



## Jornalismo Guiado por Dados: características definidoras e uma proposta de formulação do conceito

Marcelo Träsel<sup>1</sup>

**Resumo:** O jornalismo guiado por dados (JGD) compreende diversas práticas profissionais, cujo ponto em comum é o uso de bases de dados como principal fonte de informação para a produção de notícias. As práticas de JGD envolvem técnicas de reportagem assistida por computador (RAC), visualização de dados, infografia, criação e manutenção de bases de dados e a política de acesso à informação e transparência pública de governos. A partir de manuais e publicações sobre o tema, são analisados alguns dos principais elementos das práticas e rotinas produtivas ligadas ao JGD. Finalmente, é oferecida uma proposta de definição do conceito de Jornalismo Guiado por Dados.

**Palavras-chave:** jornalismo; jornalismo guiado por dados; jornalismo de precisão; reportagem assistida por computador; jornalismo computacional.

O jornalismo guiado por dados (JGD) compreende diversas práticas profissionais, cujo ponto em comum é o uso de bases de dados como principal fonte de informação para a produção de notícias. As práticas de JGD envolvem técnicas de reportagem assistida por computador (RAC), visualização de dados, infografia, criação e manutenção de bases de dados e a política de acesso à informação e transparência pública de governos. Contemporaneamente, outras expressões usadas para se referir a essa especialidade profissional são “jornalismo de dados” (GRAY, CHAMBERS e BOUNEGRU, 2012; PARASIE e DAGIRAL, 2013) ou “jornalismo computacional” (ANDERSON, 2012; DIAKOPOULOS, 2012; COHEN, HAMILTON e TURNER, 2011; HAMILTON e TURNER, 2009; LEWIS e USHER, 2013; LIMA JR., 2011). A partir de manuais e

---

<sup>1</sup> Jornalista e professor do curso de Jornalismo da Universidade Federal do Rio Grande do Sul (UFRGS). Coordenador do Grupo de Pesquisa em Jornalismo Digital (JorDi/CNPq). Coordenador da Rede JorTec. Contato: marcelo.trasel@ufrgs.br.

publicações sobre o tema, este artigo analisa alguns dos principais elementos das práticas e rotinas produtivas ligadas ao JGD. Finalmente, é realizada uma tentativa de definir um conceito de Jornalismo Guiado por Dados.

## 1. As práticas definidoras do Jornalismo Guiado por Dados

O JGD tem por objetivo a produção, tratamento e cruzamento de grandes quantidades de dados, de modo a permitir maior eficiência na recuperação de informações, na apuração de reportagens a partir de conjuntos de dados, na distribuição em diferentes plataformas (computadores pessoais, *smartphones*, *tablets*), na geração de visualizações e infografias. Principalmente, as técnicas de JGD permitem ao jornalista encontrar informação com valor noticioso em bases de dados com milhares ou milhões de registros, dificilmente manejáveis sem a ajuda de computadores. Facilitam, ainda, o trabalho de comparar diferentes bases de dados, para a produção de novo conhecimento sobre a sociedade, criando *mash-ups*<sup>2</sup> ou relatando os resultados em texto, audiovisual e imagem.

Conforme Bounegru (2012), a primeira referência a *Data Journalism* (Jornalismo de Dados), termo que derivou em *Data-Driven Journalism* (Jornalismo Guiado por Dados), foi realizada pelo programador Adrian Holovaty (2006), num artigo publicado em seu website pessoal e intitulado “Um aspecto fundamental em que os websites de jornais precisam mudar”<sup>3</sup>. Na proposta, Holovaty recomenda a incorporação de técnicas de gerenciamento de bases de dados ao cotidiano das redações, como forma de facilitar o reaproveitamento das informações coletadas no trabalho diário de reportagem.

Muito do que os jornalistas locais coletam no cotidiano é informação estruturada: o tipo de informação que pode ser picada, de forma automatizada, por computadores. No entanto, a informação é destilada numa grande bolha de

---

<sup>2</sup> Um *mash-up* é compreendido, no contexto da informática, como uma aplicação para a Web produzida pelo amálgama de duas ou mais fontes diferentes, gerando um terceiro produto. Em geral, as fontes colocam os dados voluntariamente à disposição, através de APIs (Application Programming Interface), isto é, uma abertura que permite a programadores usarem algumas funcionalidades de um determinado software. Uma lista com exemplos de mash-ups pode ser encontrada nesta URL:

<http://www.programmableweb.com/apis/directory/1?sort=mashups>.

