

Research Article

Between Objectivity and Openness—The Mediality of Data for Journalism

Frédéric Lesage* and Robert A. Hackett

School of Communication, Simon Fraser University, 8888 University Drive, V5A 1S6 Burnaby, Canada;
E-Mails: flesage@sfu.ca (F.L.); hackett@sfu.ca (R.A.H.); Tel.: +1 7787829360 (F.L.)

* Corresponding author

Submitted: 19 June 2013 | In revised form: 19 December 2013 | Accepted: 8 January 2014 |
Published: 30 January 2014

Abstract: A number of recent high profile news events have emphasised the importance of *data* as a journalistic resource. But with no definitive definition for what constitutes data in journalism, it is difficult to determine what the implications of collecting, analysing, and disseminating data are for journalism, particularly in terms of objectivity in journalism. Drawing selectively from theories of mediation and research in journalism studies we critically examine how data is incorporated into journalistic practice. In the first half of the paper, we argue that data's value for journalism is constructed through mediatic dimensions that unevenly evoke different socio-technical contexts including scientific research and computing. We develop three key dimensions related to data's mediality within journalism: the problem of scale, transparency work, and the provision of access to data as 'openness'. Having developed this first approach, we turn to a journalism studies perspective of journalism's longstanding "regime of objectivity", a regime that encompasses interacting news production practices, epistemological assumptions, and institutional arrangements, in order to consider how data is incorporated into journalism's own established procedures for producing objectivity. At first sight, working with data promises to challenge the regime, in part by taking a more conventionalist or interpretivist epistemological position with regard to the representation of truth. However, we argue that how journalists and other actors choose to work with data may in some ways deepen the regime's epistemological stance. We conclude by outlining a set of questions for future research into the relationship between data, objectivity and journalism.

Keywords: data; data journalism; mediality; regime of objectivity

1. Introduction

The recent high profile success of projects like the Guardian's *Reading the Riots* and the growing legitimacy of independent investigative organizations such as ProPublica highlight how *data*—its collection, analysis, and communication—are a major point of interest and concern in contemporary journalism. With no definitive definition for what constitutes data in journalism coupled with the existence of numerous labels for data-related journalistic practices (such as data journalism (DJ), data driven journalism, database journalism, computational journalism, data visualization) understanding data's place within journalism is problematic.

The starting point for this paper is that as the sophistication and accessibility of digital technologies for the collection, analysis and dissemination of data have become more widespread, so have the number of projects that turn to data for the production of news. Data's increasing importance within journalism raises a number of interesting questions and challenges, not least of which are the implications such data has for objectivity as one of the paradigmatic concerns of contemporary journalism. Data's meaning and value arguably stems from the extent to which it is said to be objective. But if objectivity's place within journalism is itself the source of much debate [1–3] then we must also question how data is imbued with the quality of objectivity within journalism. Instead of a history of practices like DJ or a sociological analysis of such practices, this paper draws from two different approaches—theories of mediation and journalism studies (especially political economy and media sociology approaches)—in order to question what constitutes data and how the different choices regarding its collection, interpretation and dissemination have implications for objectivity in contemporary journalism. The first part of the paper examines the connection between data and objectivity by focussing on digitally mediated data as an object used by journalists in ways that evoke socio-technical contexts in which objective data is produced—what we refer to as the mediality [4] of data. The second part of the paper delves into how the political economy of contemporary Western journalism shapes the production of objectivity [2] as a multifaceted regime. This second approach enables us to contemplate the implications that the different facets of this regime might have for data as a source of objectivity in contemporary journalism. In the final section, we put forward future research questions that build on these two approaches.

2. Data's Mediality

The term *data* is frequently applied in journalism literature as a mass noun. The Oxford English Dictionary provides two different definitions of the application of this term:

a. Related items of (chiefly numerical) information considered collectively, typically obtained by scientific work and used for reference, analysis, or calculation.

b. Computing. Quantities, characters, or symbols on which operations are performed by a computer, considered collectively. Also (in non-technical contexts): information in digital form. [5]

Both of these kinds of data have historically played a role in journalism. Journalists have long drawn on the outputs from scientific investigations as a resource for the production of news. Similarly, journalists have been developing techniques for using computers to analyse data since the late 1960s and early 1970s like precision journalism [6] and computer-assisted reporting [7]. A decade ago, scholars like Deuze ([8], pp. 8–9) pointed to the emergence of "open-source journalism" as a potential direction for new configurations of participation in journalistic practices. For Deuze, the Internet represented a new journalistic medium that afforded the opportunity to build communities of information gathering and dissemination similar to those of the open-source software community. Even more recently, Hamilton and Turner ([9], p. 2) defined computational journalism as 'the combination of algorithms, data, and knowledge from the social sciences to supplement the accountability function of journalism'. While similar in many respects to computational journalism, DJ's central preoccupation is how to produce news with data. As Bradshaw [10] puts it in the introduction to *The Data Journalism Handbook*:

'Data can be the source of data journalism, or it can be the tool with which the story is told—or it can be both. Like any source, it should be treated with skepticism; and like any tool, we should be conscious of how it can shape and restrict the stories that are created with it.' [10]

Implicit in Bradshaw's definition is that key aspects of journalistic practice and the values that underpin these practices—how to treat a source, telling stories—remain intact despite the fact that they involve the use of data. Our objective is not to determine to what extent DJ itself represents a genuine departure from its predecessors. Instead, we set out to problematize how practices and values involved in the collection, interpretation, and dissemination of data are mediated through current journalistic practice and values.

Sterne [4] uses the concept of mediality to examine how things 'evoke a quality of or pertaining to media and the complex ways in which communication technologies refer to one another in form or content' ([4], p. 9) and how these ways are articulated 'with particular practices, ways of doing things, institutions, and even in some cases belief systems'. ([4], p. 10). Building on this definition, we use mediality to ask how journalists treat data in ways that refer to forms or content of other socio-technical contexts. Conceptualising the mediality of data means problematizing

how data may at once evoke some of the symbolic and material qualities or practices taken from scientific enquiry or computation as presented in the above Oxford English Dictionary definition while also evoking the qualities and practices of news content produced and interpreted through journalistic forms with all of their political, cultural and technological baggage. Sterne's definition of mediality is useful because it highlights that we are not dealing with a *whole* medium like television, the Internet or newspapers. Data is in some ways both more specific and more abstract than such media. In order to clarify the implications of our chosen approach, we identify and develop three interconnected variable dimensions of data's mediality for journalism.

