

A web e o jornalismo de dados: mapeamento de conceitos chave

Lucas Vieira de Araújo¹

Resumo: Este artigo busca conceituar, definir o escopo e o propósito do jornalismo de dados a partir de uma visão histórica da web e dos dados como elementos básicos da informação digital. Acredita-se na relevância desse debate pela crescente difusão de tecnologias que possibilitam a coleta de dados na web e as perspectivas nesse novo cenário para o jornalismo, seja em plataformas audiovisuais até em ambiente web a partir de dispositivos móveis ubíquos. Para atingir tais objetivos, foi realizada uma pesquisa exploratória, de cunho eminentemente teórico, na qual foi utilizada bibliografia brasileira e estrangeira. Entre os resultados, percebe-se que a prática de extrair informação a partir de dados não é um recurso recente e que são necessários novos estudos para aprofundar as discussões em torno de assuntos como os limites éticos do jornalismo de dados.

Palavras-chave: Jornalismo, dados, transparência, web.

Abstract: This article seeks to conceptualize, define the scope and purpose of data journalism from a historical view of the web and data as basic elements of digital information. It is believed in the importance of this debate by the increasing diffusion of technologies that enable the collection of data on the web and prospects in this new scenario for journalism, whether in audiovisual platforms up in a web environment from ubiquitous mobile devices. To achieve these objectives, an exploratory research was conducted, eminently theoretical nature, which was used Brazilian and foreign literature. Among the results, it is clear that the practice of extracting information from data is not a new feature and that further studies are needed to deepen the discussions around issues such as the ethical boundaries of data journalism.

Keywords: Journalism, data, transparency, web.

Introdução

A ampla difusão e disseminação das mídias digitais, aliada a fatores como novas tecnologias, tem criado a impressão de que a comunicação realizada por intermédio de máquinas vive um momento único. No caso do jornalismo, percebe-se grande entusiasmo com a possibilidade de realizar reportagens utilizando-se de programas que realizam prospecção e seleção de dados e ainda ferramentas que prometem a integração entre dispositivos móveis e fixos, como aplicativos. Esse novo instrumental é chamado, mormente, de jornalismo de dados ou jornalismo digital, o qual seria uma nova técnica de produção de notícias a partir da grande quantidade de informação presente na rede de computadores, principalmente na internet (GRAY et al, 2012; BARLOW, 2015).

¹Universidade Metodista de São Paulo. E-mail: professorlucasaraujo@gmail.com.

É necessário, no entanto, pontuar que a euforia em torno das mudanças de cenário trata-se de uma evolução advinda do uso e dispersão intensiva da *web*, além de outros fatores como a infinidade de dados presentes no universo virtual. Além de informações pessoais, gerada muitas vezes por redes sociais, a internet congrega atualmente nomes e números provenientes de órgãos governamentais, os quais são extremamente valiosos para o jornalismo por serem de utilidade pública (GRAY, 2011).

Muito além de uma novidade, o jornalismo de dados parece ser o resultado de um conjunto de mudanças na sociedade, notadamente aquelas voltadas à comunicação e a disponibilização de informação. Assim, este artigo buscará compreender esse cenário a partir de fatos históricos recentes que mostrem as relações e as intenções que cercam a tecnologia e os meios de comunicação. Para isso, realizar-se-á uma discussão teórica, de cunho epistemológico, da comunicação em rede e das bases do jornalismo. Além das contribuições de autores brasileiros, este texto valer-se-á das proposições de autores estrangeiros, como Tim Berners Lee, que como criador da web lançou as bases que originaram empresas e iniciativas que mudaram a comunicação no século XXI.

História e web 2.0

Quando o físico e cientista da computação britânico Tim Berners-Lee enviou uma carta em 1989 ao CERN, organização europeia para pesquisa nuclear, ele propôs uma ferramenta que realizasse o gerenciamento de informações. Embora não imaginasse naquela circunstância que a iniciativa se tornaria a web de hoje, a preocupação maior naquele momento era evitar a perda de dados relevantes ao longo do tempo. Situação que acometia o próprio CERN e tantas outras instituições de pesquisa.

Berners-Lee partiu do pressuposto de que as pesquisas do CERN mudam conforme surgem demandas, como a chegada de novas tecnologias, e que geralmente os resultados são armazenados em forma de grandes livros, os quais não podem ser atualizados na velocidade das mudanças. Assim, argumentou o cientista, o CERN está perdendo muitas informações relevantes, o que não estaria ocorrendo apenas no centro de pesquisa europeu, mas em diversas partes do mundo, já que o local antecipa tendências pelo planeta. Ademais, a iniciativa sugerida poderia ter uma aplicação comercial (BERNERS-LEE, 1989).

Para o físico, o novo sistema deveria possibilitar que novas informações fossem atualizadas conforme as necessidades da organização, o que seria possível se não houvesse limites ao fluxo de dados nessa rede. Assim, ele usa pela primeira vez o termo *web* para referir-se à forma como seriam estruturadas as notas postas no sistema, o qual também deveriam conter *links* como referências. Dessarte, continua o físico, essa estruturação seria mais útil se fossem eliminadas hierarquias (Ibid).

Malgrado ele não tenha se referido diretamente a isso, é provável que tenha recomendado a não criação de castas para que tudo ocorresse conforme as necessidades dos usuários sem um controle direto sobre o fluxo de informações. Prova disso é que quando é referida a maneira como o próprio CERN arquivava os dados, ele faz uma analogia com árvores, pois os nomes são armazenados de forma assíncrona a partir de nós. A proposição do cientista da computação era alterar essa lógica: “era preciso um link a partir de, de e para outro nó, porque neste caso a informação não seria, naturalmente, organizada em uma árvore”². (BERNERS-LEE, 1989: 7).

