher uma amostragem para mostrarmos uma realidade?”
O Estado de São Paulo <i>Não produz</i>	“Não criamos bases de dados do nada, podemos criar bases que não estão publicadas, cruzando informações de fontes diferentes, isso a gente faz. É muito difícil a gente produzir uma base de dados a partir do nosso trabalho jornalístico, até porque nós não temos essa capacidade de coleta tão grande.”

No caso das publicações brasileiras que não incluem na rotina da redação a criação de bases, as justificativas são díspares. O Correio, que despertou para área ao ser nomeado finalista da principal premiação de Jornalismo de Dados , não produz, com regularidade, peças guiadas por dados, tampouco investe em criação de fontes de dados próprias. No Diário Catarinense, Cristian Weiss pressupõe que não há uma cultura redatorial voltada para esse tipo de trabalho – não se cria alternativas quando os dados não são disponibilizados ou são inexistentes. Para Daniel Bramatti, a recolha de informação implica esforços jornalísticos além da capacidade de coleta do Estadão Dados.

- Portugal

Tabela 7. Regularidade da produção de bases de dados nas redações portuguesas exploradas

	Entrevistados
Público <i>Não produz</i>	“Para fazer isso é preciso algum conhecimento de estatística, para fazeres um <i>dataset</i> e inferires daí uma notícia ou reportagem é preciso ter algum cuidado. Acho que ainda é preciso dar esse passo, era importante.”
Rádio Renascença <i>Não produz</i>	“Eu estou na Renascença há um ano, é um processo ainda em crescimento, mas obviamente que isso estará nos quadros.(...) É um processo que demora muito mais do que propriamente recorrer a uma instituição que já tenha a base de dados.”
Expresso <i>Não produz</i>	“Imagino que para uma redação produzir bases de dados, é preciso no mínimo ter mais do que uma pessoa, que esteja todos os dias a fazer Jornalismo de Dados . Estou sozinha e a rotina diária e quantidade de coisas que é preciso fazer e dar resposta, as vezes bloqueia.”

Nas redações portuguesas representadas, a ausência da produção dos trabalhos suportados por coletas de dados própria é sustentada, pelos jornalistas, através de três pontos: São

necessárias competências e conhecimentos inerentes para a realização desses tipos de tarefas; A produção de conteúdos baseados em dados se encontra num estágio inicial; há a necessidade de mão de obra que possibilite a realização do processo.

5.5. A autoridade das fontes baseadas em dados

Uma das principais variações entre o JBD e o jornalismo tradicional está na hierarquização das fontes utilizadas na credibilização dos conteúdos informativos. Enquanto que o Jornalismo de Dados implica, necessariamente, o tratamento de bases de dados, o tradicional possui outros caminhos alternativos. A importância dada à fonte de informação na materialização dos trabalhos jornalísticos pode ser um indicador do tipo de peças que são produzidas.

Há um consenso quanto à autoridade que os dados possuem para serem fonte única no processo, no geral, a qualidade e origem dos dados podem definir a produção de trabalhos só com recurso à base de dados. Apesar do reconhecimento, os jornalistas, na sua maioria, declaram a preferência pelo complemento de outros tipos de fontes.

A título de exemplo, na investigação conduzida por Philip Meyer, a propósito das manifestações de Detroit, enquadrar a metodologia no contexto da história é encarada como um benefício à credibilidade da peça⁹.

Para Cristian Weiss a origem dos dados pode indicar se ele é ou não “inquestionável” e cita o exemplo do Censo Escolar produzido pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira(INEP): “Esse é um dado censitário, é um dado que de facto abrange quase a totalidade da amostra. Quando a gente trabalha com outras pesquisas que são por amostragem, eu acho que os dados por si só não têm autoridade nenhuma.” Apesar de haver dados como os do Censo, o profissional do Diário Catarinense considera que “precisamos, sim, buscar as histórias na rua para ver se isso se reflete, ou quando isso se reflete e também buscar análise de especialistas” para compreender se o caminho da interpretação está correto. “Eu acho que não se pode abrir mão de fontes externas”, salienta.

⁹ Um dos tópicos da série de reportagens de Philip Meyer consistiu na apresentação dos “detalhes” da pesquisa realizada na obtenção dos dados:
<https://s3.amazonaws.com/s3.documentcloud.org/documents/2070181/detroit1967.pdf>

Na perspectiva de Juan Torres é possível fazer jornalismo só com dados, tendo em conta a origem da informação: “É possível, eu acho que o cuidado que tem que se ter, assim como a gente toma cuidado com a credibilidade da fonte humana, quando você está entrevistando você tem que saber se essa fonte tem credibilidade para falar o que está falando e de que lugar social ela está falando, qual a institucionalidade dela.”

Daniel Bramatti também é afirmativo quanto à autoridade dos dados, para o jornalista a “confiança na qualidade dos dados” pode determinar a escolha dos tipos de fontes a serem utilizados. “Se os dados estiverem bem coletados, bem organizados, é possível você fazer esse papel de entrevistador dos dados, sempre levando em conta que, da mesma forma que uma entrevista tradicional, a qualidade do seu trabalho vai depender da qualidade das perguntas. Da mesma maneira com um *dataset*, você pode olhar para a informação menos relevante e mergulhar nela, quando o *lead* pode estar logo ali do lado”, explica.

Katia Brembatti assume a possibilidade de produzir conteúdos jornalísticos sem entrevistar fonte humana e cita um trabalho produzido pelo Valor Económico sobre o Bolsa Família, que foi utilizado em suas aulas como exemplo de trabalho sem “lidar com pessoas”. No entanto, reconhece o acréscimo que o contacto com “personagens” e “especialistas” pode trazer: “Em algumas situações faço através do meu computador, só os dados se bastam, mas eu gosto dessa complementaridade, acho que ela agrega para o público.”

A confiança nos dados também é o ponto decisor para Roberto Maleson que utiliza o trabalho feito no GloboEsporte para se justificar: “Como somos nós mesmos que compilamos os dados, nós confiamos nos dados. A princípio, para nós, não precisamos entrevistar alguém, lógico que é bom sempre ter.” Para o jornalista é importante apresentar “pelo menos a metodologia” da compilação e se posiciona em relação à soberania dos dados: “Para alguns casos, resolve só a base de dados, mas se puder entrevistar a fonte humana, trazer mais história para reportagem dá uma confiabilidade melhor para conteúdo.”

Através do exemplo de um Especial sobre Reformas Ministeriais no Brasil, que colaborou, Raphael Hernandez reconhece a possibilidade de fazer matérias “só entrevistando dados”, mas recorre ao “princípio jornalístico” de ouvir as partes para

justificar a necessidade de criar contexto: “De regra, você irá precisar seja de um especialista para comentar, seja uma pessoa para te dar o outro lado daquela história.”

Marlen Couto admite que não faz parte da sua prática, mas que há informações que podem ser apreendidas só com dados: “Depende muito da situação, mas em geral, eu costumo não parar só nos dados, é preciso levar as várias interpretações possíveis daquela informação”. Para o jornalista do O Globo, a preocupação com o leitor pode levar à uma análise mais profunda: “Eu acho também que o nosso leitor, não necessariamente, ele tem essa capacidade analítica que é tão fundamental no perfil do jornalista de dados, então é preciso ter um cuidado com o leitor e propor essas interpretações.”

Ainda que o “quase sempre não” pressuponha a possibilidade, remota, de fazer trabalhos apenas com dados, Thiago Reis argumenta que “nem sempre é possível confiar na base” e que conclusões “tiradas de recortes de bases de dados” podem indicar uma inclinação, um direcionamento. “Como em qualquer peça jornalística, ouvir o maior número de fontes qualificadas é sempre o recomendado”, conclui

Para Raquel Albuquerque, “uma *dataset* não funciona sozinha” isso significa “visualizar dados” e está no âmbito da Visualização, o uso exclusivo de dados na produção de conteúdos não está no âmbito do Jornalismo de Dados : “Fazer jornalismo com base em dados é, precisamente, procurar notícias naqueles dados, procurar histórias, pessoas que dão sentido, ir à procura de especialistas que expliquem o que está nos dados”.

Rita Costa segue a mesma linha ao negar que os dados, por si só, tenham força suficiente na construção jornalística. A jornalista cita um trabalho, de sua autoria, com foco em dados que mostram a inexistência de casamentos homossexuais em mais de 50% dos concelhos país e argumenta que há detalhes que “os dados não dizem”: “não dizem o porquê e o porquê é essencial responder”.

Rui Barros, da Rádio Renascença, considera que entrevistar só dados representa uma dificuldade na atuação como jornalista, que é visível durante a etapa de escrita da peça: “Eu diria, quase sempre, que tenho necessidade de falar com alguém que é um especialista na matéria, sinto sempre a necessidade de ter outras pessoas a explicarem-me aqueles dados”.

5.6. As etapas de produção de conteúdos baseados em dados

De um modo geral, as etapas do processo jornalístico são as mesmas. As variações, na sua maioria, são resultantes da distribuição de tarefas. A visualização é a etapa onde há maior participação por parte de outros profissionais, ou através de designers, ou, nalguns casos, através de programadores/*developers*.

Nos casos da “Gazeta do Povo” e do “GloboEsporte”, onde é mais comum a produção de bases de dados a partir de coleta própria, os processos não seguem o *data pipeline*, sendo a origem dos dados e a forma como eles foram obtidos a condicionar a metodologia utilizada. Nos outros órgãos de comunicação, a etapa inicial consiste na obtenção das bases de dados a partir de entidades, levando a que haja sempre, num primeiro momento, a limpeza dessa informação.

Linguagens como *Python* e o uso do *Excel* são frequentes nas três primeiras etapas do procedimento. A visualização é, em alguns casos, uma aliada na componente analítica do trabalho. O *Tableau* é citado como referência para gerar elementos que permitam fazer inferências. Também se revelou comum, na análise, o contato com fontes humanas, com o propósito de ajudar a decifrar os dados.

Os trabalhos, na sua maioria, têm origem em perguntas, hipóteses ou assuntos e, posteriormente, são remetidos para a base de dados. Há casos, onde é prática o trabalho inverso, suportado através do monitoramento das bases de dados. No entanto prevalece a ordem ‘definição do assunto de interesse público → bases de dados como recurso para a contextualização’.

- *Brasil*

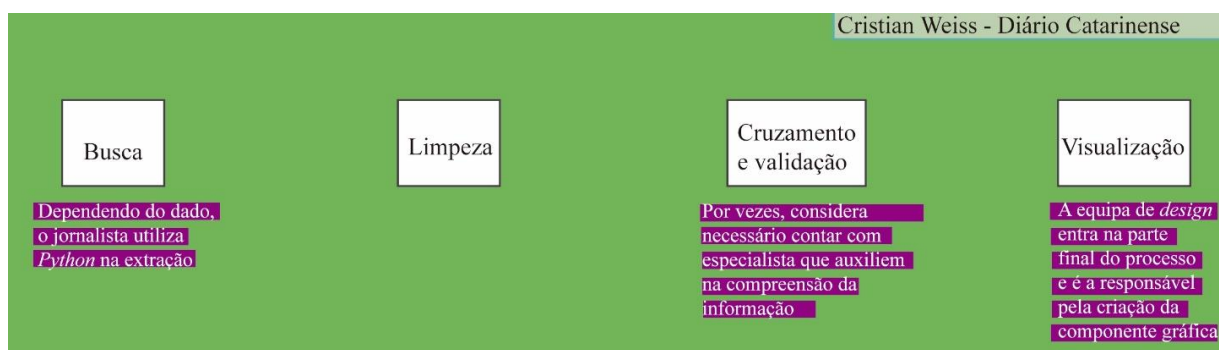


Figura 5. Etapas jornalísticas seguídas por Cristian Weiss no trabalho com dados

Cristian Weiss conta que o trabalho que realiza pode ter dois pontos de partida: ou descobre uma história e tenta “mostrar isso objetivamente em dados” ou pela “curiosidade” de aceder a bases que estão abertas. “Às vezes a história aparece antes dos dados e às vezes os dados aparecem antes da história”, conclui.

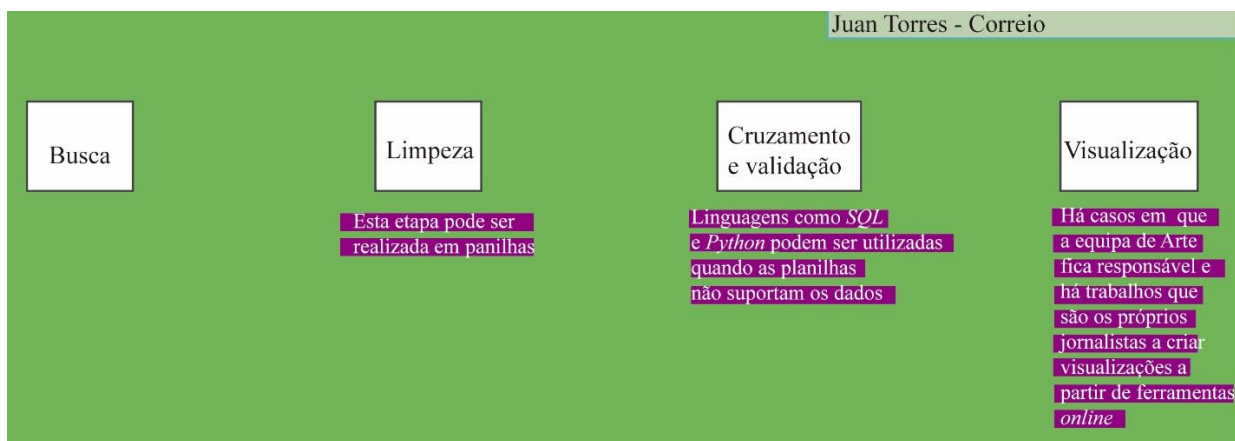


Figura 6. Etapas jornalísticas seguidas por Juan Torres no trabalho com dados

Para Juan Torres há duas formas de começar o trabalho com dados: se o jornalista não tiver a base de dados “ou parte para a coleta” ou então “desiste da matéria”; caso ele tenha a base de dados é seguir o *data pipeline*. O jornalista dá entender que a origem das matérias é semelhante aos trabalhos do jornalismo tradicional ou então resulta do acompanhamento de bases de dados.



Figura 7. Etapas jornalísticas seguidas por Daniel Bramatti no trabalho com dados

Daniel Bramatti refere que o trabalho se inicia a partir de suposições, baseadas em perguntas: “Procuramos trabalhar um pouco como cientista, formular uma hipótese diante de uma realidade, a minha hipótese diante de tal fenômeno é que ele é causado por tal coisa ou que ele está correlacionado a tal coisa, aí a gente bate os dados e checa essa hipótese”. Para os jornalistas, as etapas que seguem a busca e extração dos dados são as mais trabalhosas, uma vez que 90% do tempo “é aplicado na limpeza e análise dos dados”.

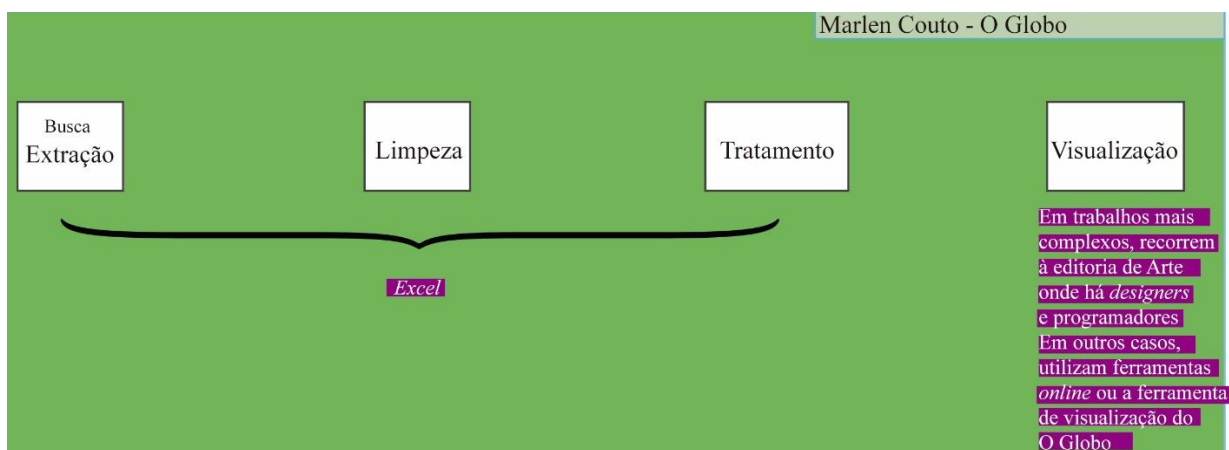


Figura 8. Etapas jornalísticas seguidas por Marlen Couto no trabalho com dados

Marlen Couto afirma que “muito da produção depende do monitoramento de bases de dados disponíveis” e que há um “acompanhamento do noticiário” para “identificar assuntos pertinentes”, que depois podem ser remetidos para Jornalismo de Dados. Para o profissional, os trabalhos surgem depois de “identificar bases de dados” e com base na “análise desses dados chegar a alguma conclusão” ou a partir de “perguntas que precisam ser respondidas”.

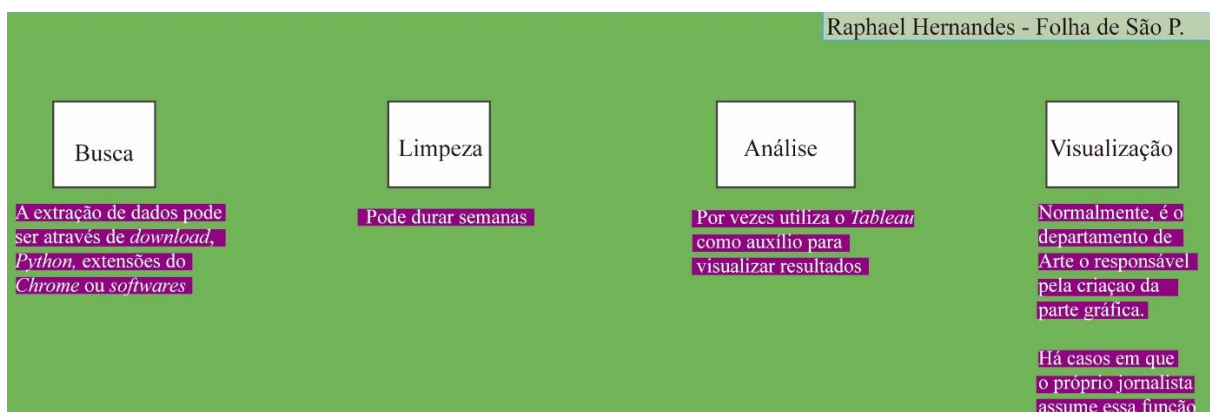


Figura 9. Etapas jornalísticas seguidas por Raphael Hernandes no trabalho com dados

Para Raphael Hernandez o início do processo se dá em três frentes: ou a partir de “uma matéria sobre um assunto x” e é necessário “um contexto” e os dados são utilizados; ou a partir do encontro com “uma base de dados ou uma base sazonal de que saiu uma novidade”; ou também através de “algum pedido da chefia”.

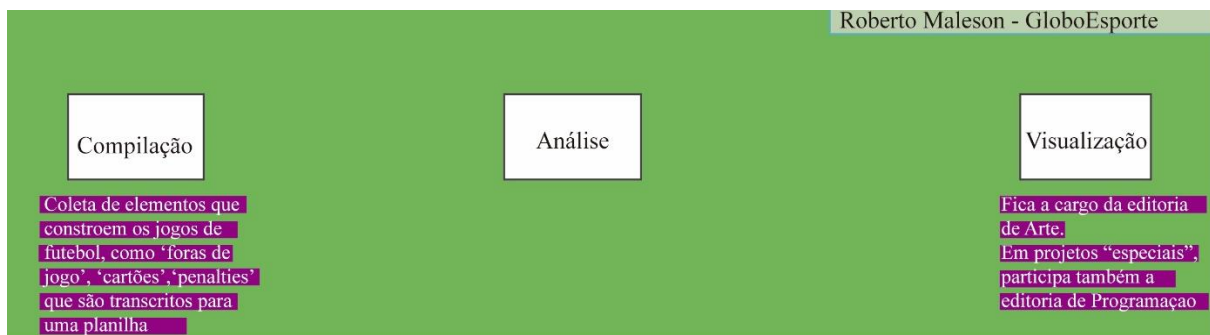


Figura 10. Etapas jornalísticas seguidas por Roberto Maleson no trabalho com dados

A rotina metodológica de Roberto Maleson – baseadas na coleta constante de dados – é causada pelo acompanhamento ‘agendado’ de eventos desportivos. Por vezes, há espaço para a produção de artigos fora dessa linha. Apesar de produzir conteúdos inspirados em trabalhos realizados em outros países, o jornalista lamenta a falta de JBD no futebol porque faz com que não haja “tantos trabalhos para se guiar” e para poder “adaptar” à realidade brasileira.

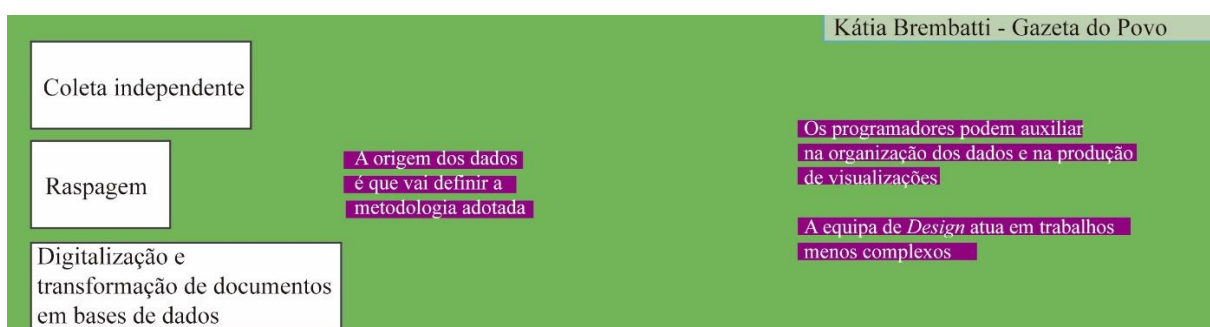


Figura 11. Etapas jornalísticas seguidas por Katia Brebatti no trabalho com dados

Kátia Brebatti considera que “o Jornalismo de Dados está, aos poucos, criando a sua própria forma, não tem um único jeito de fazer Jornalismo de Dados”. Da mesma forma, considera que os assuntos de pauta surgem “de vários modos”, a jornalista cita *releases* recebidos e a influência de trabalhos feitos por outros jornais.



Figura 12. Etapas jornalísticas seguidas por Rita Costa no trabalho com dados

Rita Costa conta que as ideias surgem a partir de sugestões ou a partir da “perceção do que pode ser feito e do que pode ser interessante” para quem lê: “Quando eu tenho a ideia, o ideal é formular um conjunto de hipóteses e perceber que dados preciso para confirmar ou desmentir essas hipóteses”. A jornalista cita ainda o exemplo das Eleições Autárquicas de 2017, onde a atualidade definiu o trabalho de Jornalismo de Dados.

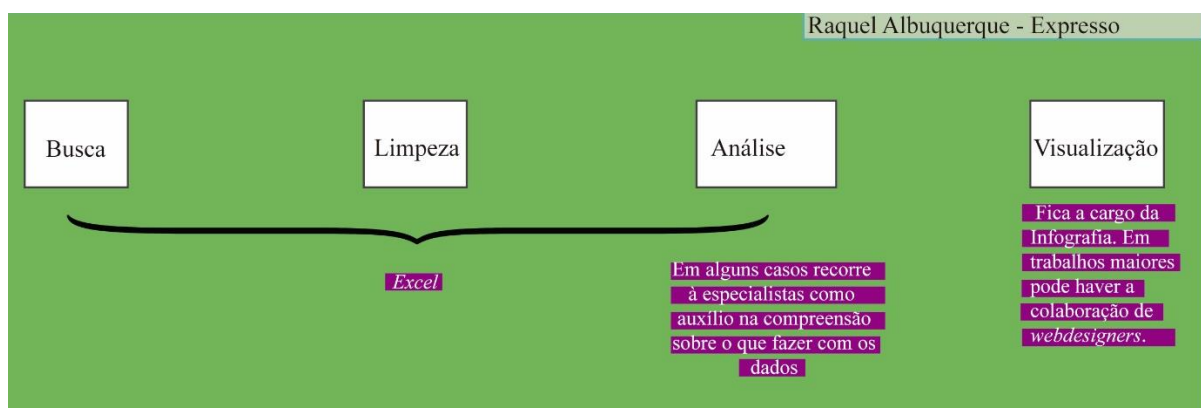


Figura 13. Etapas jornalísticas seguidas por Rita Costa no trabalho com dados

Raquel Albuquerque refere que os trabalhos podem surgir a partir de sugestões ou ter início na própria base de dados: “Tanto pode ser alguém que tenha a ideia e diz "porque não fazer uma coisa sobre isso?", então eu agarro aí, ou então cruzar com uma base de dados, uma coisa que eu nem sequer sabia que existia e o ponto de partida é precisamente

aquela base de dados”. A jornalista cita ainda os casos de incêndios em Portugal para exemplificar que há casos em que a atualidade é retratada com base em dados.



Figura 14. Etapas jornalísticas seguidas por Rui Barros no trabalho com dados

Rui Barros afirma que é “raro” o processo começar em “encontrar respostas” através uma base de dados que ele já possui: “Normalmente eu tenho ideias ou o meu editor diz que era interessante ter um [trabalho] sobre isso ou sobre [aquilo], depois faço trabalhos de jornalista, quer seja para procurar em documentos, ou ligar para instituições públicas”.

5.7. De que forma o investimento das redações é sentido

Tendo em conta o panorama atual de instabilidade financeira da imprensa a nível mundial, o surgimento de um novo formato de fazer e publicar jornalismo pode representar um motor para a recuperação da indústria.

No Brasil, a situação é exposta em números, estes baseados nas demissões contabilizadas desde 2012. Cerca de 2026 jornalistas foram demitidos, 48% de redações voltadas ao impresso e 10% do online. Em 2017, 380 casos de demissões foram registados¹⁰. Para os entrevistados – que afirmaram haver investimento – a existência de mão de obra capaz de produzir conteúdos, seja através de núcleos próprios ou jornalistas diluídos dentro da redação, é o fator de investimento mais citado.

Em Portugal – ao contrário do Brasil, o exercício pleno da profissão é dependente da posse de uma Carteira Profissional ativa – o número de jornalistas em condições de

¹⁰ “Passaralhos” é um projeto desenvolvido pela agência Volt Data Lab, que recolhe dados através da análise de notícias para registar as demissões na imprensa: <http://passaralhos.voltdata.info>

exercer a função tem diminuído, quase dois mil profissionais perderam a titularidade da carteira nos últimos dez anos, de acordo com os dados de 2017 ¹¹. Talvez com base nesse cenário atual, o principal investimento citado pelos jornalistas portugueses durante as entrevistas tenha sido a tentativa, por parte jornais, de manter um profissional dedicado a fazer trabalhos com técnicas de Jornalismo de Dados .

Quando se trata do apoio financeiro necessário para a expansão da área dentro de cada órgão de comunicação, é referido o facto dos jornais estarem a apostar na aquisição de conhecimento através de subsídios e de infraestruturas como hardwares e softwares.

- *Brasil*

No “Correio” Juan Torres afirma que não há investimento atualmente, mas um plano para que em 2018 haja a criação de “uma área ou, pelo menos, um método” para estimular o crescimento de conteúdos: “Uma metodologia para que o Jornalismo de Dados entre mais no fluxo de produção das pessoas, para que as pessoas considerem mais o Jornalismo de Dados na hora de fazer as apurações”.

Segundo Cristian Weiss, no “Diário Catarinense” os trabalhos são resultado de iniciativas individuais dos jornalistas, não há um estímulo evidente: “É importante ter essa fluência de leitura e análise de dados e a gente ainda não vê com clareza um interesse da chefia de redação, por exemplo, em investir nisso neste momento”.

No “O Globo” existe um núcleo próprio, há um jornalista em regime *trainee* e há uma ferramenta própria do jornal para facilitar a construção de visualizações. É precisamente nessa componente da visualização da informação que o jornal “tem investido muito”, segundo Marlen Couto. “Ele fez algumas mudanças até gráficas na apresentação do jornal e tem estimulado que a redação pense muito a visualização e como essa informação vai ser apresentada, como os dados vão ser apresentados”, completou.

Para Roberto Maleson ter “uma equipa de oito pessoas trabalhando com dados e entregando Jornalismo de Dados de desporto, de futebol, para o público” representa a

¹¹ <https://www.publico.pt/2017/01/12/sociedade/noticia/jornalismo-este-caminho-esta-a-levar-nos-a-autodestruicao-1757968>

aposta do jornal na área. No “GloboEsporte” há um núcleo exclusivo para o trabalho com dados, mas também a manutenção de produtos permanentes produzidos a partir de dados.

Na perspectiva de Daniel Bramatti, a atual crise que afeta também a imprensa mundial, tem condicionado o investimento do “Estadão” em todas as áreas. O jornalista acredita que a manutenção da equipa do “Estadão Dados” “em si já é um investimento”, por se tratar de algo “caro” : “Nem sempre você tem um volume grande de produção e hoje muitas coisas no jornalismo estão voltadas para o volume, volume de cliques, volume de acessos e isso é difícil conseguir com o tipo de trabalho que fazemos”.

Katia Brembatti acredita que o principal investimento do jornal está na “contratação de pessoal e tempo do jornalista”, mas cita também a aquisição de softwares e estímulo para a formação, esse último, uma preocupação recorrente: “Vamos para vários cursos, principalmente da Abraji, sempre tentar se manter atualizado sobre o que está sendo feito. Eles já me custearam várias vezes para ir, as últimas duas vezes eu mesma custeei, eles não me mandaram, mas mandaram outras pessoas, para também não ficar mandado as mesmas pessoas, eles liberaram os dias de trabalho para eu ir”.