<sup>3</sup> Tradução livre.

texto - um artigo de jornal – que não tem chance de ser reaproveitado. (HOLOVATY, 2006, s.p.)<sup>4</sup>

A alternativa a este problema, na perspectiva acima, é criar rotinas produtivas cujo objetivo seja o desenvolvimento de bases de dados estruturados sobre todos os assuntos de interesse de uma determinada redação. Desta forma, os jornalistas produziram dados estruturados independentes de formatos narrativos e suportes materiais. Esses dados livres, armazenados em bases, poderiam ser posteriormente agregados, cruzados, analisados e transformados em visualizações gráficas.

Um exemplo seria o caso de um hipotético repórter atuante nas editorias ligadas crimes ou assuntos gerais, que começasse a manter um banco de dados a respeito de incêndios ocorridos em sua cidade. A cada incêndio, poderiam ser registrados numa planilha dados como dia, horário, local, número de vítimas, brigada de incêndio responsável pelo atendimento da ocorrência, razão do sinistro conforme o resultado da perícia, idade do prédio, entre outros. A partir de uma determinada quantidade de registros, algumas tendências poderiam começar a ser percebidas. Por exemplo, ao filtrar e classificar as causas de incêndios e cruzá-las com a idade média dos prédios atingidos, o repórter poderia perceber que após um certo limite de idade podem começar a ocorrer falhas elétricas. De posse desta informação, o repórter poderia então realizar uma reportagem a respeito da prevenção deste tipo de falha, acompanhada de visualizações gráficas sobre as causas mais frequentes de sinistros na série histórica disponível em seus registros.

O cenário acima mostra como acontecimentos jornalísticos comuns e, até mesmo, triviais do ponto de vista do esforço de reportagem podem ser abordados a partir da perspectiva do JGD. Para Holovaty (2006, s.p.), qualquer acontecimento de interesse jornalístico é passível de expressão em formatos de dados estruturados. Entretanto, em geral as redações não têm políticas neste sentido e iniciativas do gênero dependem da inclinação pessoal e interesse de repórteres individualmente. Diakopoulos (2012, p.10) observa que as empresas jornalísticas realizam, inclusive como subproduto da elaboração de notícias, diversos processos informacionais de valor agregado, como validação, classificação ou verificação de dados. Entretanto, raramente estes dados são armazena-

---

<sup>4</sup> Tradução livre.

dos de forma que possam ser reutilizados noutros contextos, pela audiência, ou até mesmo comercializados em apresentações diferentes de noticiários sob demanda de pessoas físicas ou jurídicas. A base de dados sobre incêndios imaginada acima poderia, por exemplo, ser de interesse para seguradoras, imobiliárias ou para o próprio Corpo de Bombeiros.

Por outro lado, o arquivamento para posterior recuperação e reuso de notícias sempre foi uma prática nas redações, ao menos nas grandes empresas (PALACIOS, 2003, p.4). A pesquisa em documentos oficiais e manutenção de registros, escritos ou digitais, por repórteres também é uma técnica comum de apuração. A diferença, no caso do JGD, está na magnitude das bases de dados. Conforme Lev Manovich, em entrevista a Machado (2002, p.177):

Para mim, a diferença essencial entre um banco de dados de computador e as formas similares mais remotas para a organização de dados, tais como o álbum de pinturas, catálogos, arquivos, bibliotecas e enciclopédias, é que as formas remotas ainda preservam a escala humana. Elas contêm um número limitado de registros que um usuário pode acessar diretamente. Podemos virar a página de um álbum, andar dentro de um arquivo, navegar pelo interior de uma biblioteca. Em outras palavras, o corpo humano é ainda a contrapartida única da interface. Uma vez diante de milhares de registros, nós não podemos vê-los de uma vez com nossos olhos nem podemos encontrar facilmente um registro particular tão somente usando as mãos. Temos de usar técnicas de computação de busca, de combinação e de seleção. Por exemplo, nós colocamos alguns termos num campo de busca e esperamos que o computador encontre os registros adequados. Um banco de dados é tão amplo que ele não pode ser disponibilizado de uma vez, ele existe além da escala da percepção e cognição humanas.

As técnicas computacionais aplicadas pelos praticantes do JGD permitem dar conta de uma quantidade de registro que muitas vezes não poderia ser manejado por trabalhadores humanos, como foi exemplificado com a descrição do caso da investigação de despesas parlamentares pelo jornal *The Guardian* na seção 2.1.

Um segundo texto fundamental sobre JGD é “A pirâmide invertida do jornalismo de dados”<sup>5</sup>, publicado pelo jornalista Paul Bradshaw em seu weblog em 2011. O autor propõe a descrição da rotina produtiva do JGD inicialmente num processo de quatro etapas: a) compilação (*compile*); b) limpeza (*clean*); c) contextualização (*contexto*);

---

<sup>5</sup> Tradução livre.

e d) combinação (*combine*). Além dessas etapas relacionadas a rotinas produtivas específicas do JGD, há a etapa final de comunicação (*communicate*), desdobrada por sua vez em seis passos ou características (BRADSHAW, 2011a, s.p.). A descrição destas rotinas produtivas permite compreender o tipo de prática com o qual os repórteres especialistas em JGD se envolvem cotidianamente.

