

Comunicação e Ciências Empresariais

Jornalismo de dados: caracterização e fluxos de trabalho

Ana Pinto Martinho

ISCTE-IUL

ana.pinto.martinho@gmail.com

Resumo

O jornalismo de dados surge nos últimos anos como uma das grandes tendências no jornalismo, a nível mundial. Este artigo tem como objetivo clarificar o que é o jornalismo de dados na atualidade e quais são as práticas a ele associadas. Numa primeira parte deste artigo são apresentadas as raízes do jornalismo de dados, através de uma caracterização baseada em literatura diversa e na análise de exemplos. A segunda parte é focada na análise de modelos de fluxos de trabalho, publicados por vários especialistas, que mostram o processo de trabalho que leva à criação de trabalhos jornalísticos deste tipo.

Palavras-chave: Jornalismo de dados; Jornalismo Data Driven, Pirâmide de Jornalismo de Dados, Visualização de Dados.

Abstract

In the last years, data journalism emerges as one of the major trends in what regards journalism, world wide. This article aims to clarify what is data journalism today and what are the main practices associated with it. The first part of this article presents data journalism roots, characterizing it based on available literature and some example's analysis. The second part is focused on the analysis of workflow models, that show the working process that origins this kind of journalistic work, published by several specialists.

Keywords: Data Journalism, Data Driven Journalism, Data Journalism Pyramid, Data Visualization.

Introdução

A utilização de dados para o trabalho jornalístico é bastante mais remota do que pode parecer à primeira vista, principalmente numa altura em que o jornalismo de dados aparece muitas vezes como algo de novo, fruto do casamento entre a tecnologia e o jornalismo. No texto “Data journalism in perspective”⁷, Liliana Bounegru aponta 1821 como a data do primeiro exemplo de jornalismo de dados no “Guardian”:

“It is a leaked table of schools in Manchester listing the number of students who attended it and the costs per school. According to Rogers this helped to show for the time the real number of students receiving free education, which was much higher than what official numbers showed.” (Bounegru, 2012)

Os jornalistas sempre trabalharam com dados, eles são factos são dados e os factos são a base do trabalho jornalístico (Lorenz, 2010). Mas como se chegou ao jornalismo de dados praticado na

⁷ http://datajournalismhandbook.org/1.0/en/introduction_4.html

atualidade? O caminho tem sido feito de um misto de mudanças sociais, na tecnologia e no próprio jornalismo.

No início dos anos 70, no seu livro “Precision Journalism”⁸, Philip Meyer fazia a apologia da aplicação dos métodos de investigação das ciências sociais ao trabalho jornalístico, de forma a torná-lo mais preciso e objetivo. No “The New Precision Journalism”⁹, uma edição revista e aumentada do primeiro livro, Meyer defende que através destes métodos os jornalistas podem analisar e comparar dados com maior facilidade, tendo por base um conjunto de regras que sistematizam a sua investigação. Meyer fala de duas fases essenciais no trabalho de dados: a fase de *input*, em que os dados são recolhidos e analisados; e a de *output*, onde são preparados os dados para “entrar na mente do leitor” (Meyer, 1991). Citado por Bounegru (2012¹⁰), atualmente, Meyer põe mais a tónica na parte da análise e processamento dos dados:

“When information was scarce, most of our efforts were devoted to hunting and gathering. Now the information is abundant, processing is more important”.

Esta perspetiva vem marcar uma das grandes diferenças entre o que se passava no início dos anos 70 e a atualidade: a maior disponibilidade e facilidade de acesso aos dados. Se, por um lado, há mais dados disponíveis em formatos mais fáceis de trabalhar, também há um grande volume de dados ainda por disponibilizar¹¹.