2.1. *The Problem of Scale—Defining the Proportional Relations of Data in Journalism*

For Rosen [11], journalism is a response to a 'problem of scale'. People, as part of a 'self-informing populace', are unable to consider distant current events and so turn to journalism as a way of understanding what is happening in the present-day world. Rosen extends his notion of scale beyond only physical distances to encompass all of the complexity of economic, political and social systems that come with the modern condition: what he terms the 'awayness' of things. The journalist's authority, he argues, stems from being able to claim a special perspective on the awayness of things and then relate this perspective to the public. As Rosen puts it:

'I'm there, you're not, let me tell you about it.' Or: 'I reviewed those documents, you couldn't—you were too busy trying to pay the mortgage—so let me tell you what they show.' ([11], p. 30)

Contemporary texts often represent digitally mediated data as part of a similar problem of scale: the coming 'data deluge' [12], 'working with data is like stepping into vast, unknown territory' [13], or 'huge tracts' [14] of data. Digital data's mediality as a large mass evokes the unknown quantity of ones and zeros that are so often used to symbolise the digital. This problem of scale can be used to justify an authoritative journalistic role in which the journalist can answer the public's questions about data. For example, Stolte presents digital journalists as key intermediaries who can tackle 'the sheer scale' of data by making large amounts of it accessible to the public in order to enable this public to 'receive the information without being overwhelmed by it' ([15], p. 357).

But the relationship between data's scale and the journalist's authority is one that needs to be carefully considered. As Webster ([16], pp. 21–25) and Mosco ([17], p. 50) remind us in their critical examinations of digital technologies, problems of scale can often be mobilised as ideological discourses to mask deeper political and social inequities.

For Couldry and McCarthy [18], differences of scale in the media can be understood as *proportional relations* that make up the different levels of media forms and content. To understand these relations requires that we remain attentive to the multiple ways in which they are brought together. A first step towards such an understanding in the case of data and journalism involves tending to the proportional relations between data and those involved in its production, dissemination and interpretation. For example, in their case study of a series of data-related projects in a Chicago newsroom, Parasie and Dagiral [19] recount a debate between two groups of journalists regarding how to work with data. The first group of journalists treated the quantities of data as a particular kind of computational problem; a problem that could be resolved by designing the right kind of platforms for accessing and analysing data. These platforms would be designed to provide the public with individualised access to *complete* datasets at a *granular* level, allowing individuals to analyse the data to see how it affected them personally. By contrast, a second group of journalists in the newsroom emphasized the importance of providing the public with inferential statistics based on the journalists' own analysis of a sample of the data; an approach closer to social-scientific traditions of data analysis. This debate between both groups of journalists illustrates two very different perspectives on data's problem of scale, and how to resolve this problem.

The repercussions of changes in scale are not predetermined: how different actors engage in the mediation of different levels of scale are not only potential sources of inequality but also represent opportunities for alternative forms of engagement, for resistance, and for change. Parasie and Dagiral's case study highlight two very different technological and organisational options for defining the proportional relations between journalists, data and the public with very different implications for all three. Our second dimension of mediality turns to the question of how different technological and organisational configurations work together.

2.2. *Transparency Work—How the Collection, Analysis and Delivery of Digital Data Work Together as News*

To count as news, data must be subjected to processes of refinement. As our second dimension, we use *transparency work* to examine the way in which these processes of refinement are materially and symbolically ordered as part of data's production and reception. In the context of journalism *transparency* refers to making publicly available the sources, interests and methods that might influence the information presented, so that notionally, readers/viewers (as *rational* subjects) can take potential *bias* into account in their own interpretation of the account. In this case, our definition draws from science and techno-

logy studies where it is used to describe a 'process in which status, cultural and community practices, resources, experience, and information infrastructure work together' ([20], p. 257). Work to make certain aspects of data transparent, like the transparency of media forms [21] or of information systems [20], relies on social and technological standards that may have very different meanings for different people.

A basic example for illustrating transparency work for data is information visualization. Much like scientific visualizations, journalists present datasets in the form of visual diagrams that highlight the insights they wish to communicate to the public. In some cases, visualizations take the form of interactive graphics that facilitate data analysis for the general public. Interactive graphics prescribe a certain way of interacting with the datasets, making it easier for someone who is unfamiliar with data analysis to gain insights from the data. But someone who is able to conduct their own independent analysis of the datasets may interpret these same visualisations as too constraining or prescriptive.

A more complex example of transparency work with data is the provision of *raw data* as an accompaniment to a news story. For example, the Guardian's *Data Blog* [22] gives readers access to datasets online and invites readers to 'download the data' in order to conduct their own analysis. The process of making this data available to the public builds on open source principles discussed below. But this data's 'rawness' is a relative state that depends on its own refinement processes. The way in which journalists collect and format their raw data in order to present it to the public depends on a number of implicit and explicit standards, practices, and values in the same way as with information visualisation graphics. For example, the journalist may decide to clean up or format the raw data before making it available to the public. The difference between data visualizations and raw data is that providing raw data can be interpreted as an invitation to reinterpret or challenge the results of the analysis of a dataset. But while the standards for using data analysis to challenge results may be familiar to those trained in such techniques, it is unlikely to be a set of skills and knowledge that is widely available to the general public.

Transparency work does not only take place between journalists and the public. Producing news items with data also entails refinement processes among journalists. Cohen, Hamilton, and Turner, for example, deem the efforts that go into converting data from paper documents or other primary sources to be the "bothersome impediments of more interesting work" ([23], p. 71) that is possible once such primary sources have been digitised and converted into a format that can easily be analysed. Cohen, Hamilton, and Turner recommend developing more accessible methods and tools for journalists who are unfamiliar with data analysis in order to facilitate their work. These platforms would make certain aspects of data analysis transpar-

ent to novice journalists.

