

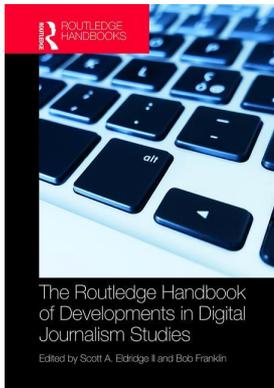
This article was downloaded by: 10.3.98.93

On: 17 Jan 2019

Access details: *subscription number*

Publisher: *Routledge*

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: 5 Howick Place, London SW1P 1WG, UK



## **The Routledge Handbook of Developments in Digital Journalism Studies**

Scott A. Eldridge, Bob Franklin

### **Data Journalism as a Platform**

Publication details

<https://www.routledgehandbooks.com/doi/10.4324/9781315270449-22>

Eddy Borges-Rey

**Published online on: 30 Aug 2018**

**How to cite :-** Eddy Borges-Rey. 30 Aug 2018, *Data Journalism as a Platform from: The Routledge Handbook of Developments in Digital Journalism Studies* Routledge

Accessed on: 17 Jan 2019

<https://www.routledgehandbooks.com/doi/10.4324/9781315270449-22>

**PLEASE SCROLL DOWN FOR DOCUMENT**

Full terms and conditions of use: <https://www.routledgehandbooks.com/legal-notices/terms>

This Document PDF may be used for research, teaching and private study purposes. Any substantial or systematic reproductions, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The publisher shall not be liable for an loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

# DATA JOURNALISM AS A PLATFORM

## Architecture, agents, protocols

*Eddy Borges-Rey*

Over the last decade, British journalism has come under increased public scrutiny. First, the phone hacking scandal (2011) and the subsequent Leveson report (2011–2012) evidenced the inability of a number of tabloid publications to act ethically and in the public interest. Then the expert–layman relationship, which granted journalism its authority, suffered severe damage when experts were discredited as a knowledgeable source of information by populist politicians in 2016. Finally, a widespread dissemination of misinformation exacerbated by social media algorithms took media distrust to unprecedented levels by the end of 2016.

During this period, while journalism struggled with its image crisis, computerized data gained prominence as “the new oil of the Digital Economy” (Yonego, 2014), and public and private actors entered a race to control and nourish this commodity. Data were also perceived as an opportunity by a small number of news organizations, who invested in data journalism (and its methodological rigor) to spearhead their efforts in positioning journalism as society’s watchdog once again.

Although a number of data/computational journalism scholars have conceptualized data journalism in more sophisticated ways (cf. Hamilton and Turner, 2009: 2; Lewis and Westlund, 2015b: 449; see Coddington, this volume, Chapter 17), at this stage I will define data journalism as the type of news reporting through which journalists engage with computerized data to inform their publics. Research on the field (Appelgren and Nygren, 2014: 40; De Maeyer et al., 2015: 441; Fink and Anderson, 2015: 475; Borges-Rey, 2016; Borges-Rey, 2017) has, nevertheless, shown that data journalism represents, to date, only a small proportion of the overall coverage by the mainstream news media, hence challenging early expectations that a methodical reporting of figures and numbers would stimulate public trust again.

Under such unstable conditions the implications of *Big Data* for journalism remain to be seen. In addition, the closure and amalgamation of a series of local news outlets has not only exposed a worsening democratic deficit but has also shifted news managers’ attention to another dimension of data: automation. While editorial boards are presented with an ever more constrained set of resources, automation is set to play an essential role in the professional newsroom of the future (Van Dalen, 2012; Clerwall, 2014; Latar, 2015; Carlson, 2015; Broussard, 2015), releasing journalists from time-consuming activities such as identifying potential tips or leads, determining whether a story is newsworthy enough, or making decisions when it comes to ethical dilemmas. While professional journalism is still confronting the complexities of data

journalism practice, scholars have made a start on what some regard as the second wave of data journalism research.

As a way of consolidating my contribution to this field over the last five years, this chapter examines the tensions and convergences that accompany the integration of data processes into professional newswork and their impact on traditional organizational news structures and cultures. It seeks to examine, in detail, the role of British contemporary data journalism to explore journalism in an increasingly data-driven society. As scholars continue providing a plethora of diverse and often competing conceptualizations of data journalism, this chapter focuses on two main epistemologies of data journalism practice (see Borges-Rey, 2017): a dominant one, whereby data are handled by journalists following axiomatic norms of news reporting; and an emerging one, whereby journalists use the methods and logics of data scientists to interact with data in innovative ways.

Drawing on a set of semi-structured interviews with data journalists, data editors, developers, graphic designers, and news managers based at the *Guardian*, the BBC, the *Financial Times*, Channel 4, Trinity Mirror Group, the *Times*, CNN, Thomson Reuters, the *Daily Telegraph*, STV, the *Scotsman*, the *Herald*, the *Detail*, Wales Online, *Irish News*, and the *Belfast News Letter*, I use Montfort and Bogost's (2009) platform studies methodology to examine data journalism as a platform that exists within an architecture (virtual and non-virtual ecosystems) that deals with agents (journalists, data, automated entities) through a series of protocols (news cultures, computational logics/cognitions, power relations/struggles).