O texto “A fundamental way newspaper sites need to change”¹²(2006), de Adrian Holovaty, faz uma das primeiras formulações daquilo que está mais próximo do que é hoje o jornalismo de dados. “So much of what local journalists collect day-to-day is structured information: the type of information that can be sliced-and-diced, in an automated fashion, by computers. Yet the information gets distilled into a big blob of text -- a newspaper story -- that has no chance of being repurposed” (Holovaty, 2006). É exatamente em 2006, ano em que Holovaty faz estas afirmações, que nasce um projeto que veio contribuir para uma maior projeção do jornalismo de dados, a WikiLeaks. Mas de facto, o projeto só se tornou mediaticamente relevante a partir do momento em que a sua informação “crua” foi “escrutinada” e tratada por jornalistas de órgãos de comunicação social de alguns dos maiores grupos de média a nível mundial. E tal começou a acontecer em 2010 com a disponibilização e mediatização dos chamados “Iraq War logs”¹³. A WikiLeaks acaba por ser o catalisador de muitas das discussões que se tornaram fulcrais para o jornalismo de dados. Questões como: que órgãos de comunicação social tinham jornalistas com competências para analisar aquele tipo e quantidade de dados; qual a legitimidade das fontes; que tipo de dados pode ser revelado, etc. O New York Times lança, em 2007 um trabalho que também ficaria na história do jornalismo de dados: “Fixing DC’s Schools”¹⁴. Como é explicado no jornal:

⁸ Esta é a denominação da primeira edição do Livro de Meyer, lançada em 1973.

⁹ Meyer relançou o livro com o nome “The new precision journalis”, a citação é do livro de 1991, disponibilizada online.

¹⁰ Primeiro capítulo do “Data Journalism Handbook”, sem página.

¹¹ Ver capítulo anterior no ponto 2.2.

¹² <http://www.holovaty.com/writing/fundamental-change/>

¹³ Um conjunto de mais 400 mil documentos sobre a guerra do Iraque.

¹⁴ <http://www.washingtonpost.com/wp-srv/metro/interactives/dcschools/#fullseries>

“In 2007, The Washington Post and Washingtonpost.com dispatched more than a dozen print and audio reporters, database analysts, videographers and photographers to investigate what was wrong with the school system in the nation's capital. The team spent months uncovering the true condition of the school system and its programs, identifying the huge obstacles to change that had defeated reforms and reformers in the past”.

Em 2009, o St.Petersbrug Times¹⁵, ganhava o prémio Pulitzer pelo seu trabalho “PolitiFact”¹⁶, uma iniciativa de *fact-checking* que, durante a campanha presidencial de 2008, mostrava se os políticos eram ou não verdadeiros nas suas afirmações, através da análise dos seus discursos. Este trabalho deu ao jornalismo de dados a força e o reconhecimento de um Pulitzer. No *Data Journalism Handbook*¹⁷ são citados mais alguns bons exemplos. Dois deles, lançados em 2010 por jornais norte-americanos, são muito interessantes: o projeto “Do no harm – Hospital Care in Las Vegas”¹⁸, do Las Vegas Sun; e o projeto “Government Employee Salary Database”¹⁹, do *Texas Tribune*. Segundo o *Data Journalism Handbook*, para realizar o trabalho “Do no harm – Hospital Care in Las Vegas”, os jornalistas do *Las Vegas Sun* tiveram de fazer um pedido de “dados públicos”²⁰ para aceder aos dados de todos os hospitais do estado do Nevada.

O trabalho realizado é muito exaustivo e pressupôs a análise de mais de 2.9 milhões de registos de faturação, que acabaram por revelar mais de 3600 casos de lesões, infeções e erros cirúrgicos que poderiam ter sido evitados.

“The journalists involved worked very hard to acquire and clean up the data. One of the journalists, Alex Richards, sent data back to hospitals and to the state at least a dozen times to get mistakes corrected,”

explicou Angélica Ramos Peralta, do jornal argentino La Nación²¹.

O trabalho teve inclusive reflexo na legislação do estado do Nevada, uma vez que originou seis novas leis. Um dos pontos a salientar no trabalho é que esta foi uma das primeiras vezes em que os dados originais foram disponibilizados para poderem ser consultados pelos leitores.

No caso do trabalho do *Texas Tribune*, “*Government Employee Salary Database*”, foi construída uma base de dados com os salários de 660 mil funcionários públicos, para que os leitores pudessem pesquisar os dados, criando os seus padrões de leitura. A base de dados permite pesquisar por nome, salário ou agência do governo. Neste caso, a novidade é a apresentação da informação num formato pouco habitual em jornalismo. Não há texto escrito pelo jornalista, é apresentada uma base de dados consultável e os leitores também podem descarregar uma folha de cálculo com os dados.