We recognize that a certain amount of transparency work is, to a greater or lesser extent, always involved in data collection, analysis and dissemination. But considering transparency work with regards to data raises questions for the politics of producing different kinds of transparency, particularly in light of the problem of scale discussed above. What values and objectives inform the decisions regarding transparency work? In the following section, we examine how 'openness', as a set of values based on the provision of access to data, represent a third dimension of data's mediality in journalism.

2.3. Openness: Extending Access to Data

It is said that files saved in the Portable Document Format (PDF) are where 'data goes to die' [24]. Such a claim is arguably exaggerated, but data journalists and programmers base it on the fact that data stored in PDF files are not as easy to access as data stored using other file formats. There currently exists a movement within a number of different institutions that emphasises making data more *open* in part by ensuring that data is not stored in these kinds of formats. A detailed discussion of the term *open data* is beyond the scope of this paper. The history of open data has close ties to the history of computing including software development. Open data's history also builds on the long-established and well-documented academic tradition of peer-review in academic research (for example, see [25] for further discussion). Movements espousing open data often subscribe to a *do it yourself* (DIY) ethos. In the context of journalism, this implies that if a reader is unconvinced or suspicious of the conclusions drawn from the data for a news story, they are given free rein to analyze the raw data themselves and draw their own conclusions. What constitutes open data for journalists is still the subject of debate but here is an example of a definition:

'structured primary information from an organization—meaning unfiltered and complete information—provided in an accessible, machine-processible, non-proprietary, license-free format' ([26], pp. 17–18).

Such definitions and the different ways in which they can be implemented as part of journalistic practice have serious implications for how people access data. For the purposes of this paper, we define *openness* as 'efforts to extend access to "data"' ([27], p. 1). This definition of openness draws inspiration from Gurstein's critical examination of open data. For Gurstein, proponents of open data tend to focus on access over other issues, resulting in an understanding of data that is isolated from other social and technological processes. While Gurstein does present a solution to this problem (discussed in Section 3.3) the provision of access to data remains a key concern

among open data enthusiasts. We use openness to examine the different ways in which this provision of access to data, as a set of values and objectives circulating in (among other contexts) academic research and computer engineering, is articulated in the context of journalism.

As an example, open data initiatives to pressure governments to provide the public with greater access to government data have meant that open data enthusiasts and journalists have historically shared an interest in openness [28]. The recent push by some news media organisations to lay bare their raw data suggests unprecedented moves to editorial openness ([3], p. 196) that extend the open data movement to journalism itself. In such cases, disclosure about the sources of data is assumed to improve accessibility, and to enable the public to make better judgments as to the trustworthiness and truth-value of news. Emphasising openness represents a qualitative shift from practices and processes whose apparent objectivity and credibility derives from *authoritative* sources, to practices and processes that ensure the openness of data. But, as we will see in the following section, what constitutes openness for journalism is still contested and may lead to diverging approaches [28]. We stress the distinction between transparent *raw data* and *open data* to highlight these different trajectories in its production and circulation.

2.4. Does Data Make Journalism More Objective?

To date, we have consciously discussed data's mediality in journalism without concern for whether or not these different dimensions have implications for data's status as a source of objectivity. The meaning of data may be familiar in the socio-technical contexts of scientific enquiry and computation but data's production, circulation and interpretation within the context of journalism cannot simply be understood as a straightforward and unproblematic transplant from these or any other contexts. The problem of scale, how transparency work takes place, and how to ensure openness are all examples of variable dimensions of data's mediality: the contingent ways in which data can be used in the context of journalism while evoking qualities and/or practices taken from empirical research or computation. While such dimensions may to a greater or lesser extent implicitly rely on data's status as objective, they do not in themselves ensure objectivity. The implications of data's mediality for its status as a source of objectivity are made all the more complicated if we consider how journalism has its own longstanding methods and technologies for producing objectivity. In the second part of this paper, we therefore turn to a multifaceted journalism studies model of the production of objectivity within journalism in order to reflect on how such a structure may in turn shape data's place in journalism.

3. Data, Journalism, and the Objectivity Regime

Objectivity in journalism, like data, is not a single, *fixed thing* but can include a range of meanings amongst different journalists in western liberal-democracies: in some cases it might refer to how journalists negate their subjectivity, in others it refers to ensuring the fair representation of opposing sides in a controversy and maintaining a sceptical approach towards all sides in a dispute, in yet others it refers to the provision of facts in order to contextualize an issue [1]. The historical sources of objectivity, and the periodization of its emergence are much debated [3]. The history of objectivity as a key concern in Anglo-American journalism can partly be attributed to the incorporation of technologies like the telegraph and photography into journalistic organisational forms like wire services in the 19th century. Mass-market advertising is also said to have greatly contributed to a declining support for a partisan press in the same period.

In this section, we explicate the *regime of objectivity* as a dominant, yet contested [29], North American [30] journalistic paradigm. As outlined by Hackett and Zhao ([2], pp. 82–88), in their conception, US journalism has been characterized by the hegemony of a discursive 'regime of objectivity' for much of the 20th century:

'The idea-complex—and set of practices—of journalistic objectivity...provide a *general model for conceiving, defining, arranging, and evaluating news texts, news practices, and news institutions.*' ([2], p. 86)

In Hackett and Zhao's view, it is a polysemic, contested and flexible idea-complex or discursive/institutional regime, with five interacting levels or elements: (1) a normative ideal (concerning both cognitive and evaluative dimensions of news); (2) an epistemology; (3) newsgathering and presentation practices, both reportorial and editorial; (4) a set of institutional relationships, such as to create the impression of journalism's autonomy from illegitimate outside pressures or internal imperatives (e.g. the separation of 'church and state' between editorial and advertising/marketing departments); and (5) an active ingredient in public discourse. The objectivity regime reinforces the journalist's claim to authority as a legitimate intermediary between the public and world events by presenting the journalist's account as *universal* and *neutral*. But objectivity as constructed through the objectivity regime also sustains what some would call a hegemonic ideology [3] that consolidates power for a few dominant actors, and for conventional social values.

Journalism is currently in profound transition, with multiple paradigms competing with the regime of objectivity, which is arguably on the wane [31]. However, digitally mediated data represents at once an opportunity for positive changes to journalism's objectivity regime and a risk that new inequities will take shape or established ones will be reinforced. It is

therefore essential that we consider how the different facets of the objectivity regime produce objectivity in order to begin to consider how such structures may enable or constrain the meaning of data.