### Platform studies for non-computational ecosystems

Since Montfort and Bogost proposed platform studies as a methodology and attempted to consolidate it through the launch of an MIT book series back in 2009, their intention was to “promote the investigation of underlying computing systems and how they enable, constrain, shape, and support the creative work that is done on them” (Montfort and Bogost, 2009: vii). In this sense, as noted by Apperley and Parikka, scholars interested in undertaking platform studies were asked to focus on three dimensions: (a) “A single platform or a closely related family of platforms”; (b) “Technical rigor and in-depth investigation of how computing technologies work”; and (c) “An awareness of and discussion of how computing platforms exist in a context of culture and society” (2015: 2).

With a manifest emphasis on computing technologies, Montfort and Bogost proposed a five-layer model that analyzes platform, code, form/function, interface, and reception/operation. Using the computational system as a means to interpret larger cultural frameworks, they remark that: “A computational platform is not an alien machine, but a cultural artifact that is shaped by values and forces and which expresses views about the world” (Montfort and Bogost, 2009: 148). In this section, I will argue that this perspective of platform studies is rather limited, and, in order to achieve its full potential, it should span beyond the technical specificity of the computational architecture. Similar criticism is voiced by Apperley and Parikka, who claim:

“*Platforms are platforms* not necessarily because they allow code to be written or run, but because they afford an opportunity to communicate, interact or sell” (Gillespie, 2010: 351, emphasis added). Platforms, then, are not just technologies but techniques that sustain interactions as well as offer an epistemological framework.

(Apperley and Parikka, 2015: 5)

Therefore, I propose an approach whereby architecture transcends the boundaries of the hardware and reaches both the virtual and material datafied ecosystems that shape our modern society.

Likewise, software, i.e., a series of agents governed by a protocol to fulfill a purpose in the system, transcends the instructions, parameters, and variables set by the code to reach the power struggle or negotiation that takes place between data journalists and data entities during the production of news.

Given that the methodology of platform studies has been deliberately left open (Apperley and Parikka, 2015), it offers the flexibility to go beyond the tensions between computational and journalistic logics, recorded by extant research on the field, to instead approach data journalism as a technoculture that engages with a variety of materialities, performativities, and reflexivities, placing, “the platform at the center of a materially grounded discourse” (Apperley and Parikka, 2015: 5–6). Consequently, I will shift the emphasis from data journalists at the epicenter of the analysis to data journalists coexisting with computerized agents in a larger ecosystem.

### Interface: the illusion of control

In the introduction, I defined data journalism as a type of news reporting through which journalists engage with computerized data to inform their publics. The intention to place the emphasis on the interaction between data journalists and data was deliberate. The idea of data as an immaterial commodity *used* by journalists is both limited and constraining, in my view. This is why perhaps the most effective way to outline the contours of data journalism as a platform is by recognizing data as an active material entity that affects the agency of data journalists. As analytical frameworks tend to depict interactivity as a human enabler, they dismiss interactivity’s material facet; that is, its capacity to enable material entities to interact as well. During my five years of data journalism research, I have realized that Miller’s (2005: 3) distinction between a “vulgar theory of mere things as artefacts” and “a theory that claims to entirely transcend the dualism of subjects and objects” is essential to expand beyond the reductionist conceptualization of data as objective evidence. This way of approaching data has been prevalent in data journalism practice in the United Kingdom and elsewhere. In this regard, data are commonly assumed by data journalists to be one of the many ingredients of the recipe we call news, hence controllable, inactive, and subordinate to human agency. Data tend to be handled as an inanimate good that functions as either evidential input for stories or, to a minor extent, as the fuel that powers data-driven outputs. Once data are found within the convoluted entanglement of datasets held by public or private institutions, they are claimed by data journalists and subsequently placed in the formulaic structure of the news item.

Yet if we understand data as the vital fluid of societal institutions and data-driven automations as “bodies without organs” (Deleuze and Guattari, 1987: 4), data transcend the boundaries imposed by the limited mind-set of *data as evidence* and become “something more than ‘mere’ matter: an excess, force, vitality, relationality, or difference that renders matter active, self-creative, productive, unpredictable” (Coole and Frost, 2010: 9). Virtually every sector of society currently deals with both media and computing software on a regular basis, generating vast amounts of data. A rising institutional dependence on database architecture and algorithmic automation ensures that society’s machinery is well oiled and working smoothly. Therefore, I argue that in order to understand how profoundly data are embedded in the structures of public bodies, governments, corporations, and civic society as a whole, data journalists must actively negotiate shares of performative power with data brokers and their datafied entities, as control shifts from traditional spheres to organizations that nourish the wealth of data. In this sense, long-standing attitudes toward discovering patterns and correlations in the data to unearth novel insights or corroborate hypotheses should be complementary to an understanding of notions of media numerical representation, modularity, automation, variability, and transcoding (Manovich, 2001) that are essential for recognizing how effectively data governs modern social life.