Raphael Hernandez conta que há apoio da “Folha” em subsídios para cursos e ressalta a existência de equipamentos próprios na redação: “Tem bastante investimento na formação, tem questão de infraestrutura também, o computador no qual eu trabalho, no qual as pessoas que trabalham mais com dados atua, eles são máquinas um pouco melhores do que o resto da redação”. Como parte da aposta na formação, a Folha de São Paulo, em parceria com o Google, anunciou um programa de *trainee* com o objetivo de “atrair programadores queiram se especializar um pouco mais em jornalismo e jornalistas que queiram aprender um pouco mais desse mundo dos dados”.

- Portugal

Para Rui Barros, o facto de ter sido contratado para assumir essa função evidencia o apoio dado à área: “Ao ter decidido apostar num jornalista de dados já é uma grande decisão, num momento em que a indústria mediática não está a contratar”. Rui Barros cita o exemplo de um encontro sobre jornalismo ocorrido na capital portuguesa, Lisboa, em que recebeu por parte do jornal a permissão de dedicar o dia ao evento, com total cobertura dos custos por parte do meio de comunicação, o que mostra também a “abertura” e “vontade” da Rádio Renascença de promover o crescimento do trabalho com dados.

Raquel Albuquerque considera que contratar uma profissional com habilidades para trabalhar com dados foi umas das formas que o “Expresso” encontrou para fomentar a área: “Eu vejo como investimento o facto de, por exemplo, na altura em que o Diário começou, em 2014, me terem contratado precisamente porque eu fazia Jornalismo de Dados”. A jornalista do “Expresso” assume, no entanto, que a “necessidade de ter notícias” e a luta “contra a voragem do tempo” acabam por fazer com que a presença de trabalhos com Jornalismo de Dados ou Investigativo sejam “diluídas”.

No Público, Rita Costa acredita que mesmo “que seja pelo reconhecimento de que isso é importante” há um investimento do jornal na área.

5.8. Constrangimentos ao crescimento do Jornalismo de Dados

O avanço do JBD a um patamar onde há o reconhecimento e a valorização como técnica eficaz na cobertura de questões públicas requer um esforço dos jornalistas para ultrapassar barreiras que colocam em causa a evolução da área.

- Brasil

Entre os fatores apontados pelos profissionais dos jornais brasileiros, predomina a ideia de que não há profissionais suficientes no mercado com as habilidades necessárias para responder às exigências do Jornalismo de Dados .

Ainda no Brasil, e no que toca à abertura dos dados, foi citada, algumas vezes, ao longo das entrevistas, a inconsistência existente na hierarquia de poderes em relação à transparência demonstrada. Cristian Weiss, Daniel Bramatti e Marlen Couto acreditam que à medida que a autoridade governamental decresce – o Governo Federal é a instância superior, seguido do Governo Estadual e do Governo Municipal – há também uma diminuição do comprometimento com as medidas de transparência.

Katia Brembatti considera que a falta de “tempo” é o principal obstáculo ao desenvolvimento do Jornalismo de Dados no Brasil. Segundo a jornalista, para “priorizar e dar tempo” a essa metodologia de trabalho, é preciso mostrar a sua relevância: “O Jornalismo de Dados precisa mostrar o impacto que ele causa, não em cliques, ele precisa mostrar o que ele muda na perceção do público, precisa mensurar coisas, que realidades foram transformadas por esse jornalismo”. Em relação à quantidade de informação disponível, a profissional acredita que há “mais informação pública disponível do que capacidade” para “processar essa informação”.

Para Cristian Weiss, a atuação do Governo perante os pedidos de acesso à informação é ineficaz porque o “poder público não está cumprindo” com a Lei de Acesso à Informação, é preciso que compreendam que, após a solicitação, devem responder “na sua integralidade” e “no prazo”. O jornalista completa que os outros pontos que têm dificultado “a expansão do desenvolvimento do Jornalismo de Dados no Brasil” são a não compreensão das chefias de redação de que “é algo inevitável e importante” de se ter nas redações e a ausência de “interesse dos jornalistas em aprenderem ferramentas e aprenderem a trabalhar com Jornalismo de Dados”.

Juan Torres acredita que a mudança exigida para se fazer JBD enfrenta obstáculos como o facto de “serem técnicas, em geral, novas e dificilmente dominadas” por profissionais mais velhos: “É uma dificuldade no Brasil, a cultura das redações que ainda estão muito amarradas ao jornalismo tradicional. O Jornalismo de Dados também pode dar origem a trabalhos impressos muito bons, mas exige uma produção digital, ele é um produto nativo digital, todas as etapas são digitais”.

Daniel Bramatti constata diversas barreiras ao Jornalismo de Dados no Brasil: há uma “dificuldade com mão de obra”, na procura por “pessoas com capacitação para essa área”; as “instâncias governamentais” não estão todas voltadas para a “cultura da transparência e também o momento que o JBD está “se desenvolvendo” coincide com o “momento em que a imprensa brasileira, e talvez a mundial, passa pela sua maior crise”.

Raphael Hernandez aponta a qualidade dos dados com o maior empecilho e cita a existência de formatos ociosos: “Muitas vezes pegamos em dados que são de um sistema que foi usado uma vez, ninguém tem o programa e temos que ficar convertendo”.

Para Marlen Couto a “qualificação dos profissionais” é uma dificuldade: “Temos uma escassez de mão de obra que esteja voltada para esse tipo de profissão, que tenha um perfil para trabalhar com Jornalismo de Dados . Acho que isso ainda é um desafio para o jornalismo brasileiro”. A adoção da “transparência” por parte de todas as instâncias governamentais também é apontada como problemática.

Roberto Maleson acredita que é necessário ter mais “pessoas que trabalhem com dados” e para atingir esse patamar é preciso que haja “mais especialização, mais cursos e eventos” sobre a área. “Falta mais incentivo das redações, as grandes redações já estão começando a ter mais espaço para o Jornalismo de Dados , às vezes é um investimento

grande para produzir dez, cinco conteúdos semanais enquanto as outras editorias estão produzindo muito mais”, afirma.

- Portugal

Em Portugal ainda há uma luta ao nível do acesso aos dados, mas também em relação à qualidade daqueles que são disponibilizados. Esta é uma das principais barreiras que os jornalistas enfrentam ao nível da produção de conteúdos, o que acaba por dificultar o crescimento da área. Em muitas situações os dados até existem e são públicos, mas os formatos, em alguns casos analógicos, em papel, inviabilizam a continuidade do trabalho. Mesmo quando são disponibilizados em formatos digitais, há certos tipos de ficheiros que não são compatíveis com as ferramentas de análise do Jornalismo de Dados, o que requer a criação de uma nova etapa no processo produtivo e, conseqüente, um aumento das horas de trabalho.

Rui Barros considera que a “grande dificuldade” está na falta de bases de dados para trabalhar: “Há muita falta de dados para trabalhar, de dados abertos, porque às vezes até há outros dados, ou estão acumulados em alguns ministérios e é muito difícil conseguí-los ou ia demorar tanto tempo em transformá-los em dados trabalháveis que seria inviável na lógica da redação”. Para o jornalista, a ausência de “pessoas com conhecimentos técnicos” na redação também limita a produção.

As dificuldades referidas por Rita Costa vão ao encontro dos aspetos referidos pelo jornalista da “Rádio Renascença”: Há “falta de acesso à informação” e aquela que se encontra disponível não está preparada “para ser utilizada, consumida por um computador”, devido ao formato. “Ainda há muitas coisas que não são medidas e que não existem em dados e isso cria alguns constrangimentos”, conclui.

Raquel Albuquerque segue a mesma linha quanto à falta de dados e ressalta o formato através do qual são muitas vezes disponibilizados: “Quando pedimos dados às instituições que têm de dar os dados, nunca te dão os dados em bruto, dão os dados trabalhados, as estatísticas trabalhadas e originalmente, o conceito de Jornalismo de Dados, não é trabalhar com dados trabalhados”. A jornalista acrescenta ainda a incompatibilidade do Jornalismo de Dados com a velocidade das redações: “Estamos todos, permanentemente, a correr, e parar para fazer um trabalho com tempo, é algo que tem muito pouco espaço”.

5.9. Quem faz Jornalismo de Dados no Brasil

“Violência barra Correios em 44% dos locais do Rio”, " 'Penduricalhos' do passado rendem a juízes R\$ 211 milhões". Essas frases representam manchetes das capas do dia 19 de fevereiro, respectivamente, do jornal Folha de São Paulo e do O Estado de São Paulo, duas das maiores publicações brasileiras. A presença de trabalhos, destacados nas capas dos jornais, evidencia o crescimento do espaço dado ao jornalismo guiado por bases de dados¹². Como tal, é importante perceber não só como esse jornalismo é feito, mas onde ele se concentra e quem assume essas funções.

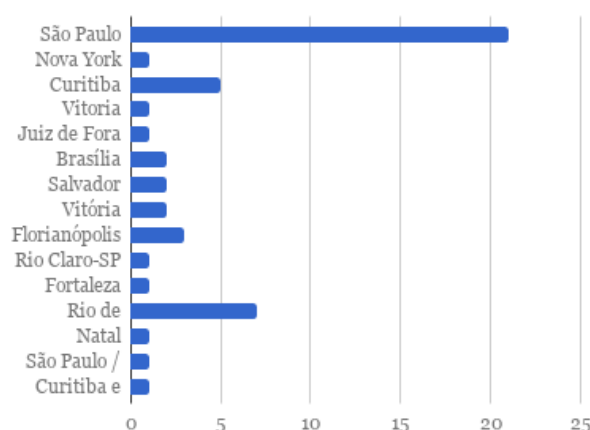


Gráfico 1. Cidade de trabalho dos jornalistas (Brasil)

Os jornais sediados nas grandes cidades trazem consigo a responsabilidade de uma cobertura jornalista descentralizada do local, com foco num alcance generalista. As duas maiores cidades do País, São Paulo e Rio de Janeiro, concentram 56% da mão de obra ativa no trabalho com dados.

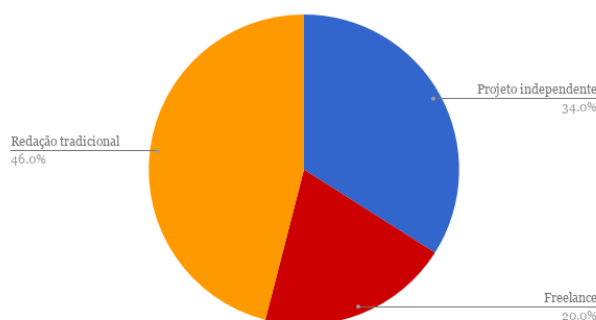


Gráfico 2. Regime editorial onde os jornalistas estão inseridos (Brasil)

¹² Tweet que faz referência às manchetes dos trabalhos originados a partir de técnicas de Jornalismo de Dados: <https://twitter.com/danielbmariani/status/965571032479358976>

A distribuição dos profissionais por tipo de projeto editorial revela a predominância de redações tradicionais, seja através de jornalistas inseridos no quadro da empresa, seja a partir da contratação de profissionais em regime *freelance*. No entanto, o número de jornalistas inseridos em propostas de jornalismo independente revela o peso das publicações autônomas na produção de trabalhos baseados em Jornalismo de Dados . De um total de 34 organizações jornalísticas mencionadas, 15 são iniciativas independentes¹³.

As redações tradicionais e os projetos independentes que foram representados pelos inquiridos contam, em 82% dos casos, com entre zero e quatro profissionais de jornalismo para a produção de peças baseadas em dados. Os jornais “GloboEsporte” e “O Estado de São Paulo” possuem cinco ou mais profissionais aptos para trabalhar com dados. Dos projetos independentes, o “Labcaos”, o “Jota” e a “Gênero e Número” também possuem cinco ou mais jornalistas com habilidades para a produção desse tipo de conteúdo.

O ritmo de produção de peças baseadas em dados, difere dos trabalhos tipo *hard news* que imperam na imprensa tradicional. Contudo, ainda que em minoria, como sugerem as respostas dos inquiridos, há projetos de nicho, redações com núcleos para esse tipo de construções jornalísticas e profissionais *freelance* que se ocupam apenas dessa função: 30% dos profissionais estão dedicados à produção exclusiva de trabalhos com técnicas de Jornalismo de Dados .

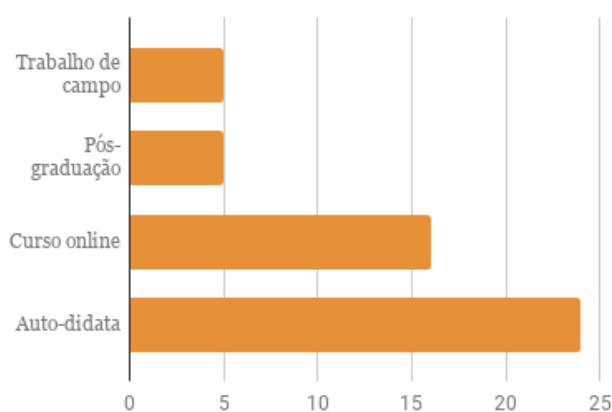


Gráfico 3. Principal método de aprendizagem adotado(Brasil)

¹³ O “Anexo 3” lista os projetos independentes que foram indicados pelos jornalistas como organização “empregadora”.

Da amostra selecionada, 90% dos inquiridos possuem formação acadêmica em Jornalismo. As competências exigidas para este fazer jornalístico que está a emergir, revelam a predominância de métodos de aprendizagem autônomos, como práticas autodidatas e cursos disciplinados em plataformas *online*.

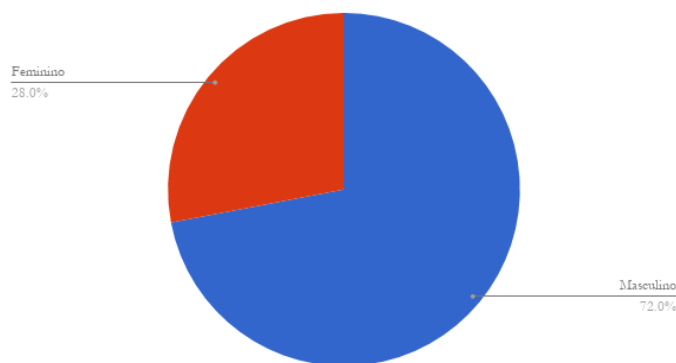


Gráfico 4. Distribuição por sexo (Brasil)

Entre os profissionais que responderam ao questionário percebe-se que os jornalistas do sexo masculino são responsáveis pelo maior fluxo de trabalhos com Jornalismo de Dados. A discrepância evidenciada não corresponde ao recente perfil publicado pelo *Volt Data Lab*, que indica que num conjunto de candidatos com interesse num programa de Bolsa-estágio em Jornalismo de Dados promovido pela organização, 67,8% são mulheres¹⁴.

Já em relação à idade dos profissionais, percebemos que os jornalistas com idade entre os 20 e os 29 anos têm em média cinco anos de experiência no jornalismo e dois anos no Jornalismo de Dados. Já os profissionais com idades compreendidas entre os 30 e os 39 anos, possuem uma média de 11 anos de carreira no jornalismo, com uma média de cinco anos no JBD. Os dois grupos etários, representam 88% do total de inquiridos. A variação entre os anos de carreira e os anos de trabalho com dados é semelhante nos dois grupos etários, o que indica que esta é uma área recente e em desenvolvimento.

¹⁴ Os dados foram coletados a partir da candidatura de 357 pessoas de mais de 100 cidades: <https://medium.com/volt-data-lab/perfil-ddj-b607651cbda>

Tabela 8. Domínio de ferramentas pelos inquiridos brasileiro

FERRAMENTA \ NÍVEL	Não domina	Básico	Intermédio	Avançado	Não respondeu
Excel	-	-	54,0%	46%	-
Open Refine	18%	38%	24%	14%	6%
Tableau	14%	24%	32%	28%	2%
ArcGis	58%	24%	8%	4%	6%

O nível de domínio do *Excel*, que varia entre intermédio e avançado, sugere a importância do *software* de gerenciamento de planilhas na resolução de tarefas intrínsecas ao JBD. O *Tableau*, presente na análise e visualização com finalidade de comunicação, figura como umas das ferramentas de maior importância para os jornalistas, na medida em que apenas 16% dos inquiridos não revelam habilidades no manuseio do aplicativo.

Os inquiridos consideraram ainda outros programas como revelantes e parte das ferramentas que dominam. Entre essas ferramentas destacam-se o *Carto* (*Software* de visualização), indicado seis vezes, *Web Scraper* (*Software* de raspagem de dados), indicado cinco vezes, *Google Sheets* (Planilhas), *DB Browser* (Editor de bases de dados), *Microsoft Power BI* (Visualização), *Infogram* (Visualização), todos indicados três vezes pelos inquiridos.

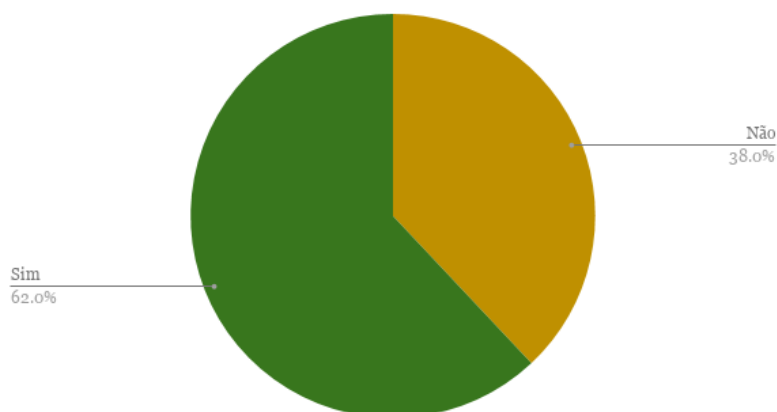


Gráfico 5. Domínio de programação pelos jornalistas (Brasil)

A programação apresenta-se como parte fundamental das rotinas jornalísticas quando o trabalho envolve dados, ou a partir da convergência com outros profissionais, ou a partir

da atuação do próprio jornalista. Mais de metade dos jornalistas inquiridos dominam alguma linguagem de programação: 81% possuem conhecimentos de *Python*, 71% de *SQL*, 52% de *HTML-CSS* e 16% referem ainda que dominam a linguagem *R*.

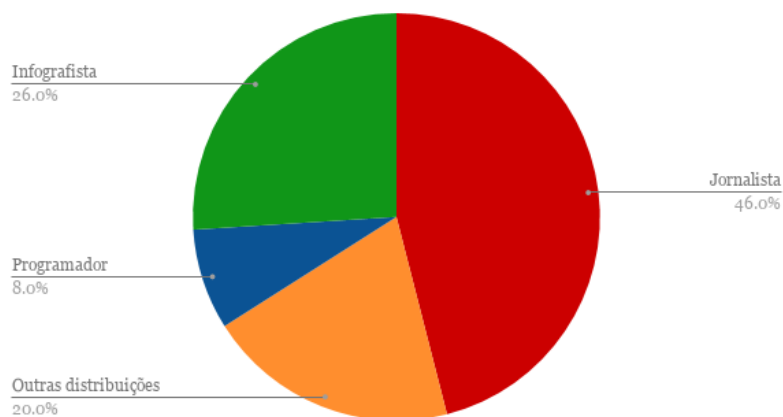


Gráfico 6. Distribuição da responsabilidade de produção de visualização (Brasil)

A visualização, etapa que se revela multidisciplinar, por fugir ao convencional trabalho do jornalista, apresenta-se como prática comum dos jornalistas na etapa de comunicação dos dados. Todos os inquiridos que optaram por outra alternativa de resposta, referiram ou que o trabalho é realizado em conjunto, ou que há condicionantes, como por exemplo, a «complexidade» do trabalho que pode definir se a visualização é criada pelo jornalista ou por outros membros da redação.

5.10. Quem faz Jornalismo de Dados em Portugal

Tal como no Brasil, o início da apreensão do conceito de JBD a partir do surgimento de projetos jornalísticos aconteceu em 2012. A iniciativa REACTION apresentou-se como o primeiro ponto de evidência do Jornalismo de Dados nas redações portuguesas¹⁵. De forma a conhecer as iniciativas que promovem os Jornalismo de Dados atualmente e quem está responsável por elas, a aplicação de questionário também foi feita aos jornalistas que atuam com dados em Portugal.

¹⁵ Foi um projeto de jornalismo computacional promovido pela Fundação para a Ciência e a Tecnologia (FCT) com a colaboração de instituições de ensino e do jornal Público: <http://blogues.publico.pt/publicolab/2012/08/14/a-abrir-caminho-no-jornalismo-computacional/>

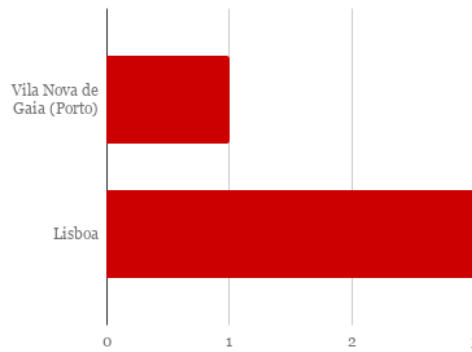


Gráfico 7. Cidade de trabalho dos jornalistas (Portugal)

Em Portugal, a produção de trabalhos com base em dados revela uma concentração nas maiores cidades do País. Lisboa, onde há um número maior de jornais generalistas, figura como a cidade com mais mão de obra na área.

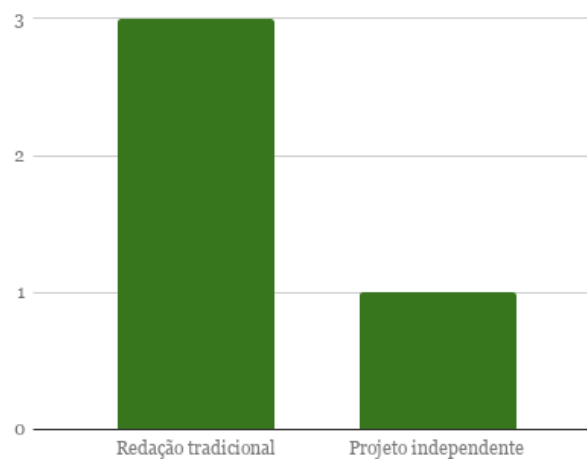


Gráfico 8. Regime editorial onde os jornalistas estão inseridos (Portugal)

É das redações tradicionais que saem a maior parte dos trabalhos de Jornalismo de Dados . Diferente do que acontece no Brasil, Portugal não aparenta estar voltado para projetos autónomos assentes na produção de JBD.

Em três das quatro organizações, há apenas um profissional destacado para a função de produzir peças com dados. Apesar de ser a principal origem dos trabalhos, não há nenhuma redação que possua um profissional dedicado exclusivamente ao Jornalismo de Dados . O inquirido que representa o projeto independente, é o único a trabalhar unicamente na produção desse tipo de conteúdo.

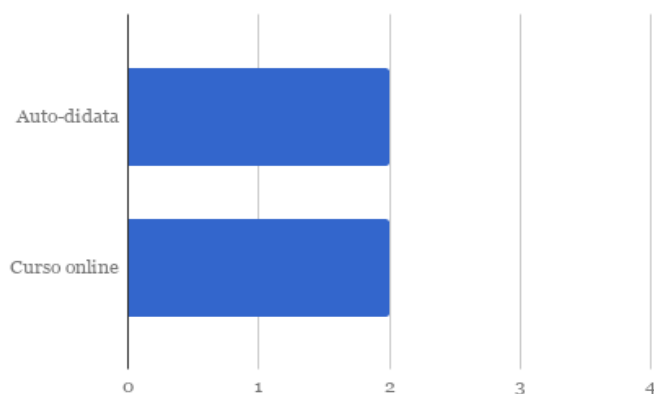


Gráfico 9. Principal método de aprendizagem adotado (Portugal)

Todos os inquiridos são formados em jornalismo e, tal como no Brasil, os jornalistas que fazem usos das técnicas de JBD em Portugal recorrem a meios autónomos para adquirirem conhecimentos que os tornem aptos a manusear grandes quantidades de dados.



Gráfico 10. Distribuição por sexo (Portugal)

Ao nível do género dos jornalistas não se verifica a mesma tendência que no Brasil. Os produtos dos trabalhos são em igual número do género masculino e feminino. Os jornalistas têm uma média de idade de 28 anos, trabalham em média há cinco anos na área e produzem trabalhos com Jornalismo de Dados há dois anos.

Tabela 9. Domínio de ferramentas pelos inquiridos portugueses

NÍVEL FERRAMENTA	Não domina	Básico	Intermédio	Avançado
Excel	-	-	2	2
Open Refine	-	2	-	2
Tableau	-	1	3	-
ArcGis	2	2	-	-

O *Excel* e o *Tableau* surgem, também em Portugal, como as ferramentas que os jornalistas mais dominam, . As respostas dos inquiridos indicam ainda que o nível de domínio dos jornalistas se situa entre o intermédio e o avançado.

Metade do número de respondentes dominam alguma linguagem de programação. Não há nenhuma linguagem comum aos dois profissionais: *JavaScript*, *HTML-CSS*, *R* e *Python* foram as linguagens referidas. Quanto à criação de visualizações, os dois jornalistas que afirmam possuir conhecimentos em programação declaram ser os responsáveis pela produção gráfica dos seus trabalhos. Os outros inquiridos referem que essa etapa do procedimento jornalístico fica a cargo de um infografista.

CONCLUSÃO

Este trabalho teve como objetivo a caracterização do estado do Jornalismo de Dados no Brasil e em Portugal sob o ponto de vista de como é feito, em quais estruturas e quem faz. Ainda que estejam nos primeiros anos de trabalho com dados, o conceito de “estágio inicial” representa realidades distintas nos dois países.

O JBD, feito de forma contemporânea, com a análise de grandes volumes de informação a partir de técnicas computacionais é um conceito novo, tanto no Brasil como em Portugal, e isso é notório quando verificamos que os anos de trabalho com técnicas de Jornalismo de Dados é semelhante na comparação entre profissionais com mais anos de profissão e os jornalistas em início de carreira.

A prática revelou-se mais comum em redações sediadas em grandes cidades, com exemplos de redações com núcleos de dados – o que pressupõe uma produção mais ativa de conteúdos. Nos jornais sediados fora dos grandes centros, no caso brasileiro, percebemos que a produção desse gênero de conteúdo disputa espaço com outros tipos de trabalhos porque demanda tempo dos jornalistas.

O Brasil aparenta possuir um ecossistema de Jornalismo de Dados estabelecido, com entidades que promovem a área a partir de congressos, cursos e fazendo a ponte na criação de *networking*. As redações investem e adotam a prática com a criação de editorias que facilitam o trabalho conjunto, através da compra de equipamentos e *softwares* e estimulando participação em fóruns sobre o JBD, em alguns casos cobrindo os custos associados.

Em Portugal, não é evidente a existência de uma cultura de Jornalismo de Dados, não há a referência de entidades que apoiem o JBD e o investimento dos jornais é maioritariamente na manutenção dos jornalistas que, nos casos estudados, trabalham individualmente na redação, com eventuais participações, mas sem uma organização definida que permita a divisão de tarefas, a trocas de ideias e uma produção mais regular.

Apesar de o Jornalismo de Dados representar uma autonomia perante a necessidade primária de recorrer a fontes humanas e oficiais, ainda é notório a dependência das redações por bases de dados disponibilizadas por órgãos públicos. Como a maioria das redações não possuem disponibilidade de tempo e profissionais para a coleta e criação de bases próprias as pautas são baseadas em informações cedidas por organizações governamentais ou ligadas a estas. No país sul-americano, há exemplos de trabalhos que

seguiram essa metodologia de coleta própria. Outro ponto de destaque em relação as bases de dados como fonte é que, enquanto no Brasil já um entendimento de que outras instâncias governamentais devem disponibilizar dados – o que acabaria por gerar peças locais –, em Portugal os jornalistas ainda sofrem para que haja uma abertura do Governo enquanto autoridade máxima e uma compreensão de que os dados devem seguir formatos legíveis.

O caráter técnico da relação jornalismo-programação, que se tornou mais explícita no JBD, limitar o trabalho jornalístico nos dois países. A ausência de programadores é um empecilho na concretização de tarefas que poderiam quebrar a complexidade de alguns trabalhos e é responsável pela desistência de algumas pautas pela carência de mão de obra capaz de suprir as necessidades de programação. Enquanto que no Brasil algumas redações possuem editorias com programadores, em Portugal esses profissionais não frequentam os ambientes redatoriais. De um modo geral, a necessidade de contar com programação tem se refletido na capacitação dos jornalistas, que demonstraram preocupação em aprender *Python*, umas das linguagens mais utilizadas em trabalhos com dados.

Apesar de o Jornalismo de Dados ser apontado como multidisciplinar, com potencial para convergência de diversos profissionais, os jornalistas entrevistados revelaram ser responsáveis pela totalidade do processo e, em alguns casos, como em projetos que requerem gráficos mais complexo, há a participação de outros profissionais. O que também se verificou, a partir dos questionários, é que há uma tendência para que as visualizações sejam criadas pelo próprio jornalistas, que já dominam linguagens de programação.

Em suma, notamos que o Jornalismo de Dados feito nos dois países segue as mesmas etapas processuais, mas os ambientes redatoriais – com mais profissionais destacados para a função e mais investimento – e o ambiente externo – com políticas de transparências mais próximas à necessidade dos jornalistas e a existência de entidades de apoio – que acabam por favorecer o desenvolvimento da especialização, fazem com que o estado geral se apresente muito mais próspero para as redações brasileiras.

Queremos ainda deixar claro que a conjuntura desta pesquisa, com a ausência de alguns tópicos, pode representar o enfraquecimento dos resultados:

- O objetivo da pesquisa não é uma comparação exaustiva sobre a prática nos dois países, não há uma medida de proporcionalidade que permita isso; mas acreditamos que é possível apresentar o estado atual de cada país e verificar qual está em um processo mais avançado de produção de JBD e o que permite que isso aconteça.

- O número de respostas obtidas no questionário português não permite estabelecer padrões que possibilitem uma compreensão sobre «Quem faz» Jornalismo de Dados em Portugal e tampouco utilizar como elemento comparativo com o questionário aplicado aos jornalistas brasileiros.

