CHAPTER 5

Is There a Press Release on That? The Challenges and Opportunities of Big Data for News Media

Anthony Cawley

Abstract This chapter reflects on the growing importance of data to journalism. It highlights the fact that data journalists are developing innovative forms of news storytelling which not only engage audiences in novel ways, but also offer greater transparency to the editorial production process. It argues, however, that data journalism finds itself caught in a long-standing tension in news industries between the dual demands on news to serve the public good (maintaining an informed citizenry and acting as a watchdog on power) and succeed as a for-profit informational commodity. The chapter suggests that, so far, data journalism aligns more comfortably with the public good objectives of news than with its commercial imperatives, which may be a factor limiting its wider practice in news organisations.

Keywords Data journalism • Data-driven news • Automated journalism • News industries • News narratives • Informational commodity • Public good • Watchdog role • Big Data • Security

A. Cawley (✉)
Department of Media and Communication, Liverpool Hope University, Liverpool, UK

A. Bunnik et al. (eds.), *Big Data Challenges*,
DOI 10.1057/978-1-349-94885-7_5

Introduction

'A rapidly changing media landscape' is a phrase that has been used so often since the mainstream take-up of internet in the 1990s that it has become a cliché. But it continues to be rolled out by academics, industry executives, and practitioners because few, if any, aspects of the media are escaping significant digital transformation. Industry and organisational structures, business models, professional roles, and ways of engaging audiences have all been shaken out of their old certainties.

Change sketches a continuous line through the media's history, but the deep imprint of recent upheavals has been extraordinary. Destabilising matters further has been the global financial crisis, which, in bringing a harsher economic climate, has threatened the commercial health, and even survival, of many media organisations. In this broad context, the news-media, through data journalism, is grappling with the implications of Big Data for society in general and for the news industry in particular.

What the implications actually are is unclear. Neither is it certain that Big Data represents, in a journalistic sense, something new. But journalism's growing attachment to data connects to unresolved questions in the centuries-long development of news. What is the core purpose of news? What are the appropriate forms in which to tell a news story? What kinds of information should journalists seek as supporting evidence? And who has the authority to decide what is newsworthy?

At the heart of these questions is a historical uncertainty: should news be classed primarily as a public good or a product? Journalists tend to adopt a public good view of news, and they see its value in maintaining informed citizenries in democratic societies and in holding the powerful to account. Managers of media companies are more likely to position news as an informational commodity to be sold for profit (Cole and Harcup 2010). In the commercial settings in which most journalism is conducted, the two perspectives often grate against each other. News output that benefits the bottom line may not support quality journalism, and vice versa.

Even as a green shoot on the media landscape, data journalism blends naturally into the public good view of news: in pitching data through inventive forms of story-telling, journalists have opportunities to strengthen their relationships with audiences and sharpen their detection of state and corporate misconduct. However, data journalism faces significant obstacles, both internal and external to the news industry, if it is to deliver on that potential.

To provide a fuller context to data journalism's emergence, this chapter will consider the practice's historical roots alongside the findings of recent academic research. Then, in a brief case study, it will outline how data journalism is emerging in the Republic of Ireland. The country's media market sustains relatively small news organisations which operate under tighter financial, editorial, and technological resource constraints than leading-edge exponents of data journalism in bigger media markets such as the United Kingdom (UK) and the USA. The Irish case illustrates how the harnessing of Big Data to journalism is uneven both within national media markets and across the international sphere and is influenced significantly by non-technological factors, including legislative frameworks, news industry commercial pressures, journalistic cultures, and professional norms. The chapter will conclude with a discussion of the challenges confronting data journalism's continued development.

A Long Time Coming: Journalism Meets Big Data

One of the hopes being pinned to Big Data is that it will help journalists craft easy-to-understand accounts of what is happening in complex and fast-changing societies. That journalism is struggling to explain the world on which it reports is a long-standing anxiety. In the 1920s, the American public intellectual Walter Lippmann argued that the political-economic system had developed beyond the comprehension of the average citizen, whose life was being impacted by 'unnamed' or 'invisible powers' managed at 'distant centres' (Lippmann 2009: 3; first published in 1927). At this point, the news-media might be expected to step forward to ease the average citizen's confusion. But Lippmann claimed that 'no newspaper reports his environment so that he can grasp it' (2009, pp. 3–4). The problem for 1920s journalism was that the complexities of the political-economic system had slipped beyond its had slipped beyond its grasp, also.