The tension between these two epistemologies can be seen at a fundamental level when data journalists acknowledge that data enable them to generate exclusives in a process whereby data journalists, while recognizing data's capacity to make their stories more accurate and rigorous, simultaneously are capable of challenging the quality of that data by looking for flaws within datasets. At a more advanced level, this power negotiation can be seen in contemporary newsrooms that depend on automations, such as web metrics, to inform editorial decision making; or scrapers, to gather and analyze large-scale datasets. These custom-made entities are normally found only in a limited number of elite news organizations that can afford such technology, or the professionals that model them. Smaller, local news organizations must normally rely on generic third-party solutions, which, though freely accessible through open-source regimes, are not always compatible with the existing architecture of these organizations' newsrooms, thus hindering the aptness of the algorithmic organism for the ecosystem in which it performs. Here, the access to the insight obtained through automated agents dictates a different power dynamic between those with access to the means to harvest data-driven insights and those who are unable to access this wealth of information because of financial constraints. An inability to program these automations in-house makes local news outlets dependent on collaborations with programmers or a few generic software solutions that are limited by the universality of their design.

This distinction between data journalists who can model the behavior of automated agents (for example, a scraper) through the design of their systems and those who have to rely on generic third-party software (say, Google Sheets) is key to understanding the importance of the interface in the power struggle between data journalists and automated systems. For instance, the transference of power from the data journalists to the automated agent becomes evident when the data journalist has to navigate, as suggested by Lister et al. (2009: 21–22), an abundance of finely tuned options offered by the software interface to maximize their perceived interaction and freedom of choice, thus increasing their apparent capacity to negotiate with the automated agent and even regain certain shares of performative power. Although the limitations of third-party software are sometimes valued by data journalists, as they prompt creative ways to circumvent these limitations through the use of other applications, it is ultimately the interface that governs the rules and levels of interaction of reporters with no coding knowledge, further evidence of how data affects the performativity of these professionals.

Understanding interactivity and interface through the prism of the computerized system can help researchers and journalists to detach themselves from the prevailing journalistic humanness that undermines data journalism's computational possibilities. Since the interface does not distinguish between digital objects and users, human agency is parametrized as another variable governed by system protocols: another command line in the code. While automated systems pervade organizational and professional cultures and logics, so do their interpretations of the protocols that govern the processes and professional practices they engage with. If the data journalists that form part of that equation wish to engage with automated agents to negotiate shares of power, they need to abandon the illusion of control and come to terms with data-driven bodies as equals to understand the ways they perceive and interpret the ecosystem. A broader understanding of data as the vital fluid of society's virtual and material infrastructures has equipped a small number of data journalists with a set of unique performativities that enable them to bend the rules of the interface and extract information from sources unavailable to data journalists anchored in the belief that data are only evidence. I will expand on this later in the chapter.

### **Form: the rules of the game**

Revisiting my introductory definition of data journalism, I have characterized the practice as a type of reporting. With this assertion, I intend to oppose enthusiastic expectations, such as those

of Tim Berners-Lee, the founder of the World Wide Web, which position data journalism as the future of journalism (Gray et al., 2012). In this section, then, I map out the form that contemporary data journalism practice takes in the United Kingdom and how it functions, taking as a point of departure the source of data and how data are gathered.

If data journalism follows a set of rules, this protocol, in the first instance, is strongly mediated by the rigid ethos of the Freedom of Information Act (FOI Act) and the Open Data movement. Although alternative ways of obtaining data, such as subscription to data feeds, data licenses, leaks, or in-house collation were mentioned, most of the data journalists I have interviewed in recent years agree that, by and large, the data they use originated within the public sector and were obtained by means of FOI requests. The act, which has been adopted with varying degrees of success in local, regional, and national administrations in the United Kingdom, has suffered from limited resources and infrastructure; differing interpretations of the legislation exhibited by civil servants; and a general public disenfranchisement with the scheme. These limitations have not only undermined the potential for openness of the scheme but also have compelled journalists to break with their overreliance on public data and to produce their stories with the help of web-scraping or collaboration with third-sector or private organizations willing to grant access to their databases. While this transition takes place, I will argue that open data largely infuse the material character of the datasets used as evidence by these professionals, at present.

In light of this tension, data journalism practice tends to be mediated by ideals of *alleged* openness and transparency while it is simultaneously subject to the bureaucracy and politics of public institutions. Data generated within the public domain tend to be perceived by data journalists as influenced by the agendas of politicians, press officers, and other spin doctors. They are also perceived to be restricted by the bureaucratic mechanisms of public bodies, which in some cases, after prolonged negotiation, provided the data in a non-machine-readable format, making it difficult to use. FOI-driven data also provide a distinctive flavor to the reports of data journalists that is generally restricted to topics limited to census, public governance, health, education, or crime, affecting not only the style of reporting but also the scope of the story.

Despite the growing expansion of open data, many data journalists have recognized that data released through this scheme is too politicized to be used effectively for journalistic purposes and see the data scraped from websites or obtained through informants or similar conventional methods to be more appropriate. Paradoxically, access to alternative sources of data or, for instance, corporate data held by data brokers does not depend on the advanced computational skills of data journalists but on the will of insiders, whistle-blowers, or leaks and similar traditional means, as it is proprietary data protected by copyright laws.

