

Cinco categorias de Jornalismo de Dados ou uma proposta para problematizar o Jornalismo a partir de dados no Brasil ¹

Fábio Vasconcellos², Leonardo Mancini³ e Carolina Bittencourt⁴

The five Data Journalism categories, or how to understand Data Journalism in Brazil

¹ Trabalho apresentado no II Seminário de Pesquisa em Jornalismo Investigativo, realizado na Universidade Anhembi-Morumbi, cidade de São Paulo, entre 02 e 04 de julho de 2015

² Jornalista e professor da ESPM Rio e da Universidade do Estado do Rio de Janeiro (UERJ). Doutor em Ciência Política. E-mail: fabio.vasconcellos10@gmail.com Twitter: @Fvas

³ Jornalista, professor, chefe do Departamento de Jornalismo da ESPM Rio e coordenador da Pós-Graduação em Jornalismo Investigativo.

⁴ Estudante do sexto período do curso de Jornalismo da ESPM Rio.



Associação Brasileira de Jornalismo Investigativo

II Seminário de Pesquisa em Jornalismo Investigativo
Universidade Anhembi-Morumbi, 02 a 04 de julho de 2015

Resumo: Neste artigo, apresentamos uma classificação que permita compreender as diferentes formas pelas quais os jornais brasileiros estão utilizando dados na produção de suas reportagens, diferenciando seu simples uso na construção da narrativa jornalista da prática que se convencionou chamar Jornalismo de Dados. Ao final, após uma reflexão teórica, apresentamos uma matriz que relaciona a prática jornalística com as competências utilizadas pelos repórteres na utilização de dados em reportagens.

Palavras-Chave: Jornalismo de Dados. RAC. Visualização de Dados.

Abstract: In this article, we present a classification that allows them to understand the many ways the Brazilian newspapers are using data in their reporting, trying to differentiate those who simply use numbers to illustrate a piece from those that are considered data journalism. At the end, after some theoretical analysis, we formulate a matrix relating the journalism practice and the skills used by the reporters while using data.

Keywords: Data Journalism. CAR. Data Visualization.

Jornalismo de Dados: alguns princípios

Em 2012, o *European Journalism Centre* e a *Open Knowledge Foundation* concluíram o primeiro guia que sistematiza um entendimento mínimo sobre o Jornalismo de Dados, uma prática que vem se disseminando em redações, blogs e sites em todo mundo. O *The Data Journalism Handbook* (Gray, Bounegru e Chambers 2012) é um documento colaborativo, que reúne exemplos e análises de especialistas de vários países. Rapidamente, o manual foi traduzido do inglês para o espanhol e para o mandarim e, em 2014, foi lançado em português, graças ao esforço da Associação Brasileira de Jornalismo Investigativo (Abraji).

O interesse crescente pelo JD pode ser explicado por duas razões. A primeira delas refere-se à disseminação da cultura de dados abertos na web, tendência que os governos em muitos países vêm seguindo para atender ao princípio democrático do *Open Government*. A segunda pode ser creditada à popularização de softwares de análise e visualização de dados, muitos deles gratuitos na internet (Flew et al, 2012).

Mas o que é exatamente o Jornalismo de Dados? Na seção do *Handbook* dedicada a responder essa pergunta, Bradshaw (2014) argumenta que “jornalismo” e “dados” são termos problemáticos. No ambiente digital, segundo o autor, os dados não seriam qualquer grupo de números reunidos numa planilha, como os jornalistas até então estavam acostumados. Documentos confidenciais, fotos, vídeos e áudios também podem ser agora descritos em números, mais especificamente em leitura binária 0 e 1. A descrição de Bradshaw não encerra, obviamente, o debate sobre em que sentido o jornalismo que utiliza esses dados se diferenciaria daquele acostumado apenas com tabelas numéricas. O autor reconhece essa limitação e argumenta que o Jornalismo de Dados se difere não pelo uso dos dados em si, mas “talvez pelas novas possibilidades que se abrem quando se combina o tradicional ‘faro jornalístico’ e a habilidade de contar uma história envolvente com a escala e o alcance absolutos da informação digital agora disponível” (Bradshaw, 2014).

Bradshaw, de certo modo, entende que o JD se definiria mais pela associação dos termos “dados” e “investigação” com um terceiro: as novas tecnologias, ou, mais precisamente, as ferramentas que permitem hoje que o jornalista possa automatizar processos, fazer associações complexas entre milhares de documentos ou produzir “infográficos envolventes”.

Anderson (2014) tem uma avaliação crítica desse argumento. Para ele, se entendermos o jornalismo pela perspectiva da busca e da reunião de informações de materiais humanos e sociais, como documentos, entrevistas e dados, então o Jornalismo de Dados tem uma história que não seguiria uma ordem cronológica de aperfeiçoamento, como sugere a visão tecnológica, mas sim de passagens irregulares e descontínuas, com mudanças na compreensão do que ele é e das técnicas utilizadas para construir conteúdos com dados. Em outros termos, Anderson se opõe à ideia de que o JD nasce a partir da evolução tecnológica e propõe que seja discutida as diferentes formas pelas quais o jornalismo utilizava dados, muitos antes do processo de aperfeiçoamento do aparato tecnológico⁵.

Embora com algumas variantes, autores brasileiros, entre eles pesquisadores e jornalistas que utilizam o JD, adotam uma visão que associa a capacidade investigativa e as novas possibilidades da tecnologia para conceituar o Jornalismo de Dados. Basicamente, as técnicas consistiriam na produção, no tratamento e no cruzamento de grande quantidade de dados, permitindo uma maior eficiência na recuperação da informação, na apuração da reportagem a partir do conjunto de dados, na circulação em diferentes plataformas e na geração de visualizações e infografias (Träsel, 2013).

Novamente, a automatização dos processos de apuração e de mensuração de dados permitiria a prática desse tipo de jornalismo (Barbosa, 2006, 2007). Lima Júnior (2011) acrescenta que o uso das Tecnologias de Informação e Comunicação (TIC) seria um elemento fundamental do que ele chama de “Jornalismo Computacional”, que ganha importância no contexto de Big Data e que seria a principal razão para uma mudança no tipo de habilidade dos futuros jornalistas, classificada pelo autor como “hacking journalism”, ou seja, a capacidade explorar tecnologias filtrando informações e colocando-as de forma visual.

