Volume 6 Number 7 September 2018

DIGITAL JOURNALISM

Routledge

Digital Journalism

Routledge Taylor & Francis Group

ISSN: 2167-0811 (Print) 2167-082X (Online) Journal homepage: http://www.tandfonline.com/loi/rdij20

Transparency, Interactivity, Diversity, and Information Provenance in Everyday Data Journalism

Rodrigo Zamith

To cite this article: Rodrigo Zamith (2019): Transparency, Interactivity, Diversity, and Information Provenance in Everyday Data Journalism, Digital Journalism, DOI: 10.1080/21670811.2018.1554409

To link to this article: https://doi.org/10.1080/21670811.2018.1554409

Published online: 02 Jan 2019.

DIGITAL JOURNALISM
https://doi.org/10.1080/21670811.2018.1554409

Routledge
Taylor & Francis Group

Transparency, Interactivity, Diversity, and Information Provenance in Everyday Data Journalism

Rodrigo Zamith

Journalism Department, University of Massachusetts Amherst, S414 Integrative Learning Center, Amherst, MA, USA

ABSTRACT

This article examines the features of day-to-day data journalism produced by *The New York Times* and *The Washington Post* in the first half of 2017. The content analysis evaluates story characteristics linked to the concepts of transparency, interactivity, diversity, and information provenance. It finds that the data journalism produced by those outlets comes from small teams, focuses on "hard news," provides fairly uncomplex data visualizations with low levels of interactivity, relies primarily on institutional sources and offers little original data collection, and incorporates just two data sources on average in a generally opaque manner. This leads to the conclusion that "general data journalism" still has a long way to go before it can live up to the optimism and idealization that characterizes much of the data turn in journalism. Instead, contemporary day-to-day data journalism is perhaps better characterized as evolutionary rather than revolutionary, with its celebrated potential to serve as a leap forward for journalism and engender greater trust in it remaining untapped.

KEYWORDS

Data journalism; data visualization; interactivity; transparency; diversity; information subsidy; data sources; trust

Data journalism has grown substantially in recent years, with many journalists and news organizations viewing it as a way to make journalism more systematic, accurate, and trustworthy (Borges-Rey 2017; Boyles and Meyer 2017). This is particularly important given the parallel developments of decreased trust in news and public institutions, public assaults on journalists and news organizations, and greater acceptance of the notion of factual plurality (Usher 2018; Waisbord 2018). Following highly public successes involving data journalists like Nate Silver and the launch of data-driven ventures by prominent news organizations such as *The New York Times* and *The Washington Post*, and digital-native outlets like *Vox*, data journalism gained a great deal of discursive currency and reputational authority (Howard 2014) – though some of that euphoria has been dampened by highly public failures over the past 3 years (Lewis and Waters 2018). Nevertheless, news organizations continue to invest (and demand skills) in data journalism, making it a rare growth area in an industry battered by economic challenges (Weber and Kosterich 2018) and drawing scholarly attention

CONTACT Rodrigo Zamith rzamith@umass.edu

to its epistemological implications (Lesage and Hackett 2014; Lewis and Westlund 2015).

While there has been "an explosion in data journalism-oriented scholarship" in recent years (Fink and Anderson 2015, 467), most of the work relies on case-study ethnographies and interviews with data journalists (e.g., Borges-Rey 2017; Boyles and Meyer 2017; Felle 2016; Lewis and Waters 2018). Those works offer insight into how data journalism is being conceptualized and practiced but provide a limited lens into what data journalistic outputs look like (Reimer and Loosen 2017). Such insight is crucial to understanding if data journalism is realizing the potential celebrated by many scholars and practitioners, and, in particular, whether it advances the journalistic ideal of transparency and online journalism's interactive affordances (Allen 2008; Karlsson 2010; Karlsson and Holt 2016). Indeed, transparency and participation – or the perception thereof via greater interactivity – have been lauded as solutions to decreasing trust in news media (Usher 2018; van der Wurff and Schönbach 2014). It lso remains unclear if data journalism has become more diverse than its predecessors (Coddington 2015) and thus able to exert greater impact on the practice and products of journalism by engaging with more topical areas and moving beyond niches (Lewis and Zamith 2017). Moreover, scholars have suggested that data journalism may harm long-term trust and fail to advance journalism's social contract by increasing dependence on institutional sources through data subsidies (Cawley 2016; Tandoc and Oh 2017). If data journalism's potential is to be valorized as a leap forward and a bulwark against declining trust, its manifestations must be closely examined.