- No Brasil, o Nexo Jornal tem se destacado e é mencionado frequentemente como produtor de Jornalismo de Dados . Ao contrário da maioria das publicações brasileiras, o Nexo produz peças com foco na visualização de dados, sem contextualizações adicionais, gerados a partir do trabalho de um cientista de dados e de um jornalista visual. O fato de não termos conseguido expor a forma e as condições da rotina de trabalho com dados na publicação – fundada em 2015 e geradora de produtos jornalísticos digitais – empobrece o conjunto dos resultados, por acreditarmos que deixa de captar uma parte importante da forma como o JBD é trabalhado no Brasil.

Durante a elaboração desta dissertação identificamos também a existência de questões que ainda não foram exploradas no campo do Jornalismo de Dados e acreditamos que a resolução dessas lacunas pode representar um avanço na fundamentação e caracterização do JBD:

- Há Jornalismo de Dados em outros formatos?

Durante a etapa de revisão bibliográfica, notamos a ausência da exploração de referências ao Jornalismo de Dados como um possível gênero multiplataforma. Sendo o jornalismo relevante, independente do modo de disseminação, é um ponto a ser explorado: identificar como é feito o Jornalismo de Dados em outros formatos, qual o seu nível e em que condicionantes difere do Jornalismo de Dados dito convencional, que está a emergir atualmente e que é predominantemente online e por vezes replicado no papel.

A título de exemplo, um caso mais concreto, citado durante esta pesquisa, mostra que na Rádio Renascença – um dos poucos órgãos de comunicação com jornalista de dados em Portugal – há a produção para outra plataforma, muito embora, como referiu Rui Barros,

a sua atuação está mais concentrada na criação de conteúdos para a publicação online e, em algumas oportunidades, há essa adaptação do discurso para o rádio.

- Visualização de dados é jornalismo?

As visualizações de dados podem ser consumidas como um produto isolado num contexto informativo ou ser uma das etapas do Jornalismo de Dados na construção e desconstrução da realidade. No entanto, no atual campo de estudos dos media, não há a definição de critérios ou abordagens que permitam configurar as visualizações de dados como autossuficiente na produção jornalística.

- O Jornalismo de Dados é explorado no contexto local e hiperlocal?

As fontes são como linhas condutoras no jornalismo. No surgimento da Reportagem Assistida por Computador, a potencial independência do jornalista perante a obrigatoriedade de fontes humanas no testemunho dos acontecimentos foi apontada como uma das vantagens da técnica. No entanto, a dependência pela disponibilização pública de dados pode limitar os processos jornalísticos. Através das componentes analisadas notamos que a ausência de dados limita a atuação – haverá entidades que se ocupam da coleta de dados mais segmentada e localizada? É importante perceber se há espaço e sob quais condições o Jornalismo de Dados é feito no contexto local e hiperlocal.

- Qual a compatibilidade entre o *fact-checking* e o JBD?

Num recente caso no Brasil, envolvendo a morte de uma vereadora, a velocidade das redes sociais e a incapacidade de contenção do que é publicado, tornou-se cenário ideal para a disseminação de correntes “pseudoinformativas” que associavam a imagem da vereadora morta à um conjunto de acontecimentos sem bases de sustentação comprovadas, mas que adquiriram proporções alarmantes. Algumas dessas *fake news*, foram combatidas com base no *fact-checking* que, por sua vez, recorreu ao Jornalismo de Dados para contrariar esses conteúdos¹⁶. Como revelado numa das entrevistas que deram suporte a este trabalho, no O Globo, há um núcleo que reúne *fact-checking* e JBD na produção de conteúdos. Torna-se relevante perceber se há padrões que permitem sustentar o vínculo entre as duas áreas e como o emprego do Jornalismo de Dados pode ser utilizado na desconstrução das *fake news*.

¹⁶ Exemplos de *fact-checking* com recurso a análise de dados: <https://aosfatos.org/noticias/nao-marielle-nao-foi-casada-com-marcinho-vp-nao-engravidou-ao-16-e-nao-foi-eleita-pelo-comando-vermelho>

- Como as universidades estão a contribuir para o desenvolvimento do Jornalismo de Dados ?

As universidades são, em alguns casos, os catalisadores que estimulam o contato entre a teoria e a prática. Contudo, as licenciaturas e graduações não foram apontadas, no conjunto das entrevistas, como estrutural no ingresso ao campo do Jornalismo de Dados . O espaço dado ao JBD nas grades curriculares e a forma como é ensinado podem dar a conhecer como o Jornalismo de Dados é explorado, por exemplo em termos de conceitualização e da valorização dada ao domínio da área.

O Jornalismo de Dados apresentou-se como uma área em desenvolvimento que, além de contar com o estímulo de iniciativas dos jornalistas, mostrou-se dependente de ações externas ao trabalho jornalístico. Ainda que a técnica represente uma metodologia nova, percebemos que a cultura de dados já é uma realidade na comunidade jornalística¹⁷. Acreditamos que, para além de trabalhos nas redações, é necessária a criação de conhecimento e a produção académica de artigos que acompanhem a dinâmica da área.

¹⁷ O “Anexo 5” demonstra como a presença da cultura de dados nas redações pode ser vista.

REFERÊNCIAS BIBLIOGRÁFICAS

- Alexandre, I. (2014). *Jornalismo de Dados : o estado da arte nos jornais generalistas diários em Portugal* (Tese de Mestrado, Universidade Nova de Lisboa). Recuperado de <https://run.unl.pt/handle/10362/13615>
- Baack, S. (2011). A new style of news reporting: Wikileaks and data-driven journalism. *Cyborg Subjects*, 1. Recuperado de <https://www.ssoar.info/ssoar/handle/document/40025>
- Baack, S. (2015). Datafication and empowerment: How the open data movement rearticulates notions of democracy, participation, and journalism. *Big Data & Society*, 2 (2). <https://doi.org/10.1177/2053951715594634>
- Bakker, P. (2012). Aggregation, content farms and huffinization: The rise of low-pay and no-pay journalism. *Journalism Practice*, 6(5–6), 627–637. <https://doi.org/10.1080/17512786.2012.667266>
- Barbosa, S. (2008). Modelo JDBD e o ciberjornalismo de quarta geração. In *III Congresso Internacional De Periodismo En La Red*. Recuperado de http://www.facom.ufba.br/jol/pdf/2008_Barbosa_RedUCMx.pdf
- Barlow, M. (2015). *Data Visualization: A New Language for Storytelling*. Sebastopol: O'Reilly Media.
- Bastos, H. (2011). Para uma história do ciberjornalismo em Portugal: das origens às múltiplas plataformas. In *1º Congresso Mundial de Comunicação Ibero-Americana*. Recuperado de <https://repositorio-aberto.up.pt/handle/10216/57427>
- Bertocchi, D. (2013). *Dos Dados aos Formatos - Um modelo teórico para o design do sistema narrativo no jornalismo digital*. (Tese de Doutorado, Universidade de São Paulo). Recuperado de <http://www.teses.usp.br/teses/disponiveis/27/27152/tde-21092015-122011/pt-br.php>
- Bounegru, L. (2012). Jornalismo de Dados em perspectiva. In J. Gray, L. Chambers & L. Bounegru (Eds.), *Manual de Jornalismo de Dados*. Recuperado de <http://datajournalismhandbook.org/pt/index.html>
- Borges-Rey, E. (2017). Towards an epistemology of data journalism in the devolved nations of the United Kingdom: Changes and continuities in materiality, performativity and reflexivity. <https://doi.org/10.1177/1464884917693864>
- Bradshaw, P. (2011, julho 7). The inverted pyramid of data journalism [Web log post]. Recuperado de <https://onlinejournalismblog.com/2011/07/07/the-inverted-pyramid-of-data-journalism/>
- Bradshaw, P. (2012). O que é o Jornalismo de Dados ? In J. Gray, L. Chambers & L. Bounegru (Eds.), *Manual de Jornalismo de Dados*. Recuperado de <http://datajournalismhandbook.org/pt/index.html>

- Bradshaw, P. (2013, setembro 30). Journalism *is* curation: tips on curation tools and techniques [Web log post]. Recuperado de <https://onlinejournalismblog.com/2013/09/30/curation-tools-tips-advice-journalism/>
- Bradshaw, P. (2015). Data Journalism. In L. Zion & D. Craig (Eds.), *Ethics for Digital Journalists: Emerging Best Practices* (pp. 202-219). Abingdon: Routledge.
- Bradshaw, P. (2017). The 'second wave' of data journalism. In J. Mair, R. Keeble, M. Lucero & M. Moore (Eds.), *Data Journalism: Past, Present and Future*. Bury St Edmunds: Abramis.
- Cairo, A. (2012). *The Functional Art: An Introduction to Information Graphics and Visualization*. Berkeley: New Riders
- Coddington, M. (2015). Clarifying Journalism's Quantitative Turn: A typology for evaluating data journalism, computational journalism, and computer-assisted reporting. *Digital Journalism*, 3 (3), 331–348. <https://doi.org/10.1080/21670811.2014.976400>
- Cox, M. (2000). The Development of Computer-Assisted Reporting. *Newspaper Division, Association for Education in Journalism and Mass Communication, Southeast Colloquium*, 2030 (305). Recuperado de <http://com.miami.edu/car/cox00.pdf>
- Crucianelli, S. (2013). ¿Qué es el periodismo de datos? *Cuaderno de periodistas*, 26, 106-124. Recuperado de <http://www.cuadernosdeperiodistas.com/que-es-el-periodismo-de-datos>
- Estadão. (2012). Estadão Dados une transparência e interatividade. Recuperado de: <http://politica.estadao.com.br/noticias/geral,estadao-dados-une-transparencia-e-interatividade,872163>
- Felle, T. (2016). Digital watchdogs? Data reporting and the news media's traditional "fourth estate" function. *Journalism: Theory, Practice & Criticism*, 17 (1), 85–96. <https://doi.org/10.1177/1464884915593246>
- Flores, J., & Salinas, C. (2013). El periodismo de datos como especialización de las organizaciones de noticias en Internet. *Correspondencias & Análisis*, 3, 15–34. https://doi.org/10.5209/rev_HICS.2014.v19.45009
- Gynnild, A. (2014). Journalism innovation leads to innovation journalism: The impact of computational exploration on changing mindsets. *Journalism: Theory, Practice & Criticism*, 15 (6), 713–730. <https://doi.org/10.1177/1464884913486393>
- Haigh, T. (2009). How data got its base: Information storage software in the 1950s and 1960s. *IEEE Annals of the History of Computing*, 31 (4), 6–25. <https://doi.org/10.1109/MAHC.2009.123>

- Hammond, P. (2017). From computer-assisted to data-driven: Journalism and Big Data. *Journalism: Theory, Practice & Criticism*, 18 (4), 408-424. <http://doi.org/10.1177/1464884915620205>
- Harrington, J. (2016). *Relational Database Design and Implementation* (4th ed.). Cambridge USA: Morgan Kaufmann.
- Heravi, B. (2017, dezembro 28). TEDx talk: How Data Journalism is changing the Newsroom [Web log post]. Recuperado de <https://medium.com/@Bahareh/tedx-talk-how-data-journalism-is-changing-the-newsroom-18353e63da88>
- Hewett, J. (2017). Collaborative learning: From CAR to data journalism and Hacks/Hackers. In J. Mair, R. Keeble, M. Lucero & M. Moore (Eds.), *Data Journalism: Past, Present and Future*. Bury St Edmunds: Abramis.
- Howard, A. (2014). The Art and Science of Data-Driven Journalism. Tow Centre for Digital Journalism - a Tow/Knight Report. <https://doi.org/10.1017/CBO9781107415324.004>
- Kayser-Bril, N., Valeeva, A., & Radchenko, I. (2015). Transformation of Communication Processes: Data Journalism. In *Proceedings of the XVI April International Academic Conference On Economic and Social Development* (pp. 414–421). Recuperado de <https://arxiv.org/abs/1605.01956>
- Knight Center. (2015). 2015 foi o ano do Jornalismo de Dados no Brasil, concluem jornalistas em seminário da ANJ e Centro Knight no Google [Web log post]. Recuperado de: <https://knightcenter.utexas.edu/pt-br/blog/00-16497-2015-foi-o-ano-do-jornalismo-de-dados-no-brasil-concluem-jornalistas-em-seminario-da-a>
- Lewis, S. C., & Westlund, O. (2015). Big Data and Journalism: Epistemology, expertise, economics, and ethics. *Digital Journalism*, 3 (3), 447–466. <https://doi.org/10.1080/21670811.2014.976418>
- Lima Junior, W. (2006). Jornalismo inteligente na era do *data mining*. *Revista Líbero*, 9 (18), 119-126. Recuperado de <https://casperlibero.edu.br/wp-content/uploads/2014/05/Jornalismo-inteligente-na-era-do-data-mining.pdf>
- Lima Junior, W. (2011). Jornalismo computacional em função da “Era do Big Data”. *Revista Líbero*, 14 (28). Recuperado de <http://seer.casperlibero.edu.br/index.php/libero/article/view/329>
- Lima Junior, W. (2012). Big Data, Jornalismo Computacional e Data Journalism: estrutura, pensamento e prática profissional na Web de dados. *Estudos Em Comunicação*, 12, 207–222. Recuperado de <http://ec.ubi.pt/ec/12>
- Loosen, W., Reimer, J., & De Silva-Schmidt, F. (2017). Data-driven reporting: An on-going (r)evolution? An analysis of projects nominated for the *Data Journalism Awards* 2013–2016. *Journalism: Theory, Practice & Criticism*. <https://doi.org/10.1177/1464884917735691>

- Lorenz, M. (2012). Por que jornalistas devem usar dados? In J. Gray, L. Chambers & L. Bounegru (Eds.), *Manual de Jornalismo de Dados*. Recuperado de <http://datajournalismhandbook.org/pt/index.html>
- Mancini, L., & Vasconcellos, F. (2016). Jornalismo de Dados : Conceito e categorias. *Revista Fronteiras – estudos midiáticos* 18 (1). Recuperado de <http://revistas.unisinos.br/index.php/fronteiras/article/view/fem.2016.181.07>
- Manovich, L. (2001). *The language of new media*. Cambridge, Mass: MIT Press.
- Martin, F. (2015). The Case for Curatorial Journalism . . . Or, Can You Really Be an Ethical Aggregator? In L. Zion & D. Craig (Eds.), *Ethics for Digital Journalists: Emerging Best Practices* (pp. 87-102). Abingdon: Routledge.
- Martinho, A. (2013). *Jornalismo de Dados : contributo para uma caracterização do estado da arte em Portugal*. (Tese de Mestrado, Instituto Universitário de Lisboa). Recuperado de <https://repositorio.iscte-iul.pt/handle/10071/8329>
- Martinho, A. (2014). Jornalismo de Dados : caracterização e fluxos de trabalho. *Revista Exedra*, 8. Recuperado de <http://www.exedrajournal.com/?p=364>
- McCandless, D. (2010, julho). *The beauty of data visualization* [Ficheiro em vídeo]. Recuperado de https://www.ted.com/talks/david_mccandless_the_beauty_of_data_visualization
- Meyer, P. (1991). *The New Precision Journalism*. Bloomington: Indiana University Press.
- Meyer, P. (2012, outubro 18). Precision Journalism and Narrative Journalism: Toward a Unified Field Theory [Web log post]. Recuperado de <http://niemanreports.org/authors/philip-meyer/>
- Palacios, M. (2014). Memória: Jornalismo, memória e história na era digital. In J. Canavilhas (Ed.), *Webjornalismo: 7 características que marcam a diferença* (pp. 89-110). Covilhã: Livros Labcom. Recuperado de http://www.labcom-ifp.ubi.pt/ficheiros/20141204-201404_webjornalismo_jcanavilhas.pdf
- Parasie, S., & Dagiral, E. (2013). Data-driven journalism and the public good: “Computer-assisted-reporters” and “programmer-journalists” in Chicago. *New Media and Society*, 15 (6), 853–871. <https://doi.org/10.1177/1461444812463345>
- Rogers, S. (2013). *Facts are sacred*. London: Faber and Faber.
- Rogers, S., Schwabish, J., & Bowers, D. (2017). *Data journalism in 2017: The current state and challenges facing the field today*. Recuperado de <https://newslab.withgoogle.com/assets/docs/data-journalism-in-2017.pdf>
- Schudson, M. (2010). Political observatories, databases & news in the emerging ecology of public information. *Daedalus*, 139 (2), 100-109. <https://doi.org/10.1162/daed.2010.139.2.100>

- Silva, D. (2014). Academia y periodismo de datos en Portugal: una relación feliz. In F. Perry & M. Paz (Eds.), *Manual de Periodismo de Datos Iberoamericano*. Recuperado de <http://manual.periodismodedatos.org/dora-santos-silva.php>
- Stavelin, E. (2014). *Computational Journalism: When journalism meets programming* (Tese de Doutorado, University of Bergen). Recuperado de <http://bora.uib.no/handle/1956/7926>
- Stoneman, J. (2015). *Does Open Data Need Journalism?* (Working paper). Oxford: Reuters Institute for the Study of Journalism. Recuperado de <http://reutersinstitute.politics.ox.ac.uk/our-research/does-open-data-need-journalism>
- Stray, J. (2014, setembro 18). The Data Journalist's Eye, An Introduction [Web log post]. Recuperado de <https://towcenter.org/the-data-journalists-eye-an-introduction>
- Susman-Peña, T. (2014). *Understanding Data: Can News Media Rise to the Challenge?* (Working paper). Washington: Center for International Media Assistance. Recuperado de <https://www.internews.org/resource/understanding-data-can-news-media-rise-challenge>
- Träsel, M. (2014). *Entrevistando planilhas: estudo das crenças e do ethos de um grupo de profissionais de jornalismo guiado por dados no Brasil* (Tese de Doutorado, Pontifícia Universidade Católica do Rio Grande do Sul). Recuperado de <http://repositorio.pucrs.br/dspace/handle/10923/6841>
- Träsel, M. (2017). Jornalismo Guiado por Dados: características definidoras e uma proposta de formulação do conceito. In *15º Encontro Nacional de Pesquisadores em Jornalismo ECA/USP*. Recuperado de: <http://sbpjour.org.br/congresso/index.php/sbpjour/sbpjour2017/schedConf/presentations?searchInitial=T&track=>
- Ullman, J. (1988). *Principles of database and knowledge-base systems*. Rockville, Maryland: Computer Science Press.
- Weber, W., & Rall, H. (2012). Data visualization in online journalism and its implications for the production process. In *Proceedings of the International Conference on Information Visualisation* (pp. 349–356). <https://doi.org/10.1109/IV.2012.65>

ANEXOS

Índice de Anexos

Anexo 1 – Guião da entrevista aos jornalistas	82
Anexo 2 – Entrevistas	83
Entrevista com o jornalista Cristian Weiss, do jornal Diário Catarinense	
Entrevista com o jornalista Daniel Bramatti, do jornal O Estado de São Paulo	
Entrevista com o jornalista Juan Torres, do jornal Correio	
Entrevista com a jornalista Katia Brembatti, do jornal Gazeta do Povo	
Entrevista com o jornalista Marlen Couto, do jornal O Globo	
Entrevista com o jornalista Raphael Hernandez, do jornal Folha de São Paulo	
Entrevista com a jornalista Raquel Albuquerque, do jornal Expresso	
Entrevista com a jornalista Rita Costa, do jornal Público	
Entrevista com o jornalista Rui Barros, da Rádio Renascença	
Anexo 3 – Questionários enviados aos jornalistas, brasileiros e portugueses, que atuam na área do Jornalismo de Dados	136
Anexo 4 – Projetos independentes	142
Anexo 5 – Sinais que indicam a emancipação de uma cultura própria	144

Anexo 1 – Guião da entrevista

1. Como começou a trabalhar com Jornalismo de Dados ?
2. Quais as características que um jornalista deve ter para trabalhar com dados?
3. Pode descrever o seu processo de produção de peças baseadas em dados?
4. Há uma estrutura pré-estabelecida para comunicar os dados?
5. É comum referir que no Jornalismo de Dados deixa-se de entrevistar pessoas e passa-se a entrevistar *datasets*. Considera que as bases de dados têm autoridade suficiente para serem a única fonte de informação de uma notícia?
6. Ao fazer a busca pelos dados, quais são as fontes que mais recorre na elaboração dos seus trabalhos?
7. Com que frequência há a criação de bases de dados na própria redação?
8. Como surge uma ideia de pauta no Jornalismo Guiados por Dados?
9. A seleção do que deve ou não ser publicado é intrínseca à linha editorial do jornal. Considera que há valores notícia no Jornalismo de Dados ?
10. Há trabalhos que envolvem programadores, designers, cientistas de dados. Como é a distribuição de tarefas na produção de trabalhos que envolvam dados na redação?
11. De que forma a programação faz parte da sua rotina de trabalho?
12. De que forma o jornal está investindo nessa especialização do jornalismo?
13. Qual é a maior dificuldade, no Brasil (ou Portugal consoante o entrevistado), para a produção de trabalhos jornalísticos guiados por dados?
14. Considera que há informação suficiente disponível?
15. Como imagina o auge do Jornalismo de Dados no Brasil (ou Portugal consoante o entrevistado)?

Anexo 2 – Entrevistas

Entrevista com o jornalista Cristian Weiss, do jornal Diário Catarinense (Florianópolis, Santa Catarina)¹⁸

Como começou a trabalhar com Jornalismo de Dados ?

Sempre falamos que às vezes fazíamos Jornalismo de Dados sem saber que estávamos fazendo Jornalismo de Dados . Foi em 2009, de facto, que eu concluí a minha graduação e comecei a trabalhar no Jornal de Santa Catarina, em Blumenau, que também é do nosso grupo. Lá fazíamos muitos trabalhos no dia a dia que muitas vezes tínhamos que coletar dados porque o poder público não tinha dado para aquela situação. Nem sabíamos que era Jornalismo de Dados , como conceito isso ainda não era tão difundido. Eu fiz várias formações, muitas formações por conta própria, em vários casos pedimos auxílio da empresa, mas nem sempre a empresa consegue nos dar auxílio para custear essas formações. Duas grandes fontes de formação, públicas e acessíveis, são a Abraji e o *Knight Center* da Universidade do Texas, que promovem diversos cursos voltados para o Jornalismo de Dados . Esses foram os primeiros que eu fiz para ter essa base inicial, depois eu fiz vários outros, muitos deles online, para me especializar em programação *Python* ou *SQL*, que são ferramentas que a gente utiliza, o *Tableau* que é a ferramenta de visualização de dados, até o próprio *Excel*, porque até hoje eu não abro mão do *Excel*, por mais que seja uma ferramenta mais simples, mas ela nos dá agilidade, como ele é mais intuitivo, é mais rápido de fazer.

Há uma estrutura pré-estabelecida para comunicar os dados?

Não, é muito incipiente. Felizmente nós temos duas pessoas que trabalham com a gente na equipa de *design*, que são duas pessoas muito interessadas em Jornalismo de Dados . As duas são muito interessadas e têm tentado mobilizar a equipa para que outros também se interessem em trabalhar com isso. Vemos até uma dificuldade do pessoal de *design* em trabalhar com ferramentas mais simples como o *Tableau*, por exemplo, que já é uma ferramenta já está disponível há mais de cinco anos para os jornalistas. Muitas vezes têm dificuldade em lidar com essas coisas, em entender quando o jornalista chega e apresenta o projeto que ele já conseguiu coletar e o que nós precisamos informar. Eu senti bastante dificuldade em alguns projetos, de que mesmo eu tendo o trabalho, que eu considero que não seria meu trabalho, ter que refinar tão detalhadamente o dado que eu vou passar para o *design*. Eu acho que nós temos que passar os nossos achados e eles trabalharem connosco para conseguirem uma melhor formatação que vai ajudá-los na exibição. Às vezes eu vejo que o repórter tem que fazer um trabalho além daquele que deveria ser a sua alçada, para facilitar a compreensão do colega da arte porque ele ainda tem dificuldade de entender essa parte da visualização de dados, de escolher o gráfico ou a representação gráfica correta para aquela história que está sendo contada ou saber priorizar, saber fazer a hierarquia dos dados. Nós fazemos uma hierarquia e eles acabam entendendo que um outro dado visualmente fica mais bonito, mas para nós não serve só

¹⁸ Entrevista realizada por chamada de voz, no dia 30 de novembro de 2017

o visualmente mais bonito, mas também a hierarquia dos dados que vai se desdobrando em detalhes.

É comum referir que no Jornalismo de Dados deixa-se de entrevistar pessoas e passa-se a entrevistar datasets. Considera que as bases de dados têm autoridade suficiente para serem a única fonte de informação de uma notícia?

Primeiro, eu acho que depende do dado, depende da fonte desse dado. Por exemplo, eu acredito que o censo escolar do INEP por ser a maior autoridade de estatística educacional do país é um dado inquestionável. Se uma secretária da educação nos diz "aquele dado está errado", é culpa da secretária de educação, porque ela é obrigada a comunicar esses dados para o INEP. Se ela tiver que questionar, ela não tem que questionar o jornalista, ela tem que questionar o INEP. Esse é um dado censitário, é um dado que de facto abrange quase a totalidade da amostra. Quando trabalhamos com outras pesquisas que são por amostragem, eu acho que os dados por si só não têm autoridade nenhuma. Mas mesmo tratando-se do censo, nós precisamos buscar as histórias na rua para ver se isso se reflete ou quando isso se reflete e também buscar análise de especialistas que já estudam esses tipos dados. Isto se houver esses estudos, muitas vezes os dados são novos e ainda ninguém teve tempo para se debruçar sobre eles ou para pelos menos nos indicar se o caminho que a gente está percorrendo por meio desses dados estão corretos. Eu acho que não pode abrir mão de fonte externas de jeito nenhum.

Ao fazer a busca pelos dados, quais são as fontes que mais recorre na elaboração dos seus trabalhos?

Como eu tenho feito nos últimos dois anos mais trabalhos na área da educação, o INEP é, para mim, a fonte primária. O INEP tem dados estatísticos do ensino fundamental, do ensino médio e do ensino superior, só não tem do ensino infantil ainda. Temos também o instituto "Ayrton Senna - Todos pela educação". Há outros institutos sérios que também fazem análises de dados que facilitam o nosso trabalho. Na área da saúde, que também tenho feito algumas coisas nesse sentido, nós observamos o DataSus, que é, talvez, a primeira grande base de dados pública aberta no Brasil, ela começou a ser aberta em 1997. Temos um histórico tremendo de informações muito detalhadas sobre doenças, causas de mortes, patologias, tratamentos. Essas grandes bases de dados para mim são muito confiáveis, podem sempre ser o nosso ponto de partida. E, de preferência, trabalhar com os microdados, que é o dado na sua forma mais bruta. Ali vamos ter a liberdade de fazer os cruzamentos e achar as respostas que queremos. Muitas vezes o poder público só disponibiliza as sínteses dos dados e os microdados não ficam públicos, aí precisamos recorrer, às vezes, à Lei de Acesso à Informação para obter esses dados mais detalhados. Na maioria das vezes temos trabalhado mais com dados públicos do que instituições privadas, mas isso vai muito da especialização do jornalista, da editoria que ele trabalha. No meu caso, por trabalhar numa editoria que faz mais reportagens na área da saúde e educação, por exemplo, recorreremos a bases públicas, que muitas vezes além de ter dados da educação pública também recorre a dados da educação privada. Normalmente partimos da base pública.

Com que frequência há a criação de bases de dados na própria redação?

Isso é algo que eu estou tentando estimular os colegas a fazerem. Quando nós fazíamos isso era no Jornal de Santa Catarina, uma redação um pouco menor, no entanto, víamos um interesse da redação em fazer esse trabalho. Aqui a cultura é um pouco diferente, no Diário Catarinense vemos a dificuldade muitas vezes do próprio repórter entender que se um dado público não está disponível, porque não tentarmos pelo menos recolher uma amostragem para mostrarmos uma realidade? Aqui a frequência é quase zero.

Como surge uma ideia de pauta no Jornalismo Guiados por Dados?

A ideia de pauta ou a gente fica sabendo de uma história que uma fonte, um entrevistado ou um colega nos comenta, isso desperta a curiosidade "onde é que eu busco dados que comprove isso", ou é na nossa própria curiosidade do dia a dia. Quando recebemos *releases* de assessorias que falam de um dado e ficamos curiosos e pegamos a base de dados só para olhar. Ou quando fazendo outra pauta entramos em um site de um órgão público ou de uma instituição de pesquisa e descobrimos que tem um dado disponível e fazemos *download* para olhar, aí descobrimos que tem um dado forte que pode render uma nova reportagem. São dois caminhos distintos, mas que no final, no processo de execução, eles se encontram. Às vezes a história aparece antes dos dados e outras vezes o dado aparece antes da história. O que eu acho importante é que o jornalismo não pode ser só dados, ele precisa ter história. Quando o dado aparece antes da história a gente precisa buscar a história depois.

A seleção do que deve ou não ser publicado é intrínseca à linha editorial do jornal. Você considera que há valores notícia no Jornalismo de Dados ?

O repertório do jornalista e a experiência que ele tem é preponderante para quando ele olhar uma base de dados, ele entender mais rapidamente o que de facto tem mais valor jornalístico naquela informação. Isso é muito subjetivo. Às vezes um colega observa aquilo como mais relevante e o outro observa outra coisa como mais relevante. O repertório de um jornalista faz olhar de um jeito e o repertório do outro jornalista faz olhar de outro, isso pode ter resultados no produto final. O repertório do jornalista é preponderante para o primeiro passo, antes mesmo de passar pelo crivo editorial, primeiro passa pelo crivo desse repórter que é a origem do processo.

Como o Jornalismo de Dados está organizado no Diário Catarinense?

