

Big Data, Jornalismo Computacional e Data Journalism: estrutura, pensamento e prática profissional na Web de dados

Walter Teixeira Lima Junior
Universidade Metodista de São Paulo, Brasil
E-mail: digital@walterlima.jor.br

Resumo

A atual configuração tecnológica da Internet está sendo desenhada a partir do aumento da velocidade de transmissão de dados, da interconexão entre máquinas computacionais com robusta capacidade de processamento e armazenamento de dados e da utilização de linguagens de programação sofisticadas, que permitem a manipulação de imensos e complexos bancos de dados. Nesse patamar tecnológico da rede, a atuação do profissional

de Jornalismo se sofisticava no sentido da aproximação com outras áreas do conhecimento humano, como o campo das Ciências da Computação. A necessidade do entendimento das lógicas computacionais possibilitam ao data journalist utilizar novas possibilidades no tratamento de informações e dados, encontrando novas abordagens para informação de relevância social, novos fatos e fontes de informação alternativa.

Palavras-chave: *big data*; jornalismo computacional; *data journalism*

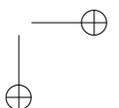
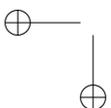
Big Data, Computational Journalism and Data Journalism: structure, thought and professional practice in the data web

Abstract

The current technological configuration of the Internet is being drawn from the increase of the data transmission speed, from the interconnection between computational machines with robust processing capacity and storage of data, and from the storage use of sophisticated programming languages, which allow the

handling of immense and complex data bases.

In this technological level of the network, the practice of the journalism professional sophisticates towards other areas of human knowledge, namely the field of Computer Sciences. The need for an understanding of computational logics will enable the data journalist the use of



new possibilities in the treatment of information and data, finding out new approaches to information of social relevance, new facts, and sources of alternative information.

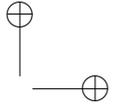
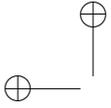
Keywords: *big data*; computacional journalism; *data journalism*

NAS ÚLTIMAS DÉCADAS, a sociedade contemporânea absorveu as Tecnologias de Informação e Comunicação (TIC's) em todos os seus segmentos. Entre os principais motivos dessa "conexão amigável", entre pessoas não especialistas em artefatos tecnológicos e as tecnologias digitais, estão à percepção de que elas trazem conforto, vantagens competitivas e podem ser obtidas com mais frequência devido à diminuição dos custos de obtenção de tais sistemas computacionais, alguns até se transformando em utensílios domiciliares e vendidos em lojas de eletrodomésticos, como por exemplo, o *Personal Computer* (PC).

Portanto, a sociedade está se apropriando das tecnologias digitais, algumas até certo tempo, utilizadas somente por governos ou grande empresas como, por exemplo, o Global Positioning System (GPS). Os resultados sociais das apropriações atingem todos os setores que utilizam as TIC's, em expansão vertiginosa no cotidiano do homem contemporâneo. O ramo de produção de conteúdo informativo de relevância social, onde o Jornalismo se situa, também está se adaptando às transformações decorrentes dessa nova configuração.

Entretanto, a adaptação tecnológica não parece que está sendo natural no campo do Jornalismo, como em outros setores importantes como, por exemplo, na Medicina. Nessa esfera da atividade humana, o desenvolvimento, a obtenção e a utilização de tecnologias de ponta são quase imperativos para a manutenção de vidas e, por consequência, da conservação do *status quo* da Medicina como atividade essencial para a sociedade. Ligada à Ciência umbilicalmente, depois que Hipócrates introduziu o pensamento científico na área, a Medicina se contrapôs ao obscurantismo advindo do curandeirismo e de outras práticas culturais seculares. Sob esse ponto de vista, as tecnologias são umas das bases da evolução das práticas da Medicina.

No campo da Comunicação Social, mas especificamente no Jornalismo, devido à abordagem deste trabalho, as TIC's foram introduzidas paulatinamente nos grupos de mídia. O primeiro setor, a ser impactado por elas, foi o da produção da notícia. Nos anos 80, os PC's começaram a ser introduzidos



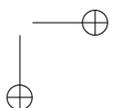
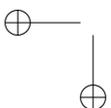
nas redações brasileiras. “O jornal Folha de S.Paulo foi o primeiro no Brasil a introduzir terminais de computador em suas redações, projeto que vinha sendo elaborado desde 1968 e foi efetivado em 1983 (Vianna, 1992)”. Entretanto, demoramos cerca de duas décadas para inserir tais equipamentos no cotidiano profissional do jornalista brasileiro. No final dos anos 60 e início dos 70 do século passado, o The New York Times estruturou o primeiro banco de dados, que foi inserido nas etapas de produção da notícia. A agência de notícias Reuters, em 1968, foi pioneira a utilizar máquinas computacionais nas conexões da sua rede interna para gerenciar a demanda de notícias recebidas.