Once data are collated, data journalists sort, clean, visualize, and make sense of the data they have gathered in order to tell a story. Early predictions suggested that data journalism would engage audiences with innovative forms of storytelling, and the familiar linear storytelling that has characterized journalism over more than 100 years would be disrupted, to some extent, by more interactive and engaging forms of informational user experience with multi-layered, multiplatform, gamified, database-linked dynamic content. This type of presentation requires users to input personal data and, after performing real-time data analysis, tells them a bespoke story about themselves. This informational experience therefore appears to be heavily mediated by the ontologies of user-interface design, user-experience design, and human-computer interaction, not only signaling the pervasiveness of computational thinking in data journalists' reflexivity but also the potential of data journalism to manage audiences' informational expectations to a granular level, offering finely tailored personal stories that challenge benchmarks such as the public interest and news values.

The tension between FOI stories and non-FOI stories, linear storytelling and informational user experience, among other factors, has led to the diversification of data journalism practice into three distinct forms:

- a. A daily, quick-turnaround, generally visualized, brief form of data journalism;
- b. An extensive, thoroughly researched, investigative form of data journalism; and
- c. A light, editorialized, entertaining, often humorous, gamified form of data journalism.

This classification not only refutes common beliefs that data journalism stories are presented by means of visualizations but also highlights the flexibility of the format to accommodate opinionated and editorialized forms of popular data journalism. For instance, the Trinity Mirror Group's discontinued data journalism experiment *Ampp3d* glamorized numbers, popularized them, and made them appealing. It used the most popular newsbeats within the publication to sell, to a very specific segment of the public, the whole data journalism venture. It used humor and satire to act as a liberal champion that often criticized the conservative government, thus protecting the disadvantaged and underrepresented. Although it performed a paradoxical form of data journalism – one that blended op-ed journalism with scientific rigor – it also showed its audiences, in very applied terms, how to correlate trends, perform data extrapolation, understand the extremely necessary context of official stats and figures and issues of sampling and population, and make sense of ordinary numbers that represent issues affecting them directly. It performed the necessary function of humanizing numbers by highlighting their social relevance. It also showed how numbers are misused and misrepresented by media and politicians alike. The story “Hey DWP! We Just Fixed Your Chart for You” (Warnes, 2014) serves as an example of *Ampp3d*'s fact-checking ethos. The original tweet released by the Department for Work and Pensions of the United Kingdom tried to illustrate how the benefit cap helped thousands move into work. However, the figures shown contradicted the proportions of the bars on the chart. By just changing the graphic proportion of the bars on the infographic, *Ampp3d*'s illustration correlates more accurately with the numeric proportions conveyed by the actual statistics. Through this type of accessible explanations, *Ampp3d* attempted to show its public how the government could spin numbers to disseminate an optimistic, yet often inaccurate, depiction of the administration. This example also illustrates the ways through which *Ampp3d* expanded its storytelling capabilities beyond the confines of the data story to reach the social media conversation triggered by it.

*Ampp3d* thus became an innovative player within the British data journalism landscape. It created a versatile transmediality that enabled its journalists to strategize (a) data infrastructures and the stories potentially concealed within them; (b) the tensions that arise between two conflicting forces – popular liberal satirical tabloid journalism and the surgical precision of numeric facts; (c) the efforts to humanize statistics and highlight its mundane facet; (d) the development of a distinctive digital persona intermediated by the charisma of the newshound; and (e) the capacity to perform the data journalist's function beyond the limits of the platform or the publication.

In sum, the specificity, the specialized knowledge base, and the disruptive nature of many of its features suggest that data journalism seemingly works better as an alternative methodology or philosophy that adapts to journalistic themes or beats – be that sports or investigative journalism – to provide both a robust backbone to stories and tools to efficiently make use of web-based knowledge infrastructures. This will be explored at a performative and reflexive level in the next section.

### Code: what makes data journalism tick?

This level – code – is potentially the one with the most challenges, as it refers to the computational code of the platform. Focusing on the comments, variable names, choices made when writing programs, and the discourses attached to them, here I explore the *epistemological*, *political*, and *discursive* codes that shape the performativity and reflexivity of data journalists rather than the computational. In order to achieve this, I have proposed a taxonomy of performative codes that

data journalists use to legitimize their knowledge claims. As data journalists model their journalistic identity, these codes outline the contours of their working practice. But, more importantly, they act as the definers used in the constant power struggle in which data journalists engage with traditional reporters to find their place in their newsrooms and journalism at large.

There is no doubt that the idiosyncratic aspects discussed earlier pervade the performativity and reflexivity of data journalists, mediating how they perceive themselves and their working procedures. In this sense, I found data journalists who felt comfortable performing more elementary forms of data analysis to produce FOI-driven stories. Equally, I found a smaller proportion that felt the need to utilize computer science methods to provide structure to data obtained through less conventional methods, such as web-scraping, thus escaping the constraints of open data regimes and the limited thematic flavor that the FOI Act scheme granted their stories. Regardless of the computational complexity of each approach, to a greater or lesser extent these performative codes were present in their newswork.