Berners-Lee argumentou, então, que o sistema de indexação por palavras-chave, já em uso, não era a mais adequada porque nem sempre as pessoas utilizavam os mesmos termos como referência. Assim, ele propôs um ordenamento em forma de hipertexto. Ele lembra, então, que o termo foi criado por Ted Nelson há muitas décadas, mais precisamente em 1950, com duas propostas diferentes. Uma delas seria interligar informações que pudessem ser lidas por humanos de forma irrestrita, a qual estaria diretamente relacionada à ideia de Berners-Lee de criar uma nova maneira de ordenar e armazenar as informações. A outra proposição de Nelson, segundo o cientista britânico, seria disponibilizar documentos em outros formatos, o que seria possível com o passar do tempo e escala.

A outra idéia, que é independente e em grande parte uma questão de tecnologia e tempo, é de documentos multimídia que incluam gráficos, voz e vídeo. Não vou discutir este último aspecto ainda mais aqui, embora eu vou usar a palavra "Hiperímídia" para indicar que não está vinculado ao texto. Tem sido difícil de avaliar o efeito de um grande sistema hiperímídia em uma organização, porque muitas vezes esses sistemas nunca foram usados em larga escala. Por esta razão, grandes quantidades de informação devem ser acessíveis usando qualquer nova informação desse novo sistema de gestão³ (Ibid: 10).

² What was needed was a link from one node to another, because in this case the information was not naturally organised into a tree. (Tradução do autor).

Assim que terminou de explicar como funcionaria seu novo sistema de armazenamento e disponibilização de informação, Berners-Lee elencou no texto diversos pré-requisitos necessários para a concretização da proposta. Entre eles:

- Acesso remoto através das redes.
 - As instalações do CERN são distribuídas, assim é necessário o acesso de máquinas remotas.
 - Heterogeneidade
 - Diferentes sistemas, oriundos de empresas, diversas devem trocar informações entre si.
 - Não Centralização
 - Sistemas de informação começam pequenos e crescem. Eles também começam isolados e em seguida, se misturaram. Um novo sistema deve permitir que os sistemas existentes sejam ligados entre si sem necessidade de qualquer controle central ou coordenação.
 - Acesso aos dados existentes
 - As bases de dados existentes devem ser acessadas em forma de hipertexto.
1. Links privados
 2. O usuário deve ser capaz de adicionar os próprios links e outros de informação pública. É preciso também ser capaz de guardar as ligações realizadas com outros usuários, assim como os humanos o fazem, em particular.
 3. Análise de dados
 4. Uma possibilidade intrigante, dada uma grande base de dados com ligações hipertexto digitadas, é que ela permite um certo grau de análise automática. É possível procurar, por exemplo, por anomalias como software indocumentados ou divisões que não contêm pessoas. É possível gerar listas de pessoas ou dispositivos para outros fins, tais como listas de discussão de que as pessoas sejam informadas de mudanças. (Ibid)

³The other idea, which is independent and largely a question of technology and time, is of multimedia documents which include graphics, speech and video. I will not discuss this latter aspect further here, although I will use the word "Hypermedia" to indicate that one is not bound to text. It has been difficult to assess the effect of a large hypermedia system on an organisation, often because these systems never had seriously large-scale use. For this reason, we require large amounts of existing information should be accessible using any new information management system. (Tradução do autor).

Todas essas características, portanto, foram gestadas desde o nascimento da *web* como rede de compartilhamento e armazenamento de informações. Isso contraria muitas ideias em torno das quais a *web* desenvolveu-se ao longo de fases. Um dos precursores dessa proposição é o norte-americano Tim O'Really, que em 2005 escreveu um artigo explicando a origem da expressão e a razão pela qual ele acredita nela, apesar das críticas que vinha recebendo. Ele argumentou, inicialmente, que o estouro da bolha de empresas pontocom em 2001, quando diversas companhias criadas devido ao surgimento e difusão da internet, foi um divisor de águas para o setor. Seguindo uma lógica evolucionista-capitalista, O'Really disse que as companhias remanescentes foram as mais fortes e preparadas, pois as mudanças estavam apenas começando.

Embora criticou algumas empresas de marketing que estariam supostamente usando o termo web 2.0 de forma inapropriada, O'Really apresentou uma lista daquilo que ele considerava a verdadeira web 2.0. Em linhas gerais, apresentou uma determinada empresa, como a Enciclopédia Britânica on-line, e a sua sucessora, a Wikipedia, ou uma prática, como a de fazer sites pessoais, e a evolução, a criação de blogs. Apesar de não se limitar a isso, a lista é um resumo daquilo que ele pensava sobre a mudança de fase na web (O'REILLY, 2005).