Assim, a cultura de apropriação tecnológica, no caso da produção do Jornalismo brasileiro, possui uma história de atraso em relação aos principais centros. O intervalo de implementação tecnológica, no que tange aos investimentos, pelos parques brasileiros muitas vezes é explicado pela diferença entre a economia estadunidense e a brasileira nas quatro últimas décadas do século XX. Porém, na atualidade, o fator não interfere tanto. Pois, aconteceu o barateamento das máquinas computacionais e de outros dispositivos tecnológicos (acesso à Internet, displays, memória entre outros). Pode-se também inserir, nesse novo momento de apropriação tecnológica, a evolução do desenvolvimento e apropriação do Software Livre (Linux), livrando as empresas e os usuários dos custos dos softwares proprietários (Windows, Adobe, Apple etc).

Uma discussão importante é se o atraso da apropriação tecnológica interfere no atual panorama da prática e, até, da pesquisa em Jornalismo. Jack Fuller (2010) afirma que o modelo de Modelo Padrão de Jornalismo Tradicional foi calcado em uma base tecnológica anterior as tecnologias digitais e conceitua:

O Modelo Padrão de Jornalismo Tradicional inclui os atributos da precisão, imparcialidade na reportagem, independência de pessoas e organizações reportadas ou afetadas pela matéria, um modo de apresentação algumas vezes chamado de objetivo ou neutro e clara rotulagem do que é fato e o que é opinião (Fuller, 2010).

Fuller (2010) acredita que o Modelo Padrão de Jornalismo Tradicional não considera que o fator de remodelação do “ambiente da informação, no qual a notícia é uma pequena parte, tem sido produzido pela tecnologia.” E, alerta aos jornalistas que no mundo atual das tecnologias digitais conectadas “há mais do que a internet (Fuller, 2010)”.



A intenção do autor é para alertar para a questão do uso doméstico das Tecnologias de Informação e Comunicação por jornalistas. Sejam eles vinculados aos grupos de mídias tradicionais, sejam nativos do mundo digital e/ou *free-lancers*. Ou seja, há necessidade de dominar tecnologias digitais mais profundamente, portanto, subir o patamar de apropriação proporcionado pela Web mais amigável (*user-friendly*), que foi denominada mercadologicamente de Web 2.0¹ por Tim O’Reilly.

Aliás, o termo Web 2.0 tem sido bastante criticado, já que a Web nasceu de forma colaborativa nas mãos de Tim Berners-Lee e seus colaboradores. Em entrevista² para IBM, disponível em formato *podcast*, Berners-Lee afirma que “ninguém sabe o que esse termo significa”.

Big Data, Pensamento Computacional e Jornalismo Computacional

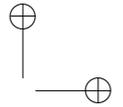
Na atual configuração tecnológica proporcionada pela Internet, estruturada pelo intermédio do aumento de velocidade de transmissão, pela evolução das máquinas computacionais com grande capacidade de processamento e armazenamento de dados, com o desenvolvimento de linguagens de programação cada vez mais amplas e que negociam de várias formas com robustos bancos de dados, a atuação profissional do Jornalismo também deve possuir outras configurações.

O Jornalismo sempre manuseou informação. Neste texto, o termo informação é conceituado como dado estruturado, que possui dependência do ambiente e tem campo semântico (Floridi, 2010). E o dado é conceituado como sendo o dado binário, que é processado e armazenado por máquinas computacionais.

O atual ambiente computacional conectado via redes telemáticas faz emergir novas possibilidades no tratamento de informações e dados no campo do Jornalismo, podendo-se reduzir, se bem utilizado, alguns dos problemas crô-

1. O’Reilly, T. (2005, setembro) What Is Web 2.0: Design Patterns and Business Models for the Next Generation of Software. O’Reilly Spreading the Knowledge of innovators. Retirado de <http://oreilly.com/web2/archive/what-is-web-20.html>.

2. Laningham (2006, julho). Interview with Tim Berners-Lee. Retirado de <http://www.ibm.com/developerworks/podcast/dwi/cm-int082206.txt>.



nicos na prática do Jornalismo, como encontrar novas abordagens, fatos e fontes.

A base tecnológica que estabelece a possibilidade de novas configurações para a prática do Jornalismo é a era denominada de *Big Data*.