⁵ Em 1821, o jornal inglês *The Guardian* apresentou uma reportagem, feita a partir de uma lista obtida de fonte não oficial, que relacionava as escolas da cidade de Manchester ao número de alunos e aos custos de cada uma. A lista ajudou a mostrar que o número de alunos que recebiam educação gratuita era muito maior do que indicava os dados oficiais (Bounegru, 2014)

A dimensão das competências

Talvez o ponto central no debate sobre o que venha a ser o Jornalismo de Dados seja a incorporação de competências pelo campo jornalístico para lidar com dados⁶. Esse tema está presente nas inúmeras tentativas de definir o JD, como aquelas lembradas até aqui, no entanto, a nosso ver, ainda de forma tangencial. O exame dessas competências, seja a partir de pesquisas do tipo etnográficas ou do exame do conteúdo e da forma do JD, pode ajudar a sair de algumas dificuldades teóricas que ainda persistem, como também a diferenciar o JD praticado pelas redações no Brasil.

Por exemplo, definir o Jornalismo de Dados pela dimensão tecnológica 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. Ora, como a internet e os softwares tornaram-se aparatos acessíveis às redações, logo todo jornalismo *com* dados seria, necessariamente, Jornalismo *de* Dados?. Retornaremos a essa questão.

Há um segundo ponto a ser problematizado: se 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 que uma variação semântica do jornalismo investigativo, incorporado às redações desde muito antes do advento da internet? Essa pode ser uma resposta, mas ainda permanece pouco explorada. Talvez o que realmente seja novo nesse cenário é a adoção da Lei de Acesso à Informação e a cultura do *Open Government* que, uma vez associada à busca online e a novos softwares de análise e visualização, amplifica o campo de investigação jornalística. O argumento é válido, mas ele implicitamente traz um segundo problema. Por esse caminho, a definição do JD escaparia do campo jornalístico, sendo, portanto, fruto de

⁶ Neste artigo, não discutimos em profundidade a diferença entre os conceitos de competência e habilidade, debate mais do campo da psicologia e da pedagogia. Assumimos genericamente que competência envolve a aplicação de habilidades do jornalismo de apurar, escrever, comunicar e, mais especificamente para o nosso caso, saber identificar, manipular, analisar e visualizar dados. Contudo, como os contextos das reportagens mudam, o jornalista precisa desenvolver competência para lidar com essas mudanças com o intuito de realizar o seu objetivo. Competência, portanto, estaria voltada para a mobilização de um conjunto de conhecimento que pode ser aplicado por meio de habilidades específicas de produzir e apresentar reportagens através de dados. Nesse sentido, um jornalista pode ter habilidade em manejar números, mas ela é continuamente aplicada para produzir reportagens descritivas sobre o índice de inflação, não como a inflação é gerada, condicionada por diversos fatores ou mesmo como pode afetar diferentes extratos de renda da população. Essa mesma reportagem pode ser apresentada num texto analítico e acompanhado de gráficos ou visualizações que permitam uma maior compreensão do assunto abordado. Neste caso, as habilidades, que são parte da competência do jornalista, foram mobilizadas e articuladas para gerar um conteúdo mais específico e detalhado sobre a inflação.

mudanças no campo político-institucional. Em outros termos, governos que não adotam a abertura de dados não criariam, ou permitiriam, o JD nos seus países?

Retornemos ao ponto sobre a questão das competências. De certo modo, a tecnologia incentiva e pode criar condições para novas práticas no processo de apuração jornalística. Mas, a nosso ver, há uma diferença tênue nessa relação. O telefone permite que seja feita uma entrevista a quilômetros de distância, alterando o contato presencial estabelecido entre um jornalista e um entrevistado por um contato não presencial. Contudo, e esse é o ponto, o telefone não modifica o lugar do jornalista como profissional dotado de competência para mobilizar um conjunto de conhecimento que permita extrair informação. Muitas vezes essa competência envolve lidar com situações que variam conforme o assunto a ser abordado, o tipo de entrevista, entre outras condições contextuais. O ponto aqui não é, portanto, o objeto técnico (o telefone) em si, mas como esse objeto pode ser explorado em diferentes situações para obter informações e produzir conteúdo jornalístico.

Como sabemos, o jornalismo sempre dependeu de algumas habilidades específicas para ser realizado. O saber coletar informações, organizá-las e produzir um texto ou imagens são exigências clássicas que fortaleceram após o processo de profissionalização na virada do século XX. Outras, contudo, surgiram ao longo desse processo e não estavam necessariamente vinculadas a utilização de tecnologias. No caso dos chamados “jornalistas investigativos”, por exemplo, há exigência de que eles desenvolvam uma atitude ativa na busca de informação, em contraposição à atitude passiva de um mero agente mediador entre instituições e a sociedade (Hunter et al 2013)⁷

Para o nosso propósito, portanto, não interessa saber se as competências são resultado da adoção de uma tecnologia, ou mesmo se existe ou não um aparato legal que subsidie a sua aplicação, mas, sim, que elas tenham sido incorporadas e como foram incorporadas pelo campo jornalístico na sua rotina de produção de notícia. Em suma, a tecnologia ou a Lei de Acesso à Informação, ou mesmo a atitude investigativa, são meios que facilitam e promovem a prática do JD, e não aquilo que o define.

⁷ “A cobertura convencional de notícias depende amplamente – e, às vezes, inteiramente – de materiais fornecidos pelos outros (por exemplo, pela polícia, governos, empresas, etc.); ela é fundamentalmente reativa, quando não, passiva. A cobertura investigativa, em contraste, depende de materiais reunidos ou gerados a partir da própria iniciativa do(a) repórter (e por isso ela é frequentemente chamada de “cobertura empreendida” – em inglês, “enterprise reporting” (Hunter et al 2013, pág. 8).

Dito isso, o que precisa ser problematizado na definição do JD é como a incorporação ou o aprofundamento de competências pelo campo jornalístico para lidar com dados altera o modo de fazer notícia ou, se preferirem, possibilita o manejo das novas habilidades. Não adianta a posse de software de análise estatística se não existir um profissional que conheça essa ferramenta, saiba como ela funciona e como ela pode atender aos diferentes contextos nos quais o jornalismo precisa extrair informação e produzir notícia a partir de dados.

Ou seja, se estamos apontando o caminho do “como” como procedimento metodológico mais adequado para pensarmos o JD, isto é, como é produzida informação a partir dessas novas técnicas de manejo de dados em contraposição a uma simples classificação “utiliza/não utiliza”, o nosso ponto diz respeito às seguintes questões de fundo: jornalistas dotados das competências para garimpar e analisar dados na internet, mas que não utilizam “infográficos envolventes”, deixam de ser jornalistas de dados? Como um corte binário entre “utiliza/não utiliza” “infográficos envolventes” se relaciona com as inúmeras variações no tipo de conteúdo que utilizam dados e outras formas de visualização menos envolventes? Ou ainda: toda utilização de dado quantitativo é necessariamente JD? Infográficos envolventes produzidos a partir de dados já tabulados e organizados segundo os critérios de instituições externas ao campo jornalístico podem ser considerados JD?