Several scholars have recently analyzed data journalism content but much of that work has focused on submissions to award programs (e.g., Appelgren 2018; Loosen, Reimer, and De Silva-Schmidt 2017; Ojo and Heravi 2018; Young, Hermida, and Fulda 2018), which represent ideal-types rather than the day-to-day works citizens are regularly exposed to (Stalph 2017; Wright and Doyle 2018). As such, the scholarly understanding of "general data journalism" (Uskali and Kuutti 2015, 87) remains limited, especially in the United States.

This paper addresses that gap in the literature by examining the features of more general, day-to-day data journalism produced by *The New York Times* and *The Washington Post* during the first half of 2017. It evaluates story characteristics linked to the concepts of transparency, interactivity, diversity, and information provenance in more than 150 data journalism articles. Overall, it finds that day-to-day data journalism is neither especially transparent nor interactive, though it is more diverse than its closest predecessor and perhaps more susceptible to a dependency on institutional sources and information subsidies. As such, it has a long way to go before it can live up to the optimism and idealization that generally characterizes the data turn in journalism.

Literature Review

As Coddington (2015) observes, computation and quantification have become increasingly important in and to journalism in recent years. Carlson (2018) adds that journalism has become more "measurable," as evidenced by the growing use of audience metrics to inform individual and organizational conceptions of audiences and, in some instances, guide editorial practices (see Zamith 2018). Moreover, data are increasingly used within news products as more journalists and news organizations embrace data journalism's goal of unearthing new stories that focus on general patterns instead of outliers (Young and Hermida 2015) and "allow the public to analyze and draw understanding from data themselves, with the journalist's role being to access and present the data on the public's behalf" (Coddington 2015, 343).

The excitement around the use of quantitative information in journalism isn't new. Data visualizations have been around since at least the eighteenth century (Tabary, Provost, and Trottier 2016) and scholars have argued that "statistics have long been a staple of daily news" (Nguyen and Lugo-Ocando 2015, 4). However, sociotechnical developments during the 1960s – and accelerated in recent years – promoted a set of practices and products organized around an emerging data-oriented logic viewed by many scholars as distinct from general journalism (Coddington 2015). Modern data journalism is typically viewed as an outgrowth of – yet now distinct from – the tradition of computer-assisted reporting, which itself is often traced back to Phil Meyer's (1973) work on precision journalism. In recent years, the explosion in the availability of data (Lewis and Westlund 2015), developments in low-cost computer hardware and software (Gray, Bounegru, and Chambers 2012), and an attitudinal shift toward journalism and the notion of "data" (Boyles and Meyer 2016; Boyles and Meyer 2017) have enhanced data journalism's standing within the field and made it more prevalent within newsrooms (Lewis and Zamith 2017).

The quantitative turn in journalism has been viewed by scholars and practitioners as having positive and negative implications for realizing journalism's broader ideals and mission. In particular, it has been argued that data journalism can promote greater *transparency* (Lewis and Westlund 2015); offer more opportunities for *interactivity* with content, newsworkers, and other citizens (Young, Hermida, and Fulda 2018); and provide more *diversity* than its specialized predecessors, both in terms of content and collaborative opportunities (Coddington 2015). Those elements have been linked to trust in news media, and as essential components to a response to declines in trust in institutions like journalism within liberal democracies (Boyles and Meyer 2016; Usher 2018). However, the prospect of greater reliance on data subsidies rather than original evidence-gathering has raised concerns about data *provenance* (Tandoc and Oh 2017), which would adversely impact data journalism's potential to engender trust.

While a growing body of work has examined data journalism content, Stalph (2017) aptly observes that most recent content analyses of data journalism have focused on entries to award competitions (e.g., Loosen, Reimer, and De Silva-Schmidt 2017; Ojo and Heravi 2018; Young, Hermida, and Fulda 2018), or self-selected ideal-types that Reimer and Loosen (2017, 95–96) note "are not likely to represent 'everyday' data journalism." Indeed, they represent "what the field itself defines as its 'gold-standard'" projects (Loosen, Reimer, and De Silva-Schmidt 2017, 1). This limits the understanding of the extent to which data journalism is realizing the potential celebrated by scholars and practitioners for it to advance journalistic values and engender trust. Of the studies that have explored more general, day-to-day data journalism, the majority cover European countries (e.g., Appelgren 2018; Knight 2015; Stalph 2017) or Canada (e.g., Tabary, Provost, and Trottier 2016).