3.1. Data and the Objectivity Regime's Normative Ideal

The normative ideals of the objectivity regime prescribe certain traits to objectivity in journalistic practice: detachment, impartiality, avoiding personal biases and interests, etc. [32]. We find that these and similar traits still apply to DJ including originality, independence, statements grounded in facts that are verified by journalists ([33], p. 187), the criteria of utility, reliability, trustworthiness ([33], p. 189) and scepticism [10]. Data provides a factual basis for analysis, attempts to minimize the risks of incorrect reporting [9], and represents the potential to counter the influence of public relations. The same 'fundamentals' of journalism are in play in DJ literature as they have been for journalists in the objectivity regime: editorial decision making, fact-checking, ethics, storytelling.

In some respects, data journalists' push for greater openness de-emphasizes certain aspects of what used to be an important form of social or cultural capital for journalists—their relationships with individual sources, their Rolodex (a pre-internet metaphor) as a semi-secret treasure chest of authorities or whistle-blowers they could employ to enhance their professional capital, and credibility. But data also depends on a greater emphasis on certain well-established ideals of the objectivity regime such as accountability. Traditional news media achieved this ideal through practices such as editorial corrections of factual errors, the interventions of ombudsmen and publication of readers' responses to stories. One of the ways in which data can be used to ensure greater accountability is through greater openness afforded by giving the public access to raw data. This type of openness draws on normative ideals from sources outside journalism and adds new ethical touchstones by enhancing the perceived validity of journalists' truth claims. The danger in such a development, however, is that it may further absolve journalists from taking responsibility for what McChesney calls the 'inescapable part of the journalism process' ([34], p. 302), namely deciding what counts as news. In cases where the public is only given access to raw data and the means to analyse it without the journalist's explicit claim of what is significant about this data, the journalist is effectively offloading the responsibility of understanding the data's significance onto the public.

3.2. Data and the Objectivity Regime's Epistemology

Part of the objectivity regime thesis posits that contemporary journalism, particularly as practiced in Anglo-American liberal democracies, depends on a compromise between a positivist faith in facts, and an

emphasis on balancing various points of view that implies an epistemological position of conventionalism, one that asserts the incommensurability of conflicting discourses [35]. At first glance, data journalists may seem to challenge positivism by taking a more conventionalist epistemological position with regard to the representation of truth. The truth-value of a story no longer depends exclusively on the stance of an individual reporter as an independent, neutral, detached, skilled observer. The collection and analysis of data in some DJ projects constitutes a collective enterprise where data collection is crowd-sourced and the analysis is participatory (for example, the Guardian's *Reading the Riots*). In such projects, news becomes iterative and dialogic: the data co-exists with the story, alongside it, and new insights gleaned from its analysis have the potential to modify the story.

Participatory forms of DJ are similar to other forms of online journalism in that they suggest a kind of postmodernist approach where journalists and the public create reality through language and interactions thereby transforming notions of truth seeking in journalism: participation and involvement trump distance and detachment ([3], p. 195). However, the ways in which data journalists implement openness may in some ways deepen the regime's positivist epistemological stance. As noted above, the provision of raw data is used to increase the perceived validity of truth claims by basing them on methods imported from scientific research and computing. That importation is an important aspect of data's mediality within journalism, and heightens the impression that the story being told is in principle empirically falsifiable (i.e., testable against empirical evidence). Just as part of the objectivity regime's epistemology was indicative of modernist journalism, data journalists' commitment to facticity means that they reproduce the incumbent *news net* [36]: reality can be described through careful, systematic analysis of data.

For Simon Rogers, the Guardian's former editor for the Data Blog and a major figure in DJ circles, the implications of this implementation of openness for epistemology remain consistent with established journalistic tradition as long as such implementation entails giving the public as much detail about the provenance of the data used to produce a news story:

'Data can be as subjective as anything else, because the choice of some types of data over others, or choice of stories, is based on my prejudices. But we have to try to be objective. There is a purity of reporting to it that is quite traditional. We put caveats in our stories about the data: Who gathered it? What do we know about how it was collected?' [37]

Others see in DJ an opportunity to improve data collection by official institutions through a combination of fact checking data and *watchdog journalism* [9]. Greater computational resources for journalists have decreased the cost associated with doing this type of

'watchdog' coverage and increased the level of public interest for political issues 'by personalizing the impact of public policies' ([9], p. 12). As digital data becomes more prevalent, journalists should extend their watchdog role to this data, recognizing that faith in official sources of data must be tempered by healthy scepticism and that with raw data must also come better indicators of its quality and provenance.

But journalism itself is not as good at extending this watchdog role to its own work with data. Implementing checks on the collection and analysis of data as part of exercising a healthy scepticism towards data relies on the very kind of social scientific epistemological traditions and expertise that are currently being challenged by programmer-journalists. It seems unlikely that reliable indicators of quality and provenance will consistently be put in place when we consider the rather limited extent to which journalists and journalist watchdogs re-examine and correct the use of incomplete or inaccurate data. For example, Messner and Garrison's [38] review of literature on journalism identifies a considerable amount of warnings to journalists about the prevalence of *dirty data* in datasets and advice on how these same journalists should deal with dirty data when writing a news item. But when the two search for instances of fact checking and/or corrections of dirty data in actual reporting, they conclude that:

'The authors are quite alarmed at the lack of attention given to [fact checking and/or corrections of dirty data] in the literature of journalism and mass communication, particularly in the literature of newsgathering. From earlier research about computer-assisted reporting, various conferences and presentations in the past decade and a half, and in discussions with professionals, it was an issue that simply remained below the research radar.' ([38], p. 97)

Finally, data journalists also run the risk of limiting their caveats to source material and to the values of the author without also including caveats as to the methodological biases and epistemological assumptions embedded in the methods used to gather the data (where *gathering* implies that the facts are lying around waiting to be collected). A simple example of such methodological bias can be suggested by the official categorization of *the unemployed* in governmental estimates of the unemployment rate. Such official statistics exclude those who involuntarily work part-time or who have given up looking for work and therefore are no longer categorized as part of the unemployed portion of the labour force.