A utilização de dados pelo jornalismo

Embora seja relativamente recente, o Jornalismo de Dados pode ser considerado parte de uma tradição iniciada na década de 60 pelo jornalista Philip Meyer (1991). Ele foi um dos primeiros a adotar o conceito de “Jornalismo de Precisão”, que consistia no uso do computador associado a métodos da ciência social para produzir reportagens com menores chances de erro⁹. O estudo de Meyer acabou por incentivar o surgimento nos anos 90 do termo Reportagem com Auxílio de Computador (RAC), ainda hoje utilizado em fóruns especializados como *o Investigative Reporters e Editors (IRE)* e a *Abraji*.

Esses novos procedimentos ajudaram os jornalistas a aprimorar o seu próprio conhecimento acerca da realidade social e política, reduzindo a dependência de fontes externas

⁹ Sobre a proximidade da prática do jornalismo com as ciências sociais ver mais em Weaver e McCombs (1980).

ao processo de produção e de análise da informação. Tudo isso ganha uma nova importância quando o contexto muda de um cenário de escassez, no qual jornalistas tinham enormes dificuldades para obter informação, para um momento de abundância, no qual o problema deixa de ser encontrar a informação, mas saber qual deve ser buscada, analisada e utilizada para subsidiar a notícia jornalística (Flew et al 2012).

Em certo sentido, o Jornalismo de Dados é uma vertente do RAC porque envolveria não apenas o uso de computadores, mas o conhecimento mínimo em estatística, sistemas computacionais e métodos das ciências sociais (Flew et al 2012; Hamilton and Tuener 2009; Gray et al 2013; Hackett 2013; Howard 2014). Coddington (2014) concorda com essa afirmativa. Para ele, o RAC, o Jornalismo de Dados e o Jornalismo Computacional, aquele mais voltado para a programação de máquinas, são categorias que se sobrepõem. Apesar disso, ele vê algumas diferenças importantes com relação à orientação desses três grupos de profissionais, e propõe uma tipologia baseada em quatro dimensões para caracterizar práticas jornalísticas.

No que diz respeito à dimensão da natureza profissional, Coddington compara as diferenças de atitude entre os jornalistas com larga experiência daqueles que buscam uma formação em rede, isto é, que procuram realizar trocas de conhecimento, inclusive com os não jornalistas. Enquanto no primeiro caso os repórteres estão mais condicionados às normas e rotinas estabelecidas por seu grupo, no segundo os profissionais buscariam formas de interação com uma maior diversidade de grupos.

A segunda dimensão analisada por Coddington se refere à transparência das técnicas e reportagens produzidas, ou seja, quão visíveis são os métodos utilizados pelos jornalistas para os leitores ou para os profissionais interessados. A terceira dimensão analisa se as reportagens são feitas a partir de amostras quantitativas orientadas ou com a utilização de estruturas mais robustas, como aquelas oriundas do Big Data. A quarta e última dimensão observa como os jornalistas compreendem o papel do seu público. Neste caso, como um público passivo ou como um público ativo, que pode e deve participar do processo de produção e compreensão da notícia.

Como conclusão, Coddington afirma que existe um fosso entre as orientações profissionais e epistemológicas da RAC, por um lado, e do Jornalismo de Dados e do

Jornalismo Computacional, por outro. Essas divisões teriam um fundo cultural a partir do qual cada um procura fazer jornalismo. O RAC surgiu de um esforço de unir as modernas técnicas das ciências sociais ao jornalismo profissional e ao jornalismo investigativo. O Jornalismo de Dados e o Jornalismo Computacional, por sua vez, são práticas que se relacionam não apenas com algumas técnicas das ciências sociais, mas com a cultura dos dados abertos.

Desse modo, o JD estaria mais próximo de uma fusão com os princípios da cultura dos dados abertos e com o trabalho computacional. Este último, por sua vez, assemelha-se ao Jornalismo de Dados, pois incorpora o princípio dos dados abertos, embora mais voltado para a programação. Seria uma prática com mais ênfase no trabalho em rede e nas trocas com outros profissionais, embora a sua técnica e seu material de trabalho sejam orientados, ao fim e ao cabo, para o Jornalismo de Dados (Coddington, 2014).

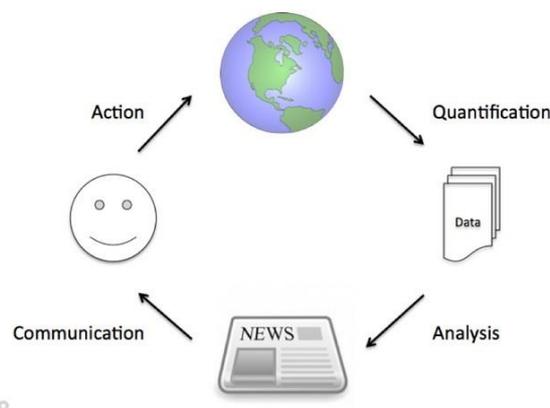
O debate sobre a cultura das três práticas jornalísticas oferece algumas pistas para se tentar conceituar o JD, embora, a nosso ver, não aprofunde sobre as suas competências. Sobre esse ponto, Stray (2014) sugere um modelo teórico bastante simples e interessante. Partindo do exemplo de um gráfico da série temporal de desemprego nos Estados Unidos, Stray problematiza os procedimentos de apuração e apresentação das informações, chegando assim a três etapas que, segundo ele, ajudariam a configurar a diferença entre uma apuração típica do jornalismo tradicional daquela classificada como Jornalismo de Dados.

Para Stray, o Jornalismo de Dados inicia-se com uma quantificação e, evidentemente, perguntas sobre essa quantificação com o objetivo de esclarecer o que é contado e como é contado. “Os dados são criados, são um registro, um documento, um artefato que apresenta um significado” (Stray 2014). Nesse sentido, um gráfico de desemprego, por exemplo, é apenas uma representação de um fenômeno e essa representação só foi possível porque houve um processo de quantificação desse fenômeno. A questão chave, diz Stray, é que existem diversos procedimentos que podem ser adotados no processo de identificação dos desempregados, bem como da organização desses dados. “O dado não é algo que existe na natureza. Os desempregados são uma coisa muito diferente de um dado de desemprego” (Stray 2014).

Partindo da premissa de que o Jornalismo de Dados inicia-se com um processo de quantificação, Stray argumenta, então, que ela se complementa com a análise, ou seja, se a quantificação transforma o mundo em dados, a análise transforma os dados em conhecimento.