A **compilação** se refere ao trabalho de coleta de dados. As bases de dados podem ser construídas pelo jornalista, através de pesquisa de campo por meio de técnicas qualitativas e quantitativas, como questionários, análise de documentos e etnografia, por exemplo, ou podem ser obtidas junto às fontes de informação – neste caso, instituições e indivíduos podem enviar planilhas contendo os dados desejados pelo repórter, ou as planilhas podem ser geradas através de *Web scraping*<sup>6</sup>. “O jornalismo de dados pode começar de duas formas: ou você tem uma pergunta que exige dados, ou você tem dados que exige investigação. De qualquer forma, a compilação de dados é o que define um ato de jornalismo de dados” (BRADSHAW, 2011a, s.p.)<sup>7</sup>. A compilação de dados numa base que possa ser submetida a processamento é a etapa fundamental do JGD, embora não seja suficiente, pois nem sempre a análise de dados revelará informações dotadas de valor-notícia. Por outro lado, em alguns casos a base de dados pode ser dotada de valor-notícia em si mesma, em especial no que se refere a serviços. Por exemplo, listagens da programação local de cinema são bases de dados tradicionalmente oferecidas pelos jornais como um serviço a seus leitores e não necessitam ser transformadas ou acompanhadas por matérias ou infográficos.

A etapa de **limpeza** visa assegurar a precisão da base de dados. Trata-se de um processo análogo ao da checagem durante a edição de uma reportagem tradicional, em termos de sua função na rotina produtiva. “Dispor dos dados é apenas o começo. Ter confiança nas notícias escondidas dentro deles significa ser capaz de confiar na qualida-

---

<sup>6</sup> *Web scraping* ou “raspagem de dados” é a prática de copiar ou transformar páginas da Web em bases de dados, usando aplicativos específicos. Com isso, é possível automatizar o processo de coleta de dados em websites que não os colocam à disposição em formatos diretamente manejáveis, como CSV, XLS ou TXT. Os dados a serem raspados podem estar integrados a um documento HTML, ou em formatos ilegíveis por aplicativos de tratamento de dados, como DOC ou PDF, ou ainda distribuídos em diversas páginas. A alternativa à “raspagem de dados” é copiar os dados manualmente, página por página, linha por linha, usando o recurso “copia e cola” para inseri-los diretamente numa planilha Excel ou similar.

<sup>7</sup> Tradução livre.

de dos dados – e isso significa limpá-los” (BRADSHAW, 2011a, s.p.)<sup>8</sup>. Há duas metas principais a serem atingidas com a limpeza da base de dados: a eliminação de erros humanos e a transformação para um formato coerente com o de outras bases de dados com as quais aquela eventualmente será combinada. Para realizar cruzamentos entre dados obtidos de duas fontes diferentes, é preciso que as duas bases possam ser mescladas num mesmo aplicativo. Assim, tanto o formato de arquivo (XLS, CSV, XML, por exemplo), quanto as formas de expressar os dados (casas decimais, notação para datas e moedas, uso de códigos ou nomes de países, por exemplo), devem seguir o mesmo padrão. Mesmo com o uso de programas específicos para a tarefa, trata-se de um trabalho maçante e demorado, mas indispensável, pois uma base de dados com erros pode levar a equívocos matemáticos e factuais graves, além de criar dificuldades para a geração de visualizações.

A fase de **contextualização** envolve uma avaliação da credibilidade da base de dados. É preciso analisar criticamente sua proveniência, a validade dos métodos de coleta, os possíveis conflitos de interesses e a competência dos pesquisadores responsáveis, entre outros aspectos. Dados não falam por si e são coletados por indivíduos ou instituições com seus próprios vieses, objetivos e trajetórias. Além disso, raramente uma série de dados produz informação relevante antes de ser contrastada com outras informações:

Por exemplo, saber o número de crimes numa cidade é interessante, mas só se torna significativo quando você o contextualiza comparando com a população, ou o número de policiais, ou os níveis de criminalidade de 5 anos atrás, ou a percepção de criminalidade, ou níveis de desemprego, e assim por diante. (BRADSHAW, 2011a, s.p.)<sup>9</sup>

Uma condição para o bom desempenho na contextualização de uma base de dados é habilidade na interpretação de estatísticas. A desenvoltura em matemática básica pode ser considerada a característica principal de um jornalista guiado por dados, pois é muito difícil identificar corretamente pautas – isto é, informação dotada de valor-notícia – em bases de dados sem a capacidade de fazer relações entre grandezas e apreciar seu significado. É preciso compreender o tipo de dados que se tem à mão, para aferir com

---

<sup>8</sup> Tradução livre.