It seems unlikely that the objectivity regime's unbalanced stalemate between positivist and conventionalist epistemologies will disappear. One of the questions raised by the use of data in journalism is how such a compromise may be reconfigured—for better or worse—by the different ways in which data is collected, analysed and presented.

3.3. Data and the Objectivity Regime's Practices

Rogers writes that DJ is at its core about 'telling the story in the best way possible' [39] rather than about flashy graphics or sophisticated interfaces. Rogers [39] goes out of his way in his definition of DJ to establish that it is an extension of traditional forms of journalism:

'If data journalism is about anything, it's the flexibility to search for new ways of storytelling. And more and more reporters are realising that. Suddenly, we have company—and competition. So being a data journalist is no longer unusual. It's just journalism.' [39]

Rogers stresses a distinction between thinking about data as a journalist and thinking about data as an analyst. This distinction seems to revolve around the continued primacy of the narrative form in the production of news and of the journalist's role as author of these news stories. Such a view is consistent with the objectivity regime in that the journalist is the one imbued with the knowledge and skills required to separate *fact* from *opinion* through the practice of news reporting. Contemporary journalists have developed design and storytelling strategies for producing interactive news items based on data visualization that ensure the kind of *narrative control* supposedly ceded to the reader because of digital media. According to Segel & Heer's [40], analysis of a sample of different kinds of narrative visualizations that include DJ news items:

'Generalizing across our examples, data stories appear to be most effective when they have constrained interaction at various checkpoints within a narrative, allowing the user to explore the data without veering too far from the intended narrative.' ([40] p. 1347)

Both Roger's definition of DJ practice and Segel and Heer's insights into storytelling techniques with data raise the question of how different techniques for the provision of openness in DJ can co-exist with transparency work for data: how to extend *access* to data while also making the insights gained from data analysis *accessible*? Gurstein suggests that while considerable good has come from (and may continue to come from) open data movements, how its proponents choose to pursue its implementation may have unintended consequences that lead to greater inequality. His critical examination of the open data movement leads him to conclude that disparities are appearing between those with access to the right kinds of technology and the knowledge to use such technology and those who do not have such technologies and/or knowledge. So while data may be open, how different actors can engage with open data varies considerably:

'Thus, rather than the entire range of potential users being able to translate their access into meaningful applications and uses, the lack of these

foundational requirements means that the exciting new outcomes available from open data are available only to those who are already reasonably well provided for technologically and with other resources.' ([27], p. 2)

For Gurstein, the processes of interpreting data and subsequently being able to make 'effective use' of this same data are just as important as ensuring access to data. He concludes that any critical analysis of *open data* has to involve questioning how and under what conditions data is contextualized and given meaning ([27], p. 4). In other words, storytelling with data or providing access to raw data cannot be understood in isolation from how those stories or that access are interpreted and in what way those who interpret the data are able to incorporate it into their lives.

One way to connect access to data with its interpretation and use is to align journalistic practice with open data movements that support a DIY approach to data. This realignment could be consistent with the current shift away from journalists having complete authority over the storytelling process and towards what Rosen [41] calls 'the people formerly known as the audience' via crowd-sourcing of data analysis and discussion forums [24]. In such cases, the journalist's role shifts to performing more administrative tasks surrounding the provision of access to data such as *curating* data, managing discussion lists, determining which of multiple blog contributions go to the top, and shaping stories into articles that span more than one publication/edition. At its best, curating data can prove to be a positive solution to the problem of scale by directing audiences to the best datasets and educating them in their use ([33], p. 190). At its least, curating data can simply be a euphemism for the management of data without analysis as discussed above.

Another way to connect openness with the interpretation and use of data may be by providing *context* for a news story. Journalists can use data to introduce more background to stories by taking the focus away from timely events towards providing greater information related to the reported event, but which lies before, after or outside the event itself. Under the strictures of objectivity in US journalism, reporters tend to shy away from providing background context partly from fear of accusations of bias: sticking to the *facts* that journalists observe themselves or that can be confirmed by *authoritative* sources, can be seen as examples of *strategic rituals* [42]. Journalists can use data in this way to move beyond the objectivity regime's event- and official-orientation. But to the extent that data journalists fail to question the assumptions embedded in datasets, or to recognize that *any* selection of a relevant context is inherently political, they may unwittingly reinforce the *frame-blindness* ([3], pp. 66–70) of the objectivity regime.

Concerns for connecting the provision of access to data with its interpretation and its effective use are not limited to the relationship between the journalist

and the public. Journalists face the same challenge in their own work. We must question to what extent journalists are able to draw attention to the flaws and particularities of the data they use to tell news stories and to what extent they recognize and respect the limits of data's portability beyond one specific news story. Many of the recent high profile examples of DJ, such as the projects listed on the Guardian's Data Blog, are the result of journalists taking a customized approach to the collection of data and its analysis based on the specific story being covered [43]. It is unclear whether such efforts can be maintained as data becomes more closely integrated into the everyday practices of news production. As the production and circulation of data become increasingly automated, relying less on offline sources, and as sources of open and/or raw data become more readily available, the participatory and bespoke (customized) approach to data gathering for individual projects may be undermined.

The stakes of the extent to which journalists are equipped and given the time to interpret and effectively use data become all the more evident when we take into account that not all types of journalism deal with the same kind of data in the same way. For example, some researchers have set out to develop a 'reporter's black box' ([44], p. 4) that would provide journalists with a set of standard query templates for working with data—a standard set of questions that journalists could use to analyze a dataset. Such standard queries are deemed particularly useful in journalistic practices that produce consistent kinds of queries from familiar datasets such as in the case of sports journalism. But standardized queries may be more problematic in the case of investigative journalism. The technical knowhow and expert knowledge needed to conduct research are perceived to be a major concern among journalists ([23] p. 70) and in such cases, the provision of user-friendly platforms for the production of news represents an interesting business proposition.