O autor explica que é nessa etapa em que o jornalismo mais se aproxima da ciência, com forte inclinação para a matemática, a estatística e a lógica. “No Jornalismo de Dados, existe profundo conhecimento técnico e específico” que permite, por exemplo, o jornalista comparar as taxas de desemprego com taxas de impostos aplicadas em diversos países, na busca de alguma explicação para o fenômeno do desemprego.

Em seguida, para Stray, é necessário que a quantificação que transforma o mundo em dados e em análise seja comunicada. “O jornalista não publica a sua história no vácuo, mas para mentes humanas e para a sociedade humana”. Deste modo, a história do desemprego incluiu um gráfico de desemprego porque é a melhor maneira de comunicar alterações na taxa de desemprego, se comparado com uma tabela de números. É essa informação que permite que o leitor ou a sociedade humana possa atuar sobre a realidade para alterá-la ou para confirmá-la. Sinteticamente, eis o modelo que definiria o JD.

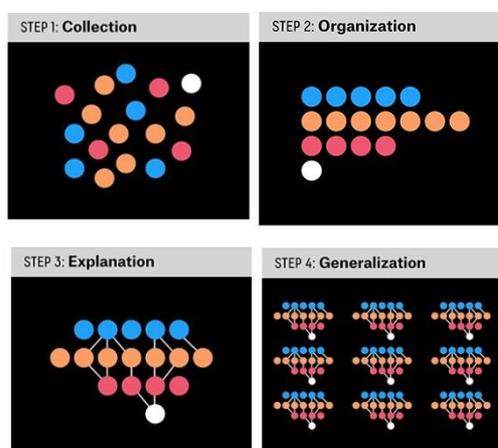


Fonte: Stray, 2014

Conhecido por suas publicações de Jornalismo de Dados, Silver (2014) apresenta também a sua visão a respeito do conjunto de procedimentos que ajudariam a determinar as competências exigidas para essa prática jornalística. Silver trabalhava no *The New York Times* até o fim de 2013, quando saiu para fundar o *FiveThirtyEight*, um site que se propõe a fazer Jornalismo de Dados. O site entrou no ar no início de março de 2014 e, na ocasião do seu lançamento, Silver apresentou o seu receituário para a prática do JD. Muito da forma de Silver

pensar o Jornalismo de Dados, e sua forte aproximação com o campo da estatística, está presente também no seu livro “O sinal e o ruído”, lançado no Brasil em 2012.

Na apresentação do site em 2014, Silver diz que Jornalismo de Dados não é apenas usar números ao invés de palavras. Segundo afirma, “o uso de números não é necessário, nem suficiente, para produzir boas obras de jornalismo” (Silver, 2014). Com essa proposição, ele argumenta que o JD envolve alguns procedimentos, entre eles: coleta, organização e exploração de dados para se obter relações que podem ser significativas entre os dados.



Fonte: Silver (2014)

Silver argumenta assim que a primeira etapa do JD é a coleta de dados ou de uma evidência que, no jornalismo tradicional, se resumiria à combinação de entrevistas e documentos ou à observação pessoal. No caso do Jornalista de Dados, existe também a coleta de informações, mas como uma combinação de pesquisas, de experimentos ou de extração de dados da internet. A segunda fase é o da organização. No jornalismo tradicional, a organização se traduz em contar uma história de maneira cronológica, por ordem de importância, da pirâmide invertida. Os jornalistas de dados, contudo, organizam as informações com a descrição estatística dos dados, ou seja, através das relações entre eles ou através da construção de visualizações.

A terceira fase corresponde ao que Silver chama de explanação. No jornalismo tradicional, essa etapa consistiria na identificação de quem, o que, onde, quando, por que e como. O jornalista de dados, por sua vez, apresenta a sua explanação a partir de técnicas

estatísticas para demonstrar ou para verificar relações entre os dados. A quarta, e última, fase proposta por Silver chama-se generalização. Essa etapa consistiria em utilizar dados e análises de eventos passados para inferir como esses eventos se comportarão no futuro. Essa técnica envolve o uso de ferramentas das ciências sociais como forma de verificação de hipóteses ou de apresentação de previsões, bem diferente do modelo de inferência praticado pelo jornalismo tradicional, que parte de um conhecimento não estatístico, mas intuitivo.

As proposições de Hirst (2013), apresentadas no seu blog da *School of Data*, uma entidade que se dedica a formar jornalistas de dados pelo mundo, parecem também bastante pertinentes para o nosso objetivo e ajuda, inclusive, a contrapor, de certo modo, o argumento de Silver sobre o peso que ele dá à habilidade com a estatística. Para Hirst (2013), o Jornalismo de Dados pode ser definido *como um output específico*, ou seja, o resultado de uma reportagem que é facilmente verificável na apresentação de gráficos e infográficos. Nesse ponto, contudo, observa que, muitas vezes, o gráfico é a parte mais fácil de ser realizada. Por trás dos gráficos houve ou deveria haver um intenso trabalho de coleta e análise de dados para que fosse possível produzir esse conteúdo.

A partir dessa primeira proposição, Hirst argumenta que o JD pode ser considerado *como uma prática que envolve um conjunto de habilidades específicas*, em outros termos, o JD seria não somente o que ele produz, mas também as habilidades que os jornalistas precisam dominar para realizá-lo. Nesse ponto, Hirst acrescenta que essas habilidades não podem ser limitadas àquelas necessárias para fazer visualizações de dados interativos na web, argumento muitas vezes usado por muitos jornalistas. Para o autor, essa proposição, que tende a considerar o JD a partir da visualização de dados, incorpora algo estranho. Ela segue uma lógica que se sustenta, por exemplo, que o áudio define o Jornalismo de Rádio, ou o que vídeo define o Jornalismo Televisivo.

Desse modo, Hirst propõe uma terceira definição: o Jornalismo de Dados é um processo cujo *output* envolve dados, mas não necessariamente ou exclusivamente, porque, embora eles possam ser utilizados na apresentação da reportagem, há momentos em que seriam apenas parte do processo de apuração. Nesse sentido, os dados podem levar o jornalista a uma história, o que os transforma em fonte. Dessa forma, o jornalista de dados tem no seu processo de trabalho uma orientação pelos dados, o que exige habilidades para identificar, para organizar e para

fazer relações entre os dados que podem ser fonte ou objeto principal da reportagem. Como se vê, Hirst faz uma defesa da necessidade de compreendermos o JD como uma associação entre uma atitude investigativa e a habilidade para trabalhar com dados, relação que tendemos a concordar em razão da sua crítica ao modelo que prega o forte apelo de visualização dos dados ou apenas habilidades específicas para lidar com programação de computadores.