This study explores these opportunities and pitfalls through the lens of two US-based news organizations: *The New York Times* and *The Washington Post*. These organizations have the resources to produce data journalism on a regular basis and are viewed as model news organizations that smaller outlets can strive to emulate as software and hardware barriers are lowered and more journalists gain the requisite skills to routinely produce data journalism (see Fink and Anderson 2015). The two newsrooms have also invested heavily in data journalism in recent years. While both outlets now compete for an international audience and focus on public-service journalism, they differ in that *The New York Times* not only has more journalists but also nearly thrice the number of digital subscribers as *The Washington Post*, as of early 2018 – though *The Post* has been closing those gaps (Atkinson 2017).

Classifying Data Journalism

There is no generally accepted definition of *data journalism* (Loosen, Reimer, and De Silva-Schmidt 2017; Lowrey and Hou 2018; Stalph 2017). Most scholars have distinguished it by examining the distinct qualities of its process (e.g., Anderson 2017; Howard 2014; Veglis and Bratsas 2017a), contending it is guided by a logic and epistemology that is different from "regular" journalism (Borges-Rey 2017; Coddington 2015). A smaller set of scholars have distinguished data journalism through the form of its content. For example, Knight (2015, 59) defines it as "a story whose primary source or 'peg' is numeric (rather than anecdotal), or a story which contains a substantial element of data or visualisation" and Lowrey and Hou (2018, 7) define it as "informational, graphical accounts of current public affairs for which data sets offering quantitative comparison are central to the information provided."

Regardless of whether data journalism is defined through its form or process, two elements stand out: (1) quantitative information should play a central role in the development or telling of the story and (2) there should be some visual representation of the data referenced in the story. Loosen et al. (2017) point to two additional elements – that "it is *frequently* 'characterized by its participatory openness'" and "*regularly* adopts an open data and open source approach" (p. 3, emphasis theirs), though there is less scholarly agreement over the necessity of those elements and evidence of their presence in even exemplary works is limited (see Borges-Rey 2017; Stalph 2017; Young, Hermida, and Fulda 2018).

Most analyses of data journalism content circumvent this definitional challenge by relying on organizations' own classification choices. For example, Stalph (2017, 6) analyzed submissions appearing on "designated data journalism landing pages of the outlets' websites that are explicitly labeled as such" and Young, Hermida, and Fulda (2018) focused their analyses on entries submitted to data journalism awards competitions. Few scholars have systematically segregated data journalism from a general pool of content themselves (cf. Knight 2015; Tabary, Provost, and Trottier 2016).

Scholars have nevertheless developed useful analytic frameworks for studying content somehow categorized as data journalism. Stalph's (2017) analytic framework examines formal characteristics (e.g., topic and number of authors), data visualizations (e.g., visualization types and level of interactivity), data sources (e.g., data provider and accessibility of data), and form and content (e.g., story format and juxtaposition). Ojo and Heravi's (2018) framework examines a range of elements, including data sources, narrative styles, degrees of interactivity, and analytical techniques. Knight (2015) suggests focusing on the complexity of data elements rather than the type of visualization used. These analytic frameworks and their operationalizations can, and have been, used to great effect in examining data journalism content and linking content attributes to broader conceptual and theoretical concerns.

Transparency

Transparency has long been viewed as an important ideal in journalism, often codified in professional codes of ethics, though its translation to a commonly enacted ritual has been limited (Singer 2007). According to Allen (2008), transparency involves "making public the traditionally private factors that influence the creation of news." It can serve a dual function of improving accountability among news actors and increasing their legitimacy among news audiences (Allen 2008). Karlsson (2010) identifies two types of transparency: disclosure transparency and participatory transparency. Disclosure transparency pertains to the degree of openness about how news is selected and produced. This includes not only visible disclosures about corrections and retractions – a ritual long employed in journalism – but also methodological notes explaining analytic choices (see Tandoc and Oh 2017). Participatory transparency pertains to the extent to which audiences are incorporated into the selection and production of news. This includes not only providing forums for discussing the news – as with the printing of letters to the editors – but also allowing audiences to annotate information that is then disseminated to other readers and integrated into a larger project (see Handler and Ferrer Conill 2016).

