

Do RAC ao JGD: evolução computacional das práticas de investigação jornalística por meio de dados¹

From CAR to DDJ: a computational evolution of practices of investigative journalism using data

André Rosa de Oliveira ²

Resumo: Há uma profusão de conceitos e siglas relacionados ao uso de computadores como aliados à produção de reportagens investigativas – destaca-se, em especial, Reportagem Assistida por Computador. Longe de esgotar este escopo, o artigo pretende traçar uma linha do tempo, desde a informatização das práticas jornalísticas até a explosão da Web, conectando o pioneirismo de Philip Meyer, chegando ao termo Jornalismo Guiado por Dados (tradução de *Data-driven Journalism*). O texto busca ainda reforçar os desafios do profissional num contexto pautado pelas práticas contemporâneas do Jornalismo Computacional.

Palavras-Chave: Jornalismo de precisão, Pensamento computacional, Reportagem Assistida por Computador, Jornalismo Guiado por Dados.

Abstract: There is a profusion of acronyms and concepts related to the usage of computers as partners to producing investigative stories - like Computer-Assisted Reporting. This paper intends to trace a timeline from the informatization of journalistic practices until the explosion of the Web, through the pioneering of Philip Meyer, getting to the term Data-driven Journalism. The text also seeks to reinforce the professional challenges in an environment guided by contemporary practices of Computational Journalism.

Keywords: Precision Journalism, Computational thinking, Computer-Assisted Reporting, Data-driven Journalism.

.....

¹ Trabalho apresentado no I Seminário de Pesquisa em Jornalismo Investigativo, realizado na Universidade Anhembi-Morumbi, cidade de São Paulo, entre 24 e 26 de julho de 2014.

² Jornalista. Docente nos cursos de Comunicação Social das Faculdades Integradas Rio Branco. Doutorando pela Universidade Metodista de São Paulo. E-mail: andrerosa.jor@gmail.com.

1. Investigação jornalística e as muitas faces da informação

Pautar, apurar, checar, redigir, editar e distribuir. Em poucas palavras e de forma simples, é possível explicar o caminho para a produção de uma notícia, base para o desenvolvimento de teorias e métodos jornalísticos. Com a mesma simplicidade, costuma-se dizer que os computadores, traduzidos em processos de informatização presentes nas mais distintas áreas, modificou a forma de se fazer jornalismo. Também é comum afirmar que, em se tratando de processos contemporâneos e em constante evolução, é fácil se perder em um labirinto de ideias futuristas – ou, pior, manter-se fixo ao encontrar um lugar envolvente.

Este artigo propõe observar a evolução dos meios digitais, especialmente a presença cada vez maior de computadores e sua interface amigável (a Web) em veículos de comunicação e escolas de jornalismo, buscando fixá-la em uma rotina de trabalho profissional própria de um tipo de jornalismo especializado: o investigativo, enraizado em um método centenário (QUESADA, 2004, p. 125). Esta discussão, envolvendo essencialmente computadores e bases de dados, pode ser balizada a partir de duas expressões, traduzidas pelas siglas RAC e JGD – mais populares em suas versões em inglês, CAR e DDJ. Respectivamente, Reportagem Assistida por Computador (*Computer-assisted Reporting*) e Jornalismo Guiado por Dados (*Data-driven Journalism*).

Diferente de outros processos rotineiros de produção noticiosa, o jornalismo investigativo, entendido aqui como uma especialização possível, procura denunciar pessoas, instituições ou empresas cujas atividades atentam contra o interesse público (QUESADA, 2004, p. 125). Investigar significa contrastar toda a informação que se manuseie, como fontes documentais; verificar todos e cada um dos dados e, especialmente, trabalhar sustentando uma sólida ética profissional (IDEM, p. 127). Há registro do jornal britânico *The Guardian* ter apontado o número de alunos matriculados nas escolas de Manchester, relacionando-o com os custos totais, sendo possível identificar quantos recebiam educação gratuita. Isso foi em 1821 (GRAY; BOUNEGRU; CHAMBERS, 2012).