<sup>9</sup> Tradução livre.

quais outras bases de dados eles podem ser comparados. Pode-se dizer que o “faro noticioso” dos jornalistas guiados por dados reside nessa etapa da rotina produtiva do JGD.

A última etapa é a **combinação** de bases de dados com o objetivo de produzir novo conhecimento. Em geral, combinam-se duas ou mais séries a partir de uma dimensão compartilhada, como o nome de um time de futebol, uma cidade, ou uma empresa. Uma das formas mais usadas de combinação de dados no jornalismo é localizar os dados de uma série num mapa. O jornal Zero Hora, por exemplo, recentemente usou dados do censo populacional brasileiro de 2010 para criar um mapa do Rio Grande do Sul em que era possível comparar a qualidade da iluminação pública, saneamento e outras características urbanas entre municípios gaúchos<sup>10</sup>. Bradshaw (2011a, s.p.) faz uma analogia entre o número de bases de dados usadas e o número de fontes numa reportagem tradicional: “Boas pautas podem ser encontradas numa única base de dados, mas frequentemente você terá de combinar duas delas. Afinal, frente à escolha entre uma notícia baseada em uma única fonte e uma notícia baseada em múltiplas fontes, qual você escolheria?”<sup>11</sup>. A analogia, porém, não é apropriada, uma vez que a combinação de séries de dados é essencial, justamente, para se chegar a uma afirmação sobre fatos sociais, a um ângulo jornalístico. Na reportagem tradicional, ao contrário, busca-se múltiplas fontes justamente para contrabalançar diferentes ângulos.

A combinação de bases de dados marca o fim da fase de apuração na rotina produtiva do JGD. Resta a etapa de publicação ou **comunicação** dos resultados, para a qual Bradshaw (2011b, s.p.) prevê seis possibilidades: a) visualização (*visualize*) em gráficos, mapas e outras formas de infografia; b) narração (*narrate*) em texto ou audiovisual, ou seja, redação de notícias tradicionais; c) comunicação social (*social communication*), isto é, o compartilhamento dos resultados com a audiência<sup>12</sup>; d) humanização (*humani-*

---

<sup>10</sup> Disponível em: <http://zerohora.clicrbs.com.br/rs/geral/conteudo/censo-2010.html>. Acesso: 25 fev. 2013.

<sup>11</sup> Tradução livre.

<sup>12</sup> O autor não explica claramente seu uso do conceito de comunicação social, mas fornece como um dos exemplos o weblog de dados do jornal britânico *The Guardian*, que coloca à disposição de seus leitores planilhas com os dados coletados ou obtidos por seus repórteres. A audiência é convidada a produzir formas alternativas de visualizar ou apresentar esses dados e pode enviar suas tentativas à equipe do weblog, que costuma publicar estes exercícios. O *Guardian Data Blog* pode ser acessado na seguinte URL: <http://www.guardian.co.uk/news/datablog>. Acesso: 25 fev. 2013.

se), através de entrevistas com indivíduos que ilustrem as informações obtidas no processamento dos dados; e) personalização (*personalise*), por meio da abertura dos produtos à interação com o leitor; e f) utilização (*utilize*), que envolve a criação de ferramentas ou serviços de informação a partir dos dados coletados.

Silver (2014, s.p.), noutro texto influente oriundo do campo profissional, considera falsa a contradição entre a abordagem qualitativa da investigação jornalística, materializada em narrativas densas encadeando uma série de anedotas, e a abordagem quantitativa da reportagem, embasada em análises de dados. Em sua opinião, é mais útil distinguir entre a investigação jornalística rigorosa, seja ela quantitativa ou qualitativa, e a investigação jornalística pouco rigorosa, anedótica:

O problema não é a falha em citar evidências quantitativas. É fazê-lo de uma forma que pode ser anedótico e ad-hoc, em vez de rigorosa e empírica, e falhar em fazer as perguntas certas aos dados. Em contrapartida, uma abundância de métodos jornalísticos são rigorosos sem ser quantitativos. A reportagem investigativa, que sintetiza várias linhas de provas para encontrar a verdade por trás de uma história, se enquadra nessa categoria. (SILVER, 2014a, s.p.)<sup>13</sup>

Entretanto, mesmo as anedotas podem ser úteis ao jornalismo rigoroso e metódico, na medida em que elas capturam sinais dos ambientes natural e cultural e estes sinais, agrupados, podem ser expressos na forma de dados e servir como base para generalizações. Esta opinião segue a do cientista político americano Ray Wolfinger que, ao ter uma afirmação factual rejeitada por um aluno durante um seminário com o argumento de que era “meramente anedótica”, retorquiu: “o plural de anedota é dados”<sup>14</sup> (SMITH, 2014). O sentido da frase é que um dado, isolado, não tem nenhum poder explicativo, ou seja, é uma anedota. Uma base de dados, porém, pode ser usada para encontrar explicações para fenômenos e realizar previsões estatísticas. A maior parte do trabalho nas redações se dedica à produção diária de anedotas que, em si mesmas, têm pouco valor explicativo. As notícias podem ser encaradas, no entanto, como instrumentos de coleta de dados qualitativos, passíveis de ser transformados em bases de dados

---

<sup>13</sup> Tradução livre.