- a *Journalistic authority*: in order to persuade their audiences of the veracity of their accounts (Broersma, 2010), journalistic authority has endured as the imperative requirement to perform within the trade, notwithstanding the importance of data and computation. Data journalists unanimously stated that the most significant professional attribute required to legitimize their accounts was journalistic skills.
- b *Numeric infallibility*: this provided quantifiable evidence to their stories, which was then reinforced by the rigor of statistical methods used during the news production process. At least one-third of the data journalists that I interviewed acknowledged having a background in a mathematics-related discipline. The remainder showed competence in basic statistical operations, such as correlation, percentage change, standard deviation, large increases and decreases, and trends over time. Data journalists were well aware of recurrent statistical errors in news reporting and how to avoid them. They were also extremely cautious with respect to their calculations, corroborating their approaches with statisticians within public bodies as to the types of equations and algorithms used to process data.
- c *Scientific rigor*: in order to remain trustworthy in the public eye, soundness and transparency in their methods was imperative.
- d *Computational neutrality*: this refers to technologies that arguably circumvent human bias while efficiently performing automated gathering, analysis, and presentation of unstructured information. Regardless of the complexity of the computing tasks performed – from number crunching and pivot-table analysis in a spreadsheet to back-end programming of computerized algorithms – involving computation throughout the news-reporting process arguably improved the sense of flawlessness in their accounts.
- e *Crowdsourced cooperation*: by providing audiences with easy access to raw data and filtering tools to make sense of these data, data journalists attempted to democratize both the databases that serve as inputs for stories and the news workflow that produces such stories.
- f *Extra- or intra-newsroom collaborations*: the absence of certain advanced computational skills and/or restricted access to certain information has compelled data journalists to embrace open-source ideals and seek internal or external collaboration in their efforts to overcome these limitations. As data journalism impregnates the news culture of professional newsrooms, specialized correspondents seek to collaborate with data journalism units to provide soundness and robustness to their stories through data methods. Collaborative projects where data journalists, developers, statisticians, and graphic designers interacted were deemed by data journalists to be very effective, combining the expertise of various disciplines to produce ground-breaking news experiences. Data journalists with rather limited technical

competences and infrastructure tended to collaborate with academics, public archivists, records officers and/or statisticians, civic organizations and open-source initiatives (such as *Hacks/Hackers* or *scrapewiki*), libraries, trusts or foundations, and third-sector organizations or private companies that contributed to the project by opening up their databases for journalistic scrutiny.

- g *Hyperlocal empathy*: this is the final and perhaps most interesting code of the taxonomy. As data journalists engaged with hyperlocal data, they connected strongly with their community and local issues. This code was shown across the board, as data journalists passionately addressed issues that include the lack of scrutiny in council meetings and reports, governmental data scarcity, power abuse, administrative bureaucracy, communal deprivation, or civic disengagement. Local data also strengthened local or national identity, as I found in the case of the devolved nations of the United Kingdom.

At a reflexive level, data journalists tend to advocate journalistic mind-sets over computational ones. Notably, as web-scrapers and similar automations become regular aspects of British data journalism, data gathered through these techniques are infused with the flexible philosophy of computerized methods, fostering, as a result, problem-solving and creative ways of finding, compiling, and understanding unstructured informational data. The adoption of such methods also provides a wider range of alternative sources of data, which allows data journalists to cover more diverse topics, thus overcoming the topical saturation of open/FOI data. In addition, this wider range of sources can help journalists expose corporate wrongdoing, placing private institutions under similar degrees of scrutiny to those experienced by public power holders.

Hence, I would argue that the optimum reflexive and performative state of data journalism practice emerges when a *nose for news* and computational cognition blend to generate a hybrid form of journalism. While a group of data journalists decides to move only within the constricted architecture of public institutions to follow the limited protocols of FOI Act regimes, another group exhibits a capacity to find stories in the back end of the largest piece of informational architecture of our times: the World Wide Web. These data journalists understand that within this architecture, protocols are highly flexible and demand a different approach from data journalists. Information is hidden in plain sight and is inaccessible to the majority of users, not because it is protected in a black box but because the public is not data literate enough to identify and untangle this information. This small number of data journalists reverse-engineer black boxes, find exploits in the system, efficiently overcome encoding errors, see structure in unstructured information, turn metadata into the variables of a spreadsheet, parse text, exhibit an awareness of the ethical and legal boundaries around database fusion or web-scraping, publish their analysis in Jupyter notebooks to enhance their accountability, seek advice in programmer communities such as Stack Overflow, and organize hackathons to team up with activists, coders, programmers, and ethical hackers to solve intricate data journalism problems.

### Operation: the invisible user

Before proceeding to the last section of this chapter and discussing the platform level of data journalism, I will review, one last time, the definition of data journalism that I provided in the introduction. Bearing in mind that data journalism is a type of news reporting through which journalists engage with computational data to inform their publics, this section will focus on the last segment of the definition, the informational experience facilitated by data journalism, and it will question whether members of the public significantly interact when they engage in this experience.