Do Jornalismo *com* Dados ao Jornalismo *de* Dados

Se a competência para dos jornalistas para manejar dados, organizá-los e analisá-los em diferentes situações e contextos dos temas abordados é um ponto de convergência entre os autores e se podemos afirmar que essa prática não nasce com a tecnologia, mas sim é facilitada por ela, resta-nos compreender o modo como o jornalismo se apropria dos dados para produzir conteúdo. Em outros termos, concordamos com Anderson (2014) quando ele argumenta que a compreensão do que seja JD muda mais em função do modo como definimos os dados (números, bits etc) e da maneira pela qual trabalhamos com eles (técnicas) do que por conta da aplicação ou não de uma tecnologia.

Nesse sentido, parece nítida a necessidade de produzirmos uma matriz que contemple o passo-a-passo defendido por Stray (2014) e por Silver (2014), mas que procure também agregar a dimensão ativa do jornalismo interessado em revelar algo para além da simples divulgação de relatórios quantitativos, como defende Hirst (2013)¹⁰. A dimensão da comunicação dos dados parece ser algo igualmente importante na construção dessa matriz. Mas não seria, a nosso ver, uma dimensão sobre a qualidade estética ou sobre o grau de complexidade de gráficos e infográficos. Embora seja desejável, por conta da sua maior comunicabilidade, a nosso ver é mais interessante, inicialmente, saber se a comunicação gráfica

¹⁰ A proposta de produzir uma matriz para pensarmos o Jornalismo de Dados praticado no Brasil faz parte da pesquisa sobre o Jornalismo de Dados e sua contribuição para o processo de *accountability* político desenvolvido pelos autores deste *paper* e financiada pela CAEPM/ESPM. Na pesquisa, monitoramos as publicações online de 43 jornais no Brasil. A matriz proposta aqui é fruto desse monitoramento, no qual percebemos uma variabilidade no uso que os jornais fazem de dados quantitativos na produção e apresentação das publicações. Muitas das reportagens não se encaixavam nos modelos de JD defendidos por autores de outros países.

sugere ou induz análises e compreensões sobre o conteúdo ou sobre o fenômeno que o jornalista apresenta na sua reportagem.

Isso não significa dizer que infografias arrojadas e criativas que acompanham as reportagens percam a sua dimensão analítica. Não se trata de um debate entre estética e conhecimento analítico. Trata-se tão somente de incorporar à análise sobre o que é o JD infografias menos criativas do ponto de vista estético ou que não demandam grande tempo de produção e programação, mas que mantêm o seu caráter analítico. Essa observação é importante porque corremos o risco de considerar JD apenas aquele conteúdo que necessariamente incorpora apresentações gráficas mais arrojadas.

Os casos apresentados no Manual de Jornalismo de Dados (Gray, Bounegru e Chambers, 2012), por exemplo, restringem a identificação do que é uma reportagem de JD, ao menos no caso do Brasil ou de países cujas empresas de jornalismo não tenham adquirido tecnologias mais avançadas. Isso porquê o Manual estabelece os parâmetros através dos quais essa prática deve ser identificada. Ao fazê-lo, define um ideal de visualização que depende sobremaneira da utilização de softwares sofisticados ou de programadores o que, de certo modo, associa o Jornalismo de Dados a um padrão estético de visualização. Em outros termos, toda produção que não apresente visualizações sequer próximas desses modelos não poderia ser considerada Jornalismo de Dados. Novamente, não estamos dizendo com isso que o aprimoramento estético da visualização dos dados não seja desejável ou importante para a sua comunicabilidade, mas que essa dimensão não pode conduzir a definição do que venha a ser JD. Se assim fosse, a reportagem e os gráficos estáticos apresentados pelo *The Guardian* na publicação de 1821 que relacionava as escolas da cidade de Manchester ao número de alunos e aos custos de cada um não poderia ser considerada um exemplo de JD¹¹.

¹¹ Neste ponto, é interessante notar que o Prêmio mundial de *Data Journalism* promovido pelo *Global Editors Networks* 2015 tenha contemplado trabalhos que utilizaram arrojadas formas de visualização de dados. O prêmio principal foi do *The Wall Street Journal*, que produziu uma reportagem sobre os efeitos das vacinas na redução das taxas de diversos tipos de doença no século XX, é um exemplo disso. Para ser realizado, o trabalho necessitou não apenas de profissionais com um conhecimento mínimo em estatística, mas também de programadores que soubessem manusear os códigos do D3, uma biblioteca virtual e gratuita que disponibiliza códigos de programação para produção de visualização de dados. Embora tenha premiado o WSJ, o GEN criou categorias distintas, nas quais procura contemplar algumas variações do tipo de Jornalismo de Dados utilizado, o que para nós indica alguma preocupação em reconhecer a variabilidade desse tipo de prática jornalística. Entre as demais categorias, chama a atenção o prêmio dedicado a grandes redações e, um segundo, dedicado a pequenas redações na utilização de dados. Há ainda um prêmio dedicado ao melhor trabalho investigativo para grandes e pequenas redações. Esta última vencida pelo site Ojo Público, do Peru, cuja reportagem não apresenta uma visualização de dados quantitativos tal qual a do WSJ. O GEN premiou também trabalho na categoria Aplicativos para grandes e pequenas redações, além do “Melhor Site de

Uma matriz com dimensões básicas para pensarmos o que é JD ajudaria, inclusive, a dimensionarmos melhor algumas questões, como, por exemplo, se todo uso de dados pelo jornalismo é necessariamente Jornalismo de Dados. Há uma diferença, a nosso ver, entre reportagens *COM* e reportagens *DE* dados. Enquanto o primeiro contemplaria reportagens que se apropriam de dados de forma ilustrativa, no segundo caso, os dados seriam a própria razão da reportagem. No primeiro caso, o dado quantitativo teria o papel de auxiliar ou de ilustrar uma reportagem; no segundo, o dado seria o próprio fundamento da pauta e a história das relações entre os dados conduziria, neste caso, a reportagem.

Outra questão importante é o papel da internet que, se não cria o JD, como defendemos, certamente incentiva o desenvolvimento de publicações que permitem diversos processos interativos entre o leitor e o conteúdo. Desse modo, como poderíamos classificar a publicação que foca apenas no recurso de visualização dos dados que, embora não traga um texto que apresente suas relações, permite que o leitor tenha uma participação ativa na interpretação desses dados? Um site, por exemplo, desenvolvido por jornalistas que busca e que organiza dados sobre o orçamento de uma prefeitura e apresenta ferramentas interativas que permitem ao leitor acompanhar os gastos do executivo municipal deveria, aos nossos olhos, ser categorizado como JD.

