

Jornalismo de Dados: conceito e categorias

Data journalism: Concept and categories

Leonardo Mancini¹
Fabio Vasconcellos²

RESUMO

Neste artigo, discutimos possibilidades de classificação do Jornalismo de Dados praticado no Brasil. Após observar como os dados quantificáveis estão presentes em narrativas jornalísticas em sites de notícias de todo o país, este artigo propõe uma classificação que inclui o que seria o Jornalismo de Dados, isto é, reportagens em que os dados guiam a construção da narrativa e que podem ser encontradas em veículos como Folha de São Paulo, O Globo e Estado de S. Paulo, e aquela que chamamos de Jornalismo Com Dados, ou seja, publicações que utilizam dados sem que esses sejam o foco da narrativa. Metodologicamente, este estudo monitorou a versão online de 48 veículos jornalísticos durante cinco meses e analisou 2.296 reportagens, com o intuito de identificar competências envolvidas nas reportagens com dados que consideramos fundamentais, entre elas, a competência investigativa (busca e tratamento próprio dos dados), a interpretativa (habilidade para contar a história e as relações entre os dados) e a competência comunicativa (a visualização dos dados como forma de ampliar a compreensão da própria história).

Palavras-chave: Jornalismo de Dados, RAC, visualização de informação.

ABSTRACT

In this article, the authors propose new ways to classify how Data Journalism is being used in Brazilian newspapers. After having observed how data is presented on news pieces in online newspapers from all of the country, this study shows there are many different approaches, ranging from what is considered the typical data journalism pieces, which can be found on Folha de São Paulo, O Globo and Estado de S. Paulo, to what is considered here as “journalism with data”, where the data is not on the center of the story. Methodologically, the research monitored 48 online newspapers during 5 months and analyzed 2.296 news pieces, in which we were able to observe the main skills needed to produce data journalism: investigative competences (data search and treatment), interpretative competences (how to tell the story and show the relations between the data) and communicative competences (data visualizations as a way to facilitate the understanding of the story).

Keywords: Data Journalism, CAR, Data Visualization.

¹ Chefe do Departamento de Comunicação Teórica e Prática da ESPM Rio. Rua do Rosário, 90, 20041-004, Rio de Janeiro, RJ, Brasil. E-mail: lmancini@espm.br

² Jornalista e professor da ESPM Rio. Rua do Rosário, 90, 20041-004, Rio de Janeiro, RJ, Brasil. E-mail: fabio.vasconcellos@espm.br

Introdução

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 (JD), uma prática que vem se disseminando em redações, *blogs* e *sites* em todo mundo. O *The Data Journalism Handbook* (Gray *et al.*, 2014) 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 chinês 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 alguns motivos. O primeiro é que o modelo de negócio das empresas jornalísticas atuais, chamado de industrial por C.W. Anderson, Emily Bell e Clay Shirky, da Escola de Jornalismo de Columbia, no estudo “Post-Industrial Journalism: adapting to the present” (Anderson *et al.*, 2013), está em transformação, indo em direção, provavelmente, a organizações com estruturas menores, mais dinâmicas e com a internet como aliada, e não inimiga, de suas práticas.

Em paralelo a esse movimento, nota-se que o JD se insere em um contexto de disseminação da cultura de dados abertos na *web*, tendência que governos de muitos países vêm seguindo para atender ao princípio democrático do *Open Government*. Com isso, o trabalho do jornalista muda de natureza. Nas palavras de (Flew *et al.*, 2012), “quando a informação era escassa, a maior parte de nossos esforços estavam voltados a caçar e reunir dados. Agora que a informação é abundante, processá-la tornou-se mais importante”.

Visto para além do campo jornalístico em si, podemos dizer que a prática do JD, ao exigir de certo modo uma maior qualificação profissional, mobiliza a discussão sobre a qualidade do papel da imprensa na dinâmica democrática. Em outras palavras, se o jornalismo pode ser visto como um importante ator no processo da *accountability* político, entendida aqui como ações que geram fluxos e contrafluxos da temática da vida pública e ainda exige do campo político justificativas para suas decisões, sua atuação institucionalizada acaba por refletir sobre a *performance* das democracias, como discutem Hallin e Mancini (2004). Nesse sentido, mudanças no *modus operandi* da produção jornalística, ou seja, a qualificação que leva, por exemplo, ao Jornalismo de Dados, implica de certo modo mudanças no tipo e na qualidade do conteúdo ofertado aos atores do espaço público democrático.