Nós estamos trabalhando já há um tempo, tentando convencer a chefia de redação que é importante unirmos não só os jornalistas do Diário Catarinense, mas os de outros veículos da nossa rede. Nós temos trabalhando no mesmo prédio, três rádios, a TV, o site G1 e vemos que as vezes há dificuldade ainda de fazer esse trabalho de integração das redações,

de maneira a dividir tarefas, fortalecer o resultado. Muitas vezes o jornalista do G1 está fazendo a mesma coisa que o repórter do Diário Catarinense. Se unirmos essas, principalmente no caso dos dados, que sempre demanda muitas horas de trabalho e muito esforço, conseguimos fazer mais cruzamentos e talvez consigamos resultados melhores para aquilo que estamos investigando por meio dos dados. E hoje não temos essa estruturação ainda, pelo que eu conheço, ainda é muito comum aqui nos jornais do sul do país. Não temos uma equipa estruturada, essa equipa trabalha com dados, eles são especialistas em dados, eles não só vão produzir as reportagens, mas também receber as demandas. O pessoal chama de "lobo solitário", é um jornalista de um lado da redação trabalhando com dados, outro de outro lado da redação trabalhando com dados, cada um fica trabalhando em seu projeto; quando tem alguma coisa leva para o editor, o editor chama o pessoal do design e aí que leva adiante. Não tem aquele trabalho integrado que acho que seria necessário desde o começo para fazer um trabalho que tivesse uma visualização melhor depois, uma entrega melhor para a audiência. Eu acho que tem que ter uma equipa integrada, que envolvesse um integrante do *design*, um editor, um repórter de dados. Ter um programador, porque precisamos muitas vezes de soluções por via da programação. Temos uma redação de mais ou menos 70 pessoas hoje e vamos dizer que cinco pessoas têm mais habilidade para trabalhar com Jornalismo de Dados e já fazem trabalhos nesse sentido.

Há trabalhos que envolvem programadores, designers, cientistas de dados. Como é a distribuição de tarefas na produção de trabalhos que envolvam dados Diário Catarinense?

Hoje não dispomos de um programador, é um pleito nosso, precisamos de um programador para soluções digitais. Temos diversas situações em que precisamos de soluções digitais e não temos e às vezes precisamos terceirizar, fazer uma contratação externa. Para esses trabalhos de dados, a raspagem dos dados é uma tarefa que precisa de alguém especialista em programação, por exemplo, e mesmo a gente tendo uma noção básica de algumas linguagens de programação, não é suficiente. Às vezes quando deparamos com um site que tem trava para extração automática, para um robô, ali precisamos de alguém que tenha conhecimentos, aí que vemos que temos um bloqueio no nosso processo, não termos um programador. Também não temos a figura de um cientista de dados na redação, até seria ótimo, mas isso é um luxo para as redações brasileiras. O que temos são os jornalistas que foram se especializando e tentando ter mais noções de estatística, de matemática, de programação para poder fazer esse trabalho. Infelizmente não contamos, dentro da equipa, com uma pessoa que tenha esse conhecimento tão avançado de análise de dados. O que procuramos fazer no nosso processo, após os nossos achados, é submeter a algum especialista, seja da universidade, seja alguma fonte nossa, para tentar comprovar, tirar a prova real daqueles nossos achados. É uma coisa que eu estou tentando estimular, para que a nossa equipa trabalhe um pouco diferente. Ainda bem que essas colegas do *design* demonstraram um interesse com dados e é colocá-las conosco no início do processo, porque isso não ocorre hoje, elas entram no trabalho no final do processo, o jornalista já fez toda a raspagem, já levantou os dados que precisa, já fez os cruzamentos e ele entrega para o *design*. Eu acho que é importante ter essa visão do designer quando estamos no início do percurso da

pauta. Eles podem nos trazer problemas e soluções que não pensamos que nós, focados na pauta ou na história, não demos conta que podia buscar um outro dado, fazer um outro cruzamento ou exibir aqueles dados de forma diferente. Muitas vezes o material chega na equipa de *design* e eles acham que o dado tinha que estar formatado de outro jeito e às vezes isso tem que voltar para o jornalista refazer o cruzamento, para que você possa organizar os dados de uma forma mais palatável, que seja mais digerível pelo leitor. Nós estamos tentando trabalhar nesse sentido, incluir a equipa de design no início do processo da pauta, acho que tem muito a contribuir.

De que forma a programação faz parte da sua rotina de trabalho?

Normalmente só vamos recorrer a essas linguagens quando já identificamos que ali pode haver uma história, até porque isso muitas vezes nos toma muito tempo, nós vamos criar um robô *Python*, precisamos saber que dado queremos obter com ele à hora que formos criar, porque senão o script vai ser totalmente diferente, depende do resultado que a gente quer obter. Por mais que já tenhamos feito vários trabalhos, somos jornalistas não programadores, não temos total fluência nessas ferramentas. Trabalhamos com várias ferramentas ao mesmo tempo. Abrimos a grande base de dados do *SQL*, extraímos só o dado que precisamos, exportamos em *CSV*, abrimos no *Excel*; se não for um dado pesado, lá dentro do *Excel* dá para fazer tabelas dinâmicas, exportamos esse resultado para o *Tableau*, por exemplo, para visualizar melhor, ainda não é visualização para publicação, mas visualização para o jornalista compreender o dado. Se formos usar o *Python* na extração de dados, primeiro devemos aprender o código que precisamos, tem o *GitHub*, que é uma rede social de programação onde encontramos muitas ferramentas prontas, muitas vezes para a nossa necessidade só precisamos adaptar alguns *scripts*. Aí, ao extrair esses dados brutos utilizamos outras ferramentas, o *R*, para poder organizar o dado e prosseguir no restante do método. Muitas vezes essa rotina da programação ela é adaptada para cada história, porque depende do dado que queremos obter. Se ele não está disponível para download, preciso fazer a extração desse dado numa página da internet. Basicamente, dá para dizer que em cada caso precisamos aprender uma metodologia diferente para extrair informação que a gente precisa.

De que forma o Diário Catarinense está investindo nessa especialização do jornalismo?

Os jornalistas que trabalham um pouco com dados, inclusive eu, estão tentando convencer a chefia da redação que é importante que nos especializemos nisso. Cada vez mais isso nos é demandado, justamente depois da aprovação da Lei de Acesso à Informação que começa a ter profusão de dados. Para a maioria das pessoas isso não significa nada porque se entra na base de dados e ela é praticamente ilegível. Então, precisa de alguém mais especializado para fazer essa tradução e vemos que ainda há uma certa resistência da nossa chefia de redação, e de outros jornais, de entender que é importante capacitar uma equipa de dados porque daqui a um tempo fará parte da nossa rotina trabalhar com matérias que envolvam dados. É importante ter essa fluência de leitura e análise de dados e ainda não vemos com clareza um interesse da chefia de redação por exemplo em investir nisso, neste momento. Até porque o nosso grupo de comunicação teve muitas transformações no ano passado, mudou de proprietário, de prédio de redação, agora temos uma nova configuração dos jornais da Rede, uma reorganização dos profissionais. Acaba

atrapalhando isso que vínhamos construindo, porque deixa ser prioridade, para priorizar as mudanças.

Qual é a maior dificuldade, no Brasil, para a produção de trabalhos jornalísticos guiados por dados?

Tem dois lados. Como a gente depende muito de bases de dados públicas, é entender que a Lei de Acesso à Informação trouxe grandes avanços para nós, de forma a gerar uma certa obrigatoriedade do poder em fornecer informação, mas tem um problema que observamos que é o seguinte: eu faço pergunta “A”, “B” e “C”, eles respondem no prazo, mas às vezes eles só respondem à pergunta “A”, nesse caso eu sou obrigado a fazer uma réplica questionando de novo o “B” e o “C”. Então perdemos muito tempo recorrendo aos processos para poder obter a informação pública que às vezes não está publicada ainda. Esse eu acho que é um problema que ocorre no Brasil inteiro, o poder público não está cumprindo integralmente com a Lei de Acesso à Informação e entender que quando a gente solicita um dado, aquela solicitação vai ter de ser atendida na sua integralidade, naquele prazo, não só aquilo que ele quer responder. E o outro lado que eu acho que tem dificultado a expansão do desenvolvimento do Jornalismo de Dados no Brasil é, como eu havia citado, as chefias de redação compreenderem que isso é algo inevitável e é importante da temos nas nossas redações e, também, o interesse dos próprios jornalistas em aprenderem ferramentas e a trabalhar com Jornalismo de Dados. Eu acho que tem essas coisas: a falta de formação, a falta de interesse e também do lado das fontes de dados, desrespeito que ainda ocorre com a Lei de Acesso à Informação.

Você considera que há informação suficiente disponível?

Não, eu acho que ainda é possível ter muito mais, principalmente no serviço público. O poder público não tem obrigação de publicar nada, agora com a Lei de Acesso à Informação ele tem a obrigação de te responder se você solicitar, mas se não solicitar, ele pode não colocar nada no site, de informação estatística. Eu acho que ainda é um avanço que precisa ser feito. Quando descemos a hierarquia da administração pública, descemos do Governo Federal para os Governos Estaduais e Municipais, vemos que o compromisso de divulgar informação é cada vez menor. Para atender a Lei de Acesso à Informação criam seus portais de transparência e colocam as informações de modo tão espalhado que é difícil obter a informação, é uma forma de mascarar, de dizer que respeita a lei, mas disponibiliza de uma forma ilegível.

Como imagina o auge do Jornalismo de Dados no Brasil?

O auge eu acho que vai demorar bastante tempo para acontecer, no mínimo uns cinco anos. Quando essa geração de agora estiver no mercado de trabalho é que a gente vai ver resultados melhores, é uma geração que já entra com uma consciência a respeito do trabalho de dados, talvez até já com uma formação, com conhecimento de como lidar com isso na prática muito mais avançado do que temos hoje nas redações. Esse auge imagino que vai ser extremamente positivo porque vamos ter uma coisa que até hoje não teve no Brasil, que é deixar de ser dependente da retórica de governantes públicos, vamos provar

e desconstruir retóricas que muitas vezes são mentirosas, eleitoreiras, com base em informação extremamente objetiva, que são os dados, sem usar a subjetividade que é discurso. Acho que vamos poder construir avanços muito importantes no Brasil, em relação à democracia, em relação ao processo eleitoral. Começaremos a cobrar por meio de dados para que as autoridades se mobilizem para fazer o seu papel.

Entrevista com o jornalista Daniel Bramatti, do jornal O Estado de São Paulo (São Paulo)¹⁹

Como começou a trabalhar com Jornalismo de Dados ?

O Estadão Dados começou em 2012, mas mesmo antes disso eu já procurava fazer matérias baseadas em dados por um interesse próprio e também por uma circunstância profissional. Fui repórter de política em Brasília, trabalhei muitos anos lá, mas em 2000 vim para São Paulo. A partir daí eu trabalhei só como editor ou subeditor e eu perdi o contato com a reportagem. Nove anos depois de eu ter deixado a reportagem, por uma circunstância aqui do jornal, eu voltei a ser repórter. Nesse momento eu já não tinha nenhum contato com as minhas fontes, já tinham deixado de ser deputados e entre outros, já tinha perdido meu caderno. Eu comecei a fazer matéria de política, não entrevistando pessoas, mas entrevistando bases de dados, olhando dados eleitorais, dados sobre as finanças dos partidos, dados sobre o desempenho nos estados, nos municípios. Foi um filão que eu fui explorando e isso acabou, naturalmente, me colocando no Estadão Dados quando a ideia foi apresentada para a direção. O que eu aprendi, aprendi fazendo ou buscando formações específicas em algumas áreas.

Quais as características que um jornalista deve ter para trabalhar com dados?

Basicamente, as características de um bom jornalista são as mesmas: saber o que é notícia, saber extrair informação, separar informação do ruído, procurar ser didático, esclarecer os seus leitores. O jornalista de dados tem que ter uma formação adicional, que é saber mexer com grandes bases de dados para saber extrair dali informações relevantes. Quando eu digo saber mexer, é saber extrair, saber limpar, saber analisar, transformar e até visualizar esses dados. Numa reportagem de dados a gente tenta fazer todas essas etapas, quando possível até à visualização.

Pode descrever o seu processo de produção de peças baseadas em dados?

Acho que 90% do nosso tempo é aplicado na limpeza e análise dos dados. Nem sempre os dados vão vir do jeito que você espera, em que você vai simplesmente abrir uma planilha e a notícia vai saltar aos seus olhos. Você tem que tirar o que não interessa, você tem que fazer contas muitas vezes, você tem que cruzar bases de dados diferentes e esse processo é trabalhoso, demorado. Temos que ter muito cuidado porque qualquer erro nesse momento vai levar a conclusões equivocadas depois. Em termos de tempo, o processo é mais demorado nessa etapa. Olhando de maneira mais ampla como é o nosso trabalho, procuramos trabalhar um pouco como cientista, formular uma hipótese diante

¹⁹ Entrevista realizada por chamada de voz, no dia 11 de dezembro de 2017

de uma realidade; a minha hipótese diante de tal fenómeno é que ele é causado por tal coisa ou que ele está correlacionado a tal coisa, aí batemos os dados e ‘checamos’ essa hipótese. Se ela se confirma, você pode ter uma matéria, se ela não se confirma você pode ter uma matéria que era o inverso do que você esperava, mas geralmente, em termos jornalísticos, você pode aproveitar das duas formas.

Há uma estrutura pré-estabelecida para comunicar os dados?

Ou publicamos reportagens em texto, em geral são acompanhados por infografias, no caso do jornal são estáticas, mas também quando possível a gente tenta publicar na internet infográficos interativos e outra maneira de trabalhar com dados, em vez de você produzir uma reportagem, constrói uma ferramenta e coloca essa ferramenta à disposição do usuário, para que ele possa descobrir os fenómenos e para que ele possa tirar suas próprias conclusões. Exemplos são ferramentas que mostram como os candidatos financiam suas campanhas, então você pega toda aquela montanha de dados no Tribunal Superior Eleitoral, de como são as doações, quem doa, quanto doa, se são empresas, se são pessoas físicas, é uma ferramenta que pode ser consultada de maneira visual na internet, isso é uma coisa que a gente fez na eleição presidencial passada.

É comum referir que no Jornalismo de Dados deixa-se de entrevistar pessoas e passa-se a entrevistar datasets. Considera que as bases de dados têm autoridade suficiente para serem a única fonte de informação de uma notícia?

Se você tiver confiança na qualidade dos dados, eu acho que é possível usar ela, se não como a única, mas como a principal fonte de informação. Se os dados estiverem bem coletados, bem organizados, é possível você fazer esse papel de entrevistador dos dados, sempre levando em conta que, da mesma forma que uma entrevista tradicional, a qualidade do seu trabalho vai depender da qualidade das suas perguntas. Da mesma maneira com um *dataset*, você pode olhar para a informação menos relevante e mergulhar nela quando o *lead* pode estar logo ali do lado.

Ao fazer a busca pelos dados, quais são as fontes que mais recorre na elaboração dos seus trabalhos?

Geralmente a gente usa dados oficiais, publicados pelo Governo, quase sempre pelo Governo Federal, em algumas instâncias a gente usa dados de Governos Estaduais e Câmaras Municipais. Tem dados que a gente usa que são do organismo oficial de estatística no Brasil que é o IBGE, que faz os censos, que publica os dados sobre inflação, sobre preços, população e algumas entidades que a gente sabe que fazem um trabalho sério, que apresentam dados que não têm o carimbo oficial, não são do Governo, mas a gente sabe que são dados bem estruturados. Por exemplo, tem uma entidade no Brasil que faz o IDH municipal, o Índice de Desenvolvimento Humano dos Municípios, a gente sabe que é uma metodologia reconhecida, então a gente usa isso sem nenhum problema.

Com que frequência há a criação de bases de dados na própria redação?

Acho que não criamos bases de dados do nada. Podemos criar bases que não estão publicadas, cruzando informações de fontes diferentes, isso fazemos. É muito difícil produzirmos uma base de dados a partir do nosso trabalho jornalístico, até porque nós não temos essa capacidade de coleta tão grande. O que a gente pode produzir são bases de dados a respeito de pesquisas eleitorais, pode coletar para uso de reportagens, não chega a ser uma base produzida pela gente, mas ela não existe fora daqui.

Como surge uma ideia de pauta no Jornalismo Guiados por Dados?

É um pouco aquele processo: você faz perguntas e procura responder essas perguntas usando dados. Uma das matérias que a gente está produzindo, o Brasil tem há alguns anos, ações afirmativas nas Universidades, quotas para alunos que sejam negros, pardos ou que sejam oriundos de escolas públicas, a gente se pergunta «Até que ponto esses programas são de facto transformadores?»; «Os alunos que entram por esse sistema de quotas eles estão entrando igualmente nos cursos mais procurados?». São perguntas que fazemos e isso pode ser numa reunião de pauta, pode ser uma elaboração conjunta ou pode ser um repórter que tem uma ideia e a ideia não precisa necessariamente sair aqui do Núcleo de Dados, a gente pode trabalhar com a redação, alguém que tem uma proposta de pauta que envolva dados, aí eles nos apresentam e a gente procura responder essas perguntas.

A seleção do que deve ou não ser publicado é intrínseca à linha editorial do jornal. Considera que há valores notícia no Jornalismo de Dados ?

Acho que isso acontece. Provavelmente alguns tipos de reportagens vão ter uma aceitação melhor, nas editorias, na direção do jornal, mas eu procuro fazer o meu trabalho sem pensar muito nisso, sem pensar 'será que a conclusão que eu vou a que vou chegar está alinhada ou contraria a linha editorial do jornal?'. Eu prefiro não pensar nisso, porque se não a gente pode ficar travado e não olhar a realidade como ela é.

Pode explicar como Jornalismo de Dados está organizado no Estadão?

Nesse momento nós somos só duas pessoas. Quando ele foi criado, começamos com três pessoas que eram fixas e mais uns colaboradores da Infografia. Já chegamos a ter cinco pessoas, das quais dois eram programadores, acho que é uma característica importante do Jornalismo de Dados , contar com programadores. Perdemos profissionais para outras empresas, não necessariamente para outras empresas jornalísticas, mas eles foram contratados por outras empresas que atuam na área de dados. Estamos tentando recompor a equipa nesse momento, mas é um momento de crise no Brasil e isso tudo aliado com a tradicional crise dos media, temos poucos recursos. Esses profissionais que a gente procura geralmente são mais caros, a nossa expectativa é termos três pessoas, sendo um programador, mas neste momento somos duas. O Estadão Dados foi criado em 2012 por iniciativa do José Roberto Toledo que é um pioneiro do uso das técnicas de reportagem com auxílio ao computador no Brasil. Ele participou num evento nos Estados Unidos, onde ele viu pela primeira vez essa integração entre programadores e jornalistas numa

redação, apresentou à direção do jornal a ideia e foi bem recebida. A partir daí, pela primeira vez no Brasil e dentro do Estadão também, fez isso de trazer programadores para trabalhar com jornalismo. Como a minha origem era da Editoria de Política, a gente durante muito tempo trabalhou muito voltado para cobertura de política, por causa da nossa formação e também porque os dados de política no Brasil, sobre eleitorado, sobre financiamento de partido, eram uma fonte muito importante de informação que ninguém tinha dado muita atenção a eles. Inclusive dados de padrões de votação não só dos eleitores, mas também padrões de votação dos parlamentares no congresso, como os partidos e os indivíduos que têm mandato se comportavam no senado, na Câmara, em relação a orientação de voto Governo, nós inclusive criamos uma ferramenta chamada “Basômetro”. A partir dessa intersecção com a política, fomos abrindo um leque e começamos a trabalhar com outras editorias, então tem trabalhos de dados com educação, com saúde, não tanto com a editoria de economia, até porque eles já têm um trabalho específico de dados. Hoje temos parcerias praticamente com o jornal inteiro, geralmente quem propõe as pautas somos nós, mas também estamos abertos a receber sugestões de pessoas de fora da nossa equipa.

Há trabalhos que envolvem programadores, designers, cientistas de dados. Como é distribuição de tarefas na produção de trabalhos que envolvam dados no Estadão Dados?

Neste momento, como nós não temos programadores e não temos designers, estamos contando com a colaboração de colegas de outras áreas. Tem o sector de Infografia do Estadão, eles têm programadores, designers, quando chegamos no teto da nossa capacidade de mexer com uma base de dados, de visualizar essa base, a gente pede auxílio para essa equipa para poder apresentar o resultado final de determinadas coisas. É só uma fase, suponho, até que a gente possa ter uma equipa mais completa. Quando é algo bem simples, conseguimos fazer por aqui. Na eleição presidencial de 2014 fizemos trabalhos e quase todos eles envolviam programação, envolviam a publicação de ferramentas interativas na internet. Neste momento, por falta de mão de obra, nós estamos impossibilitados de fazer.

De que forma a programação faz parte da sua rotina de trabalho?

O que eu uso de programação, se é que dá para chamar de programação, são alguns comandos de *SQL* para mexer com alguns bancos de dados e eu tento aprender outras linguagens. Tenho interesse em aprender *R*, tenho interesse em aprender *Python*, acho que isso é um buraco na minha formação.

De que forma o Estadão está investindo nessa especialização do jornalismo?

No momento, dadas as características da nossa crise no Brasil e da crise na imprensa, eu não vejo o Estadão fazendo investimento quase em área nenhuma, ele está tentando sobreviver. Só o fato do Estadão Dados existir, em si, já é um investimento, porque é uma espécie de luxo para um jornal ter uma equipa de dados hoje em dia, porque é algo que é

caro. Nem sempre você tem um volume grande de produção e hoje muitas coisas no jornalismo estão voltadas para o volume, volume de cliques, volume de acessos e isso é difícil conseguir com o tipo de trabalho que fazemos, conseguimos ter um interesse de uma parcela dos leitores que querem ter um conhecimento mais aprofundado de determinado assunto e isso é um nicho dos leitores. Hoje o nível de investimento está baixo, mas eu acho compreensível diante da nossa conjuntura atual.

Qual é a maior dificuldade, no Brasil, para a produção de trabalhos jornalísticos guiados por dados?

Acho que temos dificuldade com mão de obra, é difícil encontrar pessoas com capacitação para essa área. Ainda não temos uma cultura de dados em algumas instâncias governamentais, uma cultura de transparência de dados que é fundamental para o nosso trabalho, isso varia muito. No Governo federal você tem servidores dedicados à transparência, que se dedicam em publicar, em fazer portais de dados abertos, há incentivos para que essa abertura aconteça. Em outras instâncias, Governos Estaduais, Governos Municipais, a realidade é um pouco diferente, muitas vezes você pede um dado e a primeira pergunta que eles fazem é: "Para o quê que você quer?", é uma tentativa de direcionar sua pauta, ou de liberar o dado só com a condição de que você faça algo simpático, favorável à sua fonte e não é assim que trabalhamos, temos que trabalhar olhando o que o dado fala, independentemente do interesse de quem está fornecendo. São dois problemas e o terceiro seria que esse novo campo do jornalismo ele está se desenvolvendo justamente no momento em que a imprensa brasileira e, talvez a mundial, passa pela sua maior crise. Por um lado, tem a crise do Brasil, a economia não está boa e por outro tem a crise da imprensa, o nosso modelo de negócios basicamente foi derrubado. Estamos criando um novo campo de jornalismo num momento em que o jornalismo está encolhendo, as redações estão ficando menores, às vezes é incompatível você propor um núcleo de dados num jornal que não consegue nem fazer o básico do seu trabalho.

Considera que há informação suficiente disponível?

Eu acho que há bastante informação disponível e ainda pouco explorada, mas é muito provável que tenha muito mais a se abrir. Nunca vai ser o suficiente, queremos que tudo seja aberto, que tudo seja pesquisável, seja recortável. Amanhã a gente pode estar trabalhando com dados que a gente nem sabe que existem. Os dados de educação no Brasil tiveram um salto de qualidade a partir da criação das provas nas quais os alunos são avaliados e têm que apresentar um questionário demográfico, então você consegue fazer recortes dos desempenhos dos alunos com base na profissão dos pais, de gênero, idade, o que antes era impossível porque esse era um banco de dados que não existia.

Como imagina o auge do Jornalismo de Dados no Brasil?

Eu vejo com uma certa inveja as grandes redações dos Estados Unidos, eles basicamente vão incorporando os dados a todas as etapas do trabalho, quase tudo é Jornalismo de Dados em algumas publicações. O próprio editor do Vox disse que o ideal é que a pessoa

pare de dizer que é jornalista de dados e volte a dizer só que é jornalista, porque todos vão perceber que ele trabalha com dados também. Em breve será, se não impossível, muito difícil ser jornalista sem ter alguma noção de como trabalhar com dados. Empodera muito o jornalista poder ir atrás da informação que quer.

Entrevista com o jornalista Juan Torres, do jornal Correio (Salvador, Bahia)²⁰

Como você começou a trabalhar com Jornalismo de Dados ?

O meu primeiro projeto foi o “ 1000 vidas”, foi um projeto sobre homicídios em Salvador com dados da secretaria de Segurança Pública. A secretária ela oferece um boletim diário onde ela coloca todas as pessoas que são mortas na cidade da região metropolitana, dados de nome, sexo, idade e localidade em que a pessoa foi morta. Quando fizemos essa série, eu não sabia que eu estava fazendo Jornalismo de Dados . Eu particularmente sempre gostei de trabalhar com planilhas, me formei em administração, tinha essa coisa que normalmente jornalista não gosta que é trabalhar com números, planilhas e numa lógica mais cartesiana. Eu vi naquilo uma possibilidade de crescimento das reportagens e quando eu parei para olhar eu tinha uma coisa riquíssima na mão e essa série acabou sendo finalista da primeira edição do *Data Journalism Award*, que é um prêmio mundial de Jornalismo de Dados que começou em 2012 e essa série que foi feita em 2011 acabou sendo finalista. Quando eu vi que ela foi finalista de um prêmio de Jornalismo de Dados de alcance mundial, eu falei "eu estou fazendo uma coisa diferente aqui". Já tínhamos tido a repercussão da série, tinha sido muito forte, internamente ela também tinha sido finalista do prêmio *Essos*. Essa questão de ela ter sido finalista do *Data Journalism Awards*, fez com que participássemos da conferência da *GEN*, que é a organização que promove o prêmio, aí eu entrei a pesquisar e a entender o que era o Jornalismo de Dados , a conhecer melhor o que era o Jornalismo de Dados , me despertou bastante interesse e eu acabei entrando nesse mundo e me especializando, comecei a fazer muitos cursos e assistir eventos, desenvolver um *networking* nessa área. Nunca tive nenhuma pesquisa acadêmica, a minha vida acadêmica terminou na graduação, não levei adiante, mas sempre pesquisei muito e procurei me desenvolver nessa área. A maioria foram cursos *online*. Um dos cursos mais interessantes que eu participei foi um curso bem completo, em 2012 se não me engano, foi um curso *online* que era com pessoas que eram referências e o curso passou por todas as etapas do *data pipeline*, que é a metodologia de trabalhos com dados. Comecei a frequentar esses fóruns, o próprio *Knight Center* de Jornalismo das Américas, que é uma organização da Universidade do Texas e também dá cursos de Jornalismo de Dados , *online*. Conheci a Escola de Dados em 2014, foi quando ela veio para o Brasil e me envolvi, passei a integrar a equipa da Escola de Dados. Participei em *Hackathons* e eventos. Em Buenos Aires tem um evento forte que se chama *Media Party* que é da *Hacks/Hackers*, uma comunidade que une jornalistas e programadores. Eles têm uma vertente de Jornalismo de Dados muito forte. O maior encontro da *Hacks/Hackers* é em Buenos Aires, eu fui para esse encontro duas vezes, participei em um curso de mapeamento com uma bolsa na Nigéria, que foi um curso promovido pela *European Press* que fez treinamentos em vários lugares do mundo. Eu fiz muitos cursos na área e hoje damos treinamento pela Escola de Dados no Brasil todo e desenvolve projetos em Jornalismo de Dados , são projetos com a sociedade civil.

²⁰ Entrevista realizada por chamada de voz, no dia 22 de novembro de 2017

Quais são as características que um jornalista deve ter para trabalhar com bases de dados?

É difícil delinear características específicas, primeiro tem que ter disciplina para respeitar métodos. Nós jornalistas, normalmente, somos muito "rebeldes" e pouco apegados a métodos. Embora tenhamos um fluxo, para fazer apuração tem que ter um método mínimo, mas normalmente somos pouco organizados, fazemos várias coisas ao mesmo tempo, tocamos várias matérias ao mesmo tempo. Cada jornalista tem seu método, podemos fazer isso no jornalismo desde que a gente entregue um texto no final. O Jornalismo de Dados tem uma metodologia muito mais estreita, em que você tem que seguir sob pena de perder a credibilidade que é um dos grandes *plus* do trabalho de Jornalismo de Dados. O Jornalismo de Dados é uma intersecção do jornalismo com o método científico. O método científico é muito baseado na replicabilidade; na academia e na comunidade científica para você fazer um experimento e ele ser aceito pela comunidade ele tem que ser replicável. Aquele trabalho só vai ser válido se as pessoas que seguirem o mesmo caminho que você seguiu chegarem no mesmo resultado, com o Jornalismo de Dados é a mesma coisa. Eu quando faço uma matéria de Jornalismo de Dados eu tenho que dizer a metodologia que eu fiz, qual foi a base de dados que eu usei, como analisei ela, quais foram as perguntas que eu fiz para ela e apresentar aquela base de dados de maneira que todo mundo que percorra o mesmo caminho que eu percorri tem que chegar no mesmo resultado, se não chegar no mesmo resultado, aquele trabalho é invalidado. O jornalista tem que respeitar isso, se ele não respeita vai acabar colocando o trabalho dele em cheque. Por exemplo, a etapa de limpeza que é uma etapa onde você está ansioso para analisar os dados, porque nós jornalistas estamos procurando as histórias que estão naquela base de dados, então é uma etapa muito delicada. Essa ansiedade pode fazer você pular etapas importantes na parte de limpeza. Se você não fizer ela corretamente, pode comprometer todo o resto do processo. Uma das principais características do jornalista é ter essa disciplina, desse método. O *data pipeline* tem que ser respeitado, se você não seguir as etapas do *data pipeline*, não vai chegar a um resultado replicável. O jornalista basicamente tem que ter disciplina, gostar de trabalhar com números, não tem como fugir disso no Jornalismo de Dados. A programação hoje é uma incógnita, eu não sei se é necessário o jornalista de dados saber programar, por enquanto eu acho que hoje dá para participar de uma equipa de Jornalismo de Dados sem saber programar.

Qual o seu processo de produção de peças baseadas em dados?