Dito isso, entendemos que a nossa matriz precisa ser construída por um eixo vertical que parte da classificação de *Jornalismo com Dados* até o outro extremo, o *Jornalismo de Dados*. Essa matriz de algum modo limita o corte entre o que é e o que é não JD, porque permite incorporar as nuances das diversas formas pelas quais a imprensa produz conteúdo com dados quantitativos. Para isso, é necessário, no eixo horizontal, estabelecer quais atributos as reportagens apresentam. Esses atributos procuram atender aos propósitos de Stray (2014), Silver (2014), Bradshaw (2014) e Hirst (2013), ou seja, de que o JD deve buscar e apresentar relações entre os dados e, a partir daí, contar as suas histórias. Em outras palavras, esses atributos estariam vinculados à competência dos jornalistas para produzir conteúdo através de dados.

Data Journalism”, entre outras categorias. Disponível em <http://www.globaleditorsnetwork.org/programmes/data-journalism-awards/>. Acessado em 20 de junho de 2015.

Portanto, no eixo horizontal teríamos três categorias que atendem à dimensão do caráter investigativo das publicações, assim como a sua dimensão interpretativa e a dimensão comunicativa. Nesse último caso, não estamos interessados em saber se as publicações têm ou não uma alta qualidade técnica e estética, mas se existe uma comunicação gráfica dos dados que se associa com o seu conteúdo analítico de forma a enriquecer a compreensão da história contada. Em outros termos, não são meros gráficos a ocupar um espaço na página, são gráficos que contam parte da história apresentada na página. As nossas quatro categorias são assim definidas como:

Dimensão Investigativa:

Extração e/ou Organização: Procura identificar se a reportagem contempla dados que a equipe de jornalistas foi a responsável por extrair e estruturar o material bruto de alguma base e/ou foi a responsável pela produção e organização do próprio dado. Reportagens desse tipo tendem a trazer no seu enunciado indicações sobre o esforço da própria equipe ou mesmo indicações sobre o ineditismo do dado apresentado após a busca e/ou estruturação feita equipe¹².

Dimensão Interpretativa

Argumentação: Nessa dimensão busca-se identificar se existe um texto jornalístico analítico, ou seja, que procura não apenas apresentar o conteúdo da reportagem e o seu contexto, como, principalmente, se traz uma análise sobre as relações entre os dados de forma a indicar causas e/ou consequências do objeto da publicação. O texto pode ser tanto uma análise direta do jornalista, como de entrevistados que são incluídos ao longo da reportagem e que cumprem o papel de analisar as relações entre os dados trazidos pela publicação.

¹² Consideramos a “busca” como parte do processo que envolve tanto a extração (scraping), com o download de bases originais, como a simples compilação de dados já agregados, mas que foram utilizados segundo uma estruturação estabelecida pela equipe e não aqueles produzidos pela fonte original. Por estruturação, portanto, entendemos o processo pela qual a equipe cria variáveis temporais, numéricas ou categóricas para fazer desde um simples cruzamento (exemplo: série temporal das taxas de desemprego versus taxas de investimento da indústria), até estimativas sobre eventos futuros (exemplo: probabilidade de uma pessoa ser vítima de assalto no Centro de São Paulo).

Dimensão Comunicativa

Visualização Gráfica: Nessa categoria, procura-se identificar se as publicações trazem algum tipo de visualização, como gráficos ou infográficos. Avalia-se a maneira como essa visualização é utilizada de modo a permitir que o conteúdo da reportagem seja aprimorado pela comunicação visual, não no sentido estético, ainda que desejável, mas no sentido de que essa comunicação promova ou incentive a compreensão analítica da reportagem¹³.

A partir das dimensões investigativa, interpretativa e de comunicativa chegamos à nossa matriz, que busca estabelecer relações entre essas dimensões e a escala que vai do **Jornalismo com Dados** para o **Jornalismo de Dados**.

| Categorias X Níveis de JD | | Busca e/ou Elaboração própria dos dados (criação da base) | Estrutura a base (séries temporais, categorias rankings, tabelas) | Visualização dos dados (infografia) | Interpretação dos dados (texto) |
|---------------------------------|---------|--|---|--|------------------------------------|
| Jornalismo ↑ Com Dados | Nível 1 | | | | |
| | Nível 2 | | | | |
| | Nível 3 | | | | |
| | Nível 4 | | | | |
| | Nível 5 | | | | |

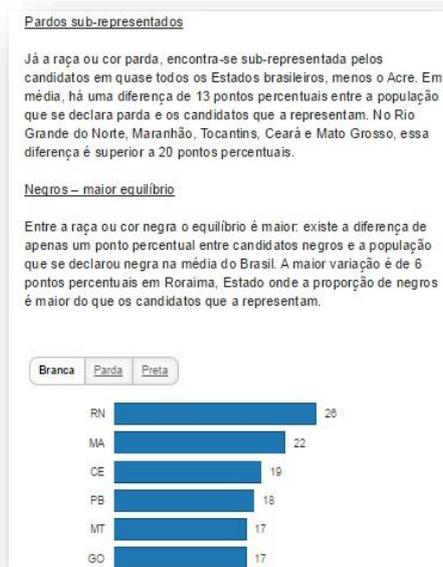
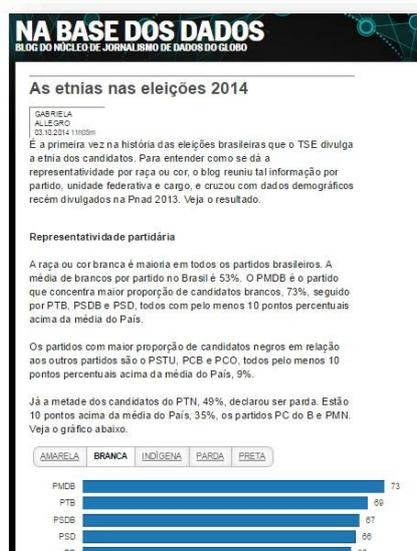
Fonte: elaboração dos autores

No Nível 1, a equipe busca e/ou produz uma base de dados primária (a partir dos microdados) ou secundária (a partir de dados agregados). Com a base, é realizada uma análise, que resulta em um texto e em uma apresentação de gráficos ou infográficos (interativos ou

¹³ No geral, todo gráfico atende a esse princípio. Por exemplo, uma reportagem sobre a taxa de desemprego numa década pode ser composta por um gráfico com a série temporal dessas taxas, ajudando o leitor a visualizar o comportamento da tendência. Contudo, para nós, importa primeiro saber se o gráfico ocupa lugar privilegiado no espaço da matéria; segundo se ele traz informações para além da série temporal (exemplo, as taxas em cada estado, ou entre homens e mulheres). Ou seja, analisa-se a utilização da visualização dos dados como um componente da narrativa da reportagem, no sentido de que ele ajuda muitas vezes a contar ou esclarecer o ponto da própria reportagem.