"*Big data*" refere-se ao conjunto de dados (*dataset*) cujo tamanho está além da habilidade de ferramentas típicas de banco de dados em capturar, gerenciar e analisar. A definição é intencionalmente subjetiva e incorpora uma definição que se move de como um grande conjunto de dados necessita ser para ser considerado um *big data*³.

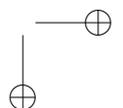
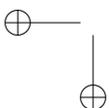
Pode-se considerar um *Big Data* a Receita Federal, o Hospital das Clínicas de São Paulo, a Companhia de Engenharia de Tráfego de São Paulo (CET), o Controle de Tráfego Aéreo, o *New York Times* entre outros. No caso da Receita Federal, por exemplo, é um *Big Data* não disponível, ou seja, com acesso livre ao público. Mas existem repositórios abertos de dados, denominados de *Open Data*, que possuem dados públicos e podem ser manuseados por quem se interessar. É importante ressaltar que o *Big Data*, na sua grande maioria, é um conjunto de dados que a cada milésimo de segundo são inseridos novos.

Vários pesquisadores estudam a quantidade de dados gerados, guardados e consumidos no mundo. As estimativas deles variam muito, mas concordam que o crescimento é exponencial. MGI estima que empresas armazenam globalmente mais do que 7 exabytes de novos dados em disk drives em 2010, enquanto consumidores guardam mais do que 6 exabytes de novos dados em PC's e notebooks. Um exabyte de dato é equivalente a mais do que quatro mil vezes a informação guardada na Livraria do Congresso dos EUA (235 terabytes).⁴

No estudo realizado pelos *Pew Research Center's Internet & American Life Project*, intitulado "*The Future of Big Data*", afirma que por volta de 2020, o uso do *Big Data* melhorará o entendimento sobre nós mesmos e sobre o mundo. O professor Jeff Jarvis, uns dos entrevistados pelos produtores do

3. Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., oxburgh, C. & Byers A. H. (2006, maio). Big data: The next frontier for innovation, competition, and productivity. McKinsey Global Institute. Retirado de http://www.mckinsey.com/insights/mgi/research/technology_and_innovation/big_data_the_next_frontier_for_innovation.

4. Idem



estudo, afirma que “há valor a ser encontrado nesses dados, valor em nossa descoberta a ser colocada em público”⁵.

O “valor” que Jarvis se refere é o mesmo “valor” encontrado por Philip Meyer, em 1967. Como consequência da evolução tecnológica no campo dos computadores (hardware) e da computação (software), o pioneiro Philip Meyer foi o jornalista que descobriu a forma de trabalhar a informação por intermédio de base de dados. Trabalhando para Detroit Free Press, Meyer produziu uma matéria sobre distúrbios raciais em Detroit. Para isso, utilizou computador mainframe para analisar a demografia dos negros na cidade. Assim, começou a “integração de computadores e a ciência social e ele foi o precursor na área de *Computer-Assisted Reporting* (CAR) (Royal, 2010)”.

Apesar de possuir o mesmo “valor”, ou seja, produzir uma nova informação a partir de dados coletados, cruzados e/ou relacionados armazenados em base de dados digitais, as habilidades do jornalista que utiliza os conceitos e ferramentas no CAR, como o uso de planilha de dados, trabalha de forma diferente do jornalista que utiliza base de dados abertas (*Open Data*) na Web ou realizando a técnica do *Scrapping*, conceito que traduzido livremente para a língua portuguesa pode ser entendido como “raspagem de dados”, geralmente feitos em documentos em formato PDF.

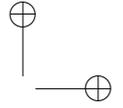
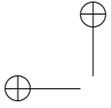
Para Paul Bradshaw, da Birmingham City University, trabalhar com planilha de dados, há 20 anos, era a única forma de obter “valor” no cruzamento de dados, “mas nós vivemos em um mundo digital agora, um mundo no qual quase tudo pode ser - e quase tudo é - descrito em números”⁶.

O mundo digital conectado que Bradshaw se refere é a Web. Composta de muitos serviços, a Web também possui um conjunto de dados que podem ser conectados, conceito denominado de *Linked Data*, que versa “sobre a utilização da Web para conectar dados relacionados, que não estavam anteriormente ligados”⁷.