A relação entre o jornalismo investigativo e a manipulação de dados por meio de computadores, ainda que seja bem próxima, não é direta: nem toda investigação jornalística que resulta em reportagens do gênero aplica metodologias e conhecimentos das ciências sociais. Tanto que o método se fortaleceu nos Estados Unidos ao final do Século XIX, mais

de cem anos antes de um computador ter sido usado para apoiar uma cobertura – o ano era 1952, quando a CBS previu os resultados das eleições presidenciais. A partir daí, no entanto, máquinas passaram a auxiliar o jornalista a simplificar, a tornar mais rápido ou a expandir seu trabalho. A proximidade entre processamento rápido de dados e ferramentas computacionais se fortaleceu a ponto de se tornar uma combinação essencial. Da mesma forma, a imprensa que se consolida nos EUA, e cuja influencia chega ao Brasil, assume a sua função social de fiscalização (*watchdog*), respaldada por esse tipo de reportagem.

Investigação jornalística demanda informação relevante e consistente, e aqui cabe uma ressalva importante a partir do significado de "informação", palavra que pode assumir muitos significados, dependendo de seu contexto. Aos olhos da ciência, o termo informação só passou a fazer sentido quando deixou de ser vaga, abstrata, imprecisa, passando a ser mensurada. Foi Claude Shannon, pai da teoria da informação, que cunhou um neologismo, a partir de *binary digits*: bits. Para ele, o problema fundamental da comunicação é reproduzir exatamente ou com a máxima similaridade um tipo de mensagem entre um ponto e outro, independente de seu significado. Isso torna difícil um conceito único de informação seja satisfatório diante de numerosas aplicações possíveis (FLORIDI, 2010).

Informação é um labirinto conceitual. Em linhas gerais, podemos definir informação em termos de dados e significado: consiste em dados estruturados, organizados a partir de um campo semântico, uma sintaxe pré-definida (um código ou linguagem) e dependência do ambiente onde se encontra (FLORIDI, 2010, p. 21). Na computação, o termo possui significados diversos a partir do seu uso (recuperação ou armazenamento de informação), e nesse sentido, os dados são binários, capazes de serem processados por computadores.

Ao perpassar o jornalismo, no entanto, a palavra adquire contornos a partir da forma como é fabricada: na visão de Adelmo Genro Filho (1987), além de "zeros e uns" contextualizados, existem procedimentos, com vistas a objetivos específicos, pautados por relações sociais produzidas pelo advento do capitalismo e dos meios industriais de difundir informações. Mesmo em reportagens mais elaboradas, esta variável não pode ser desprezada.

A notícia jornalística não pode ser considerada como uma modalidade da informação em geral. Não foi a transmissão genérica da experiência – o que sempre ocorreu em sociedade – e sim a transmissão sistemática, por determinados meios técnicos, de um tipo de informação necessária à integração e universalização da

sociedade, a partir da emergência do capitalismo, que deu origem à notícia jornalística. (GENRO FILHO, 1987)

2. Do RAC ao Jornalismo de Precisão

Por conta das práticas rotineiras, dando conta de interesses fundamentais do veículo e no imediatismo como o ponto de chegada – e não como partida, como ocorre na ciência, o método científico pouco tem a ver com o jornalístico. É por meio das técnicas de jornalismo investigativo, no entanto, que eles se aproximam. O uso de métodos de pesquisa social para relatar temas sociais objetivamente, juntando fatos aparentemente desconexos revelando seus antecedentes, impulsionou a Reportagem Assistida por Computador a partir do que convencionou chamar-se Jornalismo de Precisão – do inglês, *Precision Journalism*.

Antes de começar a usar computadores para investigar histórias e se tornar pioneiro deste método, o jornalista Philip Meyer, ainda um jovem repórter do *Miami Herald*, usou lápis e papel para criar uma base de dados, relacionando companhias de seguros e candidatos a presidência do conselho de funcionários de escolas na comunidade. O ano era 1959. Treze anos depois, em 1972, lançou seu livro sobre o uso de práticas científicas próprias das ciências sociais.