<sup>14</sup> “The plural of anecdote is data.” Tradução livre. Note-se que no inglês o substantivo coletivo *data* é o plural do latim *datum*. Em português, existem tanto a forma singular quanto a plural.



quantitativos – de fato, o *lead*, na forma narrativa pirâmide invertida, é uma estratégia de expressão de acontecimentos como dados estruturados. Um conjunto de notícias, então, poderia ser transformado em dados e deles se extrair informação. A partir dessa perspectiva, Silver (2014, s.p.) propõe a sua versão de uma rotina produtiva para o JGD, em quatro passos comparáveis aos processos tradicionais de reportagem (QUADRO 1).

### QUADRO 1

Rotina produtiva do JGD na perspectiva de Nate Silver

<b>Processo/Técnica</b>	<b>Jornalismo tradicional</b>	<b>Jornalismo guiado por dados</b>
<u>Coleção</u> : coleta de dados ou evidências.	Entrevistas, observação, produção de documentários.	Aplicação de questionários, raspagem de dados, experimentos sociais.
<u>Organização</u> : tratamento e encadeamento dos dados coletados anteriormente.	Narração (em pirâmide invertida e outras formas).	Descrição estatística, armazenamento em bases de dados relacionais, criação de visualizações e gráficos.
<u>Explicação</u> : interpretação dos resultados da organização.	Artigos de análise, jornalismo interpretativo.	Análise estatística através de testes.
<u>Generalização</u> : realização de previsões a partir da interpretação dos dados.	Jornalismo opinativo.	Verificação de hipóteses através de experimentos estatísticos.

Fonte: SILVER, 2014. Tradução livre.

Embora Silver (2014, s.p.) trace um paralelo entre as técnicas tradicionais e as técnicas do JGD em sua proposta de rotina produtiva, ele não as considera equivalentes. Em sua opinião, o jornalismo tradicional é eficaz nas etapas de coleta e organização dos dados, mas frequentemente falha em sua explicação e quase sempre realiza previsões

equivocadas. A razão para as dificuldades em explicar os acontecimentos e prever seu desenrolar seria a própria incapacidade do cérebro humano em lidar com uma quantidade grande o suficiente de informação para identificar relações causais válidas e usá-las em previsões. Em geral, as colunas e editoriais interpretam e preveem acontecimentos – por exemplo, o resultado de uma eleição – com base mais nos preconceitos e ideologia dos jornalistas do que em dados testados com rigor científico.

O ponto é que há uma negociação entre a vivacidade e escalabilidade. Narrativas noticiosas sobre eventos individuais podem ser informativas e agradáveis de ler, e elas podem ter um grande valor intrínseco, revelem ou não alguma verdade maior. Mas pode ser extraordinariamente difícil fazer generalizações sobre os acontecimentos, a menos que você pare para classificar os seus detalhes mais importantes de acordo com alguma numeração ou sistema classificatório, transformando a anedota em dados. (SILVER, 2014, s.p.)<sup>15</sup>

Deutsch e Hellerstein (2014) dividem o processo de trabalho com dados em três etapas: preparar (*prepare*), analisar (*analyze*) e aplicar (*apply*). A etapa de preparação dos dados envolve a colocação de questões, a coleta de dados, sua organização em coleções e a limpeza das bases de dados. A fase de análise se desdobra em encontrar padrões, encontrar relações, filtrar os dados, resumir os dados, calcular fórmulas e criar gráficos. Na etapa de aplicação, por fim, deve-se tomar decisões, compartilhar resultados e visualizar os dados. O interessante nesta proposta de divisão do trabalho com dados é a importância conferida à intenção do analista, evidenciada pela etapa “aplicar”, que, para além do significado da palavra tensiona o verbo no tempo presente, sugerindo um foco em ações concretas. Deste ponto de vista, a análise de dados não é um processo neutro, mas se inicia com um desejo ou demanda específica de um indivíduo ou grupo voltado a um objetivo específico pré-definido.