5. Anderson J., Rainie, L.(2012, julho). The Future of Big Data . Pew Internet Center. Retirado de http://pewinternet.org/Reports/2012/Future-of-Big-Data.%20aspx?utm_source=Mailing+List&utm_campaign=6cfef277d6-future_of_big_data_alert07_20_2012&utm_medium=email

6. The Data Journalism Handbook . European Journalism Centre. Retirado de http://datajournalismhandbook.org/1.0/en/introduction_0.html

7. Linked Data - Connect Distributed Data across the Web. Retirado de <http://linkeddata.org/>



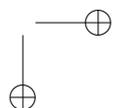
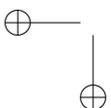
Para realizar uma prática jornalística contemporânea, o profissional deve conhecer e manipular com razoável conhecimento as terminologias e tecnologias que constituem o universo de dados digitais e as suas possíveis conexões, com o objetivo de obter “valor”, ou seja, conteúdo informativo de relevância social por intermédio de dados disponíveis na Web. Portanto, as atuais formas de obtenção e relacionamento de dados exigem requinte tecnológico, mais especializado, do profissional em Jornalismo no ambiente da Web de dados.

O entendimento mais fundamentado sobre o funcionamento das Tecnologias de Comunicação e Informação, utilizadas nos processos de produção do Jornalismo tendo como base a Web de Dados, passa por conhecer as lógicas que as criaram, como se desenvolveram e continuam evoluindo nos seus processos de inovação. Ou seja, o entendimento mais fundamentado sobre as TCI's, que estruturam sistemas e plataformas na área do Jornalismo, deve ser mais do que uma aprendizagem digital ou, ainda, de utilização de forma doméstica, termo utilizado para designar como a maioria dos usuários que se apropria das tecnologias oferecidas pela Web, por exemplo.

Desta maneira, o jornalista que pretende atuar na área da Web de Dados necessita adquirir outras habilidades, além das fundamentais ensinadas nas faculdades de Jornalismo e aprimoradas nas redações. Jeannete Wing (2006) acredita que é necessário possuir um Pensamento Computacional para extrair o potencial possível das tecnologias e criar novas possibilidades através delas. O cientista da computação aprende a pensar computacionalmente para executar computacionalmente as suas modelagens (sistemas).

O pensamento computacional é usar a abstração e decomposição quando se ataca uma grande tarefa complexa ou se desenha um grande sistema complexo. É a separação de preocupações (interesses). É escolher uma apropriada representação do problema e modelar os aspectos relevantes de um problema e fazê-los tratáveis. É usar constantes para descrever o comportamento de um sistema sucintamente e declarativamente (Wing, 2006).

No Jornalismo, por exemplo, a forma de pensar computacionalmente auxilia no entendimento de como funciona uma determinada tecnologia e quais as possíveis apropriações profissionais poder se fazer dela. Além disso, ajuda na compreensão dos processos de inovação disparados a partir da sua estrutura. Pode ser justamente a aquisição do novo conhecimento que se estabelece a ponte entre as funções do Jornalismo, na atual sociedade, e a capacidade de obter informações de relevância social por intermédio de sistemas computaci-



onais. “A computação pode promover o Jornalismo na elaboração em inovações na percepção de assuntos, análise de vídeos, personalização, agregação, visualização e produção de sentidos (Cohen; Hamilton & Turner, 2011)”.

Fred Turner e James Hamilton (2009) afirmam que existe uma especialização jornalística, que surge pelo cruzamento das áreas do Jornalismo e da Ciência da Computação. Essa formação utiliza “a combinação de algoritmos, dados e conhecimento das Ciências Sociais para suplementar a função de responsabilidade do Jornalismo (Hamilton & Turner, 2009)”. Eles a denominam de Jornalismo Computacional. Apesar de diferente das expressões Jornalismo Digital, Jornalismo Online, WebJornalismo e CiberJornalismo, o Jornalismo Computacional também possui vínculo específico com as práticas jornalísticas em ambientes digitais conectados, mas especificamente, com tratamento de dados em bases digitais.

Neste ambiente mais especializado, emerge um jornalista profissional com algumas habilidades suplementares que permite entender, na sua magnitude tecnológica, o sistema digital conectado no qual atua.

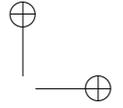
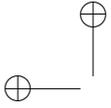
Data journalism

A definição exata de *data journalism* é uma construção difícil neste momento da evolução da atividade profissional, que possui um vetor importante de inovação: as tecnologias digitais. Juntamente com a Web, formam a base para a prática do *data journalism*, modificando a forma como a informação é captada, filtrada e publicada. Para Liliana Bounegru, do European Journalism Centre, é o “conhecimento de dados em massa”⁸. Corrobora com essa visão o inventor do WWW, Tim Berners-Lee, ao analisar que o trabalho com dados é o futuro para os jornalistas. Para ele, “os jornalistas necessitam ser especialistas em dados”⁹.