Foi o professor Everette Dennis, diante de seus alunos da Universidade de Oregon, quem empregou o termo *Precision Journalism* pela primeira vez, em 1971 – o mesmo foi adotado por Meyer logo depois. O surgimento do método coincidiu com o caso Watergate, escândalo político descoberto pelo *The Washington Post* e sinônimo de investigação jornalística, bem como o surgimento de uma fundação denominada *Investigative Reporters and Editors* – IRE³, que por sua vez, deu origem ao *National Institute for Computer Assisted Reporting* (NICAR), responsável pelo treinamento de profissionais.

Inicialmente, o jornalismo de precisão se entende como outra forma de jornalismo investigativo, uma forma em que as novas tecnologias desempenham um papel decisivo. Surgiu como uma tentativa de se trazer rigor científico ao jornalismo investigativo, desenvolvido principalmente para o uso de bancos de dados e de processamento por computador. (GALINDO ARRANZ, 2004, p. 99, tradução nossa⁴).

³ <http://www.ire.org>

⁴ Versão original: "Inicialmente, el periodismo de precisión se entiende como otra forma de hacer periodismo de investigación, una forma en la que las nuevas tecnologías juegan un papel decisivo. Surgido como un intento de aportar rigor científico al periodismo de investigación, se desarrolla fundamentalmente con el uso de los bancos de datos y su tratamiento informatizado."

Mesmo tendo completado quarenta anos, não se trata de uma prática popular no Brasil. Referência em técnicas de reportagem assistida por computador no país, José Roberto de Toledo, de O Estado de S. Paulo e um dos fundadores da Associação Brasileira de Jornalismo Investigativo (Abraji⁵), observa que estamos mal preparados para o excesso de informações: "a gente se vê surpreendido pelos fatos em vez de surpreender com fatos" (OLIVEIRA, 2010).

Uma explicação para esta percepção, corroborada por Fermin Galindo Arranz (2004, p. 102) pode estar nas palavras do pesquisador alemão Michael Kunczik. Ele toma emprestada a visão de Meyer, segundo a qual o jornalista precisa gerenciar, processar, analisar e interpretar bancos de dados, e isso requer treinamento intensivo nas técnicas. Existe um grande perigo em produzir informação incorreta em função da má interpretação. Assim, das duas, uma: ou a investigação sócio-científica não pode ser tarefa do jornalismo; ou é preciso melhorar de forma geral o nível de conhecimentos sócio-científicos básicos dos jornalistas (KUNCZIK, 1997, p. 104).

Outra explicação, observada por Lima Junior (2012, p. 209), está no atraso das adoções tecnológicas nas redações brasileiras. Enquanto os jornais norte-americanos como *USA Today*, *The Washington Post*, *Los Angeles Times* e *The New York Times* desenvolveram departamentos internos com especialistas em informática e estatística, atuaram em conjunto com profissionais e tiveram suas reportagens premiadas com o prêmio Pulitzer, a *Folha de S. Paulo* foi o primeiro jornal brasileiro a introduzir terminais na redação, já em 1983.

Atualmente, como lembra o pesquisador, houve um barateamento das máquinas computacionais e de outros dispositivos tecnológicos, ainda que o atraso persista. Mais do que isso, o atual ambiente computacional conectado faz emergir novas possibilidades no tratamento de informações e dados – e isso não se aplica apenas ao jornalismo. FLORIDI (2010, p. 12) constata um atual processo de reavaliação: de uma visão materialista, em que objetos físicos representam peça-chave, o homem assume ainda um papel metafísico, informacional. Mesmo a biologia, que se transformou a partir do microscópio, se tornou um tipo de "ciência da informação", visto que a genética também lida com informação organizada e intercomunicação de células.