## 2. O conceito de Jornalismo Guiado por Dados

Em que pese a ausência de aprofundamento teórico nas descrições de Holovaty (2006), Bradshaw (2011a) e Silver (2014) para a descrição da rotina produtiva do JGD, ela identifica as novas práticas introduzidas nas redações, além de evidenciar uma tentativa de consolidação dessas práticas no âmbito da cultura profissional jornalística. Além

---

<sup>15</sup> Tradução livre.

disso, essa descrição do processo produtivo do JGD evidencia o fato de que os repórteres passam a, cada vez mais, assumir também o papel de técnicos, ao deixarem de delegar a operação de hardware e software para especialistas e passarem eles mesmos a assumir a operação dessas tecnologias.

Por volta de 1991, Meyer afirma na segunda edição de seu livro que jornalistas interessados em RAC eram uma raça particular de repórteres, que frequentemente compravam seus próprios computadores antes mesmo de as redações os adotarem. [...] Entretanto, o reporter assistido por computador era ainda primordialmente um jornalista, não um técnico; a meta subjacente era produzir uma notícia melhor. Hoje, porém, o programador-jornalista se diferencia deste reporter de RAC ao enxergar o produto final não como uma notícia, mas como um “artefato produtivo” de “filtragem de informação”.<sup>16</sup> (LEWIS e USHER, 2013, p.605)

A partir desta percepção, os autores consideram o JGD uma fusão das ciências da computação e jornalismo, a partir da qual os programadores começam a assumir a centralidade nas redações e participar ativamente da elaboração de formatos jornalísticos como visualizações de dados, aplicativos, algoritmos de geração e recomendação de notícias, entre outros tipos de peças noticiosas derivadas da informática. Parasie e Dagiral (2013) seguem numa vertente semelhante, argumentando que os projetos de JGD dentro das redações repousam sobre o tratamento, análise e visualização de dados relacionados a uma grande variedade de domínios, como criminalidade, demografia, transportes, infraestrutura, orçamento público, meio ambiente, entre outros, e aos agentes sociais responsáveis pelo gerenciamento de ações governamentais nestas áreas, como políticos, a respeito dos quais se busca levantar informações sobre financiamento de campanhas, uso de verbas, votos, ou projetos de lei apresentados, por exemplo) em todas as esferas administrativas. Estes dados são reunidos então em “programas informáticos apresentados como constituintes intrínsecos de produtos jornalísticos”<sup>17</sup> (PARASIE e DAGIRAL, 2013, p.53).

Lima Jr. (2011), Anderson (2012), Diakopoulos (2012), Lewis e Usher (2013) e Stavelin (2014) preferem usar a expressão “jornalismo computacional”, proposta por Cohen, Hamilton e Turner (2011), em lugar de JGD. O jornalismo computacional seria,

---

<sup>16</sup> Tradução livre.

<sup>17</sup> Tradução livre.

inicialmente, “a mudança na forma como as notícias são descobertas, apresentadas, agregadas, monetizadas e arquivadas”<sup>18</sup> (COHEN, HAMILTON e TURNER, 2011, p.1). A computação, nesse caso, é vista como um fator de evolução do jornalismo, permitindo inovações como detecção automática de tópicos na rede mundial de computadores, análise de vídeos, personalização, agregação, visualização e produção de sentido. Num estudo anterior, porém, Hamilton e Turner (2009, p.2) definem o jornalismo computacional como o uso de “algoritmos, dados e ciências sociais para reforçar as funções de fiscalização do jornalismo”<sup>19</sup>. A diferença entre o conceito de jornalismo computacional e JGD, entretanto, não fica clara nos estudos destes pesquisadores. As práticas atribuídas a ambos os conceitos se sobrepõem frequentemente. Considerando-se, porém, o alerta de Meyer (1999) sobre o foco excessivo na tecnologia e sua lembrança de que o aspecto mais importante da RAC é o conhecimento gerado, não as ferramentas usadas na geração deste conhecimento, pode-se argumentar que a expressão “jornalismo guiado por dados” apresenta a vantagem semântica de direcionar a atenção para os resultados jornalísticos do processo, em detrimento da preocupação com a tecnologia em si mesma.

A partir dos elementos conceituais acima, pode-se concluir que a principal diferença entre a RAC (COX, 2002; DEFLEUR, 1997) e o JGD é a introdução da criação de algoritmos nas rotinas produtivas das redações. Embora houvesse casos, nos anos 1990, de desenvolvimento de aplicativos para a Web dotados de funcionalidades conferidas por programação, é apenas na década de 2000 que esse tipo de produto jornalístico começa a se tornar amplamente disseminado. Assim, o JGD pode ser encarado como uma evolução conceitual dos termos Jornalismo de Precisão (MEYER, 1973) e Reportagem Assistida por Computador, acompanhando Bounegru (2012, p.29):

A essa altura é senso comum que até mesmo as mais recentes práticas midiáticas têm histórias, assim como algo de novo em si mesmas. Em vez de discutir se o jornalismo de dados é completamente novo, uma posição mais frutífera seria considerá-lo como parte de uma tradição mais longa, mas respondendo a novas circunstâncias e condições. Mesmo que não haja uma diferença nos objetivos e técnicas, o surgimento do rótulo de "jornalismo de dados" no

---

<sup>18</sup> Tradução livre.