Outro marco importante de 2010 é a realização, a 24 de Agosto, da primeira mesa-redonda europeia sobre jornalismo de dados, intitulada: “*Data-driven Journalism: What is there to learn?*”. Organizada pela Universidade de Amsterdão e pelo European Journalism Centre (EJC), produziu um dos primeiros documentos em torno de uma discussão sistematizada sobre jornalismo de dados. O documento, com o mesmo título que a mesa-redonda, dá conta das intervenções no evento e contém

¹⁵ Hoje denominado “Tampa Bay Times”

¹⁶ <http://www.politifact.com/>

¹⁷ Um documento marcante na história do jornalismo de dados, (ver mais adiante neste capítulo) http://datajournalismhandbook.org/1.0/en/introduction_3.html

¹⁸ <http://www.lasvegassun.com/hospital-care/>

¹⁹ <http://www.texastribune.org/library/data/government-employee-salaries/>

²⁰ Nos EUA é possível fazer “pedidos de dados públicos” quando os dados públicos não estão disponíveis.

²¹ *Data Journalism Handbook* - http://datajournalismhandbook.org/1.0/en/introduction_3.html

alguns elementos chave para compreender o quadro e as expectativas face ao jornalismo de dados, em 2010.

Em 2012, o *European Journalism Centre* (EJC) e a *Open Knowledge Foundation*, lançam o “Data Journalism Handbook”²², um “manual” colaborativo sobre jornalismo de dados, que traz alguma luz sobre o que é o jornalismo de dados e apresenta o “estado da arte” um pouco por todo o mundo, com destaque para os Estados Unidos da América, parte da Europa e América do Sul. O *European Journalism Center* (EJC), reconhecendo a importância do jornalismo de dados, dedica-lhe área de investigação, sendo também responsável pelo site *Data Driven Journalism*.²³ Estas movimentações e acontecimentos fornecem-nos pistas acerca do “estado da arte” do jornalismo de dados. Será, então, o jornalismo de dados algo de novo? Para Bounegru (2012), esta discussão não é importante, embora ela ultrapasse uma boa parte da literatura sobre jornalismo de dados:

“Rather than debating whether data journalism is completely novel, a more fruitful position would be to consider it part of a longer tradition, but responding to new circumstances and conditions” (Bounegru, 2012).

Data journalism, database journalism ou data-driven journalism?

São três as nomenclaturas que aparecem mais recorrentemente quando se discute o que é jornalismo de dados: “data journalism”, “database journalism” e “data-driven journalism”. Alguns autores distinguem-nos, definindo o primeiro de forma mais abrangente, como o jornalismo que trata grandes quantidades de dados, o segundo aparece muitas vezes como o jornalismo centrado nos dados, por oposição ao terceiro que utiliza dados para “chegar” a histórias e enriquecê-las. Apesar destas três nomenclaturas serem as mais usuais, vamos focar-nos mais nas duas últimas, uma vez que muitas vezes a primeira é utilizada como uma moldura onde cabem as duas seguintes.

Holovaty (2006) defende que:

*“Newspapers need to stop the story-centric worldview. Conditioned by decades of an established style of journalism, newspaper journalists tend to see their primary role thusly: Collect information; Write a newspaper story. The problem here is that, for many types of news and information, newspaper stories don't cut it anymore”.*²⁴

Por oposição, Lorenz (2012) afirma: “data-driven journalism can be defined as a workflow, where data is the basis for analysis, visualization and – most importantly – storytelling”. Esta definição de Lorenz traz, por vezes, alguma confusão com outro tipo de jornalismo, nascido nos anos 90, e já com bastante trabalho feito, sobretudo nos Estados Unidos da América, o *Computer Assisted Reporting* (CAR)²⁵. Uma diferença entre estas duas formas de pensar o jornalismo de dados está no foco. O “*database journalism*” põe as bases de dados no centro do trabalho, como ferramenta essencial para disponibilizar produtos em formatos que tradicionalmente não estão ligados às práticas jornalísticas. Por exemplo, simuladores com atualizações em tempo real, filtros de dados dinâmicos, etc. Enquanto o “*data-driven journalism*” põe o foco nas histórias jornalísticas, assumindo-se como um processo que usa os dados numéricos, ou bases de dados, para chegar ao produto jornalístico final.