Diante disso, como então podemos compreender a maneira pela qual o JD tem sido utilizado no Brasil? Para discutir isso, analisamos a ainda escassa bibliografia disponível a respeito do JD, muito do que tem sido escrito foi produzido por profissionais do próprio campo jornalístico, e fizemos uma análise da versão *online* de 48 veículos jornalísticos, presentes nos 27 estados do país, durante cinco meses (de setembro de 2014 a março de 2015, considerando uma interrupção de 30 dias entre 15 de dezembro de 2014 e 15 de janeiro de 2015). Essa amostra, que representa 67% da circulação impressa verificada do período, e a coleta de informação, deram origem a uma base de dados com 2.296 reportagens que utilizam dados de alguma maneira em sua composição.

Nesse primeiro momento, portanto, a nossa problematização sobre como o JD tem sido praticado no Brasil procurou responder não sobre o que exatamente seria JD, questão que, obviamente, acompanha todo o nosso texto, mas especificamente que características ou competências necessárias podemos esperar quando lidamos com essa prática jornalística considerando o contexto brasileiro.

Jornalismo de Dados: alguns princípios

Na seção do *Handbook* dedicada a definir o que é JD, 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 em uma 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. Ele reconhece essa limitação e argumenta que o JD 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).

Temos aqui um ponto interessante. Se a utilização de dados em si, até por sua imprecisão conceitual, não são uma novidade no processo produtivo da notícia, ela se tor-

na de certo modo dispensável para caracterizar o que venha a ser o JD. Por outro lado, a investigação jornalística, vista isoladamente, nos leva ao mesmo problema. Bradshaw, de certo modo, entende que o JD se definiria mais pela associação desses dois termos 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 mesmo produzir “infográficos envolventes”.

Anderson (2015) tem uma avaliação crítica desse argumento. Para ele, não apenas o uso de dados, mas também a simples apropriação da tecnologia torna-se inadequada para definir Jornalismo de Dados. Isso porque, se entendermos o jornalismo pela perspectiva de que ele envolve a busca e a reunião de informações de materiais humanos e sociais, como documentos, entrevistas e dados, então o JD 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 se debatam as diferentes formas pelas quais reportagens utilizavam dados muito antes do processo de aperfeiçoamento do aparato tecnológico³.

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 JD. 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 “*backing journalism*”,

ou seja, a capacidade explorar tecnologias filtrando informações e colocando-as de forma visual.

A dimensão das novas competências

Talvez o ponto central no debate sobre o que venha a ser o JD seja a incorporação de novas competências pelo campo jornalístico. Esse tema está presente nas inúmeras tentativas de definir o JD, contudo, 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 praticado no Brasil, pode ajudar a sair de algumas dificuldades teóricas que ainda persistem.

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?

Essa pode ser uma resposta que, contudo, permanece ainda 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 limite desse argumento é que, se verdadeiro, o JD não seria resultado do aprimoramento do campo jornalístico, mas fruto de mudanças no campo político-institucional. Por consequência, governos que não adotam a abertura de dados não criariam – nem promoveriam – o JD nos seus países.

³ 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 (Gray *et al.*, 2014).

De certo modo, a tecnologia incentiva e pode criar condições para novas práticas no processo de apuração jornalística. 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 não presencial. Contudo, e esse é o nosso ponto, o telefone não modifica o lugar do jornalista como profissional que precisa desenvolver alguma competência para realizar entrevistas. Para além do objeto físico que simula uma relação presencial da entrevista, existe um jornalista dotado de habilidades e do interesse em extrair informações por meio do método de entrevista. O ponto aqui não é o objeto técnico em si, mas como esse objeto pode ser explorado.

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

Dizer que o jornalismo sempre dependeu de alguma competência para ser realizado não chega a ser novidade. O saber coletar informações, organizá-las e produzir um texto ou imagens, segundo os critérios do campo jornalístico, são competências clássicas exigidas do jornalismo após o processo de profissionalização crescente na virada para o século XX. Variações sobre essas competências também foram apropriadas pelos chamados “jornalistas investigativos” que, segundo Hunter *et al.* (2013), são caracterizados como aqueles que têm uma atitude ativa na busca de informação, em contraposição à atitude passiva de um mero agente mediador entre instituições e a sociedade⁴.

Dito isso, o que precisa ser problematizado na definição do JD, para nós, em especial na maneira como existe no Brasil, é como a incorporação ou o aprofundamento de algumas dessas competências, até então pouco ou quase nunca utilizadas pelo jornalismo, altera o modo de fazer notícia ou, se preferirem, possibilita o manejo das novas técnicas na construção de conteúdos. Em síntese, não adianta a posse de *software* de análise estatística se não houver um

profissional que conheça essa ferramenta, saiba como ela funciona e como ela pode atender ao propósito do jornalismo de produzir informação e conhecimento relevantes.

Por outro lado, se estamos apontando o caminho do “como” como procedimento metodológico para pensar o JD, isto é, como é produzida informação a partir dessas novas técnicas de manejo de dados, nos deparamos com uma outra questão: jornalistas dotados dessas novas competências, capazes de 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 outras variações no tipo de conteúdo que utilizam dados e outras formas de visualização desses dados?