The pace and direction of technological change also suggests that the connection between narrative and objectivity embodied in journalists' practice may undergo even more dramatic changes in the near future. Current innovations in the automation of computational processes such as online searches lead some observers to consider replacing the journalist with computational resources:

[...] 'eventually some watchdog articles will be written by algorithm in a way that would allow readers to see a customized, personalized article about how a policy problem is playing out in their neighbourhood, block or lives.' [45]

3.4. Data and the Objectivity Regime's Institutional Relationships

The objectivity regime is embedded within a set of interdependent institutions that tend towards its reproduction. These institutions include legal guarantees

provided by the state, institutions of higher learning that contribute to journalism as field of knowledge and structural arrangements within news organisations such as the separation of marketing functions from editorial functions. Much of the current literature on DJ is written from an internalist perspective ([46], pp. 2–5) in that it is frequently presented by proponents from within journalism as a way to potentially *save* it from its current state of declining credibility and economic disinvestment. It would therefore seem that data is unlikely to be used to disrupt existing institutional relationships so much as to improve and strengthen them by, for example, identifying viable business models. It is therefore of vital importance that we question to what extent the variable dimensions of scale, transparency and openness for data develop within the institution of journalism but also within related institutions like commercial enterprises and governments.

Data journalists' ties to the open data movement can in some ways serve the commercial interests of media corporations, but in a new context. Historically, the material interests behind the objectivity regime included the interests of advertisers and commercial media in generating and capturing the attention of new broad audiences in the era of the emergence of mass marketing. Other factors included the political interests of media corporations in deflecting political demands for government regulation of newspaper monopolies that were emerging by mid-20th century and in managing the media/state relationship more broadly. Also served were the occupational interests of journalists in enhancing their claims to *professional* status via the specialized skill of objective reporting ([2], pp. 60–81). Similarly, in the 21st century, collecting and interpreting data helps journalists adapt to global capitalism's information flows and to harness the potential to monetize both databanks and data analysis apps ([33], p. 191). Lorenz [47] points to examples such as the *New York Times'* custom search platform for finding and purchasing a home. DJ can help news organizations to brand themselves and to restore audience and popular trust in journalists through the provision of open data as a service and to enhance journalists' professional status in the new role of its curatorship.

Journalism in Western democracies is legally and politically protected in ways that are not available to other types of organisations or disciplines (for example, see [48] on a comparison between journalism and epidemiology and their common remit to access and publish findings from private data). Such protection extends to data journalists because, in line with the objectivity regime, journalism presents itself in terms of altruistic values such as the democratization of information. This legal protection may enable journalists to assist open data movements. Open data movements have encountered considerable resistance from local and national public institutions (see [19] for

an example of local resistance to open data in the Chicago Police Department). In a recent case study of the Obama administration's plans for a national US Open Data Program, Peled [49] shows how various departments of the US government responded to requests to implement an open data policy by various resistance tactics. Peled concludes that individual departments perceive each other to be in inter-bureaucratic competition and use data as a source of leverage between departments. An open data policy undermines such inter-departmental horse-trading. The Obama administration's early attempts to implement an open data government program failed from a civic perspective because the data made available online was considerably limited in scope and not regularly updated.

By striving for greater openness, data journalists may impose greater scrutiny of government and how it produces and provides data (the watchdog role mentioned above). But some see conflicting professional objectives between journalists and proponents of open government data. Cohen [28], for example, identifies a potential rift between people who want to produce *studies* and people who want to write *stories*. She recognizes that no matter how much people working to improve the provision of open data in government believe they are only working to increase levels of collaboration with civil society, the collection and provision of data can always be used to serve certain political or ideological interests. In such cases, it is in the journalist's interest to scrutinise and challenge such data, no matter how *open*. In such cases, proponents of open government data and journalists may find themselves in opposing camps.

Data collected independently by journalistic institutions represents another way in which data may challenge established institutional power, especially the dominance of official sources. Any particular spin on political events, for example the recent MPs' expenses scandal in Britain, can potentially be challenged by an alternative story emerging from data analysis. Collecting data may also raise the possibility for new kinds of partnerships between news organisations and other kinds of informational or media organisations as a means of providing goods and services through databases and digital platforms ([33], p. 191). What remains lacking at the moment is a critical discussion of the ethical implications of journalistic institutions collecting and storing data and what such potential collaborations may have for journalistic independence and public service.

The emphasis on open data for DJ practitioners does not necessarily mean that incumbent institutions will lose such a status. Nor does it mean that journalistic institutions are impervious to challenges from new actors. Broader civil society movements for *open government* in some respects can be interpreted as the other side of the DJ coin. But the issues raised by open data movements also come bundled within broader debates concerning intellectual property rights

regimes and the commercial interests for open data that include powerful actors like Google and some Internet service providers.

The recent case of the 2010 Wikileaks episode and the differential treatment accorded to its participants suggest the uneven power relations involved in forwarding the agenda of journalistic openness. Wikileaks' original 'pure leak strategy' ([50], p. 154) of providing all of the raw data from their Afghanistan war logs online in an attempt to provide maximum openness was met with little public fanfare. It was only once Wikileaks collaborated with *The New York Times*, *The Guardian* and *Der Spiegel* to provide a more refined analysis that the data started to gain public attention. Subsequently, Bradley Manning (now known as Chelsea Manning) was sentenced to 35 years in prison in August 2013 for leaking classified documents to Wikileaks. By contrast *The New York Times*, *The Guardian* and *Der Spiegel*, having helped to publicize the leaked material, are not facing legal retribution [51] and arguments are ongoing as to whether or not similar relative legal impunity should be afforded to Wikileaks which functioned as a middleman between Bradley and the news media.

3.5. Data and the Objectivity Regime as an Active Ingredient in the Public Discourse

This final dimension to the objectivity regime recognizes that the expectations of objectivity and the associated language for evaluating news are actively circulated among members of the public, where they shape and are shaped by the everyday lives of those who engage with news reports.

Earlier sections addressed the importance of ensuring that data's openness is not understood in isolation from the ways in which said data is interpreted and used by the public. It is therefore of vital importance that we question to what extent data is part of public discourse regarding journalism and its objectivity. For example, to what extent is the DIY ethos of open data something that is reflected in the way people engage with news reports in their everyday lives? For Natalie Fenton ([52], pp. 559–560) multiplicity and polycentricity represent characteristics of online journalism that enable journalists to offer a view of the world that is 'more contextualized, textured, and multidimensional'; a view that may challenge traditional objectivity by enabling readers to compare reports and access sources. On the other hand, she warns that behind such multiplicity can be more of the same: sophisticated marketing and commodification. Political discourse can be assimilated into entertainment, public discourse can be further homogenized, the concentration of ownership increased including the control of search engines. These risks of marketing and commodification are undoubtedly relevant to data's future place in journalism and in the wider public discourse about journalism.