²² <http://datajournalismhandbook.org/1.0/en/>

²³ <http://datadrivenjournalism.net/>

²⁴ <http://www.holovaty.com/writing/fundamental-change/>

²⁵ Ver adiante, nesta seção.

Mas o jornalismo de dados está indiscutivelmente ligado a práticas jornalísticas como o jornalismo de investigação, o jornalismo de precisão, o jornalismo de profundidade, ou o *Computer Assited Reporting* (CAR), confundindo-se, por vezes com algumas delas. Para Sandra Crucianelli²⁶, aquilo que na atualidade se apelida de jornalismo de dados²⁷ envolve todas as prática citadas, podendo envolver mais ou menos cada uma delas, dependendo do tipo de matérias que estão a ser abordadas. O aparecimento do jornalismo de dados só é possível com o amadurecimento de algumas destas práticas jornalísticas. Esclarecendo um pouco cada uma delas mais facilmente se compreenderá esta relação, contribuindo para uma clarificação do que se entende por jornalismo de dados.

Para que um trabalho jornalístico possa ser considerado de investigação são necessárias três condições (Crucianelli, 2013): que trabalhe um tema de relevância social, que alguém tenha interesse em manter segredo sobre a matéria, e que pressuponha da parte do jornalista trabalho adicional. Um trabalho de jornalismo de dados também pode, por isso, ser jornalismo de investigação. Basta olharmos para o trabalho do Las Vegas Sun sobre os hospitais no Estado do Nevada para percebermos que se trata de um caso de relevância social, que as autoridades da saúde não gostariam que viesse a público e que pressupõe trabalho que vai para além das práticas quotidianas dos jornalistas. No caso do jornalismo de profundidade, o trabalho tem as mesmas premissas que o jornalismo de investigação, exceto a premissa do segredo que aqui não se aplica. O jornalismo de precisão²⁸ baseia-se na premissa de que os jornalistas devem utilizar os métodos de investigação usados nas ciências sociais, recorrendo à investigação qualitativa e à investigação quantitativa. E, por vezes, quando se faz jornalismo de dados também são aplicados alguns destes métodos.

Crucianelli define o CAR como a utilização de recursos computacionais para a recolha e análise de dados, chegando mesmo a afirmar que quando se fala do tratamento de grandes quantidades de dados já não se deve falar de CAR, mas sim de jornalismo de dados. Mas no *Data Journalism Handbook*, Bounegru traz mais alguma luz sobre a diferença entre os dois.

“Some argue that there is a difference between CAR and data journalism. They say that CAR is a technique for gathering and analyzing data as a way of enhancing (usually investigative) reportage, whereas data journalism pays attention to the way data sits within the whole journalistic workflow.”
(Bounegru, 2012)

Aparece-nos então aqui o jornalismo de dados ligado ao processo de *workflow*, marcando-se a diferença entre esta forma de jornalismo e outras formas, através do *workflow*, através da forma como se trabalha. Face a estas formas de fazer jornalismo, quais os elementos diferenciadores do jornalismo de dados? Crucianelli (2013) salienta que o jornalismo de dados vem acrescentar novos elementos à forma como é feito o jornalismo: tratamento de grandes volumes de dados; visualização interativa desses dados; utilização da programação para desenvolver tarefas que vão desde a pura extração de dados, à limpeza dos *datasets* ou ao desenho de aplicações digitais de notícias (*news apps*). Para além destes elementos, que Crucianelli enumera, há a acrescentar um muito importante: a disponibilização dos dados crus utilizados para levar a cabo o trabalho jornalístico. Como forma de identificar mais facilmente o que serão produtos do jornalismo de dados, Crucianelli (2013) enumera “produtos” que emergem desta prática: artigos baseados em dados; visualizações interativas; conjuntos de dados abertos; aplicações digitais de notícias.

²⁶ Citando documento do MOOC sobre “Jornalismo de dados”, do *Knight Center for Journalism in the Americas*.