Um exemplo aleatório é o um artigo do pesquisador Alex Primo no qual ele discute o aspecto relacional na internet. Citando O'Really, Primo afirma sobre a web 2.0:

Trata-se de um núcleo ao redor do qual gravitam princípios e práticas que aproximam diversos sites que os seguem. Um desses princípios fundamentais é trabalhar a Web como uma plataforma, isto é, viabilizando funções online que antes só poderiam ser conduzidas por programas instalados em um computador. Porém, mais do que o aperfeiçoamento da “usabilidade”, o autor enfatiza o desenvolvimento do que chama de “arquitetura de participação”: o sistema informático incorpora recursos de interconexão e compartilhamento (2007: 2)

Não obstante o caso do pesquisador brasileiro não ilustre de forma científica o quanto o termo web 2.0 tenha se difundido, ele ilustra uma realidade que pode ser corroborada em partes pela ferramenta Google Ngram Viewer. Criada em 2010, o programa indexa palavras

ou frases curtas, a partir de uma contagem anual, encontradas em fontes impressas do período de 1800 a 2012 em diversas línguas, como inglês, francês e até chinês (WIKIPEDIA, 2015). Uma pesquisa realizada neste instrumental com a palavra web 2.0 releva fatos interessantes e desvela mitos.

O primeiro, é que o termo web 2.0 foi usado pela primeira vez em 1967, logo, o argumento de O'Really de que o termo foi criado em uma reunião entre duas empresas no início do século XXI não procede. O segundo, é que de 1977 a 1982 a palavra web 2.0 foi mais usada do que no início do ano 2000. De acordo com os cálculos do visualizador, o termo aparece em 0,0000000180% dos livros indexados naquele intervalo, ao passo que 0,0000000022% em 2002. Isso pode explicar porque O'Really acreditou que cunhou a expressão, pois ela pode ter caído em desuso até que novamente se tornasse recorrente.

O terceiro fato que chama atenção a partir da análise dos números do Google Ngram é a forte ascensão do termo web 2.0 a partir de 2005. Neste ano o visualizador marcou a presença do termo em 0,0000010242%. Já em 2008, prazo máximo indexado pela ferramenta, foi de 0,0000017891%. Alta significativa, apesar do crescimento de 2003 a 2005 também ter sido digno de anotação. Esses números, portanto, endossam o entendimento deste artigo de que jargões mercadológicos reforçam o coro acadêmico. Ademais, salientam ainda que muitas palavras não foram criadas por quem as imaginou ter feito e tampouco dizem respeito àquilo que aparentam.

Percebe-se que as características elencadas por Primo, a partir das propostas de O'Really, são basicamente as mesmas sugeridas por Berners-Lee quando pensou como a web poderia ser, mormente a participação de diversos agentes que se conectariam e compartilhariam informações, o que o britânico chamou de heterogeneidade, links privados e não-centralização. Assim, quem lançou as bases da web têm muito mais propriedade para elencar características dela do que outros que a conheceram enquanto usuário ou empresa. Desse modo, urge salientar a importância de cientistas como Berners-Lee para a discussão em torno do futuro da web como forma de evitar infortúnios de avaliação.

Inclusive, o próprio Berners-Lee em 2006 assinou um artigo com outros pesquisadores nos quais defende a tese que a web está em processo de evolução: “o desenvolvimento da Web seguiu um caminho evolutivo, o que sugere uma visão da Web em termos ecológicos”⁴

⁴The development of the Web has followed an evolutionary path, suggesting a view of the Web in ecological terms. (Tradução do autor).

(BERNERS-LEE et al, 2006: 770). Ao contrário do que possa parecer a partir de uma análise sorrateira, a alcunha ecológica usada pelos cientistas em relação à web não diz respeito a uma suposta seleção natural das melhores empresas do setor. Para os autores do artigo a ecologia da rede de computadores diz respeito a uma evolução realizada ao longo dos anos por humanos, os quais contribuíram para que a web mudasse por meio das contribuições de cada um. Exatamente em consonância com os princípios elencados por Berners-Lee como pré-requisitos para a concretização do novo sistema em 1989.

O'Really, assim como muitas pessoas que avaliam a web, a observam a partir dos exemplos de empresas que se firmaram utilizando-se das ferramentas que a web disponibiliza. Uma visão um tanto quanto superficial caso o observador paute-se não pelas funcionalidades disponíveis, mas pelos pressupostos sobre os quais foi criada a web. Ou seja, é importante uma avaliação que não seja realizada a partir da visão de usuário da tecnologia, mas daqueles que manejam o sistema. Assim, certamente será possível visualizar aspectos relevantes que tradicionalmente não são do conhecimento da maioria.

Exemplo disso são as críticas feitas pelo próprio Tim Berners-Lee a determinadas companhias. A partir de uma discussão em torno das potencialidades da web, eles dizem que é preciso haver ampla colaboração entre áreas interdisciplinares a fim que seja possível resolver problemas que há algum tempo impedem a melhora da web. Um deles é o desenvolvimento de uma web semântica, a qual traria resultados de busca mais precisos e completos caso fossem encontradas soluções para entreveros matemáticos e estatísticos.

O desafio de engenharia é permitir que os sistemas de dados desenvolvidos de forma independente sejam ligados entre si sem a necessidade de um acordo global quanto a termos e conceitos. Os métodos estatísticos, que servem para o dimensionamento de recursos de linguagem em tarefas de busca, e os cálculos de dados, que são utilizados na ampliação consultas de dados, são em grande parte baseadas em suposições incompatíveis, e unificá-los será um grande desafio⁵ (BERNERS-LEE et al, 2006: 770)

Parte do imbróglio em comento poderia ser resolvido, no entendimento de Berners-Lee e demais autores do artigo, caso muitas empresas abandonassem a prática recorrente de

⁵The engineering challenge is to allow independently developed data systems to be connected together without requiring global agreement as to terms and concepts. The statistical methods that serve for the scaling of language resources in search tasks and the data calculi that are used in scaling database queries are largely based on incompatible assumptions, and unifying these will be a major challenge. (Tradução do autor).

não disponibilizar os dados dos usuários participantes, o que contraria não apenas os primórdios da web, mas a mentalidade de existência de uma web 2.0 cujas virtudes seriam, dentre outras, a ampla colaboração e compartilhamento de informações. O fato dos cientistas lamentarem o enclausuramento de informações que poderiam contribuir para uma web, de fato, mais aberta, é prova de que os usuários e os principais entusiastas do atual modelo existente de web e de suas fases de aperfeiçoamento precisam rever essa mentalidade.