Tem duas formas de você iniciar o trabalho com dados: Uma é quando você não tem a base de dados, então você atrás dela. Quando você não tem a base de dados tem uma bifurcação, se a base de dados existe, fica mais fácil. Se a base de dados não existe, ou você desiste ou você, em alguns casos, consegue montar uma base de dados, consegue partir para a coleta de dados, que eu diria que hoje ainda é uma fronteira do Jornalismo de Dados. Dependendo da área que for essa coleta de dados necessária, há especialidades para isso. O Jornalismo de Dados, em geral, não faz isso. Tem projetos que o fazem, mas sempre pareados com pessoas de outras áreas, porque o jornalista em geral não tem esse

conhecimento para fazer essa coleta. Para não dizer que nunca fiz um projeto que envolvesse coleta de dados, tivemos uma vez um caso aqui em Salvador. Estourou uma adutora de água que abastecia a cidade, essa água vem de fora da cidade, vem toda por canos e um cano desses estourou e a cidade ficou uma semana sem água, sem abastecimento de água. A água que já estava dentro da cidade foi entregue. No final da semana, começou a faltar água nos bairros mais periféricos da cidade, nos bairros nobres não faltava. Então, nós perguntávamos para a companhia de água do governo do estado e eles diziam que não, que a cidade estava desabastecida por igual. Eles não tinham o mapeamento desses bairros, não podiam nos dizer "falta mais água no bairro X e não no bairro Y". Nós sabíamos, pelos relatos, que isso era questionável esse posicionamento, então colocamos um questionário no site do correio, perguntando para as pessoas duas coisas básicas: "Está faltando água na sua casa - sim ou não?"; e "Em que bairro você mora?". Muito rapidamente começamos a ter milhares de pessoas respondendo e a maioria das pessoas faltando água era em bairro realmente periférico. Fizemos a coleta de uma forma muito rústica, sem nenhuma metodologia de coleta. Fizemos uma coleta quase aleatória. Fizemos a matéria com muitas observações. Nesse caso, fizemos isso porque não tínhamos nenhuma outra forma. Por isso, acho que o jornalista de dados tem de se debruçar cada vez mais sobre essa questão da coleta de dados, de quando é válida, de como fazer, do que fazer nessas situações onde eu não estou conseguindo dados.

Tem essas duas questões quando você não tem: Ou parte para a coleta ou você desiste da matéria. Quando você tem a base de dados, você já entra direto, que é outro ponto onde pode começar o processo, você já vai para a parte da busca do *data pipeline* – que é basicamente busca, limpeza, análise e visualização, são essas quatro etapas. Você já consegue pular a etapa da busca porque a base de dados já está nas suas mãos. Basicamente, o meu trabalho é seguir essas etapas, ou eu começo pela busca ou pela limpeza quando já tenho a base de dados. A primeira coisa que eu faço é ver se eu consigo partir para análise dessa base de dados, para a limpeza e análise numa planilha. O primeiro ponto é analisar o tamanho dessa base de dados, se me permite um trabalho numa planilha. A maior limitação das planilhas, para mim, é o tamanho. Se eu não consigo trabalhar numa planilha, aí eu tenho que partir para outros recursos que irão envolver um pouco mais de programação, vou ter que trabalhar ou com *SQL* ou com *Python*, ou com linguagens de programação que permitam fazer análise daquela base sem ser dentro de uma planilha. Meu método é basicamente ver se consigo trabalhar em uma planilha, sim ou não; se eu conseguir, começo já a fazer as análises, ver o que estou a procura ali dentro, começo a fazer as perguntas, as vezes a pergunta já vem pronta. Às vezes você tem que dar um tempo e entrevistar mesmo uma base de dados, porque você pode descobrir coisas além daquilo que você está procurando, por vezes as matérias melhores são justamente fora daquelas perguntas iniciais. Eu dedico muito tempo à entrevista, a analisar, cruzar, ver o que dá para extrair daquela base de dados. Depois é seguir para a Visualização. Meu forte mesmo é a análise, é encontrar as histórias nas bases de dados, fazer a Visualização eu não gosto. Normalmente, na Visualização você também pode ir para o simples, que seria essas ferramentas disponíveis na internet, tipo *Infogram*, *Piktochart*. Você pode ir para uma coisa mais complexa, que seria o *Tableau*, que já trabalha com dados mais robustos. O *Tableau* também é legal muitas vezes para fazer a análise, porque ele permite que você faça análises já com Visualização e as vezes na Visualização você percebe matérias, pontos fora da curva mais rapidamente. O *tableau* é uma ferramenta que quando

a base de dados é mais robusta eu normalmente uso também para analisar e para procurar uma visualização bacana.

Acha que há uma estrutura pré-estabelecida para comunicar os dados?

Vai depender do trabalho, se a equipa de Infografia vai poder dedicar tempo. A Visualização, que é a parte onde entra a equipa de Infografia, não é uma área que eu domino muito, se eu não vou poder contar com a ajuda deles vai ter que ser uma solução básica do tipo *Infogram*, no máximo um *Tableau*. Embora hoje no Correio a parte de infografia ainda seja muito voltada para o impresso, se a matéria vai mais para o *online*, eu sei que vai depender um pouco mais de nós.

É comum referir que no Jornalismo de Dados deixa-se de entrevistar pessoas e passa-se a entrevistar datasets. Considera que as bases de dados têm autoridade suficiente para serem a única fonte de informação de uma notícia?

Depende. A diferença do Jornalismo de Dados é que a principal entrevista é a base de dados, não quer dizer que não possa ter outras. Em geral, você vai seguir para uma entrevista humana, da forma como nos fazemos o jornalismo hoje, em geral, você vai derivar para isso. Eu acho, no entanto, que sim, que o Jornalismo de Dados, que os seus produtos podem parar na base de dados, sim. O cuidado que tem que se ter, assim como tomamos cuidado com a credibilidade da fonte humana, quando você está entrevistando você tem que saber: Um, se essa fonte tem credibilidade para falar o que está falando e dois, de que lugar social ela está falando, qual a institucionalidade dela. Essas nuances também existem na base de dados. Temos que saber se ela tem credibilidade ou não – algumas bases tem mais credibilidade do que outras, pelo histórico de quem coletou os dados – e também de onde aquela base de dados partiu, ela está partindo de um instituto oficial, ela está partindo de uma organização que trabalha com ativismo específico numa área, como ela foi construída, o que está por trás daquela base de dados. Se você tem uma base de dados que você também pode desconfiar da origem dela, você vai procurar outras, que tratem sobre o mesmo tema. O Nexo fez há um tempo, uma Visualização, um aplicativo em que você coloca o seu salário; um Jornalismo de Dados que não tem nenhum texto, nenhuma apuração humana. Eles pegaram a base de dados de salários no Brasil, fizeram uma estratificação daquilo. Aquilo não tem jornalismo tradicional, não tem fonte humana entrevistada, talvez eles tenham entrevistado alguém para saber qual a base de dados mais confiável para se usar para fazer aquilo, mas não há muita necessidade ali de entrevista humana. Eu acho que aquilo é jornalismo, eu brinco que é "muito jornalismo", porque eles não te apresentam um contexto, eles te colocam dentro do contexto de forma muito customizada. É possível sim fazer jornalismo só com dados.

Ao fazer a busca pelos dados, quais são as fontes que você mais recorre na elaboração dos seus trabalhos?

Em geral, vamos atrás de fontes oficiais. De bases de dados que sejam de órgãos públicos, de órgãos que tenham credibilidade, institutos nacionais. Nós procuramos sempre qual a maior autoridade que temos que possa nos dar um dado confiável naquela história.

Com que frequência há a criação de bases de dados na própria redação?

Muito pouco, nós não fazemos isso na redação. Na redação do Correio o trabalho de Jornalismo de Dados , inclusive, já tem algum tempo que não fazemos alguma coisa. Eu saí um pouco da operação de produção de conteúdos, eu estou agora numa outra função que se chama editor de inovação, muito em parte me colocaram nesse trabalho por conta do meu trabalho com dados e tecnologia. Embora esteja crescendo nas redações no mundo a fora e no Brasil também, o mundo Jornalismo de Dados é muito pequeno ainda. Tirando as grandes redações, não existe a cultura, a estrutura para se fazer Jornalismo de Dados . O que fizemos no Correio foi muito por uma iniciativa minha de propor pautas disso quando eu estava na produção, já tem dois anos que eu estou fora disso no Correio. Inclusive quando eu estava na produção não se esperava Jornalismo de Dados , era uma coisa muito minha, porque eu era editor e incentivava que as pessoas fizessem isso ou eu mesmo fazia.

Como surge uma ideia de pauta no Jornalismo Guiados por Dados?

Ela pode surgir porque você está procurando uma pauta e quer correr atrás, ou pode surgir porque você se deparou com uma base de dados ou alguém lançou uma base de dados que não existia antes.

A seleção do que deve ou não ser publicado é intrínseca à linha editorial do jornal. Considera que há valores notícia no Jornalismo de Dados ?

Jornalismo de Dados , não deixa de ser jornalismo. Quanto mais valores notícia uma análise de dados tiver, maior vai ser a possibilidade de publicação dela. Naturalmente isso vai determinar o que vai ser publicado.

Como o Jornalismo de Dados é organizado no Correio?

Nós não temos uma estrutura de Jornalismo de Dados montada. Foi muito uma iniciativa individual minha. Hoje continua sendo através de iniciativas individuais, eu estou um pouco fora da operação, mas alguns jornalistas gostam da área e acabam correndo atrás de dados e fazendo matéria. Acho que a cultura está se expandindo, mas ainda é muito individual, depende muito de incentivos das pessoas de cima. Se eu tenho que fazer uma aplicação como a do Nexo, eu não conseguiria desenvolver, ainda que eu tivesse a base

de dados, porque eu não sei programar, eu não tenho ninguém na redação que saiba desenvolver aquilo, estamos limitados, temos limitações técnicas.

De que forma o Correio está investindo nessa especialização do jornalismo?

O Correio eu não diria, pelos menos hoje não. Eu tenho planos de criar uma estrutura lá dentro em 2018. Uma das minhas metas para 2018 é criar uma área ou um método pelo menos, já que temos uma redação pequena. Uma metodologia para que o Jornalismo de Dados entre mais no fluxo de produção das pessoas, para que as pessoas o considerem mais na hora de fazer as apurações.

Há trabalhos que envolvem programadores, designers, cientistas de dados. Como é a distribuição de tarefas na produção de trabalhos que envolvam dados no Correio?

No Correio não tem distribuição. Vamos até a parte da visualização, da visualização partimos para a equipa de Arte e Infografia. Essa é a única divisão que tem hoje no Correio. Às vezes até a visualização nós temos que fazer nessas ferramentas *online*, porque a equipe de arte está com outras demandas e não consegue. O único momento onde pode haver uma divisão, às vezes, é na arte.

De que forma a programação faz parte da sua rotina de trabalho?

A programação não faz parte da minha rotina de trabalho. Eu estou fazendo cursos de programação *online*, *Python* básico já fiz, já estou seguindo para um *Python* mais avançado agora. É uma iniciativa minha, estudar sozinho e ir atrás.

Qual é a maior dificuldade, no Brasil, para a produção de trabalhos jornalísticos guiados por dados?

Eu diria a cultura das redações. O Jornalismo de Dados é novo, as técnicas que você precisa dominar para ir através do *pipeline* são técnicas, em geral, novas e dificilmente dominadas por um público mais velho. São técnicas que envolvem computação e são pessoas mais novas, em geral, que fazem. É uma dificuldade no Brasil, a cultura das redações ainda está muito ancorada ao jornalismo tradicional. São redações, no geral, muito impressas ainda, não têm um olhar digital. O Jornalismo de Dados também pode dar origem a trabalhos impressos muito bons, mas ele exige uma produção digital. Como você vai analisar dois milhões de linhas analogicamente? Não tem como. Ele exige uma produção digital, ele é um produto nativo digital, todas as etapas são digitais. As pessoas que fazem Jornalismo de Dados precisam ser digitais. Como as redações são muito impressas ainda, isso é um obstáculo. Você vê em outros países, onde a cultura de dados está mais expandida, estou falando naturalmente de Europa e Estados Unidos, isso já é um pouco mais amplo. Além do que o Jornalismo de Dados, em geral, exige investimento, dinheiro, porque são funções e trabalhos que as redações tradicionais não

têm. Programador, *webdesigner*, analista de dados, são funções que não existiam, são pessoas que não transitavam em redações. Uma das grandes dificuldades, inclusive de equipas de Jornalismo de Dados, é envolver os desenvolvedores no jornalismo, não é uma área de trabalho lógica para eles, provavelmente eles nem sabem que o jornalismo é uma área onde eles são muito requisitados hoje. Então tem toda essa cultura que não tem nem dentro da redação nem na área de desenvolvimento, essas duas comunidades precisavam conversar mais do que conversam hoje. Eu diria que isso é um problema no Brasil.

Considera que há informação suficiente disponível?

Nós sempre temos que trabalhar para que os dados sejam muito mais abertos do que são hoje. Eu tenho certeza que tem muita história para se contar com os dados que a gente tem hoje abertos. Não quero dizer que é suficiente o que os governos estão fazendo para liberar dados, não é. Nós temos que pressionar sempre para mais, mas com certeza tem muitos bancos de dados não explorados. Não acho que a ausência de bases de dados seja uma justificativa para não se fazer Jornalismo de Dados.

Como imagina o auge do Jornalismo de Dados no Brasil?

Eu acho que hoje estamos no auge, estamos fazendo o melhor que podemos com as estruturas que cada um de nós temos em nossos veículos. Espero que amanhã possamos fazer mais.

Entrevista com a jornalista Katia Brembatti, do jornal Gazeta do Povo (Curitiba, Paraná)²¹

Como você começou a trabalhar com Jornalismo de Dados ?

Eu não sabia da existência desse termo, eu sou formada em jornalismo desde 2001 e quando eu cursei Jornalismo na Universidade Estadual de Ponta Grossa, esse termo não existia. Assim como Jornalismo Investigativo que é uma expressão tão normal no jornalismo também não estudei nada disso, para mim foi apresentado como jornalismo, não com esses nomes. Eu tomei conhecimento pela primeira vez do termo Jornalismo de Dados num congresso da Abraji. Eu fiz algumas matérias que têm perfil de Jornalismo de Dados antes disso, mas não sabia que era Jornalismo de Dados, então lá em 2006 eu já trabalhava com planilhas de *Excel*, mas de um jeito autodidata. Eu não recebi nenhum tipo de treinamento na faculdade ou no jornal para trabalhar com isso, mas eu comecei a sentir necessidade de trabalhar com isso. Eu lembro que fiz uma matéria em 2006 que traçava a correlação de números de eleitores com números de moradores das cidades, para saber as cidades que tinha mais eleitores do que moradores, o que poderia indicar uma fraude eleitoral. Eu lembro como eu fiz, que foi copiando e colando as informações de

²¹ Entrevista realizada por chamada de voz, no dia 10 de janeiro de 2018

um jeito manual, de jeito bem primitivo. Quando eu comecei a trabalhar com alguns dados de educação, comecei a trabalhar com planilhas de *Excel*, era tudo muito autodidata e sempre me preocupei com a visualização dos dados. Aqui na Gazeta do Povo temos um departamento de Infografia e sempre foi o departamento que eu fui mais próxima, porque sempre que fazia uma matéria eu ficava pensando como que eu poderia apresentar esses dados de um jeito que o leitor conseguisse entender mais facilmente. Em 2008, foi o primeiro congresso da Abraji que eu fui, eu fiz a oficina do José Roberto Toledo, que era uma introdução ao RAC, foi aí que eu tomei ciência de quem eram os principais nomes até aquele momento no Jornalismo de Dados , que existia uma categoria chamada Jornalismo de Dados , que existiam reportagens específicas de Jornalismo de Dados , foi nesse momento que eu tomei conhecimento disso.

Quais as características que um jornalista deve ter para trabalhar com dados?

São as características básicas do jornalista mesmo, você precisa ter curiosidade, precisa ter essa vontade de levar uma informação traduzida de um jeito interessante para o público. Eu brinco com os meus alunos de Jornalismo de Dados e Jornalismo Investigativo que você pode ensinar um jornalista a fazer qualquer coisa, você pode ensinar ele a cobrir um jogo de futebol, você pode ensinar ele a ser colunista social, você só não consegue ensinar um jornalista a ser jornalista investigativo e jornalista de dados. Demanda muito uma boa vontade, não tem uma pauta pronta, não tem um roteiro de agenda, não tem um *script*. Você não recebe uma pauta de Jornalismo de Dados , você produz a sua própria pauta, se não tiver iniciativa e boa vontade do jornalista, do repórter, não sai.

Pode descrever o seu processo de produção de peças baseadas em dados?

Existem estudos que mostram que existem vários tipos de Jornalismo de Dados , tem o que é o ponto de partida para uma pauta e tem aquele que é a matéria em si. Acho que o Jornalismo de Dados está, aos poucos, criando a sua própria forma, não tem um único jeito de o fazer. Eu publiquei hoje um material de base de dados que é sobre o piso dos professores do Paraná, é um material que levou vários meses de trabalho. Eu vejo como Jornalismo de Dados , mas pode ser considerado rudimentar para algumas pessoas, porque ele não é baseado em raspagem de dados, eu não consultei uma base pública de informação, tivemos que telefonar para as 399 cidades do Paraná e descobrir, perguntando para as pessoas: "Qual o piso? Qual o menor valor vocês pagam para um professor aí na sua cidade?". A partir disso montamos um *docs*, uma planilha enorme, com esses dados para conseguirmos ter um panorama da remuneração dos professores no Paraná. É Jornalismo de Dados , mas surgiu de um jeito diferente, surgiu a partir de uma *release* e para descobriremos o que tinha realmente de informação na *release*, tivemos que ligar para os lugares e formar a nossa própria planilha porque não existe no Brasil um lugar para você consultar o piso dos professores. Eu já fiz várias matérias de Jornalismo de Dados e cada uma delas tinha um perfil completamente diferente do outro. É um avanço do jornalismo conseguir enxergar para além das fontes básicas. Eu já fiz matérias de Jornalismo de Dados de várias formas, peguei uma planilha, entrei numa base de dados

específica e pensei, vou fazer uma matéria sobre isso ou vou montar uma base própria sobre aquilo. No caso dos Diários Secretos, digitamos uma planilha com 15 mil linhas a partir de 10 mil documentos. Não tínhamos um norte, entrevistando as planilhas é que surgiram as pautas da série. Eu sei que são pessoas que são bem organizadas, metódicas e sistemáticas nesse ponto. Eu costumo olhar para uma situação e pensar que esse caso seria melhor desse jeito, não há um modo único.

Há uma estrutura pré-estabelecida para comunicar os dados?

Aos poucos estamos criando modelos, não há um modelo fixo. Por exemplo, eu gosto muito dos modelos de visualização do Nexo, eu acho que a entrada do Nexo no Brasil reconfigurou toda essa questão de visualização de dados. Eu lembro que quando surgiu o Estadão Dados e o Basômetro a minha referência eram eles, hoje o tipo de dados que eu mais gosto de ver é o do Nexo, acho que essas coisas estão sendo redefinidas. Temos algumas coisas pré-estabelecidas, mas aqui, por ter um departamento de Infografia bastante ativo, eles estão sempre se perguntando como fazer diferente. Por mais que existam *templates*, há uma preocupação de como melhorar isso. Um dos motivos é que o chefe da Infografia é *designer* mas está se aproximando do jornalismo, ele vai para os Congressos da Abraji por exemplo. Ao ir para os congressos ele começa a enxergar mais como jornalista e menos como *designer* ou misturando mais as duas coisas.

É comum referir que no Jornalismo de Dados deixa-se de entrevistar pessoas e passa-se a entrevistar datasets. Considera que as bases de dados têm autoridade suficiente para serem a única fonte de informação de uma notícia?

Algumas vezes, sim. Eu mostrei, por exemplo, para os meus alunos uma matéria do Valor Económico que era sobre o Bolsa Família. Nós costumamos ver matéria de Bolsa Família com personagens, mesmo quando é redução e aumento e eu brinquei com os meus alunos: "Se vocês não querem lidar com pessoas, dá para fazer matéria" e mostrei o exemplo do Valor Económico. Dá para fazer essa matéria sem sair do computador, sem entrevistar ninguém, dá para fazer. Ficou boa a matéria? Ficou, mas eu gosto da ideia da fonte também, por mais que goste de Jornalismo de Dados, eu gosto de personagens, gosto de falar com especialistas, de ir no lugar. Em algumas situações faço através do meu computador, só os dados se bastam, mas eu gosto dessa complementaridade, acho que ela agrega para o público.

Ao fazer a busca pelos dados, quais são as fontes que você mais recorre na elaboração dos seus trabalhos?

Normalmente as informações são públicas, são oficiais. Tem também dados de organizações, se eu quero dado sobre saneamento eu vou procurar o Instituto Prata Brasil que é referência no país em saneamento, é uma ONG, mas normalmente são dados públicos. Claro que sabemos que tem muito dado escondido, dado que pedimos pela Lei

de Acesso e não recebe, dado opaco que deveria ser transparente mas que tem uma série de barreiras para você conseguir chegar no dado, mas ainda assim sabemos que não conseguiu explorar nem o que tem disponível.

Com que frequência há a criação de bases de dados na própria redação?

É menos frequente do que eu gostaria, porque costuma dar muito mais trabalho. Toma mais tempo, uma coisa é você pegar um dado pronto e trabalhar com esse dado, você economizou um percurso, você não está trabalhando com um dado exclusivo, mas dependendo da análise que você fizer ele vai ser exclusivo, mesmo sendo um dado público. Quando você produz uma base própria você tem uma informação completamente exclusiva só que toma tempo e o segundo problema é que você assume 100% da responsabilidade pela informação. Quando montamos a base do Diários Secretos, não tinha uma investigação do Ministério Público, da Polícia Federal como é o caso da Lava Jato. Se tivesse algum dado errado ali, eu não podia responsabilizar alguém, usar a Lei, porque a Lei dá algumas garantias para o jornalista quando ele usa dados oficiais; quando você monta uma base própria, você é 100% responsável por ela, se tiver algum problema, um erro de digitação que seja, você é responsável por ela também. Eu gostaria de trabalhar com mais dados próprios, mas isso tem um custo de tempo que normalmente acaba diminuindo o uso desse recurso. Por ano eu faço no mínimo duas ou três matérias com bases próprias.

Como surge uma ideia de pauta no Jornalismo Guiados por Dados?

Ela surge de vários modos, eu sou setorista de algumas áreas, quando você está numa área que você está cobrindo recorrentemente aquele assunto, você acaba percebendo coisas que possam chamar a atenção. Eu estou trabalhando em uma base de dados própria há seis meses, ela vai sair lá por abril, é um material complicado. A pauta surgiu porque eu vi uma matéria, que era uma matéria de um concurso público. A do piso mesmo, surgiu porque recebemos um *release* dizendo que 23% das prefeituras do Paraná não cumpriam a lei que determina o valor mínimo. Eu queria saber quais prefeituras eram e entrei em contato com o Ministério Público, quando eles liberaram a planilha para a gente, tinha alguns dados estranhos, algumas cidades com orçamento alto, pagando pouco, eu liguei para essa cidade, fiz o processo básico de checagem, a base do Ministério Público estava errada. Eu falei para a minha chefe que só tinha uma forma de fazermos essa matéria, que era nós mesmo ligarmos para todas as Câmaras Municipais. O jeito que surge uma matéria de Jornalismo de Dados é muito variado, às vezes é por uma influência de um outro veículo. Eu vi uma matéria na Folha de São Paulo sobre as férias dos juízes de São Paulo, eu decidi raspar os dados do Diário Oficial daqui para ver como está a situação das férias dos juízes aqui no Paraná. A fonte é muito diferente de uma matéria para outra.

A seleção do que deve ou não ser publicado é intrínseca à linha editorial do jornal. Considera que os valores notícia pautam o Jornalismo de Dados ?

Acho que os valores notícia acabam, sim, pautando o Jornalismo de Dados , não deixa de ser jornalismo. Se ele não for importante para as pessoas, se não tiver algo a ser contado, ele não vai ser só pelo especto curioso; dificilmente você vai ter Jornalismo de Dados que vai usar o valor notícia “curiosidade”. O que determina se um material é publicado ou não, normalmente, é a confiança no dado. O principal fator é: Isso é relevante? Sendo relevante, temos confiança nesses dados? Temos, então publicamos.

Como o Jornalismo de Dados está organizado dentro da redação?

Não temos um departamento de Jornalismo de Dados , mas temos um departamento de Infografia que dá apoio a todos os jornalistas. Esse departamento de Infografia é essencial aqui dentro e tem sido mais relevante ano a ano. Temos jornalistas com perfil de Jornalismo de Dados mas que não são jornalistas de dados, pessoas que não trabalham com Jornalismo de Dados no dia a dia, fazem isso eventualmente, uma matéria específica.

Há trabalhos que envolvem programadores, designers, cientistas de dados. Como é a distribuição de tarefas na produção de trabalhos que envolvam dados no Gazeta do Povo?

É bastante colaborativa, hoje trabalhamos com dois programadores, às vezes eu chego para eles com uma demanda, temos de organizar esses dados e acaba sendo construído coletivamente e eles dão sugestões que ultrapassam o trabalho deles. Para além de programar ou de criar uma visualização, eles acabam colaborando na informação que vai entregue para o público. Se é uma matéria do dia a dia, mais simples, recorremos ao *design* diretamente, quando é um trabalho mais complexo, de mais fôlego ou quando é um trabalho que demanda de programação, não necessariamente de *design*, é a pauta que define se vamos trabalhar só o jornalista, se é o jornalista com *design*, se é o jornalista com o programador ou se é o jornalista com o *design* e o programador.

De que forma a programação faz parte da sua rotina de trabalho?

Até pouco tempo atrás eu era do grupo que dizia que o jornalista não precisava programar, precisava conhecer alguém que programasse. Eu sabia o que eu queria, se era possível ou não fazer e eu sabia quem poderia fazer por mim. Eu já recorri a trabalho de *hacker* dos programadores aqui dentro do jornal. De uns tempos para cá, cada vez mais eu tenho sentido a dificuldade de não programar e tenho tentado procurar vencer essa demanda. Sinto que eu poderia otimizar e automatizar coisas simples, coisas básicas, que muitas vezes eu dependo de outra pessoa. Eu comecei a fazer um curso de *Python* e pretendo me aventurar por outras áreas da programação mais para a frente.

De que forma você vê investimento por parte da Gazeta nessa especialização do jornalismo?

Eu gostaria de ver mais investimentos da Gazeta nessa área, eu já falei inclusive algumas vezes com o diretor de redação sobre isso, que é uma área que precisaria ter mais investimentos. Normalmente, os investimentos nessa área são dois: Contratação de pessoal e tempo do jornalista. Também tem *softwares*, treinamentos, mas o grosso do investimento é tempo para os jornalistas fazerem as coisas, tirar eles do redemoinho do dia a dia e contratação de pessoal. Eu gostaria de ver mais, mas não dá para negar que tem uma estrutura, tem uma equipa de Infografia que eu já mencionei e vamos para vários cursos, principalmente da Abraji, sempre a tentar se manter atualizado sobre o que está sendo feito. Eles já me custearam várias vezes para ir, as últimas duas vezes eu mesma custeei, mas eles estavam mandando outras pessoas, eles não me mandaram, mas mandaram outras pessoas, para também não ficar mandado as mesmas, mas eles liberaram os dias de trabalho para eu ir. Há um incentivo para que a gente se mantenha atualizados nessa questão.

Qual é a maior dificuldade, no Brasil, para a produção de trabalhos jornalísticos guiados por dados?

Tempo, o Jornalismo de Dados mais complexo toma tempo, quando toma tempo você deixa de fazer outras coisas e daí tem que priorizar. Não falta *software*, não falta informação, não falta jornalistas competentes, não falta gente interessada, não falta público. O que falta é priorizar e dar tempo para isso. Temos que mostrar o impacto, o Jornalismo de Dados precisa mostrar o impacto que ele causa, não em cliques, ele precisa mostrar o que ele muda na percepção do público, precisa mensurar coisas, que realidade foram transformadas por esse jornalismo. Há uma lógica de que a internet, daqui para a frente, vai ser cada vez menos texto e cada vez mais visualização, áudios, vídeos e outros formatos.

Você considera que há informação suficiente disponível?

Existe mais informação pública disponível do que capacidade que a gente tenha de processar essa informação, de ir atrás. Não temos capacidade, em quantidade de jornalistas, em dedicação a isso, para explorar esse universo enorme de dados disponíveis.

Como você imagina o auge do Jornalismo de Dados no Brasil?

Eu imagino que essa nova cara da internet, cada vez mais interativa, cada vez mais visual, mais hiperlinkada, vai ser uma cara muito própria do jornalismo do futuro. Me parece que é mais no caminho do Nexa, me parece que é o caminho que muitos veículos vão adotar daqui para a frente.

Entrevista com o jornalista Marlen Couto, do jornal O Globo (Rio de Janeiro)²²

Como começou a trabalhar com Jornalismo de Dados ?

Eu recebi uma proposta para trabalhar no núcleo de Jornalismo de Dados do jornal O Globo. Eu entrei no programa de estágio do jornal em 2015, passei por várias editorias e no fim desse processo eu fui contratado como repórter *trainee* na editoria O País, que cobre política nacional e notícias nacionais. Nessa experiência como repórter *trainee*, eu tive a oportunidade de fazer algumas matérias, reportagens que utilizavam levantamento de dados. Acabou por se perceber que eu tinha um perfil para trabalhar com Jornalismo de Dados por essa minha capacidade de fazer análises estatísticas e cruzamentos de informações e até uma habilidade no *Excel*. O Globo passou por uma mudança no quadro de recursos humanos e acabou surgindo uma vaga nesse núcleo que já estava em funcionamento no jornal. Nesse período inicial eu não tive oportunidade de aprender. Eu já tinha feito um curso técnico de *Excel* que eu já tinha na minha trajetória acadêmica, não era nada aplicado ao jornalismo, eu já tinha certo domínio da ferramenta, mas o olhar que o Jornalismo de Dados exige eu não tinha feito um curso específico, foi fruto da minha formação embora a universidade não tenha nada na grade que seja voltado especificamente a o Jornalismo de Dados . Eu acho que eu fui construindo um pouco de um perfil mais analítico, com capacidade de analisar os dados, a minha trajetória me levou a isso, embora eu não tenha tido uma formação específica Jornalismo de Dados .

Quais as características que um jornalista deve ter para trabalhar com dados?