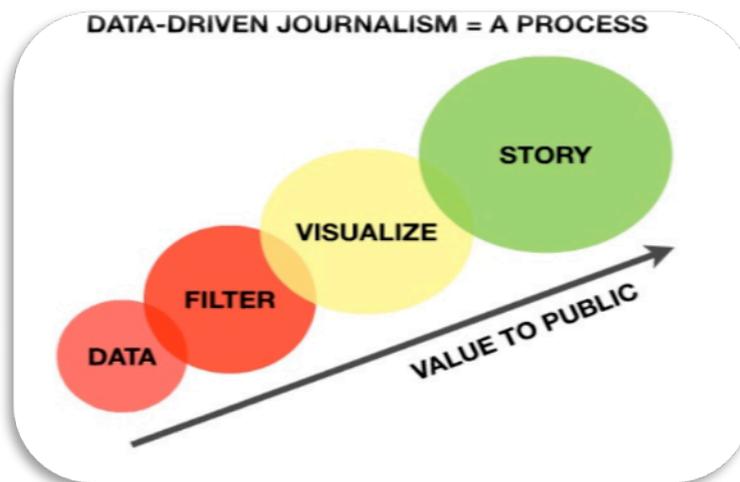
²⁷ Que Crucianelli cita como uma abreviatura de “Jornalismo de Bases de Dados”.

²⁸ Tradução de *Precision Journalism*, termo cunhado pelo livro *Precision Journalism* de Philip Meyer.

Processos

Os processos de trabalho do jornalismo de dados vão beber muito a algumas das anteriores correntes jornalísticas. As que mais influenciam esta forma de trabalho jornalístico são o «jornalismo de precisão» (Meyer) e o «jornalismo assistido por computador» (CAR- Computed Assisted Reporting). Mirko Lorenz, defende que o *data driven journalism* é um fluxo de trabalho em que os dados, à medida que vão sendo trabalhados vão ganhando mais e mais importância do ponto de vista jornalístico, sobretudo na medida em que se tornam mais e mais inteligíveis para o público. Na sua apresentação, na mesa redonda sobre Data Driven Journalism, organizada pelo European Journalism Centre, Lorenz apresentou um esquema que mostra as fases do processo de workflow do jornalismo de dados (Fig.1).

Figura 1 – Esquema Jornalismo de Dados como um Processo



Fonte: Lorenz, 2012.

São apresentadas quatro fases:

- Recolha de dados;
- Filtragem dos dados
- Visualização dos dados
- Utilização dos dados para a construção da história jornalística.

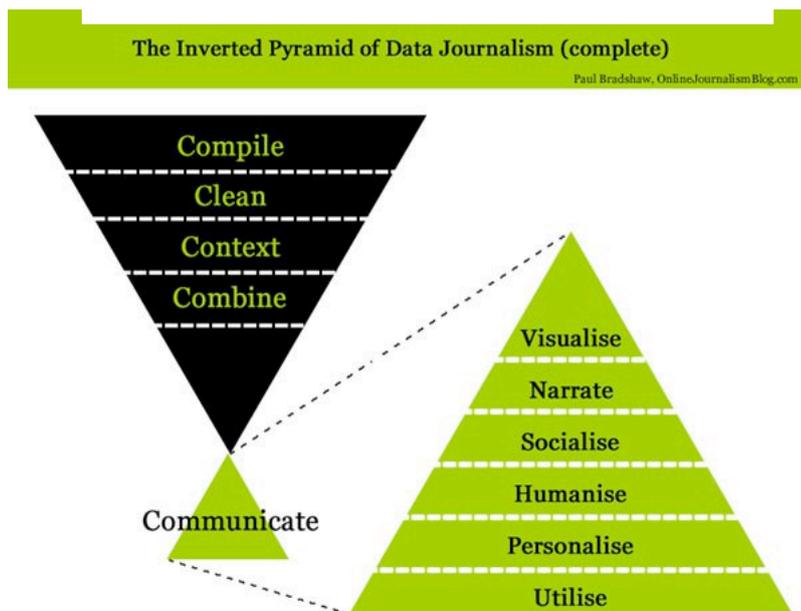
Lorenz mostra uma das mais simplistas estruturações de um trabalho de jornalismo de dados, mas dá a entender algo interessante, o aumento do valor que o trabalho tem para público, à medida que se vai avançando no fluxo de trabalho. É exatamente este interesse que é ressaltado no caso, já aqui abordado, da WikiLeaks – antes de os jornalistas pegarem nos dados e os terem tratado e tornado inteligível para o público em geral o valor para esse público era muito pouco.