Scholars have argued that the affordances of online journalism provide opportunities for disclosure transparency that are not possible with its analog counterparts (Karlsson 2010). Online journalism allows journalists to use hyperlinks to unobtrusively disclose specific sources of information and permits journalists to easily publish supplemental materials, from methodological explanations to source documents (Karlsson and Holt 2016). However, important challenges to the transparency ideal remain: journalists often prioritize other values over transparency (Plaisance and Skewes 2003); may see transparency as an intrusion on their autonomy (Singer 2007); their audiences have been found, in some instances, to perceive non-transparent articles as more credible (Tandoc and Thomas 2017); and efforts that promote transparency may have limited effects on restoring trust (Karlsson, Clerwall, and Nord 2017) – though citizens value and often demand greater transparency about the news process (van der Wurff and Schönbach 2014). Indeed, Karlsson (2010) found that transparency is often routinized into a strategic ritual that separates execution from intent by promoting a small degree of transparency (e.g., posting links) but resisting too much of it (e.g., sharing original documents). This led Karlsson to conclude that "the transparency norm has yet to make the kind of impression forecast by many scholars" (p. 543). Vos and Craft

(2017) similarly observed limits to its adoption, though they added that a paradigm shift privileging transparency is underway as the norm gains cultural capital. Computational and data-driven analytic techniques have been especially heralded for their potential to increase transparency by facilitating the enactment of best practices like sharing underlying datasets and data analysis scripts (Lewis and Westlund 2015; Zamith and Lewis 2015).

Research on data journalistic content has found that access to the underlying data is typically available in the majority of stories (cf. Young, Hermida, and Fulda 2018), with Tandoc and Oh (2017) identifying links in more than 67% of stories and Parasie and Dagiral (2012) finding that almost 90% of "database projects" offered online access to the database and nearly 70% to the "raw" data. The limited evidence about the inclusion of methodological information is mixed. While Loosen et al. (2017) found that the majority of entries to an awards competition provided information on how data were accessed by journalists, Lowrey and Hou (2018) found that only one-third of the projects they analyzed included comments about the quality of the data.

As such, the following hypothesis and research question are posited:

> *H1*: *The New York Times* and *The Washington Post* will provide access to data in more than half their data journalism stories.
>
> *RQ1*: Do *The New York Times* and *The Washington Post* typically provide supplemental methodological details in their data journalism stories?

Interactivity and Visual Complexity

Interactivity is typically viewed as a key enabler of participatory transparency as well as an affordance that distinguishes online media from its analog counterparts (Karlsson 2010; Karlsson and Holt 2016). Interactivity refers to "technological attributes of mediated environments that enable … interaction between communication technology and users, or between users through technology" (Bucy and Tao 2007, 656). Different media vehicles and devices provide distinct affordances for interactivity (Cantarero, González-Neira, and Valentini 2017), which makes optimizing for interactivity in a multi-modal news environment a difficult task. Veglis and Bratsas (2017b) point to three forms of interactivity made possible by data visualizations: transmissional (simple interactivity that offers few affordances beyond conveying additional information about elements), consultational (offering multiple views of the same data), and conversational (accepting input data that permits the user to substantially alter the visualization). The emphases on visualization and more conversational forms of interactivity are often discussed as elements that set data journalism apart from its closest predecessor, computer-assisted reporting (Coddington 2015).

Transmissional and consultational forms of interactivity have been found to enhance user enjoyment and foster favorable attitudes in general contexts, though they do not necessarily increase cognitive elaboration, knowledge acquisition, or information recall (Yang and Shen 2018). Conversational interactivity has also been found to increase loyalty toward a news organization (Lischka and Messerli 2016) and journalists have successfully leveraged that affordance to improve their reporting through crowdsourcing (Borges-Rey 2016; Handler and Ferrer Conill 2016). Despite those potential benefits, the professional culture of journalism offers formidable resistance against fully developing the ideal of interactivity (Domingo 2008) and scholars have observed that users sometimes make limited use of advanced interactivity features (Larsson 2011). Scholars have also found that journalists prefer data visualizations that have limited visual complexity (e.g., simple bar charts and line graphs) because their semiotic simplicity is perceived as being more effective, they are more accessible across platforms and screen sizes (e.g., mobile and desktop), and because the tools that allow journalists to create complex visualizations are often perceived as too complicated (Boyles and Meyer 2016; Engebretsen, Kennedy, and Weber 2018).