⁵ <http://www.abraji.org.br>

Este cenário, construído em parte graças ao desenvolvimento da Web, não permite outro caminho: há necessidade de dominar tecnologias digitais mais profundamente.

3. Da World Wide Web ao JGD

Ao lado do já mencionado Claude Shannon, o cientista norte-americano Vannevar Bush, um dos pesquisadores-líderes durante a Segunda Guerra Mundial, identificou o crescimento exponencial da informação já em 1945 e propôs uma solução conceitual: uma máquina chamada *Memex*, capaz de associar e recuperar ideias. Anos mais tarde, o físico britânico Tim Berners-Lee bebeu das ideias de Vannevar Bush, bem como Ted Nelson, que criou a expressão "hipertexto" em 1965; Paul Baran, idealizador do modelo de rede distribuída que deu origem à Internet; e Vint Cerf, que programou o protocolo TCP/IP, permitindo a qualquer dispositivo computacional, outrora incompatíveis, se conectar à esta rede.

Assim, enquanto trabalhava na Organização Europeia para a Pesquisa Nuclear (em francês, CERN) e sentia a necessidade em organizar informação perdida, em função da rotatividade de pesquisadores, desenvolveu um sistema descentralizado, onde fragmentos de texto seriam conectados, num sistema universal, abrangente, flexível e colaborativo. Escreveu os primeiros programas que funcionariam numa infraestrutura de comunicações baseada em computadores conectados, e definiu os parâmetros associado aos acrônimos URL (identificadores universais de recursos), HTTP (protocolo de transferência de hipertexto) e HTML (linguagem para marcação de hipertexto). Assim, uma vez que alguém, em qualquer lugar, disponibilizasse um documento, este poderia ser acessível por qualquer pessoa, em qualquer tipo de computador, sendo possível ainda fazer uma referência – um link – a este item, para que outros possam encontrá-lo. Estas são as únicas premissas da World Wide Web, que se espalhou pelo mundo a partir de 1991 (BERNERS-LEE, 2000, p. 36).

Ainda que tenha sido implantada por uma única pessoa, a gênese da Web envolve muitas mentes, com sonhos em comum. Sua construção permanece coletiva, baseada em regras e códigos simples, mas que adquiriu ares de complexidade diante das múltiplas apropriações. Esse conjunto de dados acumulados a cada instante por serviços públicos, portais de transparência governamentais, sites de relacionamento (como *Facebook* ou

Twitter) ou mesmo repositórios documentais controversos como o *WikiLeaks* compõe um lugar cujo tamanho vai além da habilidade comum em capturar, gerenciar e analisar, dando origem ao conceito de *Big Data* (LIMA JÚNIOR, 2012, p. 211).

Soma-se a este volume dados a popularização da Web, bem como o desenvolvimento de interfaces e ferramentas mais amigáveis. Num contexto bem diferente ao vivenciado por Philip Meyer, aplicações como *OpenRefine*⁶, que permite limpar os dados antes de avançar para análises, ou *Google Fusion Tables*⁷, plataforma para visualização de grandes conjunto de dados, faz emergir novos métodos, fontes e caminhos para descobrir, apresentar, personalizar, agregar e arquivar histórias, conectando comunidades com a informação que elas precisam. Há casos como o próprio *The Guardian*, que nos últimos anos se tornou uma plataforma aberta, encorajando o uso e reuso de seus recursos (DANIEL; FLEW, 2010).

Tais práticas, associadas ao uso de técnicas de computação nas redações desde a coleta de dados, sua curadoria e finalmente sua visualização, podem ser sintetizadas no conceito de Jornalismo Computacional:

Com um conjunto de ferramentas para reportagem, um jornalista terá capacidade para digitalizar, transcrever, analisar e visualizar padrões em documentos. A adaptação de algoritmos e tecnologia, combinada com ferramentas livres e de código aberto, vai nivelar o campo de jogo entre interesses poderosos e o interesse público, ajudando a descobrir ligações e evidências que podem desencadear investigações. Estas mesmas ferramentas também podem ser usadas por grupos de interesse público e cidadãos preocupados. (COHEN; HAMILTON; TURNER, 2011, p. 71, tradução nossa⁸)