Eu acho que o principal e cada vez mais eu tenho certeza que isso é mais difícil, não é uma capacidade meramente técnica e tecnológica de saber utilizar ferramentas disponíveis para extração, para manipulação ou visualização de dados. O grande diferencial do perfil do jornalista de dados é esse olhar para informação que está disponível através de dados, eu acho que o mais desafiador é saber o que fazer com os dados. Acho que é uma capacidade analítica, em relação aos dados, estatística e isso é muito difícil porque no jornalismo a gente não tem, ao menos na realidade brasileira, essa competência muito bem desenvolvida na universidade, nos cursos de graduação. Ao mesmo tempo também se tem um medo em relação a matemática, exatas e isso é muito comum nas universidades e no perfil de quem vai fazer jornalismo. Uma característica que eu elencaria seria essa capacidade de analisar dados estatísticos e interpretá-los, isso é muito difícil e importante nesse trabalho e acho que aí sim também a competência técnica para manipular a informação. Um terceiro ponto que eu levantaria seria uma capacidade também de apresentar esses dados. Acho que é uma dificuldade dos repórteres, principalmente, de apresentar de forma gráfica essa informação, como é que essa informação vai chegar até ao leitor, como é que vai ser visualizada. O jornalista de dados também deve ter essa preocupação e isso também não está muito no perfil de quem tem uma formação tradicional em jornalismo. Acho que acaba sendo desenvolvido na prática e quando você tem contato com outros perfis que compõem a redação, com o resto da equipe que acaba te ajudando a chegar nesse objetivo.

²² Entrevista realizada por chamada de voz, no dia 27 de novembro de 2017

Pode descrever o seu processo de produção de peças baseadas em dados?

O Núcleo de Jornalismo de Dados do O Globo ele é integrado ao grupo de *fact-checking* do jornal, então eu também estou fazendo checagem de factos, no Blog "É isso mesmo", que foi criado agora em março deste ano e tem feito checagem de declarações políticas e tem também trabalhado *Fake News* e disseminação de boatos. Olhando para dados especificamente, muito da produção depende do monitoramento de bases de dados disponíveis. Eu costumo monitorar portais públicos principalmente de divulgação de informações que sejam relevantes e sempre tentar buscar bases que estejam com informação de dados abertos dessas fontes. Isso inclusive é um problema, porque O Globo ainda tem uma dificuldade de usar programação para extração de dados, eu ainda não tenho esse perfil, ninguém daqui tem, conhecimentos de programação que tornem possíveis extrair dados da *Web*. A produção ainda está muito baseada na disponibilização das bases de dados pelos governos, pelas instituições, pelas empresas. Uma outra produção vem também do acompanhamento do noticiário, para identificar assuntos ou coberturas que sejam pertinentes, que podem também ter uma abordagem de Jornalismo de Dados, que acaba também contribuindo para uma abordagem mais jornalística do noticiário. Isso nem sempre é possível porque o Jornalismo de Dados tem um ritmo completamente diferente do factual, do jornalismo tradicional de outras editorias, mas tentamos também fazer esse exercício de ver quais são as questões que estão abertas. Por uma questão de eu estar muito no início desse processo de aprendizado em relação aos dados, eu costumo fazer muitas coisas ainda pelo *Excel*. Embora tenha outras ferramentas mais avançadas, eu faço muito do processo de extração dos dados e também a limpeza e o tratamento de dados no *Excel*. Em relação a visualização, temos usado algumas ferramentas para gerar gráfico também, temos usado o *Tableau* que é uma ferramenta gratuita para isso. Nós temos uma ferramenta do O Globo também que foi desenvolvida para esse fim, a redação consegue introduzir dados e gerar gráficos. Eu tento identificar ou perguntas que possam gerar reportagens ou analisar dados disponíveis na web e daí também gerar reportagens.

Há uma estrutura pré-estabelecida para comunicar os dados?

O Globo tem uma identidade visual, tentamos sempre adequar o gráfico, a visualização, a essa identidade visual. Temos algumas limitações tecnológicas também, dificuldade de fazer gráficos que não estejam nesses formatos mais acessíveis desses dois programas que utilizamos. São formatos mais simples, gráficos de barra, de linha, são estruturas mais simples. Quando não passa por aí, precisamos mobilizar a equipe de *Design* do jornal, isso leva um pouco mais de tempo, geralmente são projetos mais longos. Dependendo do trabalho, do que for o projeto, também mobilizamos programadores e aí também é um processo um pouco mais demorado.

É comum referir que no Jornalismo de Dados deixa-se de entrevistar pessoas e passa-se a entrevistar datasets. Considera que as bases de dados têm autoridade suficiente para serem a única fonte de informação de uma notícia?

Eu acho que muita coisa pode sim estar colocada pelos dados, mas é fundamental também fazer análises a partir de outras fontes. Os dados, por si só, eles não são necessariamente tão relevantes. É preciso ter uma análise desses dados, até para que o leitor consiga entender a relevância deles e consiga daí interpretar aquela realidade colocada. No meu processo de trabalho, não acho que os dados, por si só, bastam. Eu costumo recorrer a outras fontes que são necessárias para gerar interpretações em relação a esses dados. Depende muito da situação, mas em geral eu costumo não parar só nos dados, é preciso levar as várias interpretações possíveis daquela informação. Até porque eu acho também que o nosso leitor, não necessariamente, tem essa capacidade analítica que é tão fundamental no perfil do jornalista de dados. É preciso ter um cuidado com o leitor e propor essas interpretações.

Ao fazer a busca pelos dados, quais são as fontes que você mais recorre na elaboração dos seus trabalhos?

Depende muito das reportagens, mas eu acho que os governos ainda são a principal fonte, tanto o governo federal quanto os governos estaduais e municipais e instituições de pesquisa. Temos aqui no Brasil o IBGE e outros que estão também ligados ao governo, mas que são instituições especificamente de pesquisa, universidades, fundações de pesquisas.

Com que frequência há a criação de bases de dados na própria redação?

Dependendo também do tipo de informação nem sempre conseguimos achar a informação do jeito que queremos, precisamos montar nossas bases, nossos levantamentos. Neste momento acho que estamos num ritmo um pouco menor nesse sentido. O Globo já fez algumas iniciativas e reportagens nessa linha, eu não estava na redação nessa época. O jornal fez um levantamento que é muito emblemático, o “DNA do Congresso” tentou apresentar graficamente o perfil do Congresso Brasileiro e aí precisou criar a base de dados também, porque foi preciso cruzar informações de diferentes fontes e isso gerou uma base própria. Ainda não é muito comum no O Globo, temos dificuldade e acho que principalmente pela escassez de mão de obra, porque são levantamentos que levam um pouco mais de tempo e a nossa equipa ainda é pequena. Um projeto como esse levou dois meses. É difícil manter uma equipa destinada especificamente a levantamentos como esse, é um desafio.

Como surge uma ideia de pauta no Jornalismo Guiado por Dados?

Na minha avaliação parte de dois pontos: ou dos dados que estão disponíveis, eu acho que muita matéria surge de você identificar bases de dados, levantamentos e a partir da análise desses dados chegar a alguma conclusão ou alguma reportagem; A outra parte surge de muitas perguntas que precisam ser respondidas, ao tentar responder essas perguntas se busca dados e isso gera uma matéria.

A seleção do que deve ou não ser publicado é intrínseca à linha editorial do jornal. Você considera que há valores notícia no Jornalismo de Dados ?

Há valores notícia e eles são um pouco diferentes do noticiário tradicional na minha avaliação. Acho que a cobertura factual está muito voltada para a atualidade, por exemplo. Com a internet, com os novos media, as coisas que estão acontecendo agora têm uma certa prioridade, elas ganham uma relevância maior. No Jornalismo de Dados não necessariamente isso é possível, é muito difícil fazer reportagens sobre assuntos quentes ou que estejam em evidência nesse momento, é da natureza. Acho que o Jornalismo de Dados tem um conteúdo mais analítico, contextual, que não passa por aí. Certos princípios, valores notícias que são aplicados ao jornalismo quotidiano não são ao Jornalismo de Dados e eu acho que esse é um deles. Muitas vezes eu acho que outras coisas se mantêm, o ineditismo ainda é muito forte. Ser capaz de trazer um levantamento inédito que ninguém tem acesso, que seja exclusivo, acho que tem muita relevância, é um valor muito forte, não que não seja também no jornalismo tradicional, mas acho que chama muito a atenção.

Como o Jornalismo de Dados é organizado no O Globo?

Temos um grupo, um núcleo, uma equipa dedicada exclusivamente a isso que agora está também fazendo *fact-checking*. São dois repórteres, um editor-executivo e um editor-responsável. Isso não impede que a gente interaja com as outras editorias da redação, estamos sempre em comunicação com coordenadores, editores, repórteres e a gente também, eventualmente, faz reportagens em participação e auxiliando também essas editorias. Esse grupo foi criado em 2015, ele era formado inicialmente pelo Fábio Vasconcellos e pela Gabriela Alelgro, eles não estão mais hoje na redação. Temos hoje o um repórter *trainee* e um editor assistente.

Há trabalhos que envolvem programadores, designers, cientistas de dados. Como é a distribuição de tarefas na produção de trabalhos que envolvam dados no O Globo?

Nós tínhamos o Fábio, ele fez o doutorado em Ciências Sociais, então ele tinha o perfil de dados diferente do mercado, ele nos auxiliava nesse ponto, em relação ao ter cientista de dados na equipa. Não temos uma equipa própria de programadores, nem de *designers*. Recorremos, assim como as outras editorias, a editoria de Arte, que é onde fica essa equipa de design e programação. Precisamos, assim como as outras editorias, recorrer a eles, eles têm uma agenda própria do que é prioridade e quais são os projetos que merecem investimento e isso é muito negociado com essa equipa. Levamos sugestões, ideias de pauta e dependendo do que se tem como prioridade, do que está na agenda da equipa, isso vai se enquadrando. Quando necessitamos de algum trabalho que utilize programação, recorremos à equipa de programadores do O Globo que está atuando na

Arte, quando não precisa nós tentamos fazer. Tentamos, ao máximo, não sobrecarregar a Arte com infográficos, com visualização de dados, nós tentamos com essas ferramentas que estão disponíveis, tentamos também fazer esse processo de visualização. Recorremos à Arte quando são coisas mais complexas.

De que forma a programação faz parte da sua rotina de trabalho?

Eu estou num processo de aprendizagem de linguagem de *Python*, mas ainda não comecei a aplicar isso no meu trabalho. Isso é uma dificuldade, não temos um jornalista com esse perfil na redação hoje. Eu estou tentando aprender um pouco de linguagem de programação, eu acho que mais do que eu mesmo fazer extração de dados utilizando a programação, o objetivo para mim, neste momento, é saber mais o que pedir e o que pode ser feito com programação. Eu acho que não vou, em pouco tempo, ter um domínio necessário para executar essas tarefas.

De que forma O Globo está investindo nessa especialização do jornalismo?

Eu acho que O Globo tem investido muito em visualização, tem sido um movimento do jornal. Ele fez algumas mudanças até gráficas na apresentação do jornal e tem estimulado para que a redação pense muito a visualização e como essa informação vai ser apresentada, como os dados vão ser apresentados. Ainda estamos muito baseados no levantamento de bases de dados que estão disponíveis, públicas, processadas por instituições, por governos, ainda temos dificuldade para utilizar programação para a fazer a extração de dados, isso eu acho que é um ponto ainda a ser trabalhado.

Qual é a maior dificuldade, no Brasil, para a produção de trabalhos jornalísticos guiados por dados?

Eu acho que está na qualificação dos profissionais. Temos um desafio que é, tem sim uma dificuldade de transparência dos governos de disponibilizar informações, isso melhorou muito, principalmente no governo federal, mas os governos municipais e estaduais ainda têm muita dificuldade de disponibilizar dados em formatos que possamos manipular esses dados e fazer cruzamentos; o formato de dados abertos ainda é uma barreira. Tem uma outra dificuldade que passa pelo perfil, temos uma escassez de mão de obra que esteja voltada para esse tipo de profissão, que tenha um perfil para trabalhar com Jornalismo de Dados . Acho que isso ainda é um desafio para o jornalismo brasileiro.

Considera que há informação suficiente disponível?

Eu acho que sim, acho que há muita informação que não conseguimos dar conta, acho que temos muita informação também que não passa pelos governos, pelas instituições, mas que parte das redes sociais. Há muita informação ali que pode ser processada e que pode gerar reportagem e também fazemos isso muito pouco, processar essa informação que é gerada diariamente pelos usuários das redes sociais. É uma fonte que temos muita informação, mas que tem pouco processamento.

Como você imagina o auge do Jornalismo de Dados no Brasil?

Eu acho que o grande auge seria os jornais e os veículos não só se basearem em dados e em levantamentos que passam por instituições, acho que o auge seria os próprios jornais também produzirem informação, levantamentos com base na extração de dados.

Entrevista com o jornalista Raphael Hernandez, do jornal Folha de São Paulo (São Paulo)²³

Como começou a trabalhar com Jornalismo de Dados ?

Eu trabalhava numa televisão, em Cascavel, no interior do Paraná e eu descobri por meio de um curso da *Canvas Network*. Eu vi a disponibilidade do curso, fiz e comecei a ver a aplicabilidade daquilo no meu dia a dia e a partir daí eu comecei a estudar cada vez mais, sai de Cascavel, vim para São Paulo e trabalhei com Jornalismo de Dados . É um curso que tinha grande nomes do jornalismo global, Simon Rogers, Alberto Cairo e outros nomes gigantescos. Depois que eu fiz esse primeiro curso, *online*, eu fiz outros cursos da Abraji, depois eu comecei a me identificar um pouco mais com a programação também e peguei uns cursos específicos dessa área, basicamente tudo *online*. Como eu morava numa cidade do interior, no sul do Brasil, onde o jornalismo – inclusive saiu pesquisa sobre isso recentemente – não é tão forte, você não tem tantas opções, é mais algo localizado. Você não vai ter uma conferência ou um curso fixo de Jornalismo de Dados , você tem que usar a internet. E também, coisas de internet você aprende na internet, tem muitos recursos *online*. Eu comecei a trabalhar na RBC em 2014, no começo de 2015 foi quando eu fiz o curso, até havia alguma aplicação local, uma ou outra pauta eu ajudava na produção. Fiz o curso e comecei a ver uma ou outra aplicação. Eu saí de Cascavel, fiz um *trainee* no jornal Estado de São Paulo, lá eles têm um regime que você passa uma semana por cada lugar, eu passei pelo Estadão Dados. Acabou o *trainee* de 3 meses e fiz um outro *trainee* na Folha, voltado à Ciência e Saúde. Durante esse *trainee*, eu tentei focar no trabalho com dados, usei inclusive os dados para me vender no processo seletivo, eu era a única pessoa do grupo que tinha essas habilidades. Como estávamos trabalhando com saúde pública, o uso de dados veio muito bem a calhar. Até consegui fazer alguns trabalhos usando as habilidades que eu aprendi nesses cursos. Fiz cursos de *hacking* também, para tentar explorar o site da prefeitura de São Paulo, é extremamente desorganizado, tem muita informação, tinha uma informação que eu estava buscando e acabei descobrindo uma intranet que estava aberta e lá continha dados. A folha tem uma

²³ Entrevista realizada por chamada de voz, no dia 24 de novembro de 2017

editoria, a que eu estou agora, chamada “Audiência e dados” que trabalha com Jornalismo de Dados , mas que tem como principal função o trabalho com métricas de audiência. Se formos ver, em muitos pontos se assemelha porque é trabalho com Big Data da mesma forma, análise de dados, só que aplicado num viés um pouco mais estratégico.

Quais as características que um jornalista deve ter para trabalhar com dados?

Primeiro tem que ter um bom relacionamento com máquina. Você tem que ter essa lógica, essa noção do funcionamento das coisas, você não precisa saber programar, claro que ajuda, mas você tem que ter esse bom relacionamento, traquejo, para lidar com a máquina. Eu fiz muitos cursos de fora e muitas coisas é muito difícil aplicar para a realidade brasileira. Você faz um curso com uma base de dados do governo da Califórnia e está perfeitamente estruturado, no Brasil você muitas vezes tem que brigar via Lei de Acesso à Informação para obter algum dado que seria básico. Ou você tem um dado em formato ocioso que você tem que saber limpar ou, no mínimo, você vai ter problemas com caracter acentuado, você tem que ter bom manuseio, para saber o que vai fazer. E tem a questão da velocidade, no Jornalismo de Dados muitas vezes você tem um pouco mais tempo para trabalhar do que a redação como um todo, mas muitas vezes não é o caso. Uma votação na Câmara, na época das Eleições, que você tem que fazer as coisas com muita velocidade porque o público quer essa informação, você não pode pensar muito, você tem que saber fazer de uma maneira ágil. É preciso esse jogo de cintura: «Eu tenho este problema, como eu vou solucioná-lo?» e ter essa lógica de pensar no processo em que «eu preciso analisar esse dado e eu vou ter que fazer tal e tal passo».

Pode descrever o seu processo de produção de peças baseadas em dados?

Eu estou nessa editoria de Audiência e Dados que é uma editoria mais estratégica do jornal. No meio do ano o Folha criou o Núcleo de Inteligência Folha, essa sim é uma secção especificamente criada para trabalhar com Jornalismo de Dados , sem essa divisão com a audiência, tem duas pessoas agora e eu trabalho muito ligado a eles. O fluxo não é uma coisa fixa. Isso varia de caso em caso, tem muitos métodos que eu uso que eu sei que nunca mais vou usar. Como eu mencionei antes, você tem muitos problemas em bases de dados, principalmente com base de dados brasileiras, o que mais me consome, o que mais me dá trabalho é justamente preparar esses dados. Vou pegar alguma base, normalmente uma base de dados pública ou através da Lei de Acesso. Dependendo do fluxo, pode ser que eu esteja acedendo esse dado para ver o que eu encontro ou é um dado sazonal que acabou de sair e vamos ver o que tem de novidade ou pode ser que já tenha uma matéria que eu precise de contexto ou que alguém tem essa matéria e veio pedir ajuda. O primeiro passo é você pegar esse dado de alguma forma e essa forma varia muito, pode ser simplesmente clicar no botão de *download*, pode ser criar um robô, seja usando *Python*, seja usando extensão do *Chrome* ou algum serviço pronto, tipo o *Web Scrapper* ou o *import.io*, para tentar obter isso, ou seja, "*Scrapear*" mesmo o site. Depois vem o processo mais desgastante, que é essa limpeza e organizar essa grande massa. Recentemente fizemos uma matéria sobre violência na cidade de São Paulo com dados da Secretaria de Segurança Pública e eram dados de Boletins de Ocorrência e aquilo estava desorganizado.

Em relatórios de homicídios tinha pessoas que constava mais de uma vez, isso é esdrúxulo: se é um relatório de homicídio como a pessoa morreu mais de uma vez? Foi um processo de algumas semanas para tentar melhorar esses dados, tratar eles de alguma forma. A parte da análise também é uma coisa que varia muito do tipo de material que você tem na mão. Esse foi o caso que eu estava fazendo essa matéria junto com alguns colegas, um deles é o Daniel Mariani que é desse Núcleo de Inteligência e ajudou na limpeza dos dados e o outro que é um repórter de quotidiano, só para dar um exemplo prático, nesse caso eu montei um esquema usando o *Tableau* para poder analisar esses dados, que era uma massa grande, que precisava de uma análise mais a *doc*. Com isso eu conseguia também compartilhar com esse repórter de quotidiano e, como ele tem mais experiência na cobertura desses assuntos, ele podia também analisar esses dados à medida que eu analisava. A Folha tem um departamento de Infografia que tem um pessoal muito capacitado para trabalhar com isso, tem sistemas próprios, existem casos de eu ter feito até a parte da visualização, mas é raro. De regra, passamos esses dados de uma forma um pouco mais trabalhada, para agilizar o fluxo para esse departamento de arte que produz a visualização. É sempre em parceria com a gente, como vai ser montado, qual vai ser a melhor forma, até porque eles têm mais *know-how* para pensar e para fazer essa visualização um pouco melhor.

Há uma estrutura pré-estabelecida para comunicar os dados?

Eu acho que ainda é muito caso a caso. Você não consegue fazer Jornalismo de Dados com “Jornalismo de Dados”, você sempre vai associar a alguma seção de cultura, de saúde, de desporto, são universos muito diferentes. Você tem que ter essa maleabilidade para poder adaptar a sua narrativa para a melhor forma para aquele caso. Tem muitos casos que o trabalho de Jornalismo de Dados, por mais que ele seja grande, ele é uma peça, um elemento secundário dentro da matéria, é uma peça de contexto. Você não pode pensar que você vai ter um tratamento fixo para isso, se você vai ter casos que você vai ter, basicamente, só o dado, a tua matéria está em cima dos dados.

É comum referir que no Jornalismo de Dados deixa-se de entrevistar pessoas e passa-se a entrevistar datasets. Considera que as bases de dados têm autoridade suficiente para serem a única fonte de informação de uma notícia?

Pode ser, mas vai muito da capacidade do jornalista de dar contexto para aquilo. Eu vejo isso um pouco mais raro, mas não vou dizer que é impossível. De regra, você irá precisar seja de um especialista para comentar, seja de uma pessoa para te dar o outro lado daquela história. Se você vai fazer matéria na política, você vai ter que ouvir também a pessoa sobre quem você está falando, mesmo que a matéria seja baseada em dados. É um princípio jornalístico, você tem que ouvir sempre cada lado, também para dar contexto. Eu já fiz um especial, por exemplo, sobre reformas ministeriais no Brasil, foi na época do Estado de São Paulo, é um trabalho um pouco mais histórico, mais de pesquisa, as bases de dados em alguns casos eram jornais mesmo. Nesse caso, não tinha muito porque falar com muita gente, porque toda informação estava ali, não precisava ir muito além. Novamente, varia muito de cada caso. Sim, é possível fazer matéria só entrevistando os dados, mas você tem que ter muito critério nisso.

Ao fazer a busca pelos dados, quais são as fontes que você mais recorre na elaboração dos seus trabalhos?

Eu tenho uma base de dados de bases de dados, eu tenho mais ou menos umas 100 entradas ali, de fontes de bases de dados. Tenho também uma base de dados de ferramentas que uso recorrentemente. Depende do fluxo da matéria, preciso de um dado dado para dar contexto para “x” coisa, a primeira coisa que eu vou fazer é dar uma olhada nessa minha base de dados. De regra são dados públicos.

Com que frequência há a criação de bases de dados na própria redação?

Eu mesmo faço isso raramente, mas eu vejo isso acontecer na redação. Eu acabo fazendo pouco isso por uma questão de tempo mesmo. Por exemplo, uma matéria que saiu sobre a *Black Friday* no Brasil, o meu colega que fez essa matéria ficou, por um mês, scrapeando os sites dos principais *e-commerce* do Brasil e montou uma base de dados dele, para poder fazer essa matéria.

Como surge uma ideia de pauta no Jornalismo Guiados por Dados?

Tem casos que você está escrevendo uma matéria sobre um assunto “x” e você precisa de um contexto e usa os dados para dar esse contexto. Há casos em que você encontra uma base de dados ou vê uma base sazonal que saiu uma novidade ou que foi atualizada ou então algum pedido da chefia.

A seleção do que deve ou não ser publicado é intrínseca à linha editorial do jornal. Considera que há valores notícia no Jornalismo de Dados ?

Claro. A Folha ela não trabalha o dado pelo dado. Mesmo que você esteja trabalhando com dados, você tem que encontrar um *lead* naqueles dados. Você tem que ter algo novo, um valor notícia, uma quebra de paradigma.

Como o Jornalismo de Dados é organizado na Folha?

Temos algumas pessoas dentro das editorias que têm uma relação um pouco mais próxima com os dados, não só as pessoas que estão lá como jornalista de dados, mas que têm um pouco mais de aptidão com isso e normalmente se associam ao Núcleo de Inteligência, que está lá para fazer Jornalismo de Dados , ou às pessoas da minha editoria que também trabalham com isso. Eu mesmo me associo ao Núcleo de Inteligência, muitas vezes para desenvolver matérias. Também pode partir da gente. São essas frentes: Tem a Editoria de Audiência e Dados, que tem uma função principal estratégica, mas que também trabalha

com reportagens; tem o Núcleo de Inteligência que é voltado especificamente para Jornalismo de Dados , eles estão lá fazendo só isso; e dentro das editorias temos esses expoentes de pessoas que têm um pouco mais de domínio, que as vezes fazem algumas matérias ligadas ao uso de dados, passando por nós ou fazem entre eles. Tem também o núcleo que atende todas essas pessoas que é o da Arte. Independente do caso, na hora de produzir a visualização você passa por eles, que tem o pessoal que é mais focado no papel e o pessoal que é mais focado no *online* e tem programador.

Há trabalhos que envolvem programadores, designers, cientistas de dados. Como é a distribuição de tarefas na produção de trabalhos que envolvam dados no Folha?

De regra você tem o repórter que vai puxar e analisar esses dados e depois disso, há uma parceria com o pessoal da Arte que ajuda na produção da visualização, seja *online*, seja no papel. Tudo que vai para o papel da Folha, vai para o *online* também, não tem essa divisão. Na Arte você acaba tendo essa divisão porque o pessoal que trabalha em visualização para o *online* precisa saber programar. Temos quatro pessoas que trabalham com programação.

De que forma a programação faz parte da sua rotina de trabalho?

As bases de dados aqui no Brasil são um pouco complicadas, você tem que fazer muitas manobras. Eu não me considero programador, mas eu me desenrasco em *Python*. Eu uso muito programação para automatizar algumas tarefas mais simples. Eu não tenho o costume, por exemplo, de analisar dados em *Python*, eu estou estudando para isso agora. Mas eu gosto de utilizar outras ferramentas. De regra, eu acabo usando o *Tableau* mesmo para a análise. A programação ajuda muito, se você tem conhecimento de lógica e de programação, facilita muito. O *Python* me ajuda no dia a dia muitas vezes para isso, seja para fazer uma limpeza seja para estruturar dados.

De que forma a Folha está investindo nessa especialização do jornalismo?

Tenho visto isso crescer muito, tanto na Folha como nos profissionais dentro da Folha. Estamos, inclusive, montando um *trainee* em parceria com o Google, justamente focado em Jornalismo de Dados , tentando atrair programadores que queiram se especializar um pouco mais em jornalismo e jornalistas que queiram aprender um pouco mais desse mundo dos dados. Esse *trainee* também é aberto para pessoas que são da redação, que também passam por um processo seletivo. Eu sinto muito apoio por parte do jornal nessa área, desde cursos que eu quero e que eu preciso fazer, de receber subsídio. Amanhã e depois tem uma conferência de Jornalismo de Dados aqui no Brasil, aqui em São Paulo, vou participar como ouvinte e o jornal está subsidiando para mim. Tem bastante investimento na formação, tem questão de infraestrutura também. O computador no qual eu trabalho, no qual as pessoas que trabalham mais com dados atuam, são máquinas um

pouco melhores do que o resto da redação, tem um volume grande de processos que têm que rodar ali.

Qual é a maior dificuldade, no Brasil, para a produção de trabalhos Jornalísticos Guiados por Dados?

Qualidade dos dados. Melhorou muito, você tem muito mais disponibilidade de dados, ainda assim você ainda tem pessoas que querem barrar esse tipo de atuação da imprensa. Muitas vezes pegamos em dados que são de um sistema que foi usado uma vez, ninguém tem o programa e temos que ficar convertendo. Às vezes, não é o meu caso por que eu tenho licenças da maioria, há o uso de *software* proprietário, eu vejo alguns colegas reclamarem, isso acaba criando um problema. Tem o fator também da qualidade dos dados que estão lá dentro ou vem com muita sujeira, imputado manualmente com muitos erros de digitação, ou sem um padrão adequado. O problema maior é a qualidade do dado.

Considera que há informação suficiente disponível?

Não, nunca vai ter. Eu acho que quanto mais melhor. Tem muita coisa que eu acho que seria básica, que a gente tem que pedir via Lei de Acesso ou brigar para obter.

Como imagina o auge do Jornalismo de Dados no Brasil?

Estamos longe ainda. Vai chegar um momento que vai ser algo totalmente incorporado. Jornalismo de Dados é jornalismo simplesmente. Conforme vamos nos tornando uma sociedade cada vez mais *data-driver*, isso vai ficar cada vez mais natural e uma habilidade básica mesmo, vai ser totalmente incorporado.

Entrevista com o jornalista Roberto Maleson, do jornal GloboEsporte (Rio de Janeiro)²⁴

Como começou a trabalhar com Jornalismo de Dados ?

Eu conheci essa área no final de 2014, quando eu fiz um curso online. Tem o site *Knight Center for Journalism*, que tem vários cursos *online*, gratuitos, eu fiz um *MOOC*, que é um curso massivo de ferramentas digitais para jornalistas de dados com a Sandra Crucianelli, uma argentina. Lá eu conheci a área, a partir disso, eu gostei do assunto, do tema. Fiz em dezembro desse mesmo ano, um curso presencial da Escola de Dados, que é uma organização que tem promovido bastantes cursos, palestras, eventos, sobre o Jornalismo de Dados para poder espalhar a área dentro do jornalismo. A partir disso eu fui conhecendo o pessoal da área, fui criando um *networking* com a comunidade de dados. Fui tendo interesse, já fiz uns quatro ou cinco cursos na área pelo menos. Por conta disso que eu entrei no GE, eu entrei especificamente por isso, pelo meu conhecimento em

²⁴ Entrevista realizada por chamada de voz, no dia 5 de dezembro de 2017

dados. Entrei na equipe de dados do GloboEsporte e lá fiquei. Na minha equipa, as pessoas que trabalham com dados têm um conceito um pouco diferente do conceito que é o Jornalismo de Dados em geral. Lá o conceito era um pouco arcaico, dados seria trabalhar com tabelas de campeonatos, trabalhar com estatísticas simples de jogos. Hoje o Jornalismo de Dados faz muitas coisas, envolve programação, envolve compilação de dados, análise, limpeza. A gente tem um produto chamado Espiões Estatísticos, ele compila e faz análise de dados de futebol, o problema é que funciona tudo numa planilha de Excel que depois fica muito lento para fazer análise. Como eu sou o que está estudando na área, estou tentando ver formas, bancos de dados, para poder melhorar a nossa performance.

Quais as características que um jornalista deve ter para trabalhar com dados?