A utilização de dados pelo jornalismo

Embora seja relativamente recente, o JD pode ser considerado parte de uma tradição iniciada na década de 1960 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 erro. O estudo de Meyer acabou por incentivar o surgimento, nos anos 1990, 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 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).

⁴ “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, p. 8).

Em certo sentido, o JD 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 Turner, 2009; Gray *et al.*, 2014; Hackett, 2013; Howard, 2014). Coddington (2014) concorda com essa afirmativa. Para ele, o RAC, o JD e o Jornalismo Computacional, aquele mais voltado para a programação de máquinas, são categorias que se sobrepõem. Apesar disso, o autor vê algumas diferenças importantes com relação à orientação desses três grupos de profissionais.

Coddington propõe uma tipologia baseada em quatro dimensões para caracterizar práticas jornalísticas. No que diz respeito à natureza profissional, ele compara as diferenças de atitude entre os jornalistas com larga experiência e aqueles 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 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 aqueles oriundos do Big Data. A quarta – e última – dimensão observa como os jornalistas compreendem o papel do seu público. Nesse caso, se como um público passivo ou se 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 JD 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 JD 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 JD, 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 JD (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 práticas. 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 JD.

Para Stray, o JD 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”. Nesse sentido, para o autor, 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”.

Partindo da premissa de que o JD se inicia 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, desse modo, que essa é nessa etapa em que o jornalismo mais se aproxima da ciência, com forte inclinação para a matemática, estatística e lógica. “No JD, existe profundo conhecimento técnico e específico” que permite, por exemplo, ao jornalista comparar as taxas de desemprego com as 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

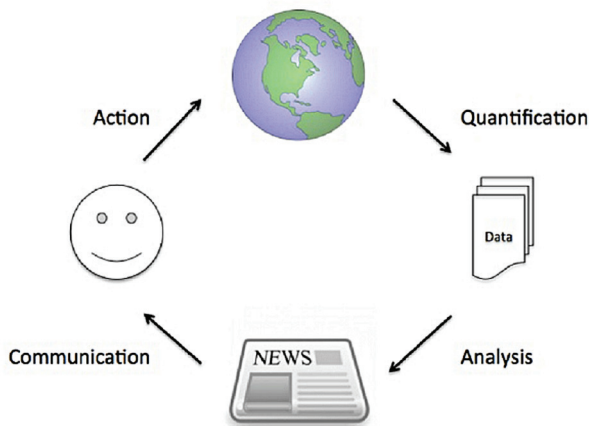


Figura 1. Modelo de JD.

Figure 1. Data Journalism model.

Fonte: Stray (2014).

no vácuo, mas para mentes humanas e para a sociedade humana”. Desse 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 confirmá-la. Sinteticamente, eis o modelo que definiria o JD.

Conhecido por suas publicações de JD, Silver (2014) apresenta também a sua definição de procedimentos que ajudariam a determinar as novas 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 JD. 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 JD e sua forte aproximação com o campo da estatística está presente no livro “O sinal e o ruído”, lançado no Brasil em 2012.

Na apresentação do *site*, em 2014, Silver diz que fazer JD não significa apenas usar números, em vez de palavras. Segundo afirma, “o uso de números não é necessário, nem suficiente, para produzir boas obras de jornalismo” (Silver, 2014). Apesar dessa 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.

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

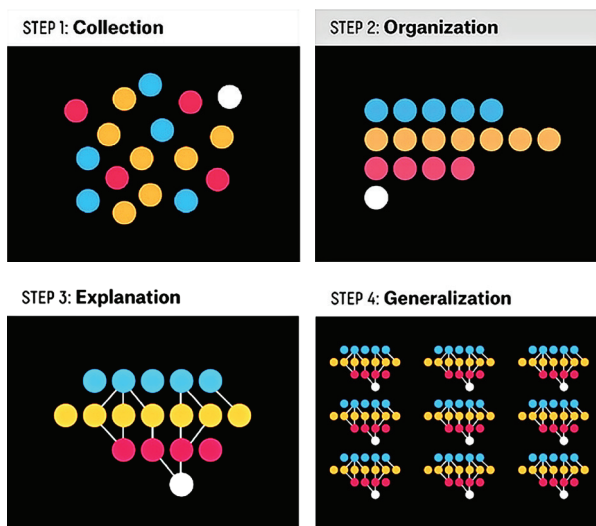


Figura 2. Procedimentos de JD.

Figure 2. Data Journalism methods.

Fonte: Silver (2014).

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, experimentos ou 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 explicação. 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 explicação 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 como forma de verificação de hipóteses ou apresentação de previsões, bem diferente do modelo de inferência praticado pelo jornalismo tradicional, que parte de um conhecimento não estatístico, mas intuitivo.