Um braço do Jornalismo Computacional começou a ser delineado a partir de 2006, quando Adrian Holovaty publicou em seu blog o texto *A fundamental way newspaper sites need to change*⁹. Fundador do *EveryBlock*, site que relacionava informações de acordo com bairros, Holovaty argumentava que, além da narrativa, a informação jornalística deveria contar com outros dados estruturados, capazes de serem compreendidos por máquinas. Mais

⁶ <http://openrefine.org>

⁷ <http://tables.googlelabs.com>

⁸ Versão original: "With a suite of reporting tools, a journalist will be able to scan, transcribe, analyze, and visualize the patterns in these documents. Adaptation of algorithms and technology, rolled into free and open source tools, will level the playing field between powerful interests and the public by helping uncover leads and evidence that can trigger investigations by reporters. These same tools can also be used by public-interest groups and concerned citizens."

⁹ <http://www.holovaty.com/writing/fundamental-change>

tarde, em 2009, o *The Guardian* publicou documentos que traziam despesas de parlamentares britânicos, em resposta ao escândalo revelado pelo concorrente *The Daily Telegraph*. O diário contou com a ajuda dos leitores para identificar os gastos com dinheiro público, num caso que se revelou uma oportunidade para o jornalismo digital (DANIEL; FLEW, 2010).

A partir destes e de outros casos, profissionais participaram de uma conferência, em agosto de 2010, organizada pelo *European Journalism Centre*, a partir da ação da pesquisadora Liliana Bounegru. O encontro, realizado em Amsterdã, ratificou a expressão *Data Journalism*, de onde deriva a expressão *Data-driven Journalism*¹⁰ (DDJ) (GRAY; BOUNEGRU; CHAMBERS, 2012), expressão traduzida para o português como Jornalismo Guiado por Dados (TRÄSEL, 2013). A expressão descreve um conjunto de práticas que usam dados para melhorar as notícias, desde o tratamento, cruzamento e recuperação de dados até a geração de visualizações e infográficos atraentes, independente da plataforma a ser publicada.

4. Relação entre RAC, JGD e o futuro

Este trabalho começa, logo no título, com uma provocação: as siglas RAC e JGD, derivadas das expressões *Precision journalism* e *Data journalism*, estão na mesma raiz: a necessidade do jornalismo em aprimorar seus métodos de investigação, encorajando estudos interdisciplinares. As duas linhas, no entanto, nascem em contextos distintos a partir de incursões feitas por jornalistas interessados em tecnologia – seja ao final dos anos 1960, com a introdução dos computadores e planilhas eletrônicas, seja a partir dos anos 2000, com a profusão de dados e ferramentas por meio da Web.

Esta distinção de contextos chega a provocar um debate: como a comunidade de pessoas autodenominadas "*data journalists*" se constituiu em plena efervescência da Web, há quem argumente de que não se trata de uma continuidade das práticas anteriores, mas sim um processo completamente diferente. Para Liliane Bonegru, mais importante do que debater a contemporaneidade das práticas, "uma posição mais frutífera seria considerá-lo como parte de uma longa tradição, mas respondendo a novas circunstâncias e condições" (GRAY; BOUNEGRU; CHAMBERS, 2012).

¹⁰ <http://datadrivenjournalism.net>

É o que pensa o jornalista Mark Briggs, que também relaciona JGD com RAC: ele recorda que o computador e os métodos de investigação acompanham o jornalismo há décadas, mas seu potencial é pouco explorado enquanto restrito ao formato impresso. Na sua visão, todo jornalista agora é um *web worker*, e neste cenário, as técnicas podem ser aprofundadas, customizadas, rastreáveis e ter vida longa (BRIGGS, 2010, p. 255).