When it comes to data, the relationship between data reporters and their publics appears to be an active one. Traditionally, journalists have been entrusted by the public to retrieve, interpret, and translate complex data that otherwise would remain cryptically inaccessible to them. This is probably the reason why many of the senior and most experienced journalists tend to perceive data (statistics) as an integral part of their trade.

I have already argued that data journalists have embraced a model of collaboration that enables them to cooperate, together with other actors, with members of their public. This type of collaboration arguably makes their news production process more transparent, open, and participatory. It is also believed that, through openness and transparency, their audiences are more aware of data journalists' methods of gathering, cleaning, and analyzing data, and, as a consequence, audiences are exposed to elementary knowledge on data processing and sorting.

Notably, regardless of data journalists' efforts to decode mathematical abstractions to make them understandable for audiences, to create visualizations in order to present figures and numbers, and to share their methods and databases for public scrutiny, empirical evidence seems to suggest that participation and engagement are not as widespread as was first anticipated.

For example, most of the data journalists interviewed who had attempted crowdsourcing expressed that in most cases calls for members of the public to participate failed to attract numbers significant enough to deem the practice successful. These journalists attributed responsibility for this lack of engagement with crowdsourcing to the common denominator of fragmented local audiences who normally prefer to consume news from newspapers, radio, or TV and tend to be disengaged from online news.

Further, numeracy, statistical literacy, or data literacy are seen as an exclusive skillset of mathematicians, statisticians, or social, computing, or data scientists, and a broader and more critical understanding of data as a cultural artifact is limited for citizens overall. As long as data literacy remains relatively low among members of the public, data journalists will report to a vacuum. Therefore, despite efforts to translate complex data and enlighten the public, data are likely to remain meaningless in the eye of the citizens, thus preventing data insights from powering civic deliberation.

Let us revisit *Ampp3d* to illustrate the effects of public disengagement from data journalism. During its two years of existence, *Ampp3d* was highly respected by data journalists at other news outlets. This proven success, able to attract some 7 million unique browsers a month (Ponsford, 2014), was discontinued by Trinity Mirror executives in 2015 (Jackson and Sweney, 2015). This was met by outcry and criticism from data journalists (Bradshaw, 2015). A Trinity Mirror spokesperson expressed that *Ampp3d* belonged to a niche area with very high costs for the volume of traffic that it produced; therefore, in order to guarantee the sustainability of their digital infrastructure, they would invest in an integrated digital team that focused on areas with more mass-market appeal (Jackson and Sweney, 2015).

Unfortunately, cases such as this call into question the actual audience appeal of data journalism and contest the effectiveness of its interactive, transparent, and open facets. It suggests that despite the expectations of cooperation, which would bring data journalists and members of the public together as companions in the quest for uncovering data insight and exposing corporate and governmental wrongdoing, in reality citizens are more likely to only come into contact with data journalists as digits in a database underpinning data-driven news pieces.

### **Platform: concluding remarks**

Data is unquestionably one of the key areas for productive development, as governments and industry alike face constant challenges in relation to the management of big data on the one hand and exploring creative and innovative ways of engaging with data for the development

of society on the other. As society continues exploring the complexities of computerized data, there is growing uncertainty around data literacy and its insertion in educational policy, in curriculum design, and life-long learning. Research has shown that data literacy levels amongst data journalists also vary. In this regard, there is a tension between two data literacies: a more instrumental one that relies on arithmetic and statistics and a more computational one that employs problem-solving and programmatic thinking. Drawing on this initial distinction, this section examines how data cultures, data awareness, data power, and critical perspectives of data, code, and algorithms mediate the data literacy of both data journalists and their publics.

It is clear that a healthy and wide-ranging data literacy will remain unreachable as long as the traditional approach of using numbers as hard facts and computers as tools for data processing persists. Following the discussion in the previous section, I argue that data literacy is essential not only for journalists to perform data journalism but, more importantly, for the public to make sense of an increasingly datafied world, where rules are dictated by automated agents and their protocols and power is mediated by data brokers.

In order to facilitate the transition from instrumental approaches to data to a more critical approach to data, it is essential that citizens are able to distinguish the main data holders and data gatherers – those responsible for granting not only access to data but also to the technologies that process it. At present, data is mediated and processed by different types of institutions with different kinds of agendas. Within public sector organizations, formal gatekeepers have significant power, able to grant or deny access to information (such as labeling a request vexatious). Furthermore, the degree of citizen access to open data held by governments is small in certain contexts (García and Gertrudix, 2011). For all these reasons, it is paramount to understand the gatekeepers who perform this mediation, the institutions within which they function, the principles and policies that guide their work, and the impact of these processes and interactions on citizen *empowerment*.

Making use of data, therefore, not only requires ever increasing rates of data literacy but also granting citizens access to data to facilitate their interaction with data infrastructures. As individuals actively engage with data, data pervade their data awareness and enhance their computational cognition. Factors such as (a) emergent forms of machine-mediated communication with smart assistants such as Siri and Alexa; (b) the prevalence of wearable tracking systems; (c) the implantation of sensors in our everyday activities, driven by a disproportionate emphasis on ‘Internet of Things’ technologies; and (d) the fast and steady development of cognitive computing not only disrupt the ways we perceive these pervasive technologies but also the ways we interact with these technologies as everyday life companions.