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

Se a competência dos jornalistas para manejar dados, organizá-los e analisá-los é um ponto de convergência entre os autores e se podemos afirmar que essa prática não nasce espontaneamente, mas é facilitada pelos usos da tecnologia da internet e pelos *softwares* estatísticos e de visualizações, resta-nos compreender o modo como o jornalismo tem se apropriado 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 Silver (2014), mas que procure também agregar a dimensão ativa do jornalismo interessado em revelar algo para além da simples divulgação de relatórios quantitativos de agentes públicos. 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 grau de complexidade de gráficos e infográficos. Embora seja desejável, por conta da sua maior comunicabilidade, nos parece mais interessante inicialmente saber se a comunicação gráfica sugere ou induz análises e compreensões sobre o conteúdo ou 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 como sendo JD apenas aquele conteúdo que necessariamente incorpora apresentações gráficas mais arrojadas.

Os casos apresentados no Manual de JD (Gray *et al.*, 2014), por exemplo, limitam a identificação do JD, ao menos no caso do Brasil ou de países cujas empresas de jornalismo não tenham adquirido tecnologias mais avançadas. Há duas questões implícitas nesse ponto.

A primeira delas é que, ao exagerar no modelo de visualização do que venha a ser JD, o Manual estabelece os parâmetros através dos quais essa prática deve ser identificada. Esse ponto está diretamente vinculado ao segundo.

Ao estabelecer um parâmetro de visualização que depende sobremaneira da utilização de *softwares* sofisticados ou de programadores, o Manual, de certo modo, associa essa prática a uma estética dos dados. Em outros termos, toda produção que não apresente visualizações sequer próximas desses modelos estaria fadada a não ser considerada JD. 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 sozinha 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 JD.

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 JD. 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 a 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 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 classifi-

cação de *Jornalismo com Dados* até o outro extremo, qual seja, *JD*. Essa matriz de algum modo limita o corte abrupto entre o que é, e o que é não *JD*, porque permite que as nuances entre as publicações possam ser contempladas. Em suma, a escala *Jornalismo com Dados* em um extremo e, em outro, *JD*, busca contemplar as diversas formas pelas quais a imprensa já produz conteúdo com dados quantitativos e como agora passou a produzir. Para isso, é necessário, no eixo horizontal, estabelecer quais atributos as reportagens apresentam. Esses atributos buscam atender ao modelo proposto por Stray (2014), Silver (2014) e Bradshaw (2014), qual seja: a importância de o *JD* buscar e apresentar relações entre os dados e, a partir daí, contar as suas histórias.

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 pelos quais a equipe de jornalistas foi responsável por extrair e estruturar o material bruto de alguma base e/ou foi 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 pela equipe⁵.

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.

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/incentive a compreensão analítica da reportagem de dados⁶.

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

No Nível 1, a equipe busca e/ou produz uma base de dados primária (a partir dos microdados) ou secundária

⁵ 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 pelo 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).

⁶ No geral, todo gráfico atende a esse princípio. Por exemplo, uma reportagem sobre a taxa de desemprego em uma 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 (alto, meio ou pé da página); 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.

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	□	□	□	□
	Nível 2	□	□	□	
	Nível 3			□	□
	Nível 4			□	
	Nível 5				

Figura 3. Matriz de classificação do JD.
 Figura 3. Data journalism classification matrix.