The question therefore remains to what extent, and

in what ways, does the public actually access, interpret and use journalistic data? In our review of the current literature, we did not encounter any material that addresses the variety of ways in which the public actually engages with data beyond the occasional DJ projects that rely on crowd sourcing data.

4. Data and Journalism: Questions for Future Research

This paper represents a critical interrogation of data, its place in journalism, and a call for scholars to fruitfully bring together insights from mediation theory and critical political economy and sociology of journalism to the study of data for journalism. We do not raise these issues in order to reject or undermine practices like DJ. Rather, we recognize data's complex and contradictory potential within (and beyond) the journalism field—a potential that in certain respects *does* have significant democratizing implications. Our objective in introducing the three variable dimensions of data's mediality and how data relates to journalism's regime of objectivity is to underline how data's future is contingent upon decisions regarding what constitutes data and the consequences of such decisions for how objectivity is produced through journalism as a set of ideals, epistemologies, practices, institutional relationships, and public discourses.

Such a future could entail placing DJ in relation to historical precedents and contemporary developments within journalism such as peace journalism [53]. DJ could improve approaches to peace journalism by strengthening the empirical basis of the cultural and structural violence that (Peace Studies scholars argue) underlies the physical violence of armed conflict; it enables researchers to more adequately explore the causes and consequences of violent conflict. For example, one could explore statistical linkages between unemployment, rising food prices, or evidence of government corruption, with outbreaks of civil unrest, like the so-called Arab Spring. Or explore the hidden costs of war (another injunction that peace journalism theory suggests for conflict reporting) by, for example, correlating spikes in domestic violence and divorce rates with the return of soldiers from war.

Based on the approach we have devised for this paper, we also suggest two sets of questions for future empirical research:

1. To what extent are roles for the collection and presentation of data within journalistic institutions consistent with those previously developed within the objectivity regime? In what ways do the definition and execution of such roles remediate practices and discourses found in scientific research?

2. How is data part of public discourse regarding the objectivity of news? In particular: (a) how does public discourse on data in journalism mediate cultures of computing; (b) how does public discourse on data in

journalism mediate cultures of scientific enquiry?

Data's objectivity, when collected, interpreted and disseminated by journalists, cannot be taken as a given. Data is technologically, organizationally and symbolically mediated through discourses and practices for its collection, representation and dissemination that evoke empirical research or computational processes as well as aspects of journalism. The inherent facticity of data is itself problematic. This paper was not written in order to resolve such a problem but as a call for tempering the claims made for data in the context of journalism, for interrogating the assumptions that come

References and Notes

1. Donsbach W, Klett B. Subjective objectivity. How journalists in four countries define a key term of their profession. *International Communication Gazette*. 1993;51(1):53–83.
2. Hackett RA, Zhao Y. *Sustaining Democracy? Journalism and the politics of objectivity*. Toronto, Canada: Garamond Press; 1998.
3. Maras S. *Objectivity in Journalism*. Cambridge, UK: Polity; 2013.
4. Sterne J. *MP3: The meaning of a format*. London, UK: Duke University Press; 2012.
5. Data. *OED*. 3rd ed. Oxford, UK: Oxford University Press; 2012. Available from: <http://www.oed.com/view/Entry/296948#eid7471943> (accessed on 12 January 2014).
6. Meyer P. *Precision journalism: A reporter's introduction to social science methods*. Bloomington, IN, USA: Indiana University Press; 1973.
7. Garrison B. *Computer-assisted reporting*. 2nd ed. Mahwah, NJ, USA: L. Erlbaum Associates; 1998.
8. Deuze M. *Online Journalism: Modelling the first generation of news media on the World Wide Web*. *First Monday*. 2001;6(10):1–16.
9. Hamilton JT, Turner F. *Accountability through algorithm: Developing the field of computational journalism*. Stanford, CA, USA: Stanford University; 2009.
10. Bradshaw P. What is data journalism? *The Data Journalism Handbook*. 2012. Available from: http://datajournalismhandbook.org/1.0/en/introduction_0.html (accessed on 1 May 2013).
11. Rosen J. The 'Awayness' Problem. *Columbia Journalism Review*. 2013;52(3):28–30.
12. The Data Deluge. *The Economist*. 2010. Available from: <http://www.economist.com/node/15579717> (accessed on 1 May 2013).
13. Lorenz M. Why Journalists Should Use Data. *The Data Journalism Handbook 10 Beta*. 2012. Available from: http://datajournalismhandbook.org/1.0/en/introduction_1.html (accessed on 1 May 2013).
14. Arthur C. Analysing data is the future for journalists, says Tim Berners-Lee. *The Guardian*. 2010. Available from: <http://www.guardian.co.uk/media/2010/nov/22/data-analysis-tim-berners-lee> (accessed on 1 May 2013).

with data as an object circulating between multiple contexts, and for a more systematic enquiry into the unstated interests that such data, as a source of objectivity, serve.

Acknowledgements

An earlier version of this manuscript was presented as a paper to the joint session on 'Creative destruction and Journalism' as part of the Communication Policy and Technology Section of the International Association for Mass Communication Research (IAMCR) conference in Dublin, Ireland, June 2013.