Mas outros autores, como Paul Bradshaw²⁹, fazem abordagens ao jornalismo de dados como um fluxo de trabalho. Num texto³⁰ de 2011, publicado no seu blog “Online Journalism”³¹, Bradshaw recupera o clássico conceito de pirâmide invertida para o jornalismo de dados (Fig. 2). Aqui são clarificadas as fases de um produto de jornalismo de dados. Bradshaw apresenta a pirâmide dividida em quatro fases essenciais relacionadas com o trabalho dos dados:

- Compilar;
- Limpar;
- Contextualizar
- Combinar.

Após esta última fase, o autor acrescenta a comunicação dos dados (que aqui significa a “descodificação” para o público), que originará uma nova pirâmide. Esta apresenta o que deve acontecer com o produto da final da pirâmide anterior: visualização, narração, socialização, humanização, personalização e utilização. E que é explicado num segundo texto³² do mesmo autor.

Figura 2 – Pirâmide invertida Jornalismo de dados



Fonte: Bradshaw 2011

²⁹ Responsável pelo MA em Jornalismo Online, da Universidade de Birmingham, autor do blogue OnlineJournalism e de vários livros sobre a área.

³⁰ <http://onlinejournalismblog.com/2011/07/07/the-inverted-pyramid-of-data-journalism/>

³¹ <http://onlinejournalismblog.com/>

³² <http://onlinejournalismblog.com/2011/07/13/the-inverted-pyramid-of-data-journalism-part-2-6-ways-of-communicating-data-journalism/>

Bradshaw acaba por especificar o fluxo apresentado por Lorenz, sem que haja contradição entre o que ambos defendem. Estes são apenas dois dos exemplos de como o jornalismo de dados é tratado enquanto fluxo de trabalho. Simon Rogers³³, por exemplo apresenta uma representação gráfica um pouco mais complexa do que é o trabalho de jornalismo de dados. Ao autor salienta, no seu artigo³⁴ sobre o fluxo de trabalho no jornalismo de dados:

“Before a dataset results in a data journalism story, there’s a whole process of sifting and finessing and generally sorting the data out. The split is roughly 70% tidying up the data, 30% doing the fun stuff of visualizing and presenting it.” (Rogers, 2013)

Rogers põe a tónica no trabalho com os dados, como sendo o que consome a maior parte do tempo e fixa grande parte da sua descrição do fluxo na parte da recolha e tratamento dos dados. Seja qual for o autor, compreende-se que os processos de trabalho no jornalismo de dados trazem novidades ao trabalho jornalístico, e também às competências necessárias para trabalhar com este tipo de “matéria-prima”.

Conclusão

O jornalismo de dados é produto da evolução do jornalismo a par da evolução tecnológica e social. Nascido da influência de várias correntes dentro do jornalismo tais como o *Computer Assisted Reporting*, o Jornalismo de Precisão ou o Jornalismo de Investigação, são cada vez mais os exemplos deste tipo de jornalismo.

Mais do que tentar chegar a uma definição única do que é o jornalismo de dados importa perceber que há, de facto, abordagens novas ao jornalismo que têm por base mudanças como: O aumento da quantidade de dados existentes em formato digital; O aumento da quantidade de dados disponíveis, como consequência das políticas de abertura de dados, tanto a nível público como privado; Os avanços tecnológicos que permitem a democratização de formas mais fáceis e rápidas de tratamento e visualização de dados; O aumento da literacia digital da população. Contemporâneo do *big data* e das tendências de abertura de dados, o jornalismo de dados vai ao encontro de alguns dos desafios que eles trazem, aproveitando o seu potencial e tornando-os mais inteligíveis e acessíveis. Seja orientado para as histórias, para os processos ou para as bases dados, é jornalismo de dados, porque a sua matéria-prima são os dados.