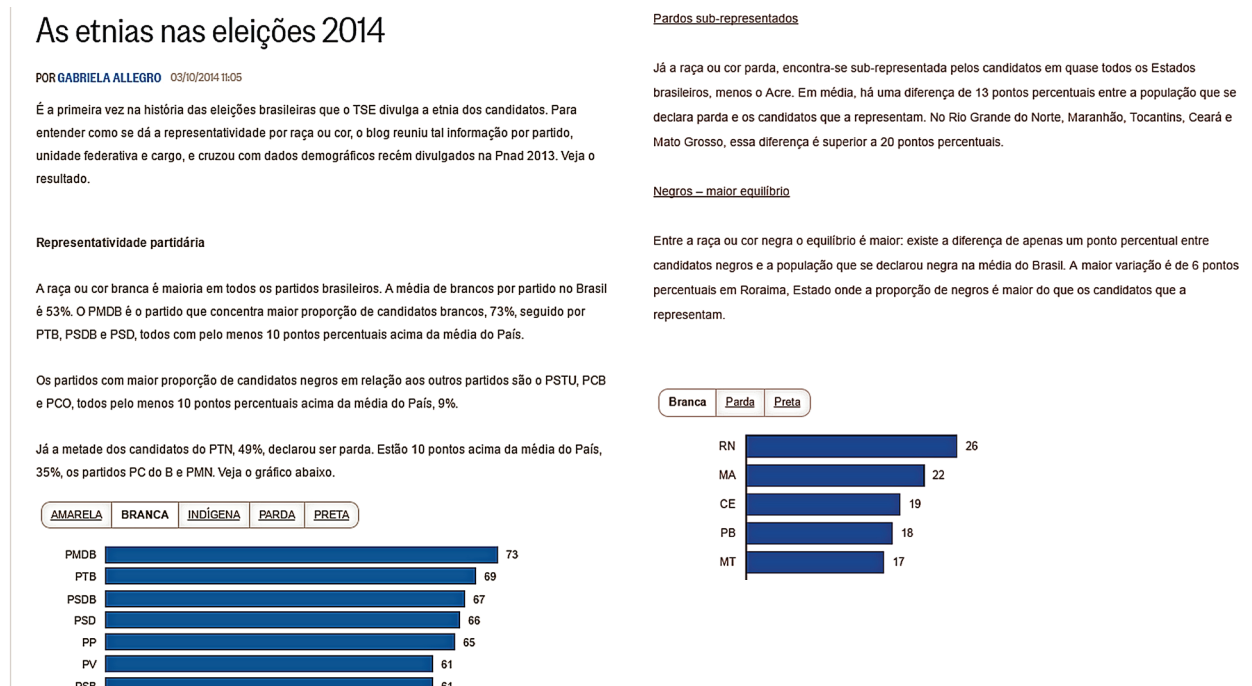


Figura 4. “As etnias nas eleições de 2014”, jornal O Globo, 3 de outubro de 2014.
 Figura 4. “The ethnicities at the 2014 elections”, published at O Globo newspaper, October 3th, 2014.

dária (a partir de dados agregados). A partir da base, é realizada uma análise, que resulta em um texto e em uma apresentação de gráficos ou infográficos (interativos ou não). Na tentativa de compreender e interpretar os dados (fenômenos), a equipe utiliza comparações estatísticas, apontando causas, consequências ou implicações. Ou seja, nesse nível, temos as competências da (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 (Figura 4) na seção especial de dados do jornal O Globo, é um exemplo de matéria classificada como nível 1. Para a sua elaboração, a equipe

Veja quanto cada investigado da Lava Jato recebeu do cartel das empreiteiras na campanha de 2014

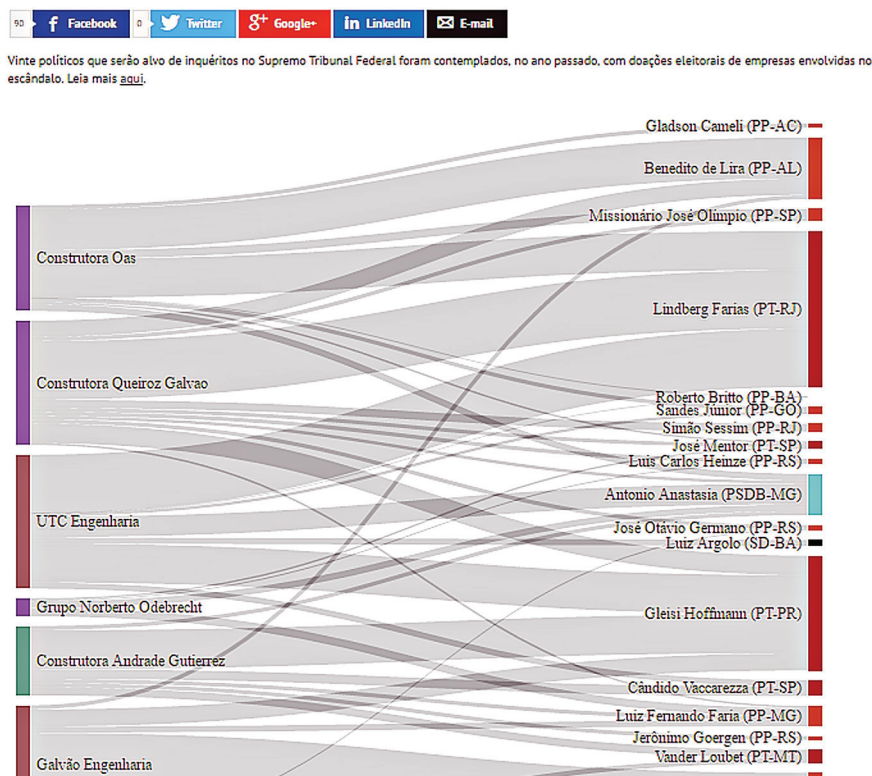


Figura 5. “Veja quanto cada investigado na Lava-Jato recebeu do cartel das empreiteiras na campanha de 2014”, jornal O Estado de São Paulo, 9 de março de 2015.