Primeiramente tem que estar disposto a trabalhar com um conjunto de dados grandes e saber entrevistar esses dados. Às vezes você pega uma base de dados com milhões de linhas, se a pessoa pega uma base dessas de milhões de linhas, às vezes é uma base suja, com data misturada com texto, palavras que são as mesmas, mas um com acento e outra sem acento, tem que limpar para poder fazer análise posterior. Se você não tem a habilidade de conseguir limpar o dado, de fazer a análise correta, você acaba não conseguindo produzir um conteúdo de qualidade. Os dados são como uma fonte humana, você tem que saber tirar o melhor deles, você não pode simplesmente assumir que são corretos, tem que olhar a metodologia por trás da compilação desses dados, isso se você estiver pegando uma base de dados pública. Você tem que ter habilidade com ferramentas de compilação de dados, de análise de dados e é ideal ter também de visualização. Hoje em dia as redações grandes têm a editoria de arte que são as que fazem as visualizações junto com a equipa de desenvolvimento, mas tem gente que é jornalista de dados por completo, que faz todas as etapas, faz tanto a raspagem de dados, para poder pegar esses dados, para compilar, depois para analisar e visualizar. O ideal é você ser completo, vai depender muito do seu tipo de trabalho.

Pode descrever o seu processo de produção de peças baseadas em dados?

A gente só trabalha com dados sobre futebol. O nosso principal produto é o “Espião Estatístico”, ele funciona da seguinte forma: temos uma planilha mãe, que é o espião, tem as planilhas para cada jogo, assistimos todos os jogos dos 20 principais clubes do Brasil, os 20 clubes que jogam a série A do Brasileiro. Vamos ter os jogos de campeonato estadual, campeonato brasileiro, libertadores, sul-americana, todos os campeonatos importantes oficiais que esses 20 clubes disputam. Se porventura tem algum clube grande que está na série B, também fazemos as planilhas dos jogos desse time. Na medida vamos fazendo a planilha dos jogos, vamos colando nessa planilha mãe. Vemos o jogo, contabilizamos remates, foras de jogo, substituições, cartões, golos, penaltis, os eventos principais que ocorrem numa partida. O futebol não é uma ciência exata, fica mais difícil transformar uma partida de futebol em um conjunto de dados, logicamente já tem isso dos *scouts*, que é contabilizar passes, desarmes, roubadas, mas para a gente vale mais

outras coisas. Nós temos o tempo que os atletas ficam no campo, a partir disso fazemos cruzamento. Por exemplo, o jogador que precisa menos tempo em campo para finalizar ou fazer um gol, temos o ranking Artilheiro da hora que é atleta que precisa menos tempo em campo para fazer um gol. A partir da análise dessa nossa compilação, conseguimos fazer várias matérias, vários produtos. Nós vamos colando tudo na planilha mãe, quando temos a planilha completa, conseguimos fazer análises, saber quem é o jogador que mais levou cartão amarelo por reclamação, a gente consegue fazer pela consulta na planilha mãe que tem todas as planilhas dos jogos. Agora no fim do Brasileiro é o período que a gente faz mais matérias, porque tem um consolidado geral, mas durante o campeonato, antes do brasileiro, durante o ano, sempre vamos soltando produtos. Hoje em dia temos um programa do Sport TV, no “Troca de Passes” toda terça e sexta-feira a gente tem um debate semanal, temos colaborações com o “Esporte Espetacular” e com o “GloboEsporte” na TV. Fazemos as análises, a parte da visualização fica toda por conta da equipa de arte do GE, não sei se daqui a um tempo iremos melhorar isso, mas por enquanto é com eles.

Há uma estrutura pré-estabelecida para comunicar os dados?

A estrutura que a gente tem é a estrutura da planilha de um jogo, a gente sabe como é o esqueleto de cada planilha de um jogo. A gente sabe que no início vamos preencher como vai ser registrado o jogo, como vai ser identificado.

É comum referir que no Jornalismo de Dados deixa-se de entrevistar pessoas e passa-se a entrevistar datasets. Considera que as bases de dados têm autoridade suficiente para serem a única fonte de informação de uma notícia?

Para alguns casos sim. Logicamente, se você vai pegar uma base de dados sobre educação, vou usar o exemplo da Farra do Fies que ganhou o Esso em 2015, lá eles tinham a base de dados da Educação e conseguiram tirar várias análises a partir dessa base de dados. Eu acho importante ter, pelo menos, a metodologia de como foi feita essa compilação e essa metodologia pode ser esclarecida com uma entrevista com quem compilou essas informações. Para alguns casos, resolve só a base de dados, mas se puder entrevistar a fonte humana, trazer mais história para a reportagem, dá uma confiabilidade melhor para o conteúdo. A questão do Jornalismo de Dados no desporto eu ainda vejo muito incipiente, o trabalho que fazemos eu não sei se tem em mais algum lugar no Brasil. Como somos nós mesmo que compilamos os dados, nós confiamos nos dados, se fosse pegar uma base de dados da CBF, se tivesse alguma base de dados boa para usar, talvez teríamos que tentar conversar com algum representante. A princípio, não precisamos entrevistar ninguém, lógico que é sempre bom ter.

Ao fazer a busca pelos dados, quais são as fontes que mais recorre na elaboração dos seus trabalhos?

A nossa compilação é própria, não buscamos em nenhum local na rede para conseguir os dados que precisamos. Como nós mesmos produzimos os dados, então não passamos por

esse problema. Se porventura precisássemos buscar dados em desporto no Brasil, acho que na CBF tem as súmulas das partidas, dos campeonatos oficiais, uma forma de extrair os dados lá seria raspar todas as súmulas e pegar algumas informações. Fugindo do desporto, tem muitas bases públicas disponíveis, a maior parte das informações estão nessas bases de dados públicas e, se não tiver, você pode pedir via Lei de Acesso a Informação.

Com que frequência há a criação de bases de dados na própria redação?

Quando tem a jornada, fazemos a planilha ou no mesmo dia ou no dia seguinte. Sempre que há jogos oficiais de um time da série A, vamos estar fazendo a planilha do jogo para poder colar na planilha mãe para fazer as análises. Esse período de final de dezembro, início de janeiro, que não tem campeonato, ficamos com pautas especiais.

Como surge uma ideia de pauta no Jornalismo Guiados por Dados?

Eu vou citar duas pautas especiais que estou produzindo agora no final de ano, uma é sobre as lesões no futebol. Cada vez que um jogador vai para o departamento médico de um clube da elite do futebol brasileiro, ele entra no registo do levantamento do departamento médico dos clubes. A partir disso, nós fazemos uma compilação, um levantamento geral para saber qual o clube perdeu mais atletas para o DM durante a temporada, qual clube menos perdeu, qual parte do corpo que mais sofre com lesões. A coxa, por ser o músculo mais exigido durante uma partida de futebol, é o que regista mais lesões, mais problemas musculares e clínicos durante uma temporada, a partir disso fazemos umas análises gerais. Outro especial que estamos fazendo, que deve ser lançado em janeiro ou fevereiro do ano que vem, é uma compilação de dados para saber qual o time que menos trocou de treinador, qual a média de permanência de um técnico no futebol brasileiro, qual o aproveitamento para ele ser demitido, quais são os fatores que influenciam. A das lesões teve uma matéria do Washignton Post, eles fizeram tipo um jogador modelo e as lesões foram representadas na parte do corpo do jogador modelo, eles tinham tipo uma bola gráfica no ombro do atleta e tinha quantas lesões tinham sido no ombro durante a temporada. Nós seguimos o mesmo modelo para o futebol, tem uma arte de um jogador modelo de futebol, tem os gráficos nas partes do corpo. Para a matéria sobre os treinadores, surgiu, não tive inspiração. Às vezes é ideia, outras vezes é inspiração. A grande questão é que Jornalismo de Dados eu vejo muito mais em economia e política. Lá fora eu vejo mais no basquete, basebol, futebol americano, no futebol vejo pouco, então não há tantos trabalhos para se guiar. Temos que olhar os outros trabalhos lá de fora, tentar ver o que se pode adaptar ao nosso trabalho, buscar alguma forma de guiar uma pauta para isso.

A seleção do que deve ou não ser publicado é intrínseca à linha editorial do jornal. Considera que há valores notícia no Jornalismo de Dados ?

Sim, temos, mais ou menos, uns princípios. Como já temos o “Espião” e ele já tem um modelo de algumas matérias e reportagens que produzimos durante o ano, já temos o

contexto editorial para publicar alguma coisa. Quando surgem coisas novas, ou assuntos que estão quentes e temos algum dado sobre ele, publicamos. Vai variar muito do que estiver sendo comentado.

Como o Jornalismo de Dados é organizado no GloboEsporte?

Temos uma editoria de dados no GE que é a “Editoria Futdados”, na verdade fazemos tudo relacionado a dados no GE, quando é especial ou qualquer parte gráfica, a editoria de Arte e a Editoria de Desenvolvimento também estão um pouco articuladas. Somos oito contando com o estagiário.

Há trabalhos que envolvem programadores, designers, cientistas de dados. Como é a distribuição de tarefas na produção de trabalhos que envolvam dados no GE?

Acho que precisamos integrar mais, eu estou tentando aprender a programar e a tentar mexer mais em visualização para ficar independente, ainda somos dependentes da equipa de Arte e Desenvolvimento, às vezes não conseguimos produzir conteúdos especiais da forma que queríamos, porque a editoria de Arte e, principalmente, de Programação está envolvida com outros projetos e não tem tempo de fazer os que queremos. Como ainda dependemos muito deles, se não têm tempo, não conseguimos fazer alguns tipos de trabalho, já esbarramos nisso. Queremos um conteúdo, como não tem mão de obra disponível para fazer tudo que o GE tem para fazer, temos que derrubar algumas pautas.

De que forma a programação faz parte da sua rotina de trabalho?

A programação só entra quando temos algum projeto especial grande, é mais para produtos com interatividade, conteúdos maiores. Eu sei que a programação serve para a raspagem de dados, até para a análise e compilação, mas ainda não usamos.

De que forma o GE está investindo nessa especialização do jornalismo?

Primeiro é tendo uma equipa de oito pessoas trabalhando com dados e entregando Jornalismo de Dados do desporto, do futebol, para o público. Também somos muito ligados ao Cartola, que é um *fantasy game* do Brasileiro, que lida muito com estatística. Tem a equipa de *Scout* que faz a compilação dos dados. A equipa do Cartola fica no mesmo local físico que nós, até o ano passado nós eramos do Cartola também. Produzimos conteúdos também pensando no Cartola, as pessoas que acompanham já têm um interesse por estatística, para a gente é um gancho para continuar produzindo conteúdo.

Qual é a maior dificuldade, no Brasil, para a produção de trabalhos jornalísticos guiados por dados?

Primeiro é ter mais gente que trabalhe com dados. Nos Estados Unidos já é muito forte, no Brasil tem muita gente que acha que não é jornalismo. Falta mais especialização, mais cursos e eventos, mais coisas que envolvam Jornalismo de Dados para poder produzir grandes reportagens. Falta mais incentivo de órgãos do jornalismo brasileiro, da Abraji, apesar que a Abraji já está com uma guinada muito forte para o dado, sempre quando tem o congresso deles, sempre tem muitas coisas voltadas a dados. Dá para ver que é uma área nova, mas já está crescendo no Brasil. Acho que falta um bom patamar para chegar ao nível dos Estados Unidos e Europa. Falta mais incentivo das redações, as grandes redações já estão começando a ter mais espaço para o Jornalismo de Dados, às vezes é um investimento grande para produzir dez, cinco conteúdos semanais enquanto as outras editorias estão produzindo muito mais.

Considera que há informação suficiente disponível?

Não, mas está melhorando. Antigamente nós não tínhamos a Lei de Acesso à Informação, pelo menos é um avanço. A partir da LAI nós começamos a ver órgãos públicos disponibilizando mais dados, acho que a transparência está começando a ser impregnada nos órgãos públicos. Ainda falta muito, tem muitos estados que não liberam dados, tem muito local que acha que os dados não devem ser públicos, aí temos que usar dos artifícios do Jornalismo de Dados, um deles é a raspagem de dados. Se alguém não quer disponibilizar, às vezes até tem no site, mas está num formato fechado, está num PDF ou escondido dentro da página, aí tem que utilizar o artifício da raspagem por programação ou por algum robô, tipo WebScrapper para poder tirar essas informações, juntar e produzir conteúdo.

Como imagina o auge do Jornalismo de Dados no Brasil?

O Jornalismo de Dados já tem um bom prestígio. Nas premiações de jornalismo vejo muitas reportagens guiadas por dados nelas e quanto mais reportagens premiadas tiver melhor será, porque a partir disso iremos ter um olhar mais atento das chefias das redações para poder investir mais em dados. Eu acho que as pessoas já estão começando a ter esse olhar, quando tiver equipas mais robustas de dados, com mais especialistas de dados dentro das redações, vamos conseguir produzir conteúdos mais relevantes e a partir disso vamos ter mais pessoas querendo produzir conteúdos com dados.

Entrevista com a jornalista Raquel Albuquerque, do Expresso (Lisboa)²⁵

Como você começou a trabalhar com Jornalismo de Dados ?

Eu conhecia muito pouco de Jornalismo de Dados e em 2012 comecei a fazer parte de um projeto de investigação em Jornalismo Computacional. A ideia era precisamente desenvolver o Jornalismo Computacional que, no fundo, é uma forma mais complexa do Jornalismo de Dados em Portugal. No fundo, o meu primeiro contacto com o Jornalismo

²⁵ Entrevista realizada presencialmente, no dia 26 de dezembro de 2017

de Dados foi nessa altura. Eu comecei a perceber o que era, estava nessa equipa que tinha vários investigadores na área das engenharias e eu estava na redação do Público, o projeto incluía isso, ter uma pessoa de jornalismo na redação do Público, que era um dos parceiros do projeto. Eu, à medida que ia descobrindo o que era o Jornalismo de Dados e como é que se fazia, ia aplicando no Público e os trabalhos iam sendo publicados. A equipa, o projeto de investigação em que eu estava já me ia dando alguma formação, porque eu ia estando com os investigadores todos, através deles fui aprendendo muita coisa. Depois fiz vários cursos online em Jornalismo de Dados, algumas coisas do *Google*, a trabalhar com ferramentas do *Google*, formações do *Excel*. Em Itália é feito todos os anos um encontro sobre jornalismo em Perugia e eles têm sempre uma espécie de formação durante uma semana em Jornalismo de Dados, a primeira coisa que fiz de formação a sério foi nessa semana. A formação que eu fui tendo, em grande medida, foi sempre eu que a procurei, seja através de cursos *online*, seja através desse tipo de Workshop mais alargado, como no de Itália. Em Portugal nunca há, para além de alguns cursos pequenos no Cenjor.

Quais as características que um jornalista deve ter para trabalhar com dados?

Em primeiro lugar ter vontade e muita paciência, não são trabalhos rápidos. Eu estou a falar verdadeiramente do Jornalismo de Dados, precisamente por ser com grandes quantidades de dados é preciso ter muita paciência, perceber logo de início que não é algo que estará pronto no dia seguinte. Eu não acho e tenho vindo a lutar contra essa ideia, de que um jornalista que trabalhe com dados é uma pessoa que sabe imenso de estatística. Um jornalista que trabalhe com dados tem que perceber algumas coisas: primeiro tem de perceber os dados que está a mexer, cada coluna, cada área que ali tens, perceber o valor de cada uma dessas coisas; e depois ter muita paciência porque muitas vezes as folhas de dados não vêm limpas, tens que passar por um processo de correção, de substituição, montes de coisas que são chatas. Às vezes tens de procurar fórmulas no *Excel* que te permitam automaticamente substituir isso ou juntar aquilo que é um trabalho que leva tempo. Segue-se o processo jornalístico. Óbvio que é preciso ter um conhecimento razoável de estatística em geral, mas não aprofundado, ou de matemática, para trabalhar com dados.

Pode descrever o seu processo de produção de peças baseadas em dados?

A obtenção dos dados é a parte mais complexa e em Portugal nota-se muito isso porque não tens assim tantos dados disponíveis de coisas que verdadeiramente interessam. Não tens muitos dados disponíveis por muitas razões, uma delas é porque não há uma grande abertura e não há essa consciência, até mesmo por entidades públicas, de disponibilizar os dados. Há muito receio, há receio de que os jornalistas usem os dados de forma errada. Se tu chegares aos dados dessa forma que é, ou encontrar ou pedir o registo que queres, o processo basicamente é: pegares naquele *Excel*, se não tiver no *Excel*, arranjar uma maneira de colocar, às vezes podem estar num *PDF*. A partir do momento em que tens os dados no *Excel*, o objetivo é primeiro perceberes o que é que tens, perceber cada coluna dessa folha, qual o significado, qual o peso que cada informação daquelas tem na

caracterização do caso. Perceberes se para todos os casos tens informação, se todas as linhas estão preenchidas ou se só estão algumas. A primeira parte do processo passa precisamente por limpar a folha que temos e perceber quais são as colunas e as informações que conseguimos efetivamente usar. O segundo passo é começar a analisar os dados que temos. Isso passa, por exemplo, por utilizar as tabelas pivô que é uma ferramenta que o *Excel* tem para fazer as contagens, para ver como evoluiu. É nesta fase que se não soubermos o que fazer convém falar com alguém e esse alguém é sempre especialistas naquela área. Se for claro e não tivermos dúvida é seguir em frente, apontar as diferentes conclusões. Quando já temos, eventualmente até uma notícia, é encontrar uma maneira de contar aquilo da melhor forma. Temos, ao mesmo tempo, a parte da infografia e da visualização de dados, dependemos um pouco da dimensão que queremos, dos recursos que temos disponível,

Há uma estrutura pré-estabelecida para comunicar os dados?

No Expresso temos a equipa de infografia, fazem infografias tanto para o papel como para o *online*. Em geral são infografias mais estáticas. Mas, de há um tempo para cá, começámos a usar ferramentas tipo *Tableau* para fazer coisas um pouco mais interativas porque o Expresso ainda tem um problema, que é não ter programadores na redação. O que conseguimos fazer é com ferramentas desse género, coisas com *Tableau* ou em que pomos os dados e alguém da infografia tem o conhecimento para dar um jeito ou eu tento fazer à minha maneira e vou pedir à Infografia ajuda com as cores, tipos de gráficos, entre outros. Para fazer visualizações interativas precisas de alguém que te ajude com a parte da programação. Um dos maiores trabalhos que fiz quando na altura estava no Público foi porque isso existia ali. O Expresso tem uma equipa de Infografia, tem uma equipa multimédia, e essa equipa já tem uma pessoa que percebe mais de *motion graphics* que já faz a parte interativa dos trabalhos de multimédia. Depois temos a equipa de programação que conseguiria pôr a funcionar coisas mais interativas, mas que trabalha para o grupo inteiro. No Expresso o retrato é um pouco esse: sobretudo infografia estática.

É comum referir que no Jornalismo de Dados deixa-se de entrevistar pessoas e passa-se a entrevistar datasets. Considera que as bases de dados têm autoridade suficiente para serem a única fonte de informação de uma notícia?

Eu diria que não, a parte interessante e o valor que o Jornalismo de Dados tem não é só explorar uma base de dados, porque isso deixa o trabalho a meio. Se estivermos a fazer uma visualização de dados é aquilo e pronto, criou-se uma ferramenta que permitiu visualizar o que está numa base de dados. Se estivermos a falar de Jornalismo de Dados, tu estás a fazer jornalismo, não estás só a visualizar dados. Fazer jornalismo com base em dados é, precisamente, procurar notícias naqueles dados, procurar histórias, pessoas que dão sentido, ir à procura de especialistas que expliquem o que está nos dados. Na minha abordagem de conceito de Jornalismo de Dados, um *dataset* não funciona sozinho, isso é visualizar dados. Fazer Jornalismo de Dados é agarrar num *dataset*, para tirar dali informação, para perceber e, em último caso, conseguir que aquilo tenha efeito. Idealmente eu acho que não.

Ao fazer a busca pelos dados, quais são as fontes que você mais recorre na elaboração dos seus trabalhos?

Eu tento procurar em sites e em instituições públicas, primeiro por uma questão de fiabilidade dos dados e depois porque há muitos dados que deveriam de facto ser públicos e não são. Eu lembro de numa altura no Público, fizemos um trabalho sobre o financiamento das campanhas autárquicas e todos os dados de financiamento de campanha são dados obrigatoriamente públicos. O problema é que é uma quantidade tão grande de informação que aquilo está tudo em papel e em corredores de arquivos no tribunal constitucional, mas aquilo é público, eles não fecham o acesso aos dados, se quisermos saber uma coisa, eles dizem: "Sim, é público, qualquer cidadão pode vir pesquisar, nós só não temos capacidade de digitalizar tudo". Eu tento procurar dados em instituições, institutos, entidades públicas, temos o peso de ter mesmo de ser público. Acho que a parte mais difícil é essa. É muito raro ter os dados, logo a partida, disponíveis. É uma busca não no sentido de encontrar, mas de contactar as instituições, de insistir e, às vezes, mandam os dados já trabalhados e não é como queremos.

Com que frequência há a criação de bases de dados na própria redação?

Não há. No Expresso há pessoas que percebem imenso de estatística e que fazem bons trabalhos, que podem não ser no conceito de Jornalismo de Dados, mas que fazem muitos trabalhos com dados. Eu imagino que para uma redação produzir bases de dados é preciso no mínimo ter mais de uma pessoa que esteja todos os dias a fazer Jornalismo de Dados. Estou sozinha e a rotina diária e quantidade de coisas que é preciso fazer e dar resposta, às vezes bloqueia. Eu acho que numa redação que não tenha um jornalista de dados é muito difícil.

Como surge uma ideia de pauta no Jornalismo Guiado por Dados?

Depende um bocado. Tanto pode ser alguém que tenha a ideia e diz "porque não fazer uma coisa sobre isso?", então eu agarro aí e penso onde eu posso buscar dados sobre isso e a quem eu posso ir pedir. Ou então cruzar com uma base de dados, uma coisa que eu nem sequer sabia que existia e o ponto de partida é precisamente aquela base de dados. Pose também alguma coisa que esteja a acontecer, ligada à atualidade, sobre o qual há pouco conhecimento. Por exemplo os incêndios, todos os anos há trabalhos sobre os incêndios com base em dados, há muitos dados em bruto.

A seleção do que deve ou não ser publicado é intrínseca à linha editorial do jornal. Considera que no Jornalismo de Dados há esse filtro com base nos valores notícia?

Sim, no Expresso temos o Papel, temos o Diário, temos o Site e temos a Revista, isso significa que são níveis de abordagem, de detalhes, de velocidade completamente diferentes. Se for uma coisa extremamente relevante e se percebermos que só eu tenho isso, mais nenhum jornal vai ter, se calhar guardamos para o papel. Caso seja alguma

coisa que toda a gente vai ter, se for mais rápida, se calhar sai no site. Mesmo em histórias com dados, como em qualquer história, há decisões editoriais, mas consoante o grau de interesse é que poderá sair ou como uma peça rápida para o site ou como uma peça mais detalhada e que sai no papel para ter mais destaque, qualquer peça com dados passa pela mesma avaliação que qualquer outra peça jornalística.

Como o Jornalismo de Dados está organizado no Expresso?

No Expresso, efetivamente, só eu é que estou como jornalista de dados, portanto, não há uma equipa. Nos últimos meses estava uma estagiária que queria muito trabalhar com dados e tem estado a fazer algumas coisas. Eu já passei por várias secções, neste momento eu faço parte da secção sociedade, faço coisa para o papel, para o diário, para o site e para a revista como todos nós no geral. Há uma equipa de Infografia que trabalha para o jornal todo e para a Revista, há uma equipa de multimédia que vai fazendo sobretudo trabalhos de vídeos. Eu estou na Sociedade sobretudo a fazer temas de sociedade, o que não invalida, por exemplo, se há necessidade de fazer um trabalho com dados de Política, em que eu faça na mesma. Infelizmente não há um programador na redação, essa é a parte que eu sinto mais distância, porque a equipa de programação trabalha para o grupo inteiro e não há proximidade de dizer para fazer um trabalho sobre isso e concretizar esse trabalho. Eu sou a única a fazer, ainda que eu não esteja todas as semanas a fazer Jornalismo de Dados, nem trabalhos com dados. Estamos longe de ter uma equipa de dados.

Há trabalhos que envolvem programadores, designers, cientistas de dados. Como é a distribuição de tarefas na produção de trabalhos que envolvam dados no Expresso?

A recolha é toda minha, o tratamento é todo meu, depois o que acontece é a infografia mais estática ou eventualmente usando algumas coisas tipo *Tableau*. No fundo o que acontece é eu converso com a Infografia e digo "era engraçado fazer isso, com essa dimensão", passo os dados e depois a Infografia trabalha essa parte. Quando estamos a falar de algo maior, e é raro conseguirmos fazer algo maior, há designers, há uma *webdesigner* dentro do jornal que também dá uma ajuda. Quando são coisas como o "Ranking das Escolas", ou mesmo maior, de vez em quando acontece, tem que se ir às tais equipas que servem o grupo inteiro, trabalham para a Sic, para Visão. Quando estamos a falar de produtos editoriais, não há disponibilidade para isso.

De que forma a programação faz parte da sua rotina de trabalho?

Da minha rotina não faz. Agora nessa pós-graduação que eu estou a fazer em Visualização de Informação a minha ideia é conseguir ser o máximo possível autónoma. Há bibliotecas de visualização de dados como *D3* e coisas do género, que eu estou a tentar aprender, principalmente para eu conseguir ter essa liberdade de já não ter que depender de um programador para fazer algo que para eles é super básico. Em termos de programação não tenho conhecimentos.

De que forma o Expresso está investindo no Jornalismo de Dados ?

É óbvio que eu vejo como investimento o facto de, por exemplo, na altura em que o Diário começou, em 2014, me terem contratado precisamente porque eu fazia Jornalismo de Dados . A partida, traduz uma aposta no Jornalismo de Dado e, no mesmo sentido, os editores, a direção e até mesmo a administração reconhecem o valor e a importância do Jornalismo de Dados . O problema é que na prática, acaba por ser muito complexo, não só por eu estar sozinha a fazer isso e ter que ter as ideias, procurar os dados, mas também ter de lutar contra a voragem do tempo, a necessidade de ter notícias e, tudo isso junto, faz com que às vezes o Jornalismo de Dados , a investigação, o jornalismo feito com mais tempo, acabe por ser diluída. Eu acho que o Expresso apostou no Jornalismo de Dados , apostou no facto de reconhecer que é uma área importante do jornalismo e que pode trazer bons trabalhos e eles têm apostado nisso, no sentido de puxar Jornalismo de Dados . Óbvio que se calhar ajudaria ter mais pessoas a fazer isso e sobretudo ter mais tempo, perceber-se que é preciso tempo para fazer um trabalho de Jornalismo de Dados .

Qual é a maior dificuldade, em Portugal, para a produção de trabalhos jornalísticos guiados por dados?

Primeiro a falta de dados e isso em geral tem a ver com o facto de culturalmente haver uma certa desconfiança na disponibilização dos dados. Mesmo quando pedimos dados às instituições, que têm de os dar, eles já vêm trabalhados e o conceito de Jornalismo de Dados não é trabalhar com dados trabalhados. Em segundo lugar, a falta de tempo. Estamos todos, permanentemente, a correr e parar para fazer um trabalho com tempo é algo que tem muito pouco espaço e isso é uma pena. Quando falamos de trabalhos com dados, que requerem andar atrás de alguém que nos dê e que depois temos que os trabalhar, é um processo que leva tempo.

Você considera que há informação suficiente disponível?

Não é falta de informação, não somos um país onde não há produção de estatística. O Jornalismo de Dados , no seu conceito original, visa interpretar e analisar dados em bruto, é esse ponto original que é muito difícil de achar.

Como imagina o auge do Jornalismo de Dados em Portugal?

Não consigo antever o que poderia ser o auge, eu imagino que o Jornalismo de Dados podia melhorar muito se nas próprias redações e nos jornais reconhecessem a necessidade de tempo para fazer esses trabalhos, mas isso não é só no Jornalismo de Dados , é o vídeo, é o multimédia e é o próprio jornalismo. O jornalismo está a viver uma crise, não é de agora, mas brutal, pessoas a serem despedidas, essa realidade não pode ser esquecida. É óbvio que se houvesse uma estrutura e algumas coisas diferentes e uma atitude em relação ao tempo mais facilitada, se por outro lado houvesse mais dados disponíveis e sobretudo uma maior capacidade de perceber que tipo de dados é que nós estamos a falar e que

poderiam estar disponíveis e se fossem cada mais frequentes esses tipos de abordagens, acho que o Jornalismo de Dados poderia de facto ser um contributo ótimo para a própria sociedade. No final objetivo é esse, conseguir fazer um retrato de várias realidades, conseguir identificar problemas, falhas, com o intuito de as melhorar.

Entrevista com a jornalista Rita Costa, do jornal Público (Lisboa)²⁶

Como começou a trabalhar com Jornalismo de Dados ?

É recente, eu trabalho no Público há cerca de quatro meses, mas já estava a investir na formação nessa área há algum tempo. Eu trabalhava numa revista de agricultura, algo pequeno, sabia que não era isso que eu queria fazer e sabia que para conseguir um trabalho na área do jornalismo tinha de diferenciar. Eu gosto dessa parte de estatística e da análise dos dados e comecei a investir, a fazer cursos *online*, até que surgiu essa oportunidade no Público. Eu estudei Jornalismo e depois da licenciatura fiz alguns cursos do *Knight Center*, eles têm algumas coisas de Visualização de Dados e Jornalismo de Dados com Alberto Cairo. Depois fiz um curso de Comunicação Visual, no ISCTE, já também com essa intenção e depois ia tentando fazer nos meus tempos livres o que eu achava que podiam ser potenciais peças para o meu portfólio. As minhas referências em Portugal eram a Raquel Albuquerque, que fazia algumas coisas e eu ia vendo no Expresso, mas mesmo assim em termos de visualização e infografias, eles não faziam muitas coisas. As minhas referências eram todas de outros países, as coisas que o New York Times faz e a Vox.

Quais as características que um jornalista deve ter para trabalhar com dados?