Já o professor da Birmingham City University, Paul Bradshaw, acredita que a resposta simples para o que é *data journalism*, mas errada, poderia ser

8. The Data Journalism Handbook . European Journalism Centre. Retirado de http://datajournalismhandbook.org/1.%200/en/introduction_4.html

9. Charles, A. (2010, novembro) Analysing data is the future for journalists, says Tim Berners-Lee. The Guardian. Retirado de <http://www.guardian.co.uk/media/2010/nov/22/data-analysis-tim-berners-lee>



que é “Jornalismo produzido com dados”. Entretanto, dois termos, Jornalismo e dados, no entender do pesquisador são “problemáticos”.

Algumas pessoas pensam que "dado" como qualquer coleção de números, muito provavelmente obtidos em uma planilha. Há 20 anos, era a única forma dos jornalistas trabalharem com dados. Mas nós vivemos em um mundo digital agora, num mundo no qual quase tudo pode ser - e quase tudo é - descrito em números¹⁰.

Jonathan Gray, da Open Knowledge Foundation, acredita que a definição *data journalism* deve ser ampla, pois é um termo recente para “descrever o conjunto de práticas que usam dados para melhorar as notícias”. No artigo “The future of data journalism”, Gray focaliza a atividade na questão das práticas. Segundo o seu entender, a definição passa por utilizar tecnologias digitais (banco de dados e ferramentas de análise desses dados) para elaborar melhores conteúdos jornalísticos, publicando conjunto de dados relevantes ao “lado das matérias e usar conjunto de dados para produzir visualização de dados interativa e aplicativos de notícias”¹¹.

No artigo “*In the age of big data, data journalism has profound importance for society*”, Alex Howard afirma que estamos vivendo na “era da cultura orientada por dados”.

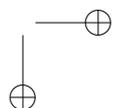
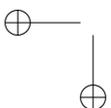
O jornalista especialista em dados pode usar a API do *Twitter* ou planilha de dados para encontrar notícias como prontamente o profissional pode usar o telefone para falar com uma fonte. Não somente isso, nós servimos muitos leitores que estão acostumados a lidar com dados todos os dias - contadores, educadores, pesquisadores e marqueteiros. Se nós quisermos capturar a atenção deles, nós precisamos falar a linguagem dos dados com autoridade¹².

Os profissionais da *British Broadcasting Corporation* (BBC), Bella Hurrell e Andrew Leimdorfer descrevem que o termo pode cobrir um leque de disciplinas e é usada de formas variadas nas organizações:

10. The Data Journalism Handbook . European Journalism Centre. Retirado de http://datajournalismhandbook.org/1.0/en/introduction_0.html

11. Gray, J. (2012, junho) The future of data journalism. Data Driven Journalism. Retirado de http://datadrivenjournalism.net/news_and_analysis/The_future_of_data_journalism

12. Howard, A. (2012, março) In the age of big data, data journalism has profound importance for society. Strata Making Data Work. Retirado de <http://radar.oreilly.com/2012/03/rise-of-the-data-journalists.html>



Permitir que o leitor descubra a informação que seja pessoalmente relevante; revelar uma história que é notável e até então desconhecida; ajudar o leitor entender melhor uma questão complexa essas categorias podem se sobrepor e em um ambiente online podem frequentemente se beneficiarem de algum tipo de visualização¹³.

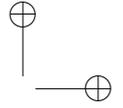
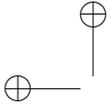
Outro grupo de mídia importante, *The Guardian*, estruturou o espaço virtual denominado *Datablog*¹⁴. Nele, conjuntos de dados produzidos pelo grupo e links para *Open Data* são disponibilizados para qualquer usuário. Também é possível acessar os produtos produzidos pelos profissionais do *The Guardian* tendo dados como base, como infográficos. Para a publicação, o *data journalism* representa a convergência de um número de campos do conhecimento que são importantes para a construção de conteúdo informativo de relevância social, possuindo como base o conjunto de dados: a pesquisa investigativa e estatística para design e programação. Para Paul Bradshaw, em artigo no *Datablog*¹⁵, a ideia de combinar essas habilidades é poderosa para contar histórias jornalísticas. Nas habilidades apontadas pelo pesquisador estão:

1. Encontrar dados: "Encontrar dados" pode envolver qualquer coisa desde possuir conhecimento de especialista e contatos para ser capaz de usar as habilidades para produzir reportagens através do computador ou utilizar o computador para ajudar nas habilidades, para alguns, possuir habilidades técnicas específicas como MySQL ou Python para reunir dados.
2. Análise de dados: Analisar dados significa a necessidade de possuir bom entendimento do jargão e o contexto mais amplo no qual os dados estão inseridos, mais a estatística - familiaridade com a planilha de dados pode ajudar a poupar muito do tempo.
3. Visualização de dados: Visualizar e converter dados têm sido historicamente responsabilidade de designers e programadores, mas com au-

13. The Data Journalism Handbook . European Journalism Centre. Retirado de http://datajournalismhandbook.org/1.0/en/in_the_newsroom_1.html

14. Datablog: facts are sacred. The Guardian. Retirado de <http://www.guardian.co.uk/news/datablog>

15. Bradshaw, P. (2010, outubro) How to be a data journalist. Strata Making Data Work. The Guardian. Retirado de <http://www.guardian.co.uk/news/datablog/2010/oct/01/data-journalism-how-to-guide>



mento do número de pessoas com experiência editorial tentando executar as duas tarefas, - particularmente por causa da ampliação da consciência do que é possível e particularmente pela diminuição de barreiras na experimentação dessas atividades.

4. Converter dados: Ferramentas tais como ManyEyes para visualização e Yahoo! Pipes para fusão de dados têm sido úteis para obter dos estudantes de Jornalismo rapidamente o entendimento de suas possibilidades.

Na visão de Jonathan Stray, o *data journalism* é “obter, reportar, ser curador e publicar dados de interesse público”¹⁶. Ele enquadra a atividade como sendo um dos braços do Jornalismo Computacional, pois utiliza as ferramentas digitais para produzir o efeito desejado. *Nesse ponto* Liliana Bounegru, reafirma a importância do uso de tais ferramentas, pois inserem o *data journalism* no ecossistema de artefatos e práticas que brotam em sites de dados e serviços, “pois as tecnologias digitais e a web estão fundamentalmente modificando a forma que a informação é publicada”¹⁷

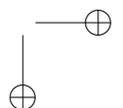
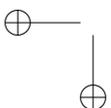
Novas fontes em função da informação democrática

Produzida mais intensamente nas últimas décadas do século XX, a consolidada teoria sobre o Jornalismo possui muito do seu embasamento referenciado na prática jornalística realizada em veículos de mídia impressos e eletrônicos. Além dessa abordagem, considera o Jornalismo como prática quase exclusiva de jornalistas que trabalhavam em organizações jornalísticas. Vide a bibliografia, quase inexistente no Brasil, sobre práticas e produção do Jornalismo por free-lancers. Nesse contexto, o conceito de fontes jornalísticas foi dividido entre pessoais ou documentais. Para Gomis (1991), as fontes são pessoas, são grupos, são instituições sociais, ou são vestígios — discursos, documentos, dados — por aqueles deixados ou construídos.

Entretanto, os dados mencionados por Gomis, na atualidade, se transformaram em digitais, podem ser acessados via sistema de dados abertos na Web

16. Stray, J. (2011, abril) A computational journalism reading list. Sítio de Jonathan Stray. Retirado de <http://jonathanstray.com/a-computational-journalism-reading-list>

17. The Data Journalism Handbook . European Journalism Centre. Retirado de http://datajournalismhandbook.org/1.0/en/introduction_4.html



e atingiram um volume impossível de mensurar, onde as dimensões são pensadas em terabyte ou petabyte. Agora, eles podem ser pesquisados, relacionados e produzir visualizações por meio de tecnologias conectadas via redes telemáticas. Os conteúdos produzidos a partir do manejo com destreza das fontes, contidas na Web de dados, adquiriram um novo patamar no processo de produção do Jornalismo.

As fontes são mais importantes para o processo de produção de notícias que o próprio jornalista, uma vez que são elas que oferecem a matéria-prima da notícia. Além disso, o jornalista renomado constrói sua carreira em função da rede de relações que mantém com suas fontes. (Lima, 2007)

A relação, entre o jornalista e as fontes, está sendo ampliada devido à profusão de novas conexões entre o profissional e os conjuntos de dados digitais. Para Mirko Lorenz, da publicação alemã *Deutsche Welle*, as múltiplas fontes favorecerem a filtragem de dados em redes sociais, por exemplo, trazendo à tona muitas informações importantes que são ignoradas pelos métodos tradicionais de apuração.