For this reason, I tend to make no distinction between the reflexivity of both data professionals and the intervening automations involved in the production of data outputs, because in my view they display similar mind (and system) processes and similar biases and perform decision making following similar patterns and stages. We have senses and a brain; they have sensors and cognitive computing. In fact, by approaching data as a material force similar to that of evolution, we can acknowledge that data and algorithms are strong determinants of our destiny as human beings.

For instance, web metrics have become a powerful aid for news editors to decide what becomes news. Algorithms measure which news stories are most liked by the public, and that insight is then used in editorial decision making to create the news agenda. It is clear then that matters of public concern, discussed every day, are nowadays strongly mediated by automated technologies unknown to the public. Therefore, it is important to understand the instrumental dynamics of data; namely the statistical operations that govern algorithmic decision making or the irrelevance of sampling for big data processing. However, by acknowledging data materiality beyond its object status, we can better understand the role of embryonic forms of AI in the production of knowledge, the negotiation of societal power, the parameterization of the world and social life, the potential remediation of humans, and algorithmic accountability.

As the effectiveness of data organizations in monetizing the insights derived from citizens' data increases, so does the power they hold over not only the individuals but also over the institutions of society. Corporations such as Google and Facebook, with a core focus on quantifying everyday life, have programmed algorithms capable of profiling and predicting people's hopes and dreams in an environment free of public or institutional scrutiny. In the past, this watchdog function was performed by news media as part of a healthy democratic society. Nowadays, news organizations seem to be unable to monitor the contemporary institutional negotiation of data power, as it arises in a scenario only accessible to actors with an understanding of the computational and legal loopholes in the system.

During my research, it was clear that data journalists were fully aware of the power dynamics driven by emergent data brokers. However, most of them felt that they could only contribute to uncovering wrongdoing within these domains if they were collaborating with beat/specialist journalists, those traditionally commissioned to cover these areas: business, technology, and science. Notably, beat reporters were perceived by most of the data journalists I interviewed as the most suitable professionals to investigate the behavior of news subjects, sources, and news events within the confines of data power arenas.

To conclude, I argue that there is an imbalance in the rhetoric, wherein big data is predominantly framed as the epicenter of contemporary innovation and the driving force of societal progress. This prevalent and tacitly neoliberal discourse appears to safeguard corporate data commodities to the detriment of overly exploited public data, making the data generated within public bodies and by citizens a profitable good to nourish. This dynamic emphasizes a major gap between the data brokers who, through a series of furtive transactions, exploit and commoditize citizens' data, and citizens who display varying degrees of awareness regarding the amount of data they generate on social media or that is generated by public bodies, and how that data is utilized by increasingly powerful data brokers. Notably, a number of non-legacy entities and civic-driven initiatives operating in platforms such as GitHub or Hacks/Hackers (see Usher, this volume, Chapter 26) and start-ups such as *scrapewiki* are now essential spaces and proficient enablers of data-driven public deliberation and civic engagement, thus fostering an ever-expanding rate of data literacy and improving public participation. Platform studies offers a novel perspective on the study of data journalism that sheds light on a series of underlying dynamics commonly unnoticed by scholarship on the field. It has the potential to make researchers aware of those entities that play a vital role in data journalism practice but are generally disregarded by human-centric approaches, with their emphasis on the attitudes and behaviors of data reporters.

### Further reading

In order to further explore some of the ideas discussed in this chapter, it is probably a good idea to review some of the titles included in Nick Montfort and Ian Bogost's MIT Press book series *Platform Studies*, as well as Thomas Apperley and Jussi Parikka's (2015) epistemic examination of the Platform Studies Methodology contained in their article "Platform Studies' Epistemic Threshold". Two seminal journal special editions that examine the relationship between journalism practice and data materialities are Seth C. Lewis and Oscar Westlund's "Journalism in an Era of Big Data: Cases, Concepts, and Critiques" (2015a) and CW Anderson and Juliette De Maeyer's "Objects of Journalism and the News" (2015).

### References

- Anderson, C. W., and De Maeyer, J. (2015) "Objects of journalism and the news." *Journalism*, 16(1), 3–9.
- Appelgren, E. and Nygren, G. (2014) "Data journalism in Sweden: Introducing new methods and genres of journalism into 'old' organizations." *Digital Journalism*, 2(3), 394–405.