<sup>19</sup> Tradução livre.

início do século indica uma nova fase, em que o grande volume de dados disponível gratuitamente online – combinado com sofisticadas ferramentas centradas no usuário, auto-publicação e ferramentas de crowdsourcing – permite que mais pessoas trabalhem com mais dados com mais facilidade do que nunca. (BOUNEGRU, 2012, p. 29)<sup>20</sup>

A partir das noções expostas acima, o conceito de Jornalismo Guiado por Dados com o qual a comunidade profissional jornalística vem operando pode ser definido da seguinte forma: o jornalismo guiado por dados é a aplicação da computação e dos saberes das ciências sociais na coleta, processamento, interpretação e apresentação de dados, com o objetivo de ampliar a função da imprensa como defensora do interesse público.

Não foi encontrada na revisão bibliográfica, ou mesmo durante a leitura de intercâmbios em fóruns eletrônicos e manifestações em websites voltados ao tema, nenhuma discussão aprofundada sobre uma possível diferença conceitual entre *data journalism* e *data-driven journalism*. A maioria dos jornalistas e até mesmo alguns pesquisadores (DAGIRAL e PARASIE, 2013 e 2012) usam os termos como sinônimos. Considerando-se a identidade no significado, parece franqueado apelar para motivações estéticas ao preferir jornalismo guiado por dados em lugar de jornalismo de dados, pois esta última expressão, em língua portuguesa, gera um cacófono.

Para Matzat (2011), o termo JGD seria o mais adequado, de fato, para se referir a essas práticas, porque ressalta o caráter dinâmico da apuração e apresentação de informações através da navegação e interação com bancos de dados online. O jornalista guiado por dados assume uma postura exploratória, característica do leitor imersivo experto descrito por Santaella (2004, p.60), que busca a informação na Internet a partir do raciocínio dedutivo, deixando-se guiar tanto pela experiência quanto por pistas disponíveis na rede para encontrar seu objetivo.

De fato, o jornalismo de dados vai um passo decisivo além: ele se baseia em registros numéricos não só como fontes de pesquisa, mas também torna os dados o tema central da história e sua apresentação. Ele reside na intersecção de três áreas: primeiro, jornalismo visual, ou infografia, segundo, multimídia e narrativas interativas, terceiro, jornalismo investigativo. (MATZAT, 2011, s.p.)<sup>21</sup>

---

<sup>20</sup> Tradução livre.

<sup>21</sup> Tradução livre.

Reelaborando este conceito oriundo do contexto profissional a partir da argumentação dos capítulos 2 e 3, entretanto, propõe-se aqui a seguinte formulação: o termo jornalismo guiado por dados se refere à aplicação da tecnociência às rotinas produtivas do jornalismo informativo, num contexto de comunicação em rede, com o fim de estabelecer reclamações mais sólidas de autoridade profissional. Essas reclamações de autoridade profissional se baseiam na observação dos procedimentos de construção da objetividade, não como mero ritual (TUCHMAN, 1993), mas com vistas, em seus melhores momentos, à objetividade disciplinar (PORTER, 1995), embora em geral apenas uma objetividade mecânica seja suficiente.

## Referências

ANDERSON, C.W.. Notes Towards an Analysis of Computational Journalism. **HIIG Discussion Paper Series**, n. 1, 2012. Disponível em: <http://ssrn.com/abstract=2009292>. Acesso: 19 abr. 2013.

BOUNEGRU, Liliana. Data Journalism in Perspective. In: GRAY et al. (orgs.). **The data journalism handbook: how journalists can use data to improve the news**. Sebastopol: O'Reilly, 2012. [Ebook]

BRADSHAW, Paul. 6 ways of communicating data journalism (the inverted pyramid of data journalism part 2). **Online Journalism Blog**, 13 jul. 2011b. Disponível em: <http://onlinejournalismblog.com/2011/07/13/the-inverted-pyramid-of-data-journalism-part-2-6-ways-of-communicating-data-journalism>. Acesso: 25 fev. 2013.

\_\_\_\_\_. The inverted pyramid of data journalism. **Online Journalism Blog**, 7 jul. 2011a. Disponível em: <http://onlinejournalismblog.com/2011/07/07/the-inverted-pyramid-of-data-journalism>. Acesso: 25 fev. 2013.

COHEN, Sarah; HAMILTON, James T.; TURNER, Fred. Computational journalism. **Communications of the ACM**, v. 54, n. 10, 2011, p. 66-71.