Cientes de que essa situação dificilmente será alterada em virtude dos interesses comerciais das grandes companhias que obtêm e armazenam dados dos usuários, os pesquisadores propõem desafios matemáticos de modelagem de dados e ainda uma discussão acerca do controle e acesso dos dados compartilhados na web. Aliás, os autores acreditam que: “a escala, a topologia, e o poder dos sistemas de informação descentralizadas, como a Web, também representam um conjunto único de desafios sociais, públicos e políticos”⁶ (BERNERS-LEE et al, 2006: 770).

História, dados e jornalismo

Assim como é preciso reavaliar visões segundo as quais a web é formada por fases, as quais são marcadas pela introdução de novas ferramentas e informações, totalmente abertas e disponíveis ao usuário, geradas pelas grandes empresas do setor, também é importante ponderar a incorporação do uso de dados no jornalismo e as implicações disso para a área. Da mesma maneira como a ampla difusão e disseminação de informações incitam as pessoas a criarem novos termos para a web, pode estar ocorrendo algo similar com a terminologia jornalismo de dados.

Segundo Barboza (2007), base de dados é um termo cunhado na década de 1960 por norte-americanos que buscavam soluções para resolver problemas de arquivo. Com o passar dos anos esses sistemas ampliaram-se e ganharam escala à medida que máquinas computadoradas passaram a utilizá-las para guardar informação. Não por acaso, base de dados se tornou repositório de informações armazenadas em computadores.

Antes, porém, de avançar na discussão em torno do uso de dados, é importante refletir sobre seu significado. Para James Gleick, dado e bit são praticamente sinônimos. Bit foi

⁶ The scale, topology, and power of decentralized information systems such as the Web also pose a unique set of social and publicpolicy challenges. (Tradução do autor).

cunhado, segundo o jornalista, por Claude Shannon ao criar a Teoria Matemática da Comunicação. Quando estudava as unidades de medida para determinar a quantidade de informação que existia em uma mensagem, Shannon teria criado o nome para referir-se a dígitos binários. Sendo estes, a menor quantidade possível de informação existente em uma mensagem emitida por um humano ou uma máquina (GLEICK, 2013).

Já Abbagnano pressupõe:

O uso filosófico estabelece dois conceitos diferentes da noção de dado: 1ª o dado é o ponto de partida da análise, isto é, a situação de que se parte para resolver um problema ou as assunções ou os antecedentes de uma inferência ou de um discurso qualquer; 2ª o dado é o ponto de chegada da busca porque é o que se obtém quando se retiram do campo de indagação preconceitos, opiniões ou superestruturas falsificadoras, permitindo que se mostre e manifeste a realidade enquanto tal (2007: 231)

Para efeito desse artigo utilizar-se-á o primeiro conceito estabelecido por Abbagnano, do dado como ponto de partida da análise, somado às colocações de Lima Júnior: “dado é conceituado como sendo o dado binário, que é processado e armazenado por máquinas computacionais” (2012: 210). Assim, o dado neste artigo sempre será aquele advindo de computadores, do qual o jornalista buscará informações que sirvam de elemento para a criação da notícia, matéria-prima do jornalismo.

Além de conceituar dados, é válido também estabelecer uma relação deste ao termo jornalismo de dados. Uma das raras bibliografias sobre o tema, o livro *Manual de Jornalismo de Dados* coloca de forma prosaica a definição da palavra: “Eu poderia responder, simplesmente, que é um jornalismo feito com dados. Mas isso não ajuda muito” (GRAY et al, 2012: 8). Diante da escassez de informação, a obra acrescenta que dado não é apenas número, mas tudo que pode ser descrito em forma numérica no mundo digital, como uma fotografia, um vídeo ou um áudio. Ainda de forma bastante simplista, os autores do livro argumentam que o grande diferencial “talvez sejam as 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” (Ibid).

Isto é, jornalismo de dados poderia ser sintetizado, nos ditames do Manual, como uma narrativa jornalística baseada em grande volume de informação digital. Por não ser o objeto desse texto, não se adentrará na discussão sobre as diferenças que cercam dado, informação e notícia, pois cada um deles tem sua peculiaridade e explicá-los retiraria a possibilidade de uma discussão relativamente elaborada sobre os objetivos desse texto. De qualquer maneira, é importante salientar esses aspectos porque existem diferenças significativas entre eles.

O livro *Manual de Jornalismo de Dados* não é exceção quanto à dificuldade em definir o que é jornalismo de dados. A obra *Ferramentas digitais para jornalistas* (2010), que figura entre as escassas contribuições para a área, não faz qualquer menção ao que seria jornalismo de dados. Embora esteja nítida a preocupação em destacar instrumentais para a realização de reportagens a partir de dados coletados na web, lamenta-se não haver nenhuma definição da nova técnica que o livro almeja apresentar.