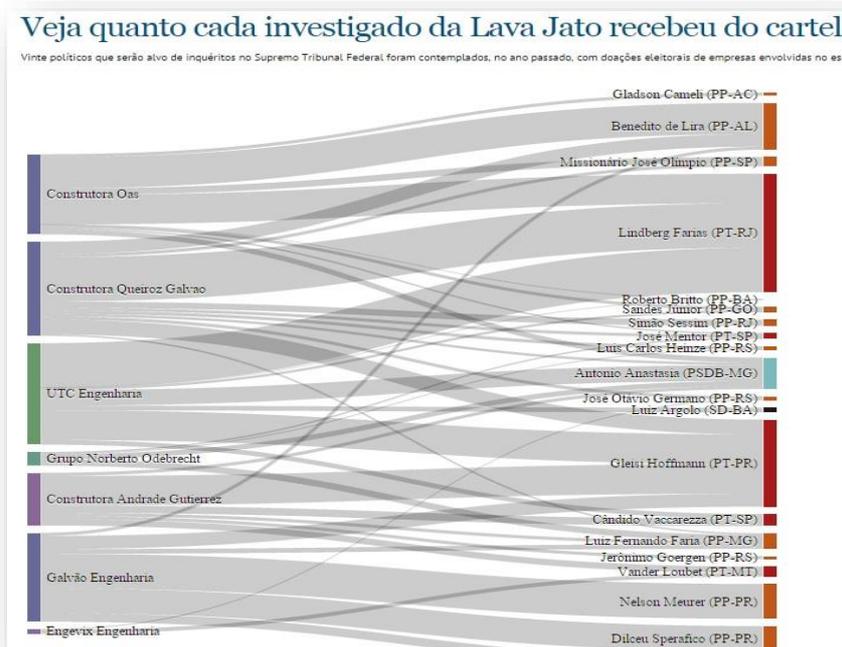
não). Na tentativa de compreender e interpretar os dados (fenômenos), a equipe utiliza comparações estatísticas, apontando causas, consequências ou implicações. Ou seja, neste nível temos as habilidades de 1) extração dos dados, 2) estruturação, 3) análise e 4) visualização.

A reportagem “As etnias nas eleições 2014”, publicada no dia 3 de outubro de 2014 na seção especial de dados do jornal O Globo, é um exemplo de matéria classificada como nível 1. Para a sua elaboração, a equipe reuniu dados divulgados pelo TSE a respeito da etnia dos candidatos daquele ano por partido, unidade federativa e cargo, e cruzou com informações demográficas divulgadas na Pnad 2013. Em seguida, as informações foram organizadas em gráficos interativos, além de um texto ter sido elaborado para ajudar a explicar as informações para os leitores.



No nível 2, a equipe busca uma base de dados de alguma instituição, ou produz uma própria, e a estrutura. Novamente, a forma como processam o banco fica bem clara no acabamento da reportagem. Diferentemente do nível 1, aqui a reportagem somente apresenta uma visualização dos dados. Não há uma análise textual das causas e consequências do fenômeno apresentado, deixando para o leitor a tarefa de cruzar informações e interpretar os dados. Em resumo, neste nível temos 1) extração, 2) estruturação e 3) visualização.

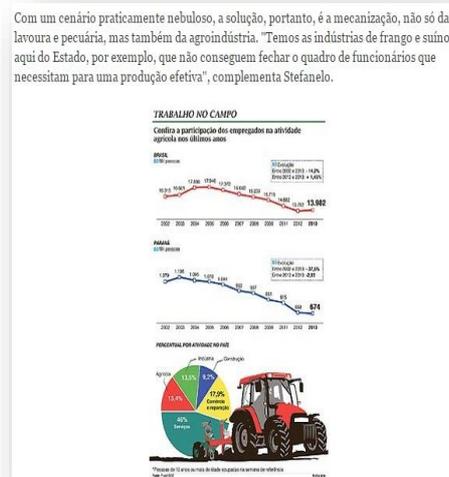
Encontramos exemplos desse nível em muitas matérias do jornal O Estado de São Paulo, principalmente nas que não estão em sua seção especial de Jornalismo de Dados, o Estadão Dados. Um exemplo de reportagem de nível 2 é a “Veja quanto cada investigado na Lava-Jato recebeu do cartel das empreiteiras na campanha de 2014”, publicada pelo jornal no dia 9 de março de 2015. Para a elaboração dessa matéria, a equipe extraiu os dados do financiamento de campanha na base do TSE e cruzou os nomes dos políticos investigados na operação Lava-Jato da Polícia Federal e do Ministério Público Federal. Há uma clara preocupação com a visualização dos dados, mas a publicação não faz um análise posterior sobre a relação desses dados. Nesse nível, poderíamos citar ainda como exemplo o “Basômetro”, do Estadão Dados, que consiste na visualização interativa sobre o comportamento dos partidos nas votações no Congresso Nacional.



No nível 3, encontram-se as reportagens que se baseiam em dados ou estudos produzidos e apresentados por uma instituição. A principal diferença é que a equipe não é responsável pela extração e produção da base dos dados, fazendo somente o trabalho de mediação entre a instituição que fez o estudo e o público leitor. Essas publicações tendem a

seguir as agendas de divulgação dos institutos de pesquisa. Neste nível, há apresentação de dados por meio de gráficos ou infográficos. Há também uma análise feita de forma direta, isto é, pelo próprio autor do texto, ou por meio de especialistas, que são entrevistados e aparecem no texto explicando causas e consequências do fenômeno apresentado. Desta forma, neste nível temos: 1) Análise e 2) Visualização dos dados.

A matéria “Escassez de mão de obra prejudica agricultura”, publicada no dia 11 de outubro de 2014 pela Folha de Londrina, é um exemplo de reportagem nível 3. Nela, o jornal apresenta dados da Pnad 2013, divulgados dias antes, sobre as características dos trabalhadores no campo e avança sobre um possível impacto na mão de obra. Ou seja, além dos dados foi realizada também uma análise desses números a partir de uma entrevista com um especialista no assunto. Além disso, foi construído um gráfico com uma série temporal que justificava e explicava o lide da reportagem.