The majority of the scholarship has found that data journalistic content typically includes some data visualizations (Loosen, Reimer, and De Silva-Schmidt 2017; Parasie and Dagiral 2012; Stalph 2017; Tabary, Provost, and Trottier 2016). However, the scholarship sheds limited light on the number of visualizations present in the average story, with Stalph (2017) observing an average of just two visualizations. Notably, scholars have typically found low levels of interactivity with limited visual complexity, with Appelgren (2018) going so far as to argue that most projects they studied were highly paternalistic and offered only an "illusion of interactivity [that] replaces real interactivity" (p. 15). Indeed, the visualizations are typically simple charts that either lack any interactivity at all (Knight 2015; Stalph 2017) or primarily offer only transmissional affordances (Loosen, Reimer, and De Silva-Schmidt 2017). Such findings are not universal, however. Ojo and Heravi (2018) found more advanced forms of interactivity in their analysis of award-winning entries.

The following hypothesis and research question are posited in light of this evidence:

> *H2*: *The New York Times* and *The Washington Post* will feature data visualizations with (a) low levels of interactivity and (b) low levels of visual complexity.

> *RQ2*: How many data visualizations are featured, on average, in *The New York Times*' and *The Washington Post*'s data journalism?

Diversity and Data Provenance

Scholars have long been interested in the notion of diversity in news, which broadly refers to the "dispersion of representation" (Voakes et al. 1996, 585) of content, sources, and media exposure. Of particular note is content diversity, which focuses on the ideas, perspectives, and topics advanced within a news product (Voakes et al. 1996), though it may also include the workforce (Napoli 1999). According to Coddington (2015), one of the key elements distinguishing data journalism from its predecessors of computer-assisted reporting and precision journalism is its decoupling from investigative journalism and integration into broader journalistic practices. Put differently, data journalism is conceptualized as being more diverse in terms of its content, bringing data analysis techniques to areas largely neglected by computer-assisted reporting, such as entertainment and sports. Additionally, scholars have emphasized the potential for interdisciplinarity within data journalism (Anderson 2017; Howard 2014), noting that skills gaps must often be addressed through collaboration (Borges-Rey 2016; Owen 2017). Such diversity in content and among practitioners is needed to elevate the impact of data journalism, moving it from a specialty to a general practice with increased social currency and greater ability to reshape professional norms and boundaries (Boyles and Meyer 2017; Lewis and Zamith 2017; Weber and Kosterich 2018). As Cawley (2016) observes, one of data journalism's greatest challenges lies in gaining wider acceptance among news professionals and audiences.

Scholars have also long been interested in the provenance of information – where it originates – and have been especially drawn to examinations of information subsidies, or prepackaged information provided for free by organizations and public relations practitioners in order to influence news coverage (Gandy 1982). It has been argued that increasing pressures on journalists to do more with less, and faster, only increases dependencies on subsidies (Lewis, Williams, and Franklin 2008). Scholars have cautioned that the proliferation of data journalism in particular may make journalism even more vulnerable to a dependence on a new type of information subsidy – data – that can be exploited due to news organizations' inability to collect their own large datasets (Cawley 2016; Tandoc and Oh 2017). This is especially concerning in light of the limited data literacy and analytic prowess found in some newsrooms (Appelgren and Nygren 2014; Borges-Rey 2017) and a problematic mythology around "data" that emphasizes its neutrality and objectivity (Lesage and Hackett 2014). That, in turn, may result in greater manipulation of news media and influence on publics, and ultimately impose a heavy social cost as "the quality of information in a democratic society is steadily impoverished" and trust eroded (Lewis, Williams, and Franklin 2008, 43). Moreover, scholars have routinely found a general predilection toward government sources in journalism, which raises another set of epistemological concerns (see Reich 2009).

Research on data journalistic content has found that the majority of content focuses on "hard news" topics, such as politics and world affairs (Knight 2015; Loosen, Reimer, and De Silva-Schmidt 2017; Stalph 2017; Young, Hermida, and Fulda 2018). Stories are typically produced by one or two journalists, rather than teams (Stalph 2017; Young, Hermida, and Fulda 2018), though Loosen et al. (2017) found that stories submitted to an awards competition had an average of five people on the team. Notably, scholars have routinely found that journalists rely on publicly accessible data from institutional sources, and especially from governmental sources (Knight 2015; Lowrey and Hou 2018; Loosen, Reimer, and De Silva-Schmidt 2017; Parasie and Dagiral 2012; Tabary, Provost, and Trottier 2016; Young, Hermida, and Fulda 2018). The use of self-collected data is rare.