Por isso, o *data journalism* é tão importante. Obtendo, filtrando e visualizando o que está acontecendo além do que os olhos podem ver, tem um crescimento de valor. O suco de laranja que você consome de manhã, o café que você prepara - na economia global de hoje possui conexões invisíveis entre esses produtos, outras pessoas e você. A linguagem dessa rede é o dado: pequenos pontos de informação que não são frequentemente relevantes sozinhos, mas massivamente importantes quando vistos do ângulo certo¹⁸.

Obter, filtrar e dispor para visualização informações que estão “escondidas” do público, mas muitas das vezes em poder de um pequeno grupo de pessoas, que trabalham em empresas privadas, governos, instituições etc. Portanto, o *data journalism* é importante para a sociedade, pois na velocidade que grande quantidade das informações são transmitidas na atualidade, a função do *data journalism* também é combater a assimetria da informação, ou seja, que a informação estruturada seja transmitida para um número cada vez maior de pessoas.

No contexto do Big Data, o movimento denominado *Open Data* trabalha para que governos e instituições estatais disponibilizem os seus dados de

18. The Data Journalism Handbook . European Journalism Centre. http://datajournalismhandbook.org/1.0/en/introduction_1.html

forma aberta e de fácil acesso na Web, para que possam ser trabalhados pelos cidadãos, assim ajudando a fiscalização das ações dos governos, principalmente no que tange aos gastos públicos.

A fim de manter o governo responsável por suas ações, cidadãos devem conhecer sobre essas ações. Para isso acontecer, eles devem insistir que o governo atue de forma aberta e transparente possível. No século XXI, isso é colocar os dados disponíveis de forma online e com fácil acesso. Se o dado do governo é colocado disponível em formatos úteis e flexíveis, cidadãos podem ser capazes de utilizar modernas ferramentas em software livre para lançar luz sobre as atividades do governo. Tais ferramentas inclui *mashups*, que realçam conexões escondidas entre diferentes conjuntos de dados e o *crowdsourcing*, que faz o leve trabalho de peneiras através de montanhas de dados por intermédio de milhares de olhos sobre um conjunto particular de dados. (Brito, 2008)

O *data journalist* atua nessa área produzindo informações estruturadas a partir do cruzamento de dados através de bases de dados online fornecidas por governos, como *data.gov*¹⁹, dos EUA.

Para Vivek Kundra, ex-chefe do Governo Eletrônico estadunidense, *Open Data* é uma valiosa contribuição de aplicação no campo da democracia, pois ajuda a combater “a corrupção governamental, melhorar a responsabilidade e melhorar os serviços do governo; mudar para um governo aberto, transparente e participativo”. Nesse ambiente pode se criar novos modelos de Jornalismo fornecendo informação de relevância social através de conjuntos de dados abertos.

Ao permitir que qualquer pessoa se aprofunde em fontes de dados e encontre informação que é relevante para ele, bem como verificar as afirmações e desafiar as suposições comumente recebidas. O *data journalism* efetivamente representa a democratização em massa das ferramentas, técnicas e metodologias que eram anteriormente utilizadas por especialistas (repórteres investigativos, cientistas sociais, estatísticos, analistas e outros especialistas)²⁰.

Diferente do modelo adotado tradicionalmente para o exercício da profissão, com o ensino sendo obtido através de curso universitário e a prática realizada nas organizações midiáticas, o *data journalism* se expande na Web e

19. DataGov. Retirado de <http://www.data.gov/>

20. The Data Journalism Handbook . European Journalism Centre. http://datajournalismhandbook.org/1.0/en/introduction_4.html

pode ser apropriado por qualquer usuário que deseja aprendê-lo. Pertencendo ao movimento de cultura livre (*Open Source*), as práticas podem ser adquiridas por intermédio de cursos livres online, websites especializados em ensinar e debater técnicas, encontros denominados de *Hack Day* e concursos, como o realizado pela Mozilla²¹ - *Knight-Mozilla OpenNews*, que possui a *Knight Foundation* como parceira.

Knight-Mozilla OpenNews é sobre construir um ecossistema para ajudar o Jornalismo obter sucesso na Web aberta. É sobre produzir uma nova geração de soluções na Web que resolvam problemas reais nas notícias. É sobre ajudar comunidades de desenvolvedores e jornalistas como eles fazem, aprendem e inventam juntos. É sobre organizar bolsistas e códigos.