Figura 5. “How much each one of the suspects investigated by the Car-Wash Operation received by the cartel in 2014”, published at O Estado de São Paulo newspaper, March 9th, 2015.

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, nesse nível, temos (1) extração, (2) estruturação e (3) visualização.

Encontramos exemplos desse nível em muitas matérias do jornal O Estado de S. Paulo, principalmente nas que não estão em sua seção especial de JD, 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” (Figura 5), 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 uma análise posterior sobre a relação desses dados.

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

11/10/2014
Escassez de mão de obra prejudica agricultura
 Nem os melhores salários e a capacitação contínua dos trabalhadores rurais tem segurado as famílias no campo

A valorização do agronegócio brasileiro, com ótimo preço médio das commodities, exportação em alta de alguns segmentos e a grande representatividade no Produto Interno Bruto (PIB) nacional apontam para uma rentabilidade efetiva no meio rural nos últimos anos. Por outro lado, uma preocupação dos produtores vem se arrastando e se tornou praticamente insustentável: não há mão de obra para trabalhar nas propriedades rurais. Nem os melhores salários e a capacitação contínua dos trabalhadores rurais para operar máquinas e equipamentos agrícolas tem segurado as famílias no campo. Os números da Pesquisa Nacional por Amostra de Domicílios (Pnad) de 2013, divulgado pelo Instituto Brasileiro de Geografia e Estatística (IBGE) no mês passado, apontam para um caminho sem volta. Se em 2002 eram 16,3 milhões de pessoas trabalhando no campo em todo o território nacional, no ano passado esse número sofreu retração de 14,2%, para 13,9 milhões de brasileiros atuando na atividade agrícola. De 2012 para 2013, houve um pequeno fôlego para o segmento, com incremento de 1,45%. Já no Paraná, a queda é contínua desde 2003: uma incrível diminuição de 1,1 milhão de pessoas para 674 mil pessoas atuantes no meio rural, algo em torno de 40%.

Com um cenário praticamente nebuloso, a solução, portanto, é a mecanização, não só da lavoura e pecuária, mas também da agroindústria. "Temos as indústrias de frango e suíno aqui do Estado, por exemplo, que não conseguem fechar o quadro de funcionários que necessitam para uma produção efetiva", complementa Stefanelo.

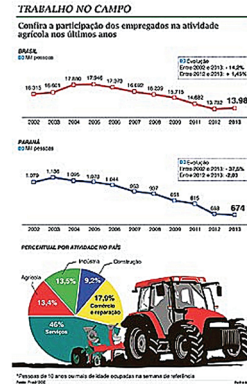


Figura 6. “Escassez de mão de obra prejudica agricultura”, Folha de Londrina, 11 de outubro de 2014.
 Figura 6. “Lack of workforce slows down agriculture”, published by Folha de Londrina newspaper, October 11th, 2014.

Veja os números da Pnad 2014

A Pesquisa Nacional por Amostra de Domicílios (Pnad) 2014 mostra piora no mercado de trabalho, com alta do desemprego, aumento da informalidade e freio no crescimento da renda, mas também revela que a desigualdade no país caiu pelo 10º ano seguido. Navegue pelos principais temas no gráfico interativo abaixo:

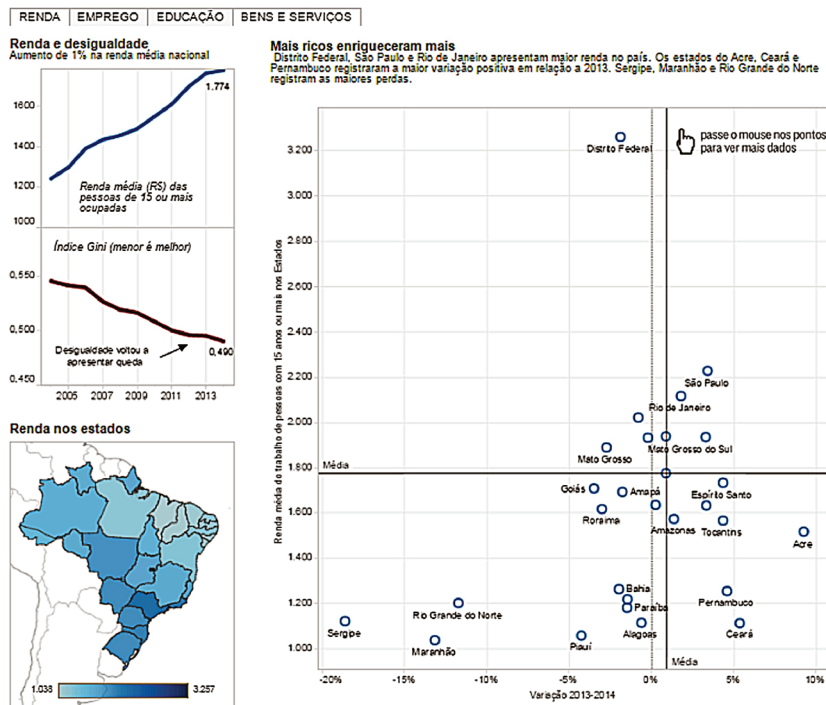


Figura 7. “Os números da Pnad”, O Globo, 19 de setembro de 2014.
 Figura 7. “PNAD statistics”, published by O Globo newspaper, September 19th, 2014.