Ter muita disponibilidade mental para trabalhar no *Excel* e nesses programas, o dia inteiro. Às vezes eu sinto que eu passo o dia inteiro a olhar para tabelas do *Excel*. Ser muito analítico e ter muita atenção ao detalhe, porque se estás a fazer uma história baseada em números, se te enganas em algo a tua história muda, é preciso muito ter esse cuidado.

Pode descrever o seu processo de produção de peças baseadas em dados?

Há duas formas diferentes, ou eu tenho a ideia ou alguém diz: “Olha, precisamos disso”. Quando eu tenho a ideia, o ideal é formular um conjunto de hipóteses e perceber que dados preciso para confirmar ou desmentir essas hipóteses. Vou à procura dos dados e, normalmente, se forem coisas de Portugal é bastante difícil. As coisas têm evoluído em termos de disponibilização dos dados, mas ainda há um longo caminho a percorrer. Normalmente os formatos que eu estou mais à vontade é no *Excel* ainda, mas estou a investir em aprender a programar em *Python*, acho que é importante. Quando abro o *Excel*, algo muito importante é limpar as tabelas e ordenar aquilo de forma a que depois eu possa fazer a análise. Depois, normalmente, uso muito o *Tableau* para visualizar as tendências nos gráficos e ajudam, em termos analíticos, a perceber se há ali alguma

²⁶ Entrevista realizada presencialmente, no dia 29 de novembro de 2017

história ou não. Depois de perceber o que descobri nos dados, falar com especialistas é crucial, ver o que está por trás dos dados, isso é fundamental. É muito raro uma história viver só de números, às vezes acontece, mas não acho que seja o melhor cenário. Lá no Público não sou eu que faço as infografias, nem as visualizações, temos uma equipa que é bastante boa a fazer isso, normalmente passo-lhes os dados e discutimos a melhor forma de colocar a informação, eles a parte técnica fazem. Outra parte, quando não há os dados, é importante percebermos que em muitos casos nós podemos pedir, há a Lei do Acesso aos documentos administrativos e nós podemos exigir o acesso aos dados.

Há uma estrutura pré-estabelecida para comunicar os dados?

Não, há muita gente que já faz isso lá, mas de forma informal. Estar a fazer uma história e perceber que há a necessidade de ir buscar dados que a suportem. Há histórias mais completas em números e que a intenção é a mesma, utilizar os dados e analisar os dados.

É comum referir que no Jornalismo de Dados deixa-se de entrevistar pessoas e passa-se a entrevistar datasets. Considera que as bases de dados têm autoridade suficiente para serem a única fonte de informação de uma notícia?

Eu acho que não, podes estar a olhar para um conjunto de dados, se não fores aplicar isso à realidade nunca vais saber o que aquilo significa. Como é que isso se traduz na vida das pessoas? Acho que os dados são muito importantes, mas nós não podemos descorar as pessoas. Há uns tempos escrevi um trabalho sobre o facto de nunca ter havido casamentos homossexuais em 50% dos concelhos do país, mas em que isso se traduz na vida das pessoas? Depois fui falar com casais homossexuais, com pessoas que trabalham no interior do país, porque de facto ainda há um grande preconceito em relação ao casamento homossexual nessas zonas do interior, mais rurais e isso os dados não dizem, não dizem o porquê e o porquê é essencial responder.

Ao fazer a busca pelos dados, quais são as fontes que você mais recorre na elaboração dos seus trabalhos?

Os institutos de estatísticas oficiais, INE, Eurostat, gabinetes de estatísticas de outros países. Nos casos dos fogos, recorríamos aos dados que vinham do *Copernicus*, do satélite. Depende do tema, porque dependendo do tema há mais dados ou menos dados disponíveis. Agora há um site muito bom do Governo que é o *dados.gov*, onde eles têm muitos *datasets*, que basicamente existem em outros sítios, mas que ali estão mais *clean* e são mais fáceis de utilizar. Eu gosto muito de utilizar dados geográficos, porque acho que também podem contar uma história e aí temos o portal de dados georeferenciados da Câmara de Lisboa.

Com que frequência, no Público, há a criação de bases de dados na própria redação?

Ainda não estamos nesse ponto, até porque acho que para fazer isso é preciso algum conhecimento de estatística, para fazeres um *dataset* e inferires daí uma notícia ou reportagem é preciso ter algum cuidado. Acho que ainda é preciso dar esse passo, era importante.

Como surge uma ideia de pauta no Jornalismo Guiados por Dados?

Muita coisa vem da minha percepção do que pode ser feito e do que pode ser interessante para o leitor. Outras coisas surgem pela atualidade. Quando foi altura das Autárquicas, fizemos muitas coisas de análise e análises históricas e isso surge simplesmente porque é atualidade. Na altura dos fogos também fizemos muitas coisas. Há duas vertentes: quando fazes uma reportagem essa noção de que esse tema pode ser interessante para o leitor ou a atualidade.

A seleção do que deve ou não ser publicado é intrínseca à linha editorial do jornal. Considera que há valores notícia no Jornalismo de Dados ?

Acho que os valores são os mesmos, não há diferença. Tens de ter cuidado com as fontes.

Como está organizado o Jornalismo de Dados no Público?

A única pessoa que faz, normalmente, Jornalismo de Dados sou eu. É algo muito embrionário, no Público e em todas as outras redações. O objetivo é aumentar as equipas e que isso seja uma parte estruturante do trabalho diário.

Há trabalhos que envolvem programadores, designers, cientistas de dados. Como é a distribuição de tarefas na produção de trabalhos que envolvam dados no Público?

Lá, em termos de programadores não temos ninguém, temos para o site, mas para reportagens não há nenhum programador que nos ajude. Em relação aos *designers*, há muito isso de trabalhar os dados em conjunto e perceber o que e o que não faz sentido colocar, como é que determinar visualização pode ajudar o leitor a perceber ou a complementar a notícia.

De que forma a programação faz parte da sua rotina de trabalho?

Ainda é uma parte muito básica da minha rotina. Eu nunca fiz um trabalho em que eu só utilizei o *Python*, mas o que eu tento fazer é, quando tenho um trabalho que eu sei que dá para usar, vou tentando ganhar prática.

De que forma o Público está investindo nessa especialização do jornalismo?

Acho que bastante. Todos os jornais que percebam que este é o caminho, que investir nos conteúdos online e interativos é caminho, vão investir nessa área, mesmo que seja pelo reconhecimento de que isso é importante. Além de mim, há outras pessoas que se estão a formar na área.

Qual é a maior dificuldade, em Portugal, para a produção de trabalhos jornalísticos guiados por dados?

A falta do acesso à informação e muita da informação que está disponível não está feita para ser utilizada, consumida por um computador. Por exemplo, os dados da Comissão Nacional de Eleições, aquilo é escabroso, eu passei horas e horas a limpar aqueles dados para conseguir fazer alguma coisa com eles. Ainda há muitas coisas que não são medidas e que não existem em dados e isso cria alguns constrangimentos. Há trabalhos incríveis que se fazem nos Estados Unidos e eram impossíveis de fazer cá porque simplesmente esses dados não estão disponíveis.

Você considera que há informação suficiente disponível?

Não é que não haja informação suficiente, há muita coisa que se pode fazer, mas acho que não há como poderia haver. Eu tenho a certeza que existe muita coisa que não é tornada disponível, ficam nas empresas e organizações.

Como você imagina o auge do Jornalismo de Dados no Portugal?

Idealmente como se calhar é agora em outros países, nos Estado Unidos. Eu acho que nós vamos evoluir e caminhar para o que se faz em outros países, não sei o que será o auge, mas idealmente ter dados abertos que nós possamos utilizar sem constrangimentos, equipas multidisciplinares com programadores e *webdesigners* para podermos fazer coisas incríveis.

Entrevista com o jornalista Rui Barros, da Rádio Renascença (Vila Nova de Gaia, Porto)²⁷

Como você começou a trabalhar com Jornalismo de Dados ?

Eu licenciiei-me em Ciências da Comunicação na Universidade do Minho. Segui a área de Ciências e Tecnologias no ensino secundário, então já tinha uma certa formação em matemática. Entrei no curso em 2012, precisamente no ano do Edward Snowden. Comecei a seguir exemplos do que era feito lá fora, o Guardian, o New York Times e foi aí que comecei a ganhar algum interesse. Na altura de chegar ao estágio, já no final do meu mestrado, fui para a Renascença, falei que gosto dessa área e que queria fazer um bocado disso, comecei a mostrar alguns trabalhos e isso mostrou que havia algo de diferenciador e foi assim que eu comecei a trabalhar na área. Eu aprendi online, fui

²⁷ Entrevista realizada por chamada de voz, no dia 30 de novembro de 2017

seguinto o que os outros andavam a fazer, o meu primeiro contacto com ferramentas foi um dos cursos do *Knight Center*, depois fui seguindo, vendo tutoriais.

Quais as características que um jornalista deve ter para trabalhar com dados?

Primeiro é preciso ser jornalista, não no sentido de formação, no sentido de saber fazer perguntas, precisas sempre ter um foco jornalístico e saber fazer perguntas é essencial. Saber fazer perguntas não chega porque tens esse lado muito técnico que é preciso explorar, perceber de tipos de ficheiros, o que é um *CSV*, porque é diferente de um *XLS*. Aprender sobre estatística é essencial, acho que a área de Visualização de Dados também é necessária, principalmente se trabalhares em uma redação pequena, que não tenha uma pessoa específica só a fazer visualização de dados.

Pode descrever o seu processo de produção de peças baseadas em dados?

Uma coisa que eu notei, é um problema grave em Portugal, há muitas poucas bases de dados de acesso livre. As instituições públicas ainda têm um problema muito grave em disponibilização de dados, quando disponibilizam estão em *PDF*. É raro o meu processo começar com «tenho uma base de dados sobre isso, vou tentar encontrar respostas». Normalmente eu tenho ideias ou o meu editor diz que era interessante ter um sobre isso ou sobre aquilo, depois faço trabalhos de jornalista, quer seja para procurar em documentos ou ligar para umas instituições públicas. Temos o problema que é a maior parte das bases de dados estão em formatos que não são muito práticos de trabalhar, mesmo alguns que estão em *Excel* estão muito sujos, tenho sempre que limpar os dados. Normalmente uso o *Open Refine* para fazer essa limpeza ou simplesmente o *Excel*, quando é algo mais rápido. Depois na análise, como eu faço coisas que têm sempre uma certa urgência de publicação, acabo sempre por fazer uma análise mais rápida no *Excel*, quando o *Excel* já é muito limitado recorro ao *R* ou ao *Pandas*, que é uma biblioteca de *Python*, aí já são casos em que tenho mais tempo para poder analisar os dados. Quando já tenho um *insight*, vem o processo que eu considero que é o mais doloroso, vem o momento de escrever a história e eu sinto uma certa dificuldade, eu costumo dizer que é passar do cérebro de engenheiro para o cérebro de jornalista, que é: Já tens as conclusões dos teus dados, mas como está num *mindset* muito matemático, tenho uma certa dificuldade, os dados para mim já são conclusões banais. Mas sim, depois passo para o processo de escrita e, normalmente, passo para o processo de Visualização de Dados em que habitualmente sou eu que programo as minhas infografias. Na Renascença só sou eu a trabalhar com dados, vou tendo uma ajuda do nosso *designer* ou de outros jornalistas. Normalmente programo as infografias e faço eu as minhas próprias Visualizações de Dados. Para isso, tenho a minha ideia e costumo fazer um rascunho em papel e a seguir programo, normalmente, uso a biblioteca *D3* para programar as minhas infografias.

Há uma estrutura pré-estabelecida para comunicar os dados?

Eu sou jornalista do online, mas nós somos uma rádio. Claro que essa Visualização de Dados depois não passa para a rádio, mas a conclusão do meu trabalho de dados passa.

Eu tenho sempre a preocupação, normalmente sempre com ajuda de alguém que também seja jornalista de rádio, de comunicar aquele processo para a pessoa que está a fazer rádio.

É comum referir que no Jornalismo de Dados deixa-se de entrevistar pessoas e passa-se a entrevistar datasets. Considera que as bases de dados têm autoridade suficiente para serem a única fonte de informação de uma notícia?

Eu acho que não, ao contrário dos meus colegas, o meu processo normalmente começa sempre por entrevistar a base de dados e eu entrevisto e tiro conclusões, mas no momento de escrever a peça, tenho dificuldade em escrever só com base na base de dados. Isso não é linear, mas quem sou eu às vezes para tirar conclusões quando nem sou especialista na matéria. Eu diria, quase sempre, que tenho necessidade de falar com alguém que é um especialista na matéria. Sinto sempre a necessidade de ter outras pessoas a explicarem-me aqueles dados. Há um outro tipo de história, que eu gosto mais de fazer, que é o momento em que o Jornalismo de Dados é invisível. Por exemplo, a Renascença fez diversas reportagens, na altura das eleições, sobre o país, sabíamos que íamos falar sobre a questão do arrendamento local, então eu analisei uma base dados sobre *Airbnb*, conclui qual era a rua no Porto com mais *Airbnbs* e a seguir fizemos reportagem com pessoas sobre o alojamento local, o leitor quando vê não sabe que houve um trabalho de dados profundo, de geolocalização e tudo mais. Houve um trabalho prévio que está quase ali subentendido.

Ao fazer a busca pelos dados, quais são as fontes que você mais recorre na elaboração dos seus trabalhos?

A Pordata é sempre o sítio onde eu começo, só que os dados da Pordata são sempre insuficientes, é sempre para dar uma ideia. Depois acabo por fazer pesquisas nas instituições públicas em concreto e às vezes não há dados ou quando há estão em bases de dados que é preciso fazer *scrape*. Cada trabalho é um trabalho, não acho que tenho fontes recorrentes, para além da Pordata, INE e Eurostat. Acho que a maior parte acaba por ser fontes governamentais, mas há esse problema, a maior parte delas não disponibiliza dados. Eu diria mais instituições públicas.

Com que frequência há a criação de bases de dados na própria redação?

Criação pura, nunca fizemos. Eu estou na Renascença há um ano, isso é um processo ainda em crescimento, mas obviamente que isso estará nos quadros. Trabalhar uma base de dados que já está recolhida é muito mais fácil do que perder o tempo a formar a base de dados. É um processo que demora muito mais do que propriamente recorrer a uma instituição que já tenha a base de dados.

Como surge uma ideia de pauta no Jornalismo Guiados por Dados?

Há aquela altura que eu chego ao meu editor e digo que “era interessante fazer alguma coisa sobre isso” ou está em agenda este tema e vamos fazer um trabalho sobre isso, ou vou procurar uma base de dados ou já tenho uma base de dados. Outras vezes é o editor que diz que o tema é interessante, vamos procurar e tentar fazer. Há uma terceira, que não acontece só no Jornalismo de Dados, como eu sou obrigado todos os dias a aprender uma coisa nova no sentido que o campo está a mudar sempre, é preciso sempre estar em cima do que os outros estão a fazer lá fora. Obviamente quando vejo um trabalho do Guardian, penso que se calhar é interessante fazermos também uma coisa sobre o mesmo tema para cá, isso também influencia.

A seleção do que deve ou não ser publicado é intrínseca à linha editorial do jornal. Considera que há valores notícia no Jornalismo de Dados ?

Sim, é jornalismo, por isso que todos os princípios do jornalismo normal estão intrínsecos. Obviamente que eu só vou fazer um trabalho sobre determinado tema porque está na agenda ou então porque tem valor notícia. No caso do incêndio de Pedrogão, na Renascença, eu fui analisar uma base de dados sobre incêndios e concluí que o de Pedrogão foi o 10º mais mortal desde 1900, uma base de dados que obtive, eu só dei essa notícia porque tem valor notícia. Há um conjunto essencial em todo o tipo de jornalismo que me permitem transformar esta primeira análise que eu fiz a base de dados, transformá-la numa notícia. Acho que os princípios são os mesmos.

Como o Jornalismo de Dados está organizado na Rádio Renascença?

A aposta no Jornalismo de Dados na Renascença é um projeto relativamente recente, eu fui para lá, apesar de não fazer exclusivamente, também faço informação dita normal, com o intuito de fazer crescer o projeto. Eu estou integrado na redação Web, estou lá dedicado a fazer Jornalismo de Dados. Ou eu ou um editor faz uma proposta de trabalho e a seguir eu faço a recolha e a análise de dados e a limpeza se for necessário. Normalmente tenho sempre esses pontos de contato com o editor, eu fiz a análise dados, tirei conclusões, falo com o editor, aí está a importância da experiência jornalística, dizer-me que estou a esquecer-me desse fator ou disto. Exclusivamente com dados sou eu, isso não quer dizer que eu trabalhe sozinho com dados, muitas das vezes trabalho com jornalistas de uma área, que percebem da área, para me ajudarem a perceber os dados. Saber limpar uma base de dados é relativamente fácil, agora conseguir interpretar o que os valores dizem, já é preciso ter um conhecimento na área e conto sempre com esse conhecimento da redação.

Há trabalhos que envolvem programadores, designers, cientistas de dados. Como é a distribuição de tarefas na produção de trabalhos que envolvam dados na Rádio Renascença?

Normalmente, o trabalho de análise de dados passa por mim, o processo de elaboração e programação da infografia passa por mim e com a ajuda de um *designer*, porque infelizmente eu trabalho no Porto e o *designer* em Lisboa, não é aquele contato direto, de

trabalhar lado a lado. Normalmente eu faço uma primeira versão, envio-lhe e ele diz que se calhar não usava essa cor, que o tipo de letra fica péssimo, que mudava isso e fazia assim e vamos trabalhar assim em versões. Depois, depende do tipo de trabalho a relação com a informática, quando são trabalhos mais simples em que é só preciso fazer um *embed*, eu só faço o upload para o servidor da Renascença. Quando são trabalhos que envolvem mais construções especiais de multimédia, aí já há uma relação com a informática.

De que forma a programação faz parte da sua rotina de trabalho?

Tenho conhecimentos de programação que me ajudam no meu dia a dia, não sou um programador. Fala-se sempre das equipas interdisciplinares, da importância de ter um *designer*, um programador e um jornalista a falarem. Eu acho isso essencial, eu gostava de ter um programador e um designer só em dados e não a fazerem tudo que fazem na Renascença, que é imenso. Por exemplo, recentemente eu tive um caso em que eu precisava desesperadamente de uma base de dados que não havia, eu tive que escrever um *Scraper* em *Python*, fui eu que programei e o *Scraper* fez o trabalho e mesmo a minhas infografias são programadas por mim. Eu aprendi tudo sozinho, a perceber como funciona e a usar muito código aberto que está no *GitHub*, tenho uma rotina, mas não me considero programador.

De que forma a Renascença está investindo nessa especialização do jornalismo?

Não querendo estar a fazer um autoelogio, acho que a Renascença está a ser pioneira em apostar em Jornalismo de Dados em Portugal ao ter decidido apostar num jornalista de dados já é uma grande decisão, num momento em que a indústria mediática não está a contratar. Para além disso, dentro das possibilidades, há sempre a abertura, ainda recentemente tivemos o encontro em Lisboa da *Global Editors Network* e a Renascença não só me deu o dia, como também pagou as despesas para estar em Lisboa, acho que mostra uma certa vontade da Rádio em apostar na área.

Qual é a maior dificuldade, em Portugal, para a produção de trabalhos jornalísticos guiados por dados?

Bases de dado, acima de tudo. Sinto que há muita falta de dados para trabalhar, de dados abertos, porque às vezes até há outros dados, ou estão acumulados em algum ministério e é muito difícil consegui-los ou ia demorar tanto tempo em transformá-los em dados trabalháveis que o trabalho seria inviável na lógica da redação. Acho que essa é a grande dificuldade. Depois, enquanto leitor de Jornalismo de Dados, eu sou leitor do Guardian, do FiveThirtyEight, do New York Times, do Washington Post que têm equipas de dados, pessoas a trabalhar em dados, olhamos para as fichas técnicas e são três, quatro pessoas; na Renascença é uma pessoa. Isso dá nas certas limitações porque não temos um número de pessoas com conhecimentos técnicos. A minha formação é jornalismo e por isso muitas das vezes sinto que não tenho conhecimento de estatística ou de programação para fazer aquilo.

Considera que há informação suficiente disponível?

Acho que há informação, mas muitas das vezes essa informação está ‘intrabalhável’.

Como imagina o auge do Jornalismo de Dados em Portugal?

O Guardian deve ter sido das primeiras redações a apostar na área e se não tivessem sido as revelações do *Wikileaks*, o Jornalismo de Dados não teria se desenvolvido tanto. Acho que ainda não tivemos uma grande história que envolvesse dados em Portugal, acho que quando isso acontecer, uma grande revelação, as restantes redações vão perceber um bocado mais a importância e vão apostar mais na área. Não estou a criticar as universidades, porque até acho que as universidades têm estado mais desenvolvidas do que as próprias redações. Eu posso não ter tido uma cadeira em específico, mas por exemplo tive uma disciplina onde fiz só Jornalismo de Dados, acho que quando as universidades também incorporarem mais ou uma cadeira ser mais frequente, obviamente que isto vai despertar a atenção.

Anexo 3 – Questionários enviados aos jornalistas que atuam na produção de conteúdos através de Jornalismo de Dados , no Brasil e em Portugal

Brasil

Perfil: Quem faz Jornalismo de Dados no Brasil

Para o efeito de obtenção do grau mestre em Jornalismo pela Universidade Nova de Lisboa, este questionário tem como objetivo traçar o perfil do profissional e ajudar a compreender como é feito o Jornalismo de Dados no Brasil.

Qualquer dúvida ou sugestão de melhoria, peço que me contactem para o seguinte e-mail: isabellamouc@gmail.com

1. 1. Sexo

Mark only one oval.

- Feminino
 Masculino

2. 2. Cidade onde trabalha

3. 3. Idade

4. 4. Onde você trabalha?

Mark only one oval.

- Redação tradicional
 Projeto independente
 Freelance

5. 5. Na redação, quantos profissionais trabalham com você na área do jornalismo de dados?

Se a sua resposta foi "Freelance", passe para a pergunta número 6

Mark only one oval.

- 0
 1
 2
 3
 4
 5 ou mais

6. 6. Qual a sua organização empregadora?

Freelance: Qual organização mais solicita seus trabalhos?

7. 7. Você é formado(a) em jornalismo?

Mark only one oval.

- Sim
 Não

8. 8. Há quantos anos trabalha com jornalismo?

9. 9. Há quantos anos trabalha com jornalismo de dados?

10. 10. Como você adquiriu competências em jornalismo de dados?

Mark only one oval.

- Auto-didata
 Pós-graduação
 Curso online
 Trabalho de campo

11. 11. Com que regularidade produz peças com recurso a base de dados?

Responda a quantidade e, separada por uma barra (/), a frequência temporal (semana, mês ou ano)

12. 12. Na redação, você só produz peças com recurso ao jornalismo de dados?

Mark only one oval.

- Sim
 Não

13. 13. Quantas horas trabalha, em média, por dia?

14. 14. Ferramentas que domina (sendo 0 não domino, 1 básico, 2 intermédio e 3 avançado)

Mark only one oval per row.

	0	1	2	3
Excel	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
OpenRefine	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
ArcGis	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
TableauPublic	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

15. 15. Outra (de 0 a 3): Refira a ferramenta e indique, entre parênteses, o nível.

16. **15. Você domina alguma linguagem de programação?**

Mark only one oval.

- Sim
 Não

17. **16. Se sim, qual/quais?**

Check all that apply.

- Python
 SQL
 JavaScript
 HTML-CSS
 Other: _____

18. **17. Quem é responsável por criar as visualizações de dados nos seus trabalhos?**

Mark only one oval.

- Eu, enquanto jornalista
 Um programador
 Um infografista
 Other: _____

19. **18. Que outros profissionais conhece que trabalhem na área do jornalismo de dados?
(Coloque os nomes separados por vírgulas)**

Obrigada pela disponibilidade!

Perfil: Quem faz Jornalismo de Dados em Portugal

Para o efeito de obtenção do grau mestre em Jornalismo pela Universidade Nova de Lisboa, este questionário tem como objetivo traçar o perfil do profissional e ajudar a compreender como é feito o Jornalismo de Dados em Portugal.

Qualquer dúvida ou sugestão de melhoria, peço que me contactem para o seguinte e-mail: isabellamouc@gmail.com

1. 1. Sexo

Mark only one oval.

- Feminino
 Masculino

2. 2. Cidade onde trabalha

3. 3. Idade

4. 4. Onde você trabalha?

Mark only one oval.

- Redação tradicional
 Projeto independente
 Freelance

5. 5. Na redação, quantos profissionais trabalham com você na área do jornalismo de dados?

Se a sua resposta foi "Freelance", passe para a pergunta número 6

Mark only one oval.

- 0
 1
 2
 3
 4
 5 ou mais

6. 6. Qual a sua organização empregadora?

Freelance: Qual organização mais solicita seus trabalhos?

7. 7. Você é formado(a) em jornalismo?

Mark only one oval.

- Sim
 Não

8. 8. Há quantos anos trabalha com jornalismo?

9. 9. Há quantos anos trabalha com jornalismo de dados?

10. 10. Como você adquiriu competências em jornalismo de dados?

Mark only one oval.

- Auto-didata
 Pós-graduação
 Curso online
 Trabalho de campo

11. 11. Com que regularidade produz peças com recurso a base de dados?

Responda a quantidade e, separada por uma barra (/), a frequência temporal (semana, mês ou ano)

12. 12. Na redação, você só produz peças com recurso ao jornalismo de dados?

Mark only one oval.

- Sim
 Não

13. 13. Quantas horas trabalha, em média, por dia?

14. 14. Ferramentas que domina (sendo 0 não domina, 1 básico, 2 intermédio e 3 avançado)

Mark only one oval per row.

	0	1	2	3
Excel	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
OpenRefine	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
ArcGis	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
TableauPublic	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

15. Outra (de 0 a 3): Refira a ferramenta e indique, entre parênteses, o nível.

16. 15. Você domina alguma linguagem de programação?

Mark only one oval.

- Sim
 Não

17. 16. Se sim, qual/quais?

Check all that apply.

- Python
 SQL
 JavaScript
 HTML-CSS
 Other: _____

18. 17. Quem é responsável por criar as visualizações de dados nos seus trabalhos?

Mark only one oval.

- Eu, enquanto jornalista
 Um programador
 Um infografista
 Other: _____


**19. 18. Que outros profissionais conhece que trabalhem na área do jornalismo de dados?
(Coloque os nomes separados por vírgulas)**

Obrigada pela disponibilidade!


Anexo 4 – Projetos independentes

Brasil

 Abraji	É a Associação Brasileira de Jornalismo Investigativo, promove o jornalismo e também produz conteúdos com técnicas de Jornalismo de Dados , como o projeto CTRL + X.
 Aos Fatos	Projeto jornalístico que atua na checagem do discurso público.
 Código urbano	Iniciativa com foco na abertura de dados sobre questões da rede pública de São Paulo.
 Conectedados	Projeto de cobertura jornalística da cidade de Joinville, Santa Catarina. Tem como suporte, o Jornalismo de Dados .
 E & P Brasil	Publicação com foco na cobertura sobre energia no Brasil, com destaque para o petróleo, gás natural e energia elétrica.
 Farol Reportagem	Iniciativa centrada na transparência, nos dados públicos e nos direitos humanos, com maior cobertura para assuntos do Estado de Santa Catarina.
 Gⁿ Gênero e Número	Projeto independente que se dedica ao debate sobre gênero e equidade.
 Impacto.Jor	Iniciativa que utiliza dados para medir o impacto do jornalismo na vida das pessoas.
 InfoAmazonia	Plataforma independente que utiliza dados na contextualização do estado da floresta Amazónia.
 Jota	Organização independente que atua na cobertura de assuntos sobre o poder judiciário.
 LabCaos	Iniciativa que se dedica ao trabalho de verificação e validação da informação a partir de <i>fact-checking</i> e Jornalismo de Dados .
 Livre.Jor	Agência de notícias que propõe uma construção jornalista a partir de dados de origem pública; maior predominância na cobertura de assuntos do Estado do Paraná.
 Marco Zero	Iniciativa independente de Jornalismo Investigativo com interesse na cobertura de assuntos relacionados ao semiárido nordestino, urbanismo e relações de poder.
 Operação Serenata de Amor	Projeto que utiliza a análise de dados na fiscalização de dados públicos.

	Volt Data Lab	Iniciativa independente que utiliza técnicas de análise e comunicação de dados na construção de narrativas e produtos jornalísticos.
---	----------------------	--

Portugal

	Frames	É uma <i>startup</i> que sugere e promove a incorporação de dados, numa componente mais gráfica, em artigos jornalísticos.
---	---------------	--

Anexo 5 - A emancipação de uma cultura própria

A existência e convenção de novos termos para referir processos, tarefas e produtos a partir da implementação de técnicas de Jornalismo de Dados na produção de conteúdos de carácter informativo, pode ser considerado um ponto relevante na consagração do Jornalismo de Dados e também um indicador de que a área já faz parte da cultura das redações.

Alguns termos referidos ao longo das entrevistas:

Planilhas: é considerado o ambiente natural de exploração dos dados, onde muitas vezes é possível aceder e ler bases de dados. Tornou-se parte integrante da rotina jornalista em projetos guiados por dados.

Linhas, células: o *Excel* é o software base no tratamento e análise dos dados, como tal, é recorrente na linguagem jornalística, em se tratando de trabalhos com dados, referir elementos que compõem o programa.

Extração: A fase inicial do trabalho do jornalista sofreu alterações e este termo é referente ao contacto com as fontes, que neste caso pode ser estabelecido a partir de um ‘clique’.

Mineração: Diz respeito à etapa de análise e consiste na identificação de tendências e padrões nos dados.

“Dados sujos”, “base suja”: Termo utilizado em referência às bases de dados que apresentam contradições, como elementos repetidos e elementos com caracterização irregular.

Limpeza: é o processo de correção das bases de dados sujas. Na maioria das vezes é feita através da eliminação de dados não trabalháveis e da substituição de dados inconsistentes.

Raspagem: Pode ser um dos modos de extração dos dados, quando estes não estão em formatos compatíveis com o trabalho jornalístico.

Robô: Os robôs são estruturados através de linguagens de programação, como *Python* ou *R*, e permitem a criação de processos automatizados que possibilitam a extração e coleta de dados.

Hacker: Há uma consciência de que outras metodologias de acesso aos dados podem ser utilizadas.