No nível 4, estão classificadas as publicações que se baseiam em dados ou estudos produzidos e fornecidos por instituições sem que haja, no entanto, esforço de análise ou interpretação por parte dos jornalistas. Ou seja, não identificamos as habilidades de extração, estruturação e análise, apenas a de visualização dos dados para que o leitor faça a sua própria interpretação. A publicação “Os números da Pnad”, atualizada no 19 de setembro de 2014 no site do jornal O Globo, é um exemplo. Nela, estão organizados em gráficos interativos os

de causas ou consequências entre os dados) e a dimensão comunicativa (a centralidade da visualização do dado, compreendido aqui como um componente que ajuda o leitor a entender por imagens as relações entre os dados). Na nossa matriz, estabelecemos no eixo vertical uma escala que vai do extremo *Jornalismo com dados* ao *Jornalismo de Dados*. Trata-se de uma tentativa não só de conceituar o JD através de diferenciações implícitas na forma de produzir, como também, a partir do conteúdo das publicações.

Entendemos que a escala é um modelo mais adequado de compreensão do JD por dois motivos centrais: o primeiro é que ela contempla de forma mais harmoniosa as diferentes nuances existentes entre as publicações da imprensa que utilizam dados quantitativos, sem entrar no mérito de uma discussão sobre a sua qualidade estética. A segunda é que a matriz contempla dimensões que, a nosso ver, são imprescindíveis para estabelecer uma diferença entre a produção de um conteúdo que nasce e se estrutura a partir dos dados quantitativos e a produção de um conteúdo que utiliza dados quantitativos. Ora se em ambos os casos há a utilização de dados, o mesmo não podemos dizer sobre a competência que as equipes mobilizaram para produzir o conteúdo com dados. Enquanto no primeiro, o trabalho é limitado à utilização de dados, no segundo caso, envolveria não somente manipular dados, como identificar histórias através das relações entre os dados nos seus diferentes contextos.

Referências

ANDERSON, C.W. (2015). “Between the Unique and the Pattern: Historical Tensions in our Understanding of Quantitative Journalism.” forthcoming in a special issue of Digital Journalism, Journalism in an Era of Big Data: Cases, Concepts, and Critiques, Seth Lewis (ed).

_____; **BELL, E.**; **SHIRKY, C.** Post-Industrial Journalism: Adapting to de presente. Columbia Journalism School, 2013. Disponível: http://towcenter.org/wp-content/uploads/2012/11/TOWCenter-Post_Industrial_Journalism.pdf

BARBOSA, Suzana. O que é Jornalismo Digital em Base de Dados. Trabalho apresentado ao Grupo de Trabalho “Estudos de Jornalismo”, do XV Encontro da Compós, na Unesp, Bauru, SP, em Junho de 2006.

_____. Jornalismo Digital em Base de Dados (JDBD): um paradigma para produtos jornalísticos digitais dinâmicos. Tese de doutorado. Universidade Federal da Bahia, 2007.

BRADSHAW, Paul. O que é Jornalismo de Dados, in Manual de Jornalismo de Dados, 2014. Disponível em http://datajournalismhandbook.org/pt/introducao_0.html

- COHEN, S.; HAMILTON, J. e TURNER, F.** Computational journalism: How computer scientists can empower journalists, democracy's watchdogs, in the production of the news in the public interest. *Communication of the ACM*, vol. 54, n.10. 2011.
- FLEW, T.; SPURGEON, C.; DANIEL, A. e SWIFT, A.** The promise of computational journalism. *Journalism Practice*, vol. 6, n. 2, pp 157-171, 2012.
- GRAY, Jonathan; BOUNEGRU, Liliana e CHAMBERS, Lucy.** Manual de Jornalismo de Dados, 2014. Disponível em <http://datajournalismhandbook.org/pt/>
- HACKETT, K.** Journalism data age. *Quill Magazine*. Setembro pp 35-38, 2013.
- HAMILTON, JAMES T. and TURNER, FRED.** Accountability Through Algorithm: developing the field of computational journalism. Report from Center for Advanced Study in the Behavioural Sciences, julho 27-31, 2009.
- HEAVER, David H; McCOMBS, Maxwell.** Journalism and Social Science: a new relationship. *Public Opinion Quarterly* Vol. 44:477-494, 1980.
- HIRST, Tony.** Several takes on defining Data Journalism, 2013. Disponível em <http://schoolofdata.org/2013/06/11/several-takes-on-defining-data-journalism/>. Acessado em 20 de novembro de 2014.
- HOWARD, Alexander Benjamin.** The art and Science of data-driven journalism. Tow Center Digital Journalism, Columbia Journalism School, 2014.
- HUNTER, Lee Mark; HANSON, Nils; SABBAGH, Rana; SENGERS, Drew Sullivan; SVITH, Flemming Tait e THORSEN, Pia.** A investigação a partir de histórias: um manual de jornalismo investigativo. Uneco, 2013.
- JUNIOR, Walter Teixeira Lima.** Jornalismo Inteligente (JI) na era do Data Mining. Trabalho apresentado no 2º Encontro Nacional de Pesquisadores em Jornalismo. Salvador, 2004.
- _____; Jornalismo Computacional em função da Era do Big Data. Paper 9º Encontro Nacional de Pesquisadores em Jornalismo. Rio de Janeiro, novembro, 2011.
- KNIGHT, Megan.** Data journalism in the UK: a preliminary analysis of form and content. *Journal of Media Practice*, Vol. 16, No. 1, 55–72, 2015
- LEWIS, Seth C; NIKKI Usher.** 2013. "Open Source and Journalism: Toward New Frameworks for Imagining News Innovation." *Media, Culture and Society* 35: 602–619. doi:10.1177/0163443713485494
- LORENZ, M.** Por que jornalistas devem usar jornalismo de dados? In: Manual de Jornalismo de Dados. GRAY, J.; BOUNEGRU, L.; CHAMBERS, L. (org), 2013
- MEYER, Philip.** The new precision journalism. Indiana University Press, 1991.
- SILVER, Nate.** What the Fox Knows, 2014. Disponível em <http://fivethirtyeight.com/features/what-the-fox-knows/>. Acessado 20 novembro de 2014.
- _____. O Sinal e o Ruído. Intrínseca, Rio de Janeiro 2013.
- STRAY, Jonathan.** The Data Journalist's Eyes, An Introduction. Disponível em (<http://towcenter.org/blog/the-data-journalists-eye-an-introduction/>) Acessado em 20/11/2014
- WINACHT, Stefan; SPILLER, Ralf.** Datenjournalismus in Deutschland. *Journal of communication science Publizistik*, vol. 59/4, pp. 411-433, 2014.