Outra “tipologia” é o Moradias Arroio, ao lado do “porto seco”, na CIC. É vila pequenina e artificial, nascida em meio àqueles entroncamentos de rodovias próprias da região industrial. A exemplo do Corbélia, parece um dormitório. Ainda não há associação de moradores. O comércio só se desenha. Na falta de um representante, os moradores apontam o ex-craque Calita - Carlos Marques da Luz, 78 anos - como o ilustre que faz do Arroio um lugar de verdade. Ele exibe suas glórias nos campos, guardados numa pasta de recortes de jornal. “Pede para o Calita mostrar a pasta”, grita um. Calita mostra. Depois conta que vivia no Bigorrihlo. Paciência. E que varre a frente a calçada todos os dias. É um cuidado que não dispensa.

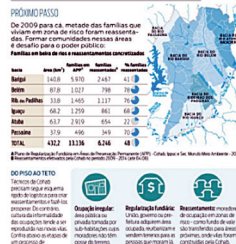


Figura 8. “Cultura da favela persegue os reassentamentos urbanos”, Gazeta do Povo, 25 de setembro de 2014.
 Figura 8. “Shantytown culture resist at urban resettlement”, published by Gazeta do Povo newspaper, September 25th, 2014.

não é responsável pela extração 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. Nesse 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. Dessa forma, nesse nível, temos: (i) Análise e (ii) Visualização dos dados.

A matéria “Escassez de mão de obra prejudica agricultura”, publicada no dia 11 de outubro de 2014 pela Folha de Londrina (Figura 6), é um exemplo de reportagem nível 3. Nela, o jornal apresenta dados da Pnad 2013, que tinha sido divulgada fazia pouco tempo, em relação aos trabalhadores no campo. Em conjunto com os dados, também foi realizada 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 competências 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 (Figura 7), é um exemplo. Nela, estão organizados em gráficos interativos os diversos números divulgados pela pesquisa Pnad 2014, sem que exista qualquer tipo de texto que sirva para contextualizar as informações apresentadas para o público. Não houve um esforço maior de busca e análise com outras bases de dados, mas a reprodução de forma gráfica dos dados da Pnad.

Finalmente, no nível 5, classificamos as matérias que utilizam dados ou números a partir da divulgação de indicadores econômicos, financeiros ou sociais. Os dados são repassados pelos institutos, mas não existe uma maior preocupação com a análise (causas, relações e consequências) e com a exploração dos potenciais da interpretação quantitativa. Eventualmente, alguma visualização de dados é desenvolvida, mas tem uma função mais ilustrativa. São reportagens com menor aprofundamento de todas as categorias elencadas na matriz do JD. Nessa categoria, estão também as publicações curtas, que citam dados, mas não exploram os dados contextual ou visualmente; ou mesmo reportagens mais longas, mas que nitidamente utilizam dados de forma tangencial.

A reportagem “Cultura da favela persegue os reassentamentos urbanos”, publicada no dia 25 de setembro de 2014 pelo site da Gazeta do Povo (Figura 8), é um exemplo de notícia classificada como nível 5. Isso porque a publicação não tem com foco principal os dados que são apresentados na matéria, que dizem respeito aos

reassentamentos realizados na região. Os dados são apenas um apoio, um pano de fundo para ilustrar a história dos assentamentos.

Conclusão

Neste texto, procuramos problematizar a definição do JD. Para isso, recorremos a autores que apresentaram algumas linhas gerais sobre essa prática. O nosso argumento baseia-se na seguinte questão: se dados é termo problemático na definição do JD, em razão da sua imprecisão conceitual, o mesmo ocorre pela análise que leve em conta apenas a dimensão tecnológica que toma conta de muitas redações. No primeiro caso, a imprecisão decorre do fato de que dados quantitativos não são uma novidade na produção de conteúdo das redações; no segundo ponto, a questão diz respeito à visão determinística, ou seja, uma vez adotada a tecnologia caminhamos necessariamente para a prática do JD?