COX, Melisma. The development of Computer-Assisted Reporting. In: Southeast Colloquium for the Association for Education in Journalism and Mass Communication, Chapel Hill. **Proceedings...** Columbia: AEJMC, 2000.

DEFLEUR, Margaret H. **Computer-Assisted Investigative Reporting: development and methodology**. New Jersey: Mawah, 1997.

DEUTSCH, Amit; HELLERSTEIN, Joseph. Making sense of data. **Google**, mar. 2014. Curso online. Disponível em: <https://datasense.withgoogle.com/course>. Acesso: 27 mar. 2014.

DIAKOPOULOS, Nicholas. Cultivating the landscape of innovation in computational journalism. **City University of New York**, abr. 2012. Disponível em: [http://cdn.journalism.cuny.edu/blogs.dir/418/files/2012/04/diakopoulos\\_whitepaper\\_systematicinnovation.pdf](http://cdn.journalism.cuny.edu/blogs.dir/418/files/2012/04/diakopoulos_whitepaper_systematicinnovation.pdf). Acesso em: 10 mai. 2013.

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

- HAMILTON, James; TURNER, Fred. Accountability Through Algorithm: Developing the Field of Computational Journalism. **Stanford**, 2009. Disponível em: <http://www.stanford.edu/~fturner/Hamilton%20Turner%20Acc%20by%20Alg%20Final.pdf>. Acesso em: 9 dez. 2013.
- HOLOVATY, Adrian. A fundamental way newspaper sites need to change. **Adrian Holovaty**, 6 set. 2006. Disponível em: <http://www.holovaty.com/writing/fundamental-change>. Acesso: 17/02/2013.
- LEWIS, Seth; USHER, Nikki. Open source and journalism: toward new frameworks for imagining news innovation. **Media, culture & society**, v. 35, n. 5, p. 602-619, 2013.
- LIMA JR., Walter T. Jornalismo computacional em função da “Era do Big Data”. **Revista Líbero**, v. 14, n. 28, p. 45-52, 2011.
- MACHADO, Irene. Banco de dados como gênero na linguagem das novas mídias: as formulações de Lev Manovich. **Galáxia**, v. 2, n. 3, 2002.
- MATZAT, Lorenz. Datenjournalismus. In: **Dossier Open Data**. Bonn: Bundeszentrale für politische Bildung, 2011.
- MEYER, Philip. **Precision Journalism: a reporter’s introduction to social science methods**. Bloomington: Indiana University Press, 1973.
- \_\_\_\_\_. The Future of CAR: Declare Victory and Get Out! In: PAUL, Nora (org.). **When nerds and words collide: reflections on the development of Computer-Assisted Reporting**. St. Petersburg: Poynter Institute, 1999.
- PALACIOS, Marcos. Ruptura, continuidade e potencialização no jornalismo online: o lugar da memória. In: PALACIOS, Marcos; GONÇALVES, Elias (Org.). **Modelos de jornalismo digital**. Salvador: Calandra, 2003.
- PARASIE, Sylvain; DAGIRAL, Éric. Data-driven journalism and the public good: “Computer-assisted-reporters” and “programmer-journalists” in Chicago. **New Media & Society**, nov. 2012.
- \_\_\_\_\_. Des journalistes enfin libérés de leurs sources? Promesse et réalité du “journalisme de données”. **Sur le journalisme**, v. 2, n. 1, 2013.
- PATTERSON, Thomas. **Informing the news: the need for knowledge-based journalism**. Nova York: Vintage Books, 2013.
- PORTER, Theodore. **Trust in numbers: the pursuit of objectivity in science and public life**. Princeton: Princeton University Press, 1995.
- SANTAELLA, Lucia. **Navegar no ciberespaço: o perfil cognitivo do leitor imersivo**. São Paulo: Paulus, 2004.
- SILVER, Nate. What the fox knows. **FiveThirtyEight**, 17 mar. 2014a. Disponível em: <http://fivethirtyeight.com/features/what-the-fox-knows/>. Acesso: 17 mar. 2014.
- SMITH, David. The plural of anecdote is data, after all. **Revolutions**, 6 abr. 2011. Disponível em: <http://blog.revolutionanalytics.com/2011/04/the-plural-of-anecdote-is-data-after-all.html>. Acesso: 17 mar. 2014.
- STAVELIN, Eirik. **Computational journalism: when journalism meets programming**. Tese (Doutorado em Ciências da Informação e Estudos de Mídia), Faculdade de Ciências Sociais, Universidade de Bergen, Bergen, 2014.

TUCHMAN, Gaye. A objetividade como ritual estratégico: uma análise das noções de objetividade dos jornalistas. In: TRAQUINA, Nelson (org.). **Jornalismo**: questões, teoria e estórias. Lisboa: Vega, 1993.