When Lippmann was lamenting the limitations of print journalism almost ninety years ago, he was not, of course, referring to Big Data. Unwittingly, though, he was turning in that direction when he spotted the tendency of capitalistic societies to produce greater volumes of specialised information. In a decade in which the word 'media' itself was still a novelty (Briggs and Burke 2005), Lippmann believed journalism had already been overwhelmed by the scale and sophistication of capitalism's informational structures and was failing in its informed-citizenry and watchdog roles.

Capitalism's complexity would escalate through the twentieth century. In parallel, as the decades slipped by, news evolved to populate newer media forms: radio, television, and the internet. But this progress was not necessarily matched by an enhanced capacity to communicate the complexities of capitalistic societies as they attracted labels such as 'post-industrial society' or 'information society'. The global financial crisis that surfaced in the late 2000s illustrated the difficulties news media face in trying to report on advanced capitalism in a meaningful manner to the public. As the crisis deepened, even specialist business journalists were found to lack the skills, expertise, and resources properly to investigate, understand, and critique the intricate products and models emerging from the financial sector (Doyle 2006; Tambini 2010; Fahy 2014; O'Brien 2015).

Journalism's limited explanatory and critical capacities can be tied to news industry structures and commercial demands. Producing original journalistic content, especially from longer-term investigative projects, is time consuming, labour intensive, expensive, and risky (not all investigations will lead to a publishable story). News industry economics have evolved to favour cheaper, reliable, and predictable sources of information, including public relations material and regular journalistic 'beats' such as parliaments and courts. A contemporary of Lippmann's, the social commentator Upton Sinclair, recognised in the 1920s the eagerness of American newspapers to republish as news the information supplied by the government and the business sector (Sinclair 2003; first published in 1920). These actors were motivated to feed material to journalists to gain favourable publicity in the news, or, at a minimum, to dull criticism. Some sixty years later, Oscar Gandy (1982) called the practice an 'information subsidy', in the sense that governments, corporate actors, and public relations professionals were subsidising the cost of news production for commercial media organisations.

Lewis et al. (2008) argue that the news media's dependence on supplied material is likely to increase in the future. One reason is that newsroom resources are being hollowed out as news organisations try to carve out sustainable business models in the digital economy. Also, given that governmental and corporate actors tend to maintain efficient public relations units, the supply of pre-packaged 'news' is unlikely to run dry. Within this supply-demand relationship is, perhaps, a point of added interest to data journalists. The core activities of governmental and corporate actors increasingly rely on the gathering, processing, and manipulation of Big Data. To date, however, data sets have fallen outside the conventional provision to journalists of press releases, prepared statements, and verbal briefings.

Academic Views of Journalism and Big Data

The academic literature on data journalism remains 'relatively sparse' (Franklin 2014, p. 485). However, media scholars are directing greater attention to developments in the area as more journalists rely on 'data-processing tools' to highlight social issues and expose wrongdoing (Parasie 2015, p. 364). Specifically, media scholars are trying to gauge the disruptions and continuities between data journalism and traditional journalistic values, practices, and ethics. Attracting significant research interest is the potential of data journalism to break free of the news industry's obsession with immediacy and allow stories to unfold over longer periods of time (Anderson 2015).

From the public's perspective, the most visible impact of data journalism has been the increase in news stories told through infographics, interactive visualisations, and probability models (Lewis and Westlund 2015). In a move to increase editorial transparency, some data journalists have been making their data sets publicly available to facilitate audience engagement and analysis (Lewis and Westlund 2015). Their actions are a marked departure from conventional news production, which traditionally has been closed to outside scrutiny.

The recent turn to data signals a broadening of the kinds of information that journalists consider legitimate to underpin news stories. The shift has historical precedent. Newspapers used to publish stories gathered almost exclusively from information from documents (Anderson 2015). Only in the late nineteenth century did the interview emerge in Anglo-American and Irish journalism as a technique to gather first-hand information from people (Anderson 2015).

Evaluating data journalism through a historical lens reveals that the principles underlying it are decades old (Anderson 2015). Having surveyed infographics in 1940s American newspapers, Anderson concluded that 'data aggregation and information visualisation have been part of journalism for a long time' (2015, p. 351). In a similar vein, Parasie and Dagiral (2013) point out that the meshing of electronic data and journalism emerged, through computer-assisted reporting, in the 1960s. They suggest that, in its modern guise, data journalism could contribute to democracy in three key ways: through increased journalistic objectivity, the offering of new tools to keep a check on government actions, and increasing citizens' participation.