The following hypothesis and research question are therefore posited:

> *H3*: The *New York Times*' and *The Washington Post*'s data journalism will typically (a) involve either one or two authors, (b) focus on "hard news" topics, and (c) rely primarily on governmental data sources.

> *RQ3:* How many data sources are listed, on average, in *The New York Times*' and *The Washington Post*'s data journalism?

Method

This study used quantitative content analysis to evaluate data journalism appearing on the websites of *The New York Times* and *The Washington Post* over a 6-month period. Because of the nature of these two organizations, the findings of this study should not be taken to represent all news organizations. They are better viewed as indicative of elite, well-resourced, and highly professionalized news organizations. The unit of analysis was the news article.

Sampling Strategy

A two-stage sampling strategy was used. First, a listing of all articles promoted by both organizations' graphics teams (e.g., @nytgraphics and @postgraphics) between January 1 and June 30, 2017 was obtained through the Twitter API. A series of computer scripts were developed to acquire all tweets during that time range; identify all of the links and extend any shortened ones (e.g., bit.ly); and extract just the unique links to a page within the news organization's domain (e.g., nytimes.com).

Second, all 493 of the identified links were manually reviewed by the author and a human coder to segregate data journalism from other articles. Building on the conceptualizations of data journalism in the scholarly literature (e.g., Howard 2014; Knight 2015; Veglis and Bratsas 2017a), data journalism was operationally defined as a news item produced by a journalist that has a central thesis (or purpose) that is primarily attributed to (or fleshed out by) quantified information (e.g., statistics or raw sensor data); involves at least *some* original data analysis by the item's author(s); and includes a visual representation of data.[1] Intercoder reliability testing was performed using 53 articles, or 10.8% of the sample, and resulted in a Krippendorff's alpha coefficient of 0.80, exceeding the generally accepted minimum bound for the chance-corrected reliability measure (Krippendorff 2011).

Instrument and Reliability

This study's instrument focused on three conceptually linked dimensions: *transparency*, *interactivity and visual complexity*, and *diversity and data provenance*. The measures were drawn from existing frameworks (e.g., Knight 2015; Loosen, Reimer, and De Silva-Schmidt 2017; Ojo and Heravi 2018; Stalph 2017) and refined based on an initial qualitative assessment of the content and subsequent coder training sessions involving two student coders who were blind to the study's hypotheses. All variables reported in this study had a Krippendorff's alpha coefficient above 0.75, again exceeding the suggested minimum bound (Krippendorff 2011).

Transparency. Two variables were measured to assess this dimension: level of access to data and the inclusion of supplemental methodological details. Level of access to data was measured by reviewing all links in the story to see if there was either direct or indirect access to all of the referenced data, some of the referenced data, or none of the referenced data. The inclusion of supplemental methodological details was evaluated on a yes/no basis by seeing if there was additional information provided about the data source(s), data collection method(s), or analysis through an external link (e.g., to a methodology article), pop-up, or dedicated area on the page.

Interactivity and Visual Complexity. Three variables were measured to assess this dimension: the number of data visualizations, the type of interactivity, and the number of quantitative variables used. The number of data visualizations involved tabulating the number of *distinct* visualizations appearing in the article, with a visualization becoming distinct when it was set apart from other visualizations in order to communicate something different. (For example, two side-by-side maps would be considered a single visualization if the objective was to make a comparison.) Type of interactivity was measured through three categories on a present/not present basis: transmissional (e.g., playing and pausing a visualization), consultational (e.g., filtering its contents), and conversational (e.g., personalizing it). While prior work has generally focused on visualization types (e.g., bar graph, pie chart, map, etc.), pretesting found such categorizations were not useful as increasingly creative visualizations were used that could not be reliably or validly classified using existing typologies. Instead, the number of quantitative variables featured in the visualization (excluding a "value" variable) was measured with three options: univariate (low complexity), bivariate (medium complexity), and multivariate (high complexity).[2]

Diversity and Data Provenance. Four variables were measured to assess this dimension: number of authors, main story topic, the number of data sources, and types of data sources. Number of authors was measured by counting the number of individuals credited with a byline. Main story topic was measured by evaluating the topical references and language within the story, with only one (main) topic coded for. "Hard news" topics included: defense and national security, economy, education, judicial and legal affairs, policy, and politics. "Soft news" topics included: culture and entertainment, health, sports, and society. The number of data sources was measured by tabulating the amount of unique data sources listed either at the bottom of an article or at the bottom of a visualization featured in the article. Those data sources were then coded for on a present/not present basis at the story level for the following types: business and industry, educational and academic, government, non-profit group, self-collected, and other news organization.