Rodrigues (2015) afirma que o jornalismo de dados começou a se desenvolver a partir do final da década de 1960, quando profissionais da imprensa como Philip Meyer iniciaram um debate sobre a necessidade de o jornalismo utilizar métricas quantitativas, sobretudo a partir de base de dados, para obtenção de informações mais precisas e menos baseadas em crenças pessoais. Vale ressaltar que os dados utilizados por Meyer não eram advindos da web ou de qualquer outra fonte digital. O jornalista coletava dados a partir de pesquisas realizadas por ele a partir de métodos científicos. A relação entre a prática hoje cunhada como Jornalismo de Dados e o que Meyer fez na década de 1970 estabeleceu-se “com o passar dos anos, a incorporação de dados numéricos em matérias jornalísticas ficou mais recorrente e ganhou ares de legitimação do discurso” (RODRIGUES, 2015, p. 337).

Desde o final da década de 1960, Meyer publicou estudos defendendo seu ponto de vista sobre o *modus operandi* do jornalismo. Em 1968 ele escreveu seu primeiro artigo sobre o tema *Truth in Polling* e em 1973 ele publica *Precision Journalism*, no qual sintetiza suas proposições. O livro foi relançado posteriormente em 1991 com o título *The New Precision Journalism* com um enfoque ainda maior na necessidade de o jornalista basear-se em metodologias técnicas e científicas de apuração da notícia e no uso de dados para a realização de reportagens. Desde o princípio, Meyer acreditava que os dados eram imprescindíveis para o trabalho jornalístico. Como os computadores pessoais não tinham se popularizado

(CHRISTENSEN, 1993), Meyer utilizou a coleta de dados a partir de pesquisa empírica, geralmente com questionários impressos, entrevistando as pessoas para obter informações.

O que diferenciava o jornalismo de precisão de Meyer do trabalho realizado atualmente pelo jornalismo de dados é o uso do computador como ferramenta de apuração jornalística. Intrinsecamente, não existem divergências entre elas porque ambas as práticas baseiam-se na busca por nomes, números e outros tipos de mensagens que estão presentes na realidade. Partindo da premissa do jornalismo como o relato dos fatos de interesse público à sociedade (TRAQUINA, 2001), compreende-se as origens em comum e os fundamentos que sustentam o jornalismo de precisão de Meyer e o jornalismo de dados atual.

Outro fato incontestável dessa constatação está na obra mais atual de Meyer. O jornalista criou uma lista com seis recomendações básicas para o profissional de imprensa apropriar-se e divulgar adequadamente das informações empíricas coletadas. Esses tópicos endossam a mentalidade segundo a qual saber o que fazer com os dados é a essência do jornalismo:

1. Reúna-o. Querendo ou não você nunca tentará imitar cientistas em seus métodos de coleta de dados, você pode lucrar se souber alguns de seus truques. É sempre bom lembrar, como o professor H. Douglas Price disse-me em Harvard na primavera de 1967, que "os dados não vêm da cegonha."
2. Armazene-o. Jornalistas à antiga armazenam dados em pilhas de papel em suas mesas, em cantos de seus escritórios, e, se eles são realmente bem organizados, em grampo-arquivos. Computadores são melhores.
3. Recupere-o. As ferramentas do jornalismo de precisão podem ajudá-lo a recuperar dados que você mesmo recolheu e armazenou, dados que alguém armazenou, ou ainda dados que alguém armazenou por motivos completamente alheios a seu interesse.
4. Analise-o. Análise jornalística muitas vezes consiste em apenas triagem para encontrar e listar os desvios interessantes, mas também pode envolver pesquisas para o nexos de causalidade implícita, para os padrões que sugerem que fenômenos diferentes variam juntos por razões interessantes.

5. Reduza-o. Redução de dados tornou-se tão importante no jornalismo como a coleta de dados. Uma história boa notícia é definida pelo que deixa de fora, bem como o que inclui.
6. Comunique-o. Um relatório não lido ou não entendido é um relatório desperdiçado.⁷ (MEYER, 1991, p. 33)

Tal listagem é extremamente importante porque mostra o quanto os princípios do jornalismo de dados realizado hoje se baseiam nos pressupostos defendidos por Meyer no começo da década de 1990. Tal como ocorreu com as ideias seminais de Berners-Lee para a criação de uma rede no final da década de 1980 e o suposto surgimento da web 2.0 de O'Reilly no começo deste milênio, há certa inconsistência em creditar uma nova fase do jornalismo a partir dos dados. A técnica de captação de informação mudou em decorrência da disponibilização do computador e da criação da web como repositório de dados. O jornalismo, porém, permanece baseado na premissa de que é preciso saber o que coletar, onde coletar, porque coletar, quando coletar e como divulgar. Tal qual defendeu Meyer, usar corretamente os dados é a quintessência do jornalismo de qualidade.

Os jornalistas britânicos Jonathan Gray e Paul Bradshaw não adentram a questões teóricas como Meyer, mas contribuíram de forma substantiva para o jornalismo de dados ao discutir determinados aspectos básicos da prática. Bradshaw escreveu em 2010 um artigo intitulado *How to be a data journalist* no qual assume um papel pedagógico para explicar o que seria apropriado os profissionais da área fazerem para praticarem a nova técnica. Antes de dar dicas que considera importantes, Bradshaw faz alguns apontamentos:

⁷1. Collect it. Whether or not you ever try to emulate scientists in their data-collection methods, you can profit from knowing some of their tricks. It is always worth remembering, as Professor H. Douglas Price told me at Harvard in the spring of 1967, that "data do not come from the stork."

2. Store it. Old-time journalists store data on stacks of paper on their desks, in corners of their offices, and, if they are really well organized, in clip-files. Computers are better.