The roots of data journalism may be deeper than expected, but the practice still holds the potential to spark radical transformations in the news industry. For instance, Carlson highlights the emergence of 'automated

journalism', where human reporters largely are displaced from news production and algorithms recast data sets as news narratives (2015, p. 416). He identifies as an example *The Washington Post*'s Truth Teller app, which runs real-time accuracy checks on politicians' speeches (http://truthteller.washingtonpost.com/). Such initiatives, Carlson suggests, are part of a wider 'movement of Big Data from newsgathering aid to the production of news itself' (2015, p. 417). For this reason, he argues, automated journalism calls into question traditional notions of news work and what constitutes journalistic authority if human editors and journalists are sidelined.

Lewis and Westlund caution that, amid all the 'hype' and 'hope' associated with Big Data, it is impossible to know what will happen, but it is becoming increasingly apparent that '*something* important is changing' (2015, p. 448). From the perspective of public good, to unlock the potential of data journalism would require news organisations to invest in additional, and unfamiliar, editorial resources. But doing so would go against the grain of the news industry over the last few years, as the push has been to reduce costs in newsrooms. From the perspective of news as a commodity, however, Big Data may afford media companies improved opportunities for revenue growth in at least two ways: first, through helping to streamline their business practices, and, second, through refining their understanding of audience and advertiser preferences when offering news products and services (Lewis and Westlund 2015).

Data Journalism in Ireland

Irish newspapers and broadcasters have been 'slow' to embrace data journalism, and much of the initial energy has been generated not by professional newsrooms but by independent bloggers (Lenihan 2015). This may be a symptom of the discomfort that data journalism causes to traditional newsmodels in Ireland, which have tended to place a higher value on reporting events and personalities than processes and long-term developments. Support for data journalism also requires ambition, innovation, and investment at a time when, in the context of a small but highly globalised domestic media market, Irish news organisations have been experiencing tough trading conditions. Figures from industry representative group News Brands Ireland could be considered to be an industry lobby group, but I want to avoid using that term because of its negative connotations (NBI 2015). In the same period, the market value of print and broadcast advertising in Ireland has also declined significantly, with growth in digital advertising being too low to compensate (Indecon 2014).

The dominant players in Ireland's news market are the public service broadcaster RTE and the legacy newspaper publisher Independent News and Media (INM). Neither organisation has developed a significant profile in data journalism. The resources available to RTE's newsroom have suffered through the downturn in advertising (which, alongside a licence fee, provides the bulk of the broadcaster's funding). INM, which publishes one of Ireland's most popular news websites, independent.ie, has adopted a 'digital first' editorial structure (INM 2015). However, this refers to the priority platform on which news content is published, not to a data-driven news agenda.

Newspaper and website publisher *The Irish Times* announced an initiative in early 2015 to 'put data journalism to the foreground' of the organisation's news offering (The Irish Times 2015). Irish Times Data highlighted the potential for data journalism to tell stories related to emigration, health, politics, and the economy in compelling and accessible formats. The production team makes its source data publicly available, but conceded that filtering and structuring data-sets and running visualisations was 'labour intensive' and required a range of journalistic and technical skills (The Irish Times 2015). Given the resource pressures squeezing Irish media organisations, the labour and technical costs associated with data journalism may be restricting its wider practice.

Globally, groups such as Open Knowledge have been campaigning for more transparent government and freer public access to the data sets that state institutions accumulate. Irish data journalism is developing in a legislative environment where Freedom of Information Act provisions are quite restrictive (Lenihan 2015). However, the government has made some efforts to increase public availability of data, notably through the establishment of Open Government Partnership Ireland (data.gov.ie) as a portal to the data sets of 85 state organisations and bodies. The facility is consistent with open government initiatives elsewhere, including the UK's portal at data.gov.uk. Internationally, organisations such as the World Bank, the World Health Organisation, and the European Union have been making extensive databanks publicly available.

Concluding Comments

Big Data itself is on the expanding list of social and technological complexities that the news-media is expected to explain to the public. Already, outlets such as *The Guardian* in the UK have provided vigorous coverage of Big Data's emergence and controversies, with a keen focus on

the legitimacy of how state and corporate actors gather and use data on people's public and private activities. *The Guardian's title* is also a pioneer in the practice of data journalism in the UK. In these regards, the news organisation is applying its traditional informed-citizenry and watchdog roles to Big Data and is pushing the boundaries of data journalism innovation. Doing likewise in another of the world's larger and richer media markets, the USA, are *The New York Times*, *The Washington Post*, and the foundation-funded ProPublica.

Their efforts and achievements are the exception rather than the norm, however. Data journalism faces considerable challenges if it is to become a well-established, industry-supported genre of news. Like the implications of Big Data for the media generally, the challenges facing data journalism are not all apparent yet. But, at this stage, the transparency of externally-held data sets is a priority. To adequately extend their public service role to Big Data, news organisations will need greater knowledge of the data sets that state and corporate actors possess and why they possess them (including the configuration of the algorithms processing the data). Further, news organisations will require formal and accountable mechanisms to access state-held data sets when it is appropriate and necessary to do so in the public interest.