Findings

A total of 159 data journalism articles were published by the two outlets in the first 6 months of 2017. Of those, 42.1% ($n = 67$) were produced by *The New York Times* and 57.9% ($n = 92$) were produced by *The Washington Post*. Because a purposive (non-random) sample was drawn and it was analyzed in its entirety, descriptive statistics are used exclusively in reporting the findings.

Transparency

The first hypothesis posited that more than half of the data journalism stories produced by the two outlets would provide access to the supporting data. H1 was soundly rejected. The vast majority of stories (87.4%, $n = 139$) did not provide any clear avenue for downloading any of the data used in the article, while 8.2% ($n = 13$) linked to some of the data and just 4.4% ($n = 7$) linked to all of it. *The Washington Post* did not link to any data in 82.6% ($n = 76$) of stories, linked to some data in 9.8% ($n = 9$) of stories and linked to all of the data in 7.6% ($n = 7$) of stories. *The New York Times* was even less transparent, failing to link to any of the data in 94.0% ($n = 63$) of articles, linking to some of the data in 6.0% ($n = 4$) of articles, and not once linking to all of the data.[3]

RQ1 asked whether the outlets typically provided supplemental methodological information. Only 16.4% ($n = 26$) of stories featured additional information about the data source(s), data collection method(s), or analysis through an external link, pop-up, or dedicated area on the page. That figure was lower for *The New York Times* (11.9%, $n = 8$) than *The Washington Post* (19.6%, $n = 18$).

Interactivity and Visual Complexity

The second hypothesis focused on the visual complexity and interactive affordances of the data journalism produced by the two outlets. H2(a) posited that they would typically feature visualizations with low levels of interactivity. This hypothesis was supported as the vast majority of stories (81.8%, $n = 130$) did not include an interactive visualization and an additional 8.8% ($n = 14$) included just a transmissional level of interactivity. A total of 7.5% ($n = 12$) of articles had a consultational level of interactivity and just 1.9% ($n = 3$) had a conversational level of interactivity. The lack of interactivity was common to both *The New York Times* (82.1%, $n = 55$) and *The Washington Post* (81.5%, $n = 75$) (Figure 1).

H2(b) posited that the two outlets would typically feature visualizations with low levels of visual complexity. This hypothesis was partly supported as 71.1% ($n = 113$) of stories included a low-complexity, univariate-type data visualization. However, 75.5% ($n = 120$) included a medium-complexity, bivariate-type visualization, and just 17.0% ($n = 27$) of stories included a complex, multivariate-type visualization. Broken down by organization, *The New York Times* was slightly more likely than *The Washington Post* to include a univariate-type data visualization (71.6%, $n = 48$ vs. 70.7%, $n = 65$) and less

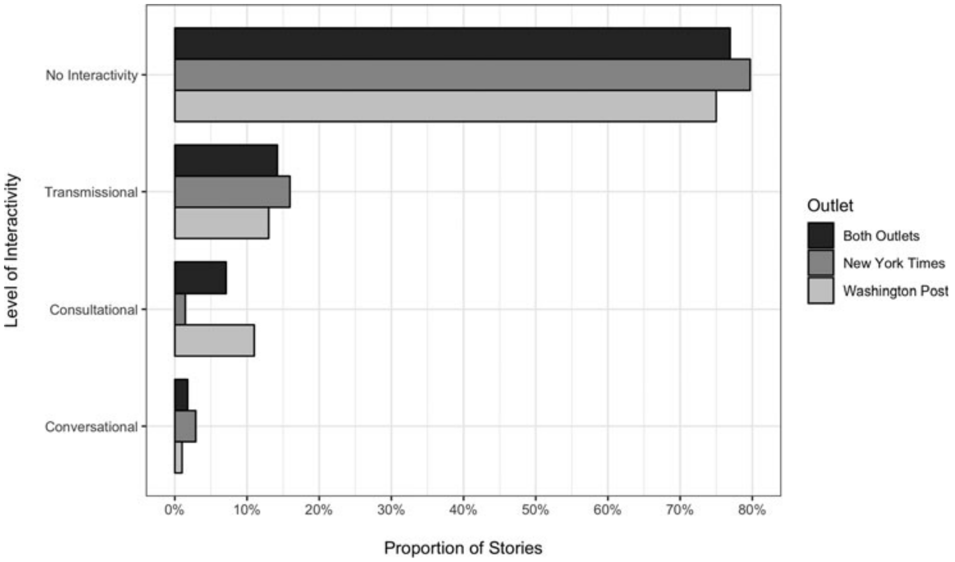

Figure 1. The proportion of stories with interactive elements. Figures may exceed 100% because a single story could have multiple types of interactivity.