A nosso ver, uma forma de melhor entender o que é e como se faz JD é construir uma matriz que leve em consideração três dimensões, que dizem respeito às competências que o jornalismo pode ou não adotar: a dimensão investigativa (atuação proativa na busca de dados e revelações), a interpretativa (capacidade ou interesse em expor relações 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 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. 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 as competências que a equipe empreendeu para produzir o

conteúdo. Em suma, na nossa matriz, importa não exclusivamente o uso de uma técnica, mas como a técnica pode ser utilizada pela equipe para produzir conteúdo jornalístico.

Referências

- ANDERSON, C.W. 2015. Between the Unique and the Pattern: Historical Tensions in our Understanding of Quantitative Journalism. In: S. LEWIS (ed.), *Digital Journalism, Journalism in an Era of Big Data: Cases, Concepts, and Critiques*. *Digital Journalism*, 3(3):321-330.
<http://dx.doi.org/10.1080/21670811.2014.976407>
- ANDERSON, C.W.; BELL, E.; SHIRKY, C. 2013. Post-Industrial Journalism: Adapting to the Present. Columbia Journalism School, 2013. Disponível em: http://towcenter.org/wp-content/uploads/2012/11/TOWCenter-Post-Industrial_Journalism.pdf. Acesso em: 01/12/2014.
- BARBOSA, S. 2006. O que é Jornalismo Digital em Base de Dados. In: Encontro Anual da Compós, XV, Bauru, 2006. Disponível em: http://www.academia.edu/15328433/O_QUE_%C3%89_JORNALISMO_DIGITAL_EM_BASES_DE_DADOS_1. Acesso em: 12/05/2016.
- BARBOSA, S. 2007. *Jornalismo Digital em Base de Dados (JDBD): um paradigma para produtos jornalísticos digitais dinâmicos*. Salvador, BA. Tese de Doutorado. Universidade Federal da Bahia, 331 p.
- BRADSHAW, P. 2014. O que é Jornalismo de Dados. Manual de Jornalismo de Dados. Disponível em: http://datajournalismhandbook.org/pt/introducao_0.html. Acesso em: 15/05/2015.
- CODDINGTON, M. 2014. Clarifying Journalism's Quantitative Turn. *Digital Journalism*, 3(3):331-348.
<http://dx.doi.org/10.1080/21670811.2014.976400>
- FLEW, T.; SPURGEON, C.; DANIEL, A.; SWIFT, A. 2012. The promise of computational journalism. *Journalism Practice*, 6(2):57-171. <http://dx.doi.org/10.1080/17512786.2011.616655>
- GRAY, J.; BOUNEGRU, L.; CHAMBERS, L. 2014. Manual de Jornalismo de Dados, 2014. Disponível em: <http://datajournalismhandbook.org/pt/>. Acesso em: 01/12/2014.
- HACKETT, K. 2013. Journalism Data Age. *Quill Magazine*, 5:35-38.
- HALLIN, D.; MANCINI, P. 2004. *Comparing Media Systems: Three Models of Media and Politics*. Cambridge, Cambridge University Press, 342 p.
<http://dx.doi.org/10.1017/CBO9780511790867>
- HAMILTON, J.T.; TURNER, F. 2009. Accountability Through Algorithm: developing the field of computational journalism.

Report from Center for Advanced Study in the Behavioural Sciences. Disponível em: <http://web.stanford.edu/~fturner/Hamilton%20Turner%20Acc%20by%20Alg%20Final.pdf>. Acesso em: 05/05/2016.

HOWARD, A.B. 2014. The art and Science of data-driven journalism. *Digital Journalism*, 3(3):331-348.

HUNTER, L.M.; HANSON, N.; SABBAGH, R.; SENEGERS, D. S.; SVITH, F. T.; THORSEN, P. 2013. A investigação a partir de histórias: um manual de jornalismo investigativo. Unesco. Disponível em: <http://unesdoc.unesco.org/images/0022/002264/226456POR.pdf>. Acesso em: 12/05/2015

MEYER, P. 1991. *The new precision journalism*. Indiana, Indiana University Press, 273 p.

SILVER, N. 2014. What the Fox Knows, Disponível em: <http://fivethirtyeight.com/features/what-the-fox-knows/>. Acesso em: 01/12/2014.

STRAY, J. 2014. The Data Journalist's Eyes, An Introduction, Disponível em <http://towcenter.org/blog/the-data-journalists-eye-an-introduction/>. Acesso em: 12/05/2015

TRÄSEL, M. 2013. Jornalismo guiado por dados: relações da cultura hacker com a cultura jornalística. In: Encontro Anual da Compós, XXII, Bahia, 2013. *Anais...* Universidade Federal da Bahia. Disponível em: http://www.academia.edu/3136931/JORNALISMO_GUIADO_POR_DADOS_rela%C3%A7%C3%B5es_da_cultura_hacker_com_a_cultura_jornal%C3%ADstica. Acesso em: 12/05/2016.

Submetido: 26/11/2015

Aceito: 29/03/2016