- Apperley, T. and Parikka, J. (2015) "Platform studies' epistemic threshold." *Games and Culture*, 1, 22.
- Borges-Rey, E. (2016) "Unravelling data journalism: A study of data journalism practice in British newsrooms." *Journalism Practice*, 10(7), 833–843.
- Borges-Rey, E. (2017) "Towards an epistemology of data journalism in the devolved nations of the United Kingdom: Changes and continuities in materiality, performativity and reflexivity." *Journalism* [online first ahead of print]. Retrieved from <http://journals.sagepub.com/doi/abs/10.1177/1464884917693864>
- Bradshaw, P. (2015) "The legacy of Ampp3d, UsVsTh3m and RowZed." *The Online Journalism Blog (UK)*. Retrieved from <https://onlinejournalismblog.com/2015/05/13/the-legacy-of-ampp3d-usvsth3m-and-row-zed/>
- Broersma, M. (2010) "Journalism as performative discourse. The importance of form and style in journalism." In V. Rupaar (ed.), *Journalism and Meaning-making: Reading the Newspaper*. New York, NY: Hampton Press (pp. 15–35).
- Broussard, M. (2015) "Artificial intelligence for investigative reporting: Using an expert system to enhance journalists' ability to discover original public affairs stories." *Digital Journalism*, 3(6), 814–831.
- Carlson, M. (2015) "The robotic reporter: Automated journalism and the redefinition of labor, compositional forms, and journalistic authority." *Digital Journalism*, 3(3), 416–431.
- Clerwall, C. (2014) "Enter the robot journalist: Users' perceptions of automated content." *Journalism Practice*, 8(5), 519–531.
- Coole, D. and Frost, S. (eds.). (2010) *New Materialisms*. Durham, NC: Duke University Press Books.
- Deleuze, G. and Guattari, F. (1987) *A Thousand Plateaus*. Translated by Brian Massumi. Minneapolis, MN: The University of Minnesota Press.
- De Maeyer, J., Libert, M., Domingo, D., Heinderyckx, F. and Le Cam, F. (2015) "Waiting for data journalism: A qualitative assessment of the anecdotal take-up of data journalism in French-speaking Belgium." *Digital Journalism*, 3(3), 432–446.
- Fink, K. and Anderson, C. W. (2015) "Data journalism in the United States: Beyond the 'usual suspects'." *Journalism Studies*, 16(4), 467–481.
- García, F. and Gertrudix, M. (2011) "Naturaleza y características de los servicios y los contenidos digitales abiertos." *CIC Cuadernos de Información y Comunicación*, 16, 115–138.
- Gray, J., Chambers, L. and Bounegru, L. (2012) *The Data Journalism Handbook: How Journalists Can Use Data to Improve the News*. Sebastopol, CA: O'Reilly Media.
- Hamilton, J. T. and Turner, F. (2009, July) "Accountability through algorithm: Developing the field of computational journalism." *Report From the Center for Advanced Study in the Behavioral Sciences, Summer Workshop* (pp. 27–41).
- Jackson, J. and Sweney, M. (2015) "Trinity Mirror's UsVsTh3m and Ampp3d thought to be facing axe as jobs set to go." *Guardian (UK)*. Retrieved from [www.theguardian.com/media/2015/may/13/trinity-mirrors-usvsth3m-and-ampp3d-thought-to-be-facing-axe-as-jobs-set-to-go](http://www.theguardian.com/media/2015/may/13/trinity-mirrors-usvsth3m-and-ampp3d-thought-to-be-facing-axe-as-jobs-set-to-go)
- Latar, N. L. (2015) "The robot journalist in the age of social physics: The end of human journalism?" In G. Einav (ed.), *The New World of Transitioned Media: The Economics of Information, Communication, and Entertainment (The Impacts of Digital Technology in the 21st Century)*. Cham: Springer International Publishing (pp. 65–80).
- Lewis, S. C. and Westlund, O. (2015a) "Journalism in an era of big data: Cases, concepts, and critiques." *Digital Journalism*, 3(3), 321–330.
- Lewis, S. C. and Westlund, O. (2015b) "Big data and journalism: Epistemology, expertise, economics, and ethics." *Digital Journalism*, 3(3), 447–466.
- Lister, M., Dovey, J., Giddings, S., Grant, I. and Kelly, K. (2009) *New Media: A Critical Introduction*. Abingdon: Routledge.
- Manovich, L. (2001) *The Language of New Media*. Cambridge, MA: The MIT Press.
- Miller, D. (2005) *Materiality*. Durham, NC: Duke University Press.
- Montfort, N. and Bogost, I. (2009) *Racing the Beam: The Atari Video Computer System*. Cambridge, MA: The MIT Press.
- Ponsford, D. (2014) "Mirror rewards Coles and Belam with new jobs in wake of UsVsTh3m and Ampp3d successes." *Press Gazette (UK)*. Retrieved from [www.pressgazette.co.uk/mirror-rewards-coles-and-belam-new-jobs-wake-usvsth3m-and-ampp3d-successes/](http://www.pressgazette.co.uk/mirror-rewards-coles-and-belam-new-jobs-wake-usvsth3m-and-ampp3d-successes/)
- Van Dalen, A. (2012) "The algorithms behind the headlines: How machine-written news redefines the core skills of human journalists." *Journalism Practice*, 6(5–6), 648–658.
- Warnes, S. (2014) "Hey DWP! We just fixed your chart for you." *Ampp3d (UK)*. Retrieved from [www.mirror.co.uk/news/ampp3d/hey-dwp-just-fixed-your-4584836](http://www.mirror.co.uk/news/ampp3d/hey-dwp-just-fixed-your-4584836)
- Yonago, J. T. (2014) "Data is the new oil of the digital economy." *Wired*. Retrieved from [www.wired.com/2014/07/data-new-oil-digialeconomy/](http://www.wired.com/2014/07/data-new-oil-digialeconomy/)