3. Retrieve it. The tools of precision journalism can help you retrieve data that you collected and stored yourself, data that someone else stored with a user like you in mind, or data that someone else stored for reasons completely unrelated to your interest, perhaps with no earthly idea that a journalist or public user would ever be interested in it.

4. Analyze it. Journalistic analysis often consists of merely sorting to find and list the interesting deviances. But it can also involve searches for implied causation, for patterns that suggest that different phenomena vary together for interesting reasons.

5. Reduce it. Data reduction has become as important a skill in journalism as data collection. A good news story is defined by what it leaves out as well as what it includes.

6. Communicate it. A report unread or not understood is a report wasted. You can make a philosophical case that, like the sound of a tree falling in the forest, it does not exist at all. (Tradução do autor).

Jornalismo de dados é enorme. Eu não quero dizer 'enorme' como na moda - embora se tenha tornado que nos últimos meses - mas 'enorme' como em 'incompreensivelmente enorme'. Ela representa a convergência de um número de campos que são significativos em seu próprio direito - a partir de pesquisa investigativa e estatística para design e programação. A idéia de combinar as habilidades de contar histórias importantes é poderosa - mas também intimidante (BRADSHAW, 2010).

O jornalista e professor da *City University in London* diz também que existem diferentes formas de fazer jornalismo de dados, as quais fariam partes de um quebra-cabeças. A primeira delas seria encontrar os dados, o que poderia ser feito por meio de sistemas de gerenciamento de banco de dados como MySQL ou por linguagens de programação como Python. Bradshaw recomenda que o jornalista tenha conhecimento especializado para realizar certas tarefas.

A segunda parte do quebra-cabeça seria criar compreender os dados, o que Bradshaw chama alegoricamente de interrogatório. Ele sugere noções de estatística para compreensão de planilhas, material básico neste novo setor do jornalismo baseado em números e outras informações compiladas em forma não-textual. A terceira seria criar maneiras adequadas de mostrar os dados ao leitores, no caso de um veículo de comunicação impresso ou pela internet. O professor diz que essa atribuição mormente recai sobre programadores e designers, no entanto é preciso que jornalistas adentrem essa seara. A quarta e última seria criar uma compreensão dos dados, o que Bradshaw chama de Mashing dados. Para tanto, o jornalista recomenda o uso de ferramentas como ManyEyes ou Yahoo Pipes (BRADSHAW, 2010).

A partir dessas exortações, Bradshaw lembra que o jornalismo de dados deve partir sempre de elementos básicos do próprio jornalismo, como o de contar histórias, porém, no caso específico do jornalismo de dados, procurar fazê-lo por intermédio de números. Outra abordagem sugerida é de começar a reportagem sempre a partir de uma pergunta, a qual seria respondida ao longo da daquela. Todos esses aspectos são classificados como importantes por manuais e outras obras que se valem do jornalismo, *o que demonstra o caráter pioneiro do trabalho de Bradshaw.*

Contudo, uma das principais contribuições do professor e de outros jornalistas ao tratar de jornalismo de dados tenha sido outra. Eles tocam em um aspecto muito mais relevante que sugestões de como iniciar uma notícia. O professor comenta uma preocupação de vários setores da sociedade britânica pela disponibilização de dados por parte de fontes governamentais, o que no entendimento dele e de diversos setores da sociedade civil organizada é imprescindível para o exercício do bom jornalismo (BRADSHAW, 2010).

Embora Bradshaw não afirme, existem muitos entes preocupados em forçar fontes governamentais e não-governamentais a divulgar dados e outras informações públicas necessárias ao exercício da cidadania. Jonathan Gray, um dos jornalistas que trabalhou na elaboração do livro *Manual de Jornalismo de Dados*, o qual foi traduzido para diversas línguas, como o português, posteriormente, recorda que em 2006 o jornal *The Guardian* lançou uma campanha para que as entidades públicas tornem os dados abertos. Nesta mesma reportagem de 2006, Gray afirma que Tim Berners-Lee e outros cientistas endossaram a proposta por acreditarem na importância dela para o desenvolvimento da web (GRAY, 2011).

Gray salienta que em pouco mais de meia década, desde o início do século XXI, prefeituras, Banco Mundial, Comissão Europeia e tantas outras instituições supranacionais aderiram à proposição de serem mais transparentes. No entanto, o jornalista afirma que ainda existem muitos desafios a serem superados.

No início deste ano houve relatos de que *Data.gov* terá seu financiamento cortado. No Reino Unido, há preocupações de que o ameaçadoramente intitulado "Public Data Corporation" pode significar que uma quantidade crescente de dados seja bloqueada e vendida para aqueles que podem dar ao luxo de pagar por isso. E na maioria dos países ao redor do mundo a maioria dos documentos e conjuntos de dados ainda é publicado em condições legais ambíguas ou restritivas, que inibem a reutilização. Cortes de gastos do setor público e medidas de austeridade em muitos

países vão tornar ainda mais difícil para os dados abertos tornarem-se prioridades⁸ (GRAY, 2011).

No caso brasileiro, a afirmação de Gray torna-se factível de variados modos. Pelo aspecto positivo, há a criação e promulgação da Lei 12.527 em 18 de novembro de 2011, chamada de lei de acesso à informação. Nos ditames do *caput* da legislação, fica estabelecido:

Regula o acesso a informações previsto no inciso XXXIII do art. 5º, no inciso II do § 3º do art. 37 e no § 2º do art. 216 da Constituição Federal; altera a Lei no 8.112, de 11 de dezembro de 1990; revoga a Lei no 11.111, de 5 de maio de 2005, e dispositivos da Lei no 8.159, de 8 de janeiro de 1991; e dá outras providências (BRASIL, 2011).