15. Stolte Y. Journalism and Access to Data. *Schwerpunkt*. 2012;36(5):354–358.
16. Webster F. *Theories of the Information Society*. 3rd ed. London, UK: Routledge; 2006.
17. Mosco V. *The Digital Sublime: Myth, power, and cyberspace*. Cambridge, MA, USA: MIT Press; 2004.
18. Couldry N, McCarthy A. Orientations: Mapping MediaSpace. In: Couldry N, McCarthy A, editors. *MediaSpace: Place, scale and culture in a media age*. London, UK, and New York, NY, USA: Routledge; 2004. pp. 1–18.
19. Parasie S, Dagiral É. Data-driven journalism and the public good: "Computer-assisted-reporters" and "programmer-journalists" in Chicago. *New Media and Society*. 2013;15(6):853–871.
20. Star SL, Bowker GC, Neumann LJ. Transparency beyond the Individual Level of Scale: Convergence between information artifacts and communities of practice. In: Bishop AP, Van House NA, Battenfield BP, editors. *Digital Library Use: Social practice in design and evaluation*. Cambridge, MA, USA: MIT Press; 2003. pp. 241–269.
21. Bolter JD, Grusin R. *Remediation: Understanding new media*. Cambridge, MA; USA: MIT Press; 2000.
22. Datablog, *The Guardian*. Available from: <http://www.guardian.co.uk/news/datablog> (accessed on 1 January 2014).
23. Cohen S, Hamilton JT, Turner F. Computational Journalism. *Communications of the ACM*. 2011;54(10):66–71.
24. Rogers S. Open data journalism. 2012. Available from: <http://www.guardian.co.uk/news/datablog/2012/sep/20/open-data-journalism> (accessed on 1 May 2013).
25. Willinsky J. The Unacknowledged Convergence of Open Source, Open Access, and Open Science. *First Monday*. 2005;10(8):1–19.
26. Coleman J. Open data: "there's an app for that". *Journal of professional communication*. 2011;1(1):17–21.
27. Gurstein M. Open data: Empowering the em-

powered or effective data use for everyone? *First Monday*. 2011;16(2).

28. Cohen S. Shared Values, Clashing Goals: Journalism and open government. *XRDS*. 2011;18(2):19–22.

29. As we will show below, the regime's dominance has arguably been unraveling due to different pressures in recent years but it is unclear what kind of formation is taking its place.

30. While the paradigm is found for the most part in the United States and in Canada, we can also point to parts of Western Europe and historically colonial states such as Australia where permutations of this regime are also found.

31. "The current status of objectivity in journalism is complicated. In the US, objectivity was a characteristic of journalism's mid-20th century "high modernist" period, one that has arguably been eclipsed since the 1980s by a "postmodern" paradigm characterized by a multichannel mediascape, profit-oriented conglomerate ownership, deregulation, the commodification of the public sphere, the displacement of "serious" news by infotainment, the unfolding impacts of the internet, and an epistemological relativism that rejects the possibility of objectivity." Hackett RA. Objectivity in Reporting. In: Donsbach W, editor. *Concise Encyclopaedia of Communication*. Malden, MA, USA: Blackwell; forthcoming.

32. McQuail D. *Media Performance: Mass communication in the public interest*. London, UK: Sage; 1992.

33. Daniel A, Flew T. The Guardian reportage of the UK MP expenses scandal: A case study of computational journalism. *Communications Policy & Research Forum*. Sydney, Australia: Network Insight Institute; 2010. pp. 186–194.

34. McChesney RW. The Problem of Journalism: A political economic contribution to an explanation of the crisis in contemporary US journalism. *Journalism Studies*. 2003;4(3):299–329.

35. Hackett RA. Objectivity in Reporting. 2008. Available from: http://www.communicationencyclopedia.com/subscriber/tocnode.html?id=g9781405131995_chunk_g978140513199520_ss1-1 (accessed on 1 May 2013).

36. Tuchman G. The News Net. *Social Research*. 1978;45(2):253–276.

37. Quote taken from Robert A. Hackett's interview with Simon Rogers on 29 October 2012.

38. Messner M, Garrison B. Journalism's 'dirty data' below researchers' radar. *Newspaper Research Journal*. 2007;28(4):88–100.

39. Rogers S. Data Journalism at the Guardian:

What is it and how do we do it? 2011. Available from: <http://www.guardian.co.uk/news/datablog/2011/jul/28/data-journalism> (accessed on 1 May 2013).

40. Segel E, Heer J. Narrative visualization: Telling stories with data. *IEEE Transactions on Visualization and Computer Graphics*. 2010;16(6):1139–1148.

41. Rosen J. The People Formerly Known as the Audience. 2006. Available from: http://archive.pressthink.org/2006/06/27/ppl_frmr.html (accessed on 1 May 2013).

42. Tuchman G. Objectivity as Strategic Ritual: An examination of newsmen's notions of objectivity. *American Journal of Sociology*. 1972;77(4):660–679.

43. Rogers S. Free our Data! In: Lorenz M, editor. *Data-Driven Journalism: What is there to learn?* Amsterdam, The Netherlands: European Journalism Centre; 2010.

44. Cohen S, Li C, Yang J, Yu C. Computational Journalism: A Call to Arms to Database Researchers. Presented at the 5th Biennial Conference on Innovative Data Systems Research, Asilomar, CA, USA, 9–12 January 2011.

45. Hamilton JT. Tracking toxics when the data are polluted. *Nieman Reports: Nieman Foundation for Journalism at Harvard* 2009. Available from: www.nieman.harvard.edu/reportsitemprint.aspx?id=100933 (accessed on 1 May 2013).

46. Anderson CW. Towards a sociology of computational and algorithmic journalism. *New Media and Society*. 2013;15(7):1005–1021.

47. Lorenz M. Status and Outlook for Data-Driven Journalism. In: Lorenz M, editor. *Data-Driven Journalism: What is there to learn?* Amsterdam, The Netherlands: European Journalism Centre; 2010.

48. Westrin C-G, Nilstun T. The ethics of data utilisation: A comparison between epidemiology and journalism. *BMJ*. 1994;308:522–523.

49. Peled A. When transparency and collaboration collide: The USA Open Data Program. *Journal of the American Society for Information Science and Technology*. 2011;62(11):2085–2094.

50. Beckett C, Ball J. *Wikileaks: News in the networked era*. Cambridge, UK: Polity; 2012.

51. Abrams F, Benkler Y. Death to Whistle-Blowers? 2013. Available from: http://www.nytimes.com/2013/03/14/opinion/the-impact-of-the-bradley-manning-case.html?_r=0 (accessed on 1 May 2013).

52. Fenton N. News in the Digital Age. In: Allan S, editor. *Routledge Companion to News and Journalism*. Florence, KY, USA: Routledge; 2010. pp. 557–567.

53. Lynch J, McGoldrick A. *Peace Journalism*. Gloucestershire, UK: Hawthorn Press; 2005.