likely to include either a bivariate-type visualization (73.1%, $n = 49$ vs. 77.2%, $n = 71$) or a complex, multivariate-type visualization (13.4%, $n = 9$ vs. 19.6%, $n = 18$).

RQ2 asked about the number of data visualizations featured on the average data journalism story. The median number of visualizations for both outlets together was 4; when disaggregated *The New York Times* had a slightly higher median (4) than *The Washington Post* (3).

Diversity and Data Provenance

The third hypothesis focused on content diversity and data origin. H3(a) posited that the data journalism produced by the two outlets would be primarily produced either by a single author or a two-person partnership. That hypothesis was supported as 78.6% ($n = 125$) of the stories fit that criteria, with 39.6% ($n = 63$) of the stories being sole-authored and another 39.0% ($n = 62$) having two authors. *The Washington Post* was more likely to feature sole-authored work (45.7%, $n = 42$) and less likely to feature dual-authored work (34.8%, $n = 32$) than *The New York Times* (31.3%, $n = 21$ and 44.8%, $n = 30$, respectively), though the small number of authors was typical for both organizations (Figure 2).

H3(b) posited that posited that both outlets would focus on "hard news" topics. That hypothesis was also supported as 68.5% ($n = 100$) of stories focused on a "hard news" topic. That figure was higher for *The Washington Post* (74.4%, $n = 64$) than *The New York Times* (60.0%, $n = 36$) (Figure 3).

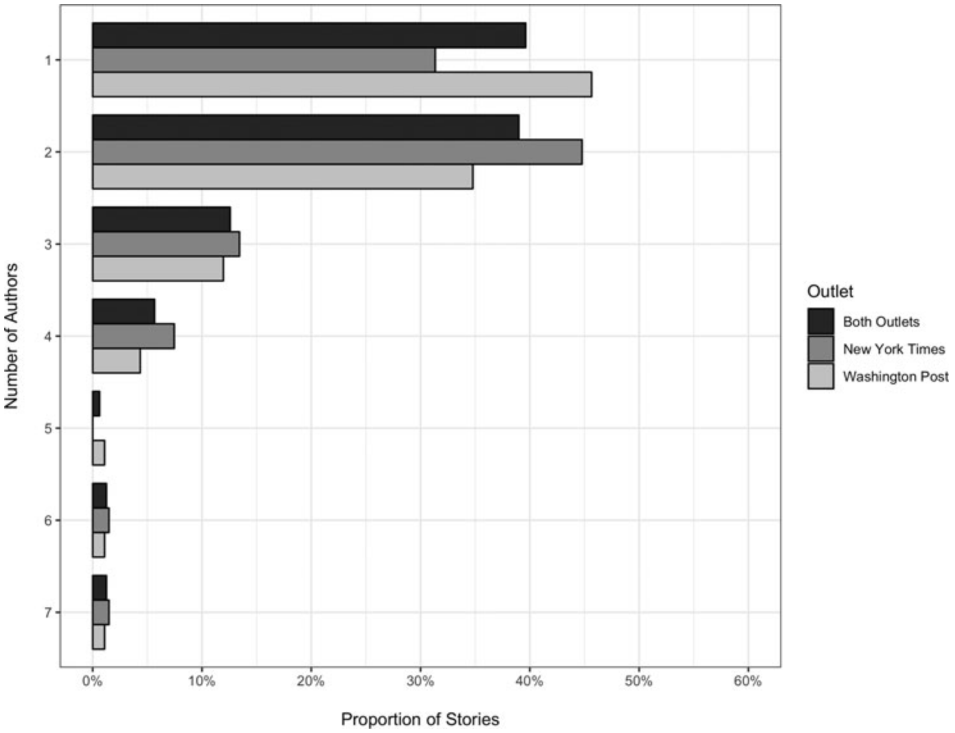

Figure 2. The proportion of stories with one to seven credited authors in