Na prática, a legislação obrigou os entes públicos das diversas esferas, quais sejam, municipais, estaduais e federais, a divulgarem informações até então inacessíveis à população, como o gasto com os vencimentos de servidores. A medida também forçou empresas públicas, fundações, autarquias e outras instituições vinculadas ao setor governamental a publicar suas normas, realizar consultas públicas virtuais e demais atos administrativos até então inalcançáveis à população, o que representou grande avanço em relação ao panorama anterior.

No entanto, o aspecto negativo disso, conforme vaticinado por Gray, é o fato de que a lei de acesso à informação é constantemente vilipendiada pelas entidades e órgãos que deveriam zelar por ela. Seja por omissão, negligência ou sob o argumento difuso de confidencial, o cidadão, veículos de comunicação ou qualquer outro ente que busque determinadas informações não terá o seu direito respeitado (LEALI, 2015; GONZALES, 2015). Infelizmente, a realidade é a mesma em quaisquer dos níveis, notadamente no interior do país, e também nas entidades públicas ligadas ao governo (SCHUINSKI, 2014).

⁸Earlier this year there were reports that Data.gov will have its funding slashed. In the UK there are concerns that the ominously titled "Public Data Corporation" may mean that an increasing amount of data is locked down and sold to those who can afford to pay for it. And in most countries around the world most documents and datasets are still published with ambiguous or restrictive legal conditions, which inhibit reuse. Public sector spending cuts and austerity measures in many countries will make it harder for open data to rise up priority lists. (Tradução do autor).

Isso significa que para o jornalismo de dados ser realizado de forma satisfatória é preciso, inicialmente, uma disponibilização de informações na web, pois a prática se concretiza à medida que profissionais dispõem-se a coletar dados, selecioná-los, criar uma visualização adequada para ele e disponibilizá-lo da maneira mais adequada possível de acordo com o público consumidor de notícia.

Não é o jornalista que cria a informação, ele coleta-a e transforma-a em reportagem por meio de recursos textuais, audiovisuais, infográficos ou outras. Logo, a matéria-prima do jornalista continua sendo a mesma desde os primórdios. O que está mudando é a forma como ela é disponibilizada e, no caso específico do jornalismo de dados, o leque de opções tecnológicas usadas para ter acesso a elas e criação de novas narrativas, já que esta nova técnica requer um repertório específico de conhecimento de informática e técnicas corretalatas.

Silveira (2010, 2015) converge com Gray (2011, 2012) e Bradshaw (2010) e endossa as colocações de Meyer (1991). No entanto, Silveira parte de uma premissa mais radical. Ele estabelece uma relação entre Jornalismo de Dados e a cultura hacker por acreditar “que o universo hacker seja tão ou mais complexo que o universo do jornalismo” (2015, p. 34). Baseado nos estudos antropológicos de Colemann e Golub (2008) a cerca dos três tipos de prática hacker, o autor realiza um retrospecto histórico para salientar que ambos são originários de valores liberais, os quais foram levados ao extremo pela cultura hacker ao pregar o individualismo sem limites e o compartilhamento de informações. O jornalismo, pelo contrário, teria “se escondido sob a falácia da imparcialidade” (SILVEIRA, 2015, p. 35). Assim, o jornalismo teria se afastado das suas atribuições naturais de defesa dos interesses humanos.

Silveira cita o caso das denúncias de espionagem do ex-analista da NSA (agência de segurança norte-americana) Eduard Snowden e do jornalista e ativista Julian Assange criador do site Wikileaks para exemplificar que foi preciso a interferência de hackers para muitas informações importantes à sociedade se tornarem de conhecimento público. Outrossim, Silveira defende que “sem a lógica hacker, o jornalismo de dados pode ser apenas uma versão mais sofisticada do velho jornalismo de grupo e dos interesses corporativos”. Isto é, o jornalismo de dados, no entendimento do autor, não só deve abandonar totalmente a busca

pela imparcialidade como precisa utilizar-se de métodos ortodoxos e até contrários à lei em nome da informação de interesse público.

Considerações finais

Mais que exercício de predição, o jornalismo de dados representa, de fato, uma nova possibilidade do profissional de imprensa coletar dados para realizar satisfatoriamente seu trabalho. Em um momento em que se questiona a necessidade de humanos para produzir uma notícia, haja vista as máquinas que já estão a fazê-las, é irônico constatar as diversas possibilidades que se aventam para os seres humanos jornalistas a partir de computadores e programas, pois o jornalismo de dados se realiza a partir do trabalho conjunto de robôs e pessoas. Ademais, nota-se que o jornalismo de dados abre novas possibilidades ao exercício da profissão, mas não representa uma total ruptura com os padrões convencionais.

Isto posto, o jornalismo de dados tem um grande desafio pela frente haja visto os dilemas que a maior disponibilidade de informação e a sociedade conectada a dispositivos transmídia e ubíquos trouxeram. Novos estudos teóricos e práticos sobre limites éticos, formas de coletar, tratar e divulgar a informação e a prática como um todo do jornalismo de dados são importantes e necessários para a profissão.

Referências

ABBAGNANO, Nicola. **Dicionário de filosofia**. São Paulo: Martins Fontes, 2007.