A evolução destes processos não escondem as dificuldades que sempre acompanharam o jornalismo: o acesso às fontes. Tanto no passado quanto no presente, o desafio de driblar silos informativos em favor de histórias transparentes permanece. Na Espanha, este obstáculo deu origem ao termo "jornalismo de filtragens" ("*periodismo de filtraciones*") (QUESADA, 2004, p. 129), prática perigosa de quem direciona informações ou documentos, ocultando interesses que o jornalista desconhece e, por consequência, se vê distante da credibilidade.

Dessa forma, o jornalista (e em especial quem optar por este caminho especializado) precisa conciliar sua capacidade criativa, técnicas de investigação, tratamento e análise de dados e conhecimento tecnológico para se transformar – e assim como a Web, o caminho é duro. Nos últimos anos, Tim Berners-Lee, ao lado de muitos outros pesquisadores, vem se debruçando no que definiu como Web Semântica. Resumidamente, refere-se à capacidade de computadores reconhecerem dados. Uma organização, por exemplo, pode criar uma base denominada "*mean-diurnal-temperature*" (temperatura diária média), enquanto outra faz o mesmo, denominando-a "*daily average-temp*" – é a mesma informação, com nomes diferentes. Trata-se de um desafio: continuamente, bases de dados são produzidas e disponibilizadas em rede por diferentes grupos e empresas, com as mais variadas intenções.

Num futuro ainda incerto, "a Web Semântica, como a Web já faz, fará com que muitas coisas antes impossíveis pareçam óbvias" (BERNERS-LEE, 2000, p. 198). Com o jornalismo, não será diferente. Não se trata de se voltar totalmente aos computadores, como se fossem a solução dos problemas: eles podem ser capazes de fazer a maior parte do trabalho braçal necessário, auxiliando nossas mentes em, intuitivamente, encontrar saídas por meio do labirinto.

Referências bibliográficas

BERNERS-LEE, T. **Weaving the Web**. New York: HarperCollins, 2000.

BRIGGS, M. **Journalism Next**. Washington, DC: CQ Press, 2010.

COHEN, S.; HAMILTON, J. T.; TURNER, F. Computational journalism. **Communications of the Association for Computing Machinery**, v. 54, n. 10, p. 66-71, 1 out. 2011.

DANIEL, A.; FLEW, T. The Guardian Reportage of the UK MP Expenses Scandal: a Case Study of Computational Journalism. **Communications Policy and Research Forum**, n. November, 2010.

FLORIDI, L. **Information: a very short introduction**. New York: Oxford University Press, 2010.

GALINDO ARRANZ, F. Propuesta de periodización histórica y evolución conceptual del Periodismo de Precisión. **Estudios sobre el Mensaje Periodístico**, n. 10, p. 97-112, 2004.

GENRO FILHO, A. **O Segredo da Pirâmide: Para uma teoria marxista do jornalismo**. Porto Alegre: Tchê! Editora, 1987.

GRAY, J.; BOUNEGRU, L.; CHAMBERS, L. **The Data Journalism Handbook: How Journalists Can Use Data to Improve the News**. [S.l.] The Open Knowledge Foundation, European Journalism Centre, 2012.

KUNCZIK, M. **Conceitos de Jornalismo: Norte e Sul**. São Paulo: Edusp, 1997.

LIMA JUNIOR, W. T. Big Data, Jornalismo Computacional e Data Journalism: estrutura, pensamento e prática profissional na Web de dados. **Estudos em Comunicação**, n. 12, p. 207-222, 2012.

OLIVEIRA, A. R. DE. Aproximação necessária entre a academia e o mercado. **Revista CoMtempo**, v. 2, n. 2, p. 1-7, 2010.

QUESADA, M. Periodismo de investigación: una metodología para el periodismo especializado. In: FERNANDEZ DEL MORAL, J. (Ed.). **Periodismo Especializado**. Barcelona: Editorial Ariel, 2004. p. 123-143.

TRÄSEL, M. **Jornalismo guiado por dados: relações da cultura hacker com a cultura jornalística**. XXII Encontro Anual da Compós, 2013