O *Hack Day* é uma atividade, geralmente, que possui a duração de 24 horas e propõe aos participantes (sozinhos ou organizados em equipes) que realizem um projeto tendo como base um desafio. Os encontros podem ser realizados por grupos, como o *Hacks/Hackers*²² ou por grupos de mídia²³, que neste caso fornecem sua base de dados para o desenvolvimento de produtos a partir delas. Em função de existir encontros de hackers há muitos anos, com público formado por programadores/desenvolvedores, com a possibilidade de produzir conteúdo informativo de relevância social por intermédio de base de dados abertas na Web emergiu o termo *hacker journalist*, que pode ser considerado sinônimo de *data journalist*.

Considerações finais

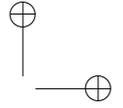
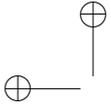
A atual sociedade se apropria das Tecnologias de Comunicação e Informação como nunca na sua história. As inovações tecnológicas possibilitaram o barateamento e o fácil uso de diversas plataformas digitais, muitas destas fazendo parte do cotidiano, seja no campo pessoal ou profissional.

A expansão vertiginosa da TIC's também atingiu o setor de produção de conteúdo informativo de relevância social. O Jornalismo está tentando se

21. Knight-mozilla OpenNews. Retirado de <http://www.mozillaopennews.org/>

22. Hacks/Hackers. Retirado de <http://hackshackers.com/about/>

23. (December, 2011) Hack Day. New York Times. Retirado de <http://www.nytimes.com/marketing/timesopen/hackday.html>



adaptar às transformações resultantes da configuração tecnológica digital, que modifica a forma que a informação é obtida, filtrada e distribuída.

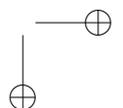
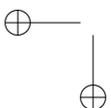
Assim, o Jornalismo enfrenta uma crise quando confronta o Modelo Padrão de Jornalismo Tradicional com atual configuração tecnológica proporcionada pela Internet, que está sendo potencializada pelo aumento de velocidade de transmissão de dados, pela evolução das máquinas computacionais (aumento de capacidade de processamento e armazenamento de dados), com a multiplicação e consolidação de linguagens de programação cada vez mais amplas, permitindo transacionar dados de diversas formas.

Nessa estrutura tecnológica, que emerge o *Big Data*, a atuação profissional do Jornalismo também deve possuir outras configurações. Ele deve adquirir habilidades técnicas/tecnológicas que proporcionem transformar-se em data jornalista. Entre algumas das suas especialidades estão: encontrar dados em bancos e documentos digitais disponíveis na Web; capacidade de analisar dados através de softwares para tal fim e produzir visualização de dados que foram obtidos e relacionados.

Esse conjunto de práticas, que usam dados para melhorar as notícias, proporciona a descoberta de novas fontes, fatos jornalísticos e angulações, que ajudam na disseminação de informações relevância social, fortalecendo o desenvolvimento de uma sociedade mais justa e igualitária, portanto, mais democrática.

Referências

- Brito, J. (2008) Hack, mash, & peer: crowdsourcing government transparency. *The Columbia Science and Technology Law Review*, IX, 119 – 157.
- Cohen, S., Hamilton, J. T.& Turner F. (2011). Computational Journalism. *Communications of the ACM*, 54(10), 66-71.
- Floridi, L. (2010). Information: a very short introduction. New York: Oxford University Press.
- Fuller, J. (2011). What is happening to news: the information explosion and the crisis in Journalism. Chicago: The University of Chicago Press.
- Gomis. L. (1991). Teoria del periodismo; como se forma el presente. Barcelona: Paidós.



- Hamilton, J. T. & Turner, F. (2009). Accountability Through Algorithm: Developing the Field of Computational Journalism. *A report from Developing the Field of Computational Journalism*, a Center For Advanced Study in the Behavioral Sciences Summer Workshop, July 27-31. Recuperado em 20 de outubro de 20112, de <http://www.stanford.edu/~fturner/Hamilton%20Turner%20Acc%20by%20Alg%20Final.pdf>
- Lima, W. T., Junior (2007). Bases conceituais visando a criação de UML (Unified Modeling Language) para pesquisa e validação qualitativa de fontes de informação jornalística (*Relatório de Pós-doutorado*), São Bernardo do Campo, Programa de Pós-graduação em Comunicação Social, Universidade Metodista de São Paulo.
- Royal, C. (2010). The Journalist as Programmer: a Case Study of The New York Times Interactive News Technology Department. *Anais do International Symposium in Online Journalism*, The University of Texas at Austin, Austin, TX, USA.
- Vianna, R. P. A. (1992). *Informatização da imprensa brasileira*. São Paulo: Edições Loyola.
- Wing, J. M. (2006). Computational Thinking. *Communications of the ACM*, 49(3), 33–35.