A challenge internal to the news industry hinges on data journalism gaining wider acceptance and support amongst news professionals. Since the late nineteenth century, journalists' professional expectations have developed around notions of news being *about* people, produced by speaking *to* people. This 'social practice' of journalism is a far cry from the technical task of interrogating patterns across multiple data sets (Carlson 2015, p. 417). Warming to the idea of data as news will, for many journalists, take time. Further, and resonating with the discussion in Zwitter's contribution to this volume, news media organisations will need to refresh and perhaps redraw the ethical frameworks and professional guidelines underpinning their corporate and editorial behaviour in an increasingly data-driven and networked world. Another challenge stems from the commercial structures of news production, which are calibrated to attract audiences and advertisers. To receive robust industry support, data journalism will need to justify itself as an informational commodity. To this point, arguably, the public good value of data journalism is more evident than its commercial value.

References

Anderson, C. W. (2015). Between the unique and the pattern: Historical tensions in our understanding of quantitative journalism. *Digital Journalism, 3*(3), 349–363.

Briggs, A., & Burke, P. (2005). *A social history of the media: From Gutenberg to the internet*. Cambridge: Polity Press.

Carlson, M. (2015). The robotic reporter: Automated journalism and the redefinition of labour, compositional forms, and journalistic authority. *Digital Journalism, 3*(3), 416–431.

Cole, P., & Harcup, T. (2010). *Newspaper journalism: Key texts*. London: Sage.

Doyle, G. (2006). Financial news journalism: A post-enron analysis of approaches towards economic and financial news production in the UK. *Journalism, 7*(4), 433–452.

Fahy, D. (2014). A limited focus? Journalism, politics and the Celtic Tiger. In M. O'Brien & D. O'Beachain (Eds.), *Political communication in the Republic of Ireland* (pp. 129–146). Liverpool: Liverpool University Press.

Franklin, B. (2014). The future of journalism. *Journalism Studies, 15*(5), 481–499.

Gandy, O. (1982). *Beyond agenda setting: Information subsidies and public policy*. New York: Ablex.

Independent News and Media. (2014). Annual Report. Retried April 4, 2016, from http://inmplc.com/news/uploads/INM_Annual_Report_2014_for_web.pdf

Indecon. (2014). Economic analysis of the advertising market in Ireland 2013. Retrieved July 9, 2015, from http://www.dcenr.gov.ie/NR/rdonlyres/D5003758-B55B-446B-985C-6A131369B3D2/0/EconomicAnalysisofAdvertisingMarketIndeconReport.pdf

Lenihan, H. (2015). Digital journalism: Reinventing news for 21st century. Retrieved July 7, 2015, from http://www.irishtimes.com/opinion/data-journalism-reinventing-news-for-21st-century-digital-society-1.2119160

Lewis, S., & Westlund, O. (2015). Big Data and journalism: Epistemology, expertise, economics and ethics. *Digital Journalism, 3*(3), 447–466.

Lewis, J., Williams, A., & Franklin, B. (2008). A compromised fourth estate? *Journalism Studies, 9*(1), 1–20.

Lippmann, W. (2009). *The phantom public*. New Jersey: Transaction Publishers.

News Brands Ireland. (2015). Retrieved July 9, 2015, from http://newsbrandsireland.ie/data-centre/circulation/

O'Brien, M. (2015). The Irish press, politicians, and the Celtic Tiger economy. In S. Schifferes & R. Roberts (Eds.), *The media and financial crisis: Comparative and historical perspectives* (pp. 75–86). New York: Routledge.

Parasie, S. (2015). Data-driven revelation? Epistemological tensions in investigative journalism in the age of 'Big Data'. *Digital Journalism, 3*(3), 364–380.

Parasie, S., & Dagiral, E. (2013). Data-driven journalism and the public good: 'Computer-assisted-reporters' and 'programmer-journalists' in Chicago. *New Media and Society, 15*(6), 853–871.

Sinclair, U. (2003). *The brass check: A study of American journalism*. Urbana and Chicago: University of Illinois Press.

Tambini, D. (2010). What are financial journalists for? *Journalism Studies, 11*(2), 158–174.

The Irish Times. (2015). Welcome to Irish Times data. Retrieved July 7, 2015, from http://www.irishtimes.com/news/ireland/irish-news/welcome-to-irish-times-data-1.2089414