BARBOZA, Suzana. **Jornalismo digital em base de dados (JDBD): um paradigma para produtos jornalísticos digitais dinâmicos**. 319 f.. Tese (doutorado) - Universidade Federal da Bahia, Faculdade de Comunicação, 2007.

BARLOW, Mike. **Data Visualization: A New Language for Storytelling**. Sebastopol: O'Reilly Media Inc, 2015.

BERNERS-LEE, Tim. Information Management: A Proposal. **History**, v. 198, n. 9, 1989. Disponível em: <http://www.w3.org/History/1989/proposal.html>. Acesso em 24 de setembro de 2015.

BERNERS-LEE, Tim et al. Creating a Science of the Web. **Science**, v. 313, n. 5788, p. 769-771, 2006.

BRADSHAW, Paul. How to be a data journalist. **The Guardian**. 1 out. 2010. Disponível em : <http://www.theguardian.com/news/datablog/2010/oct/01/data-journalism-how-to-guide>. Acesso em: 1 de agosto de 2015.

BRASIL. Lei 12.527, 18 de novembro de 2011. Regula o acesso a informações previsto no inciso XXXIII do art. 5º, no inciso II do § 3º do art. 37 e no § 2º do art. 216 da Constituição Federal; altera a Lei nº 8.112, de 11 de dezembro de 1990; revoga a Lei nº 11.111, de 5 de maio de 2005, e dispositivos da Lei nº 8.159, de 8 de janeiro de 1991; e dá outras providências. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2011/lei/112527.htm. Acesso em: 12 de setembro de 2015.

COLEMAN, E. Gabriella; GOLUB, Alex. Hacker practice Moral genres and the cultural articulation of liberalism. **Anthropological Theory**, v. 8, n. 3, p. 255-277, 2008.

CHRISTENSEN, Clayton. The Rigid Disk drive Industry: A History od Commercial and Technological Turbulence. **Business History Review**, n. 67, Inverno, p. 531-588, 1993.

CRUCIANELLI, Sandra. **Ferramentas Digitais para jornalistas**. Austin: University of Texas, 2010.

GONZALES. Amélia. Pesquisa revela que Lei de Informação não é respeitada em obras para as Olimpíadas. **G1**. 6 jul 2015. Disponível em: <http://g1.globo.com/natureza/blog/nova-etica-social/post/pesquisa-revela-que-lei-da-informacao-nao-e-respeitada-em-obras-para-olimpiadas.html>. Acesso em: 20 de setembro de 2015.

GLEICK, James. **A informação: uma história, uma teoria, uma enxurrada**. São Paulo: Companhia das Letras, 2013.

GRAY, Jonathan. The future of open data? **The Guardian**. 11 out. 2011. Disponível em: <http://www.theguardian.com/news/datablog/2011/oct/11/open-government-data-camp-2011-free-our-data?newsfeed=true>. Acesso em 25 de setembro de 2015.

GRAY, Jonathan et al. **Manual de jornalismo de dados**. Março, 2012. Disponível em: <http://datajournalismhandbook.org/pt/index.html>. Acesso em: 27 ago 2015.

LEALI, Francisco. Lei de acesso é desrespeitada por órgãos do governo. **O Globo**. 10 mai 2015. Disponível em: <http://oglobo.globo.com/brasil/lei-de-acesso-desrespeitada-por-orgaos-do-governo-16110382>. Acesso em: 21 de setembro de 2015.

LIMA JÚNIOR, Walter Teixeira. “Big Data, Jornalismo Computacional e Data Journalism: estrutura, pensamento e prática profissional na Web dados”. **Estudos em Comunicação**, nº 12, p. 207-222, dez. 2012.

MEYER, Philip. **Truth in Polling**. New York: Columbia Journalism Review, Summer, 1968.

_____. **Precision Journalism**. Bloomington: Indiana University Press, 1973

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

O'REILLY, Tim. **What is web 2.0**. 30 de setembro de 2005. Disponível em: <http://www.oreilly.com/pub/a/web2/archive/what-is-web-20.html?page=1>. Acesso em: 25 ago 2015.

PRIMO, Alex. O aspecto relacional das interações na Web 2.0. In: **E-Compós (Brasília)**, v. 9, p. 1-21, 2007.

RODRIGUES, Kelly De Conti. Jornalismo de Dados: influência da construção narrativa no agendamento midiático. In: **Ciências da Comunicação: Circularidades Teóricas e Práticas**

Acadêmicas. (Org.s) BULHÕES, Marcelo; MORAIS, Osvando J. Sarapuí: OJM Casa Editorial, 2015.

SILVEIRA, Sérgio Amadeu. A informação quer ser livre. In: **Revista de Jornalismo ESPM**, nº 12, ano 4, jan-mar, p.30-37, 2015.

_____. Ciberativismo, cultura hacker e o individualismo colaborativo. **Revista USP**, n. 86, ano 3, jun-ago, p. 28-39, 2010.

SCHUINSKI, Rodrigo Menegat. Gestores públicos de Ponta Grossa ainda desrespeitam Lei de Acesso à Informação. **Portal Comunitário**. 9 set 2014. Disponível em: <http://www.portalcomunitario.jor.br/index.php/cidade-cidadania/3897-gestores-publicos-de-ponta-grossa-ainda-desrespeitam-lei-de-acesso-a-informacao>. Acesso em: 1 de setembro de 2015.

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

WIKIPEDIA. **Google Ngram Viewer**. Disponível em: <https://en.wikipedia.org/wiki/N-gram>. Acesso em: 11 de setembro de 2015.