A importância de perceber os fluxos de trabalho do jornalismo de dados é grande. As redações começaram nos últimos anos a compreender a sua importância. Olhando para alguns dos autores que descrevem estes fluxos de trabalho podemos estabelecer que as principais fases são: Obter/Extraír; Limpar/Normalizar/Enriquecer; Analisar/Combinar; e Visualizar os dados. Tendo em conta que, como salienta Lorenz (2010), à medida que vamos avançando no processo o interesse para o público aumenta. De salientar também que uma das características do jornalismo de dados é que grande parte do trabalho é efetivamente feito com os dados. Segundo Rogers (2013), 70% do tempo dedicado a um trabalho de jornalismo de dados é gasto com os processos relativos à recolha, tratamento, limpeza e análise dos dados, e os restantes 30% são empregues na visualização e na preparação das histórias que daí derivam.

³³ Jornalista especializado em jornalismo de dados, escritor e palestrante.

³⁴ Rogers, Simon, “Data journalism workflow”, 2013, retirado da internet em <http://simonrogers.net/2013/01/27/a-data-journalism-workflow/> (21/11/2014)

No final, apesar das várias abordagens aos fluxos de trabalho de jornalismo de dados, citadas neste trabalho, podemos concluir que elas não se contradizem entre si, antes se complementam. E que apesar de o trabalho com os dados representar uma grande parte do tempo gasto com este tipo de trabalho jornalístico, os autores todos consideram que o valor do trabalho aumenta à medida que se torna mais inteligível para o público, porque esse é um dos fins últimos dos jornalistas, decodificar a informação para que seja compreendida pelo público e que seja no interesse do público.

Referências bibliográficas

- Bradshaw, P. (2011). *The Inverted Pyramid of Data Journalism*. Disponível em: <http://onlinejournalismblog.com/2011/07/07/the-inverted-pyramid-of-data-journalism/> [2014, 21 novembro]
- Bradshaw, P. (2011). *The Inverted Pyramid of Data Journalism, part two: 6 ways of Communicating Data Journalism*. Disponível em: <http://onlinejournalismblog.com/2011/07/13/the-inverted-pyramid-of-data-journalism-part-2-6-ways-of-communicating-data-journalism/> [2014, 21 novembro]
- Castells, M. e Cardoso, G. (2006). *A Internet e a Sociedade em Rede*. Disponível em: http://cies.iscte.pt/linhas/linha2/sociedade_rede/ [15 Dezembro 2011]
- Grey, J., Bounegru, L. e Chambers, L. (2012). *The data journalism handbook*. Disponível em: <http://datajournalismhandbook.org/>
- Holovaty, A. (2006). *A fundamental way newspaper sites need to change*. Disponível em: <http://www.holovaty.com/writing/fundamental-change/> [2014, 21 novembro]
- Martinho, A. P. (2013). *Jornalismo de Dados: Contributo para uma Caracterização do Estado da Arte em Portugal*. Dissertação conducente ao grau de mestre em Comunicação Cultura e Tecnologias de Informação, ISCTE-IUL.
- Meyer, P. (1991). *The new precision journalism*. Disponível em: <http://www.unc.edu/~pmeyer/book/> [2014, 18 novembro]
- Rogers, S. (2013). *Data journalism Workflow*. Disponível em: <http://simonrogers.net/2013/01/27/a-data-journalism-workflow/> [2014, 21 novembro]
- Sirkkunen, E., Aitamurto, T. e Lehtonen, P. (2011). *Trends In Data Journalism*. Disponível em: http://virtual.vtt.fi/virtual/nextmedia/Deliverables2011/D3.2.1.2.B_Hyperlocal_Trends_In%20Data_Journalism.pdf [2014, 17 novembro]
- Stolte, Y. (2012). Journalism and Access to Data – The Phone Hacking Scandal, WikiLeaks and the Public Interest. *Datenschutz Und Datensicherheit - DuD*, volume 36 issue 5, 354-358
- Yiu, C. (2012). *The big data opportunity*. Disponível em: <http://www.policyexchange.org.uk/images/publications/the%20big%20data%20opportunity.pdf> [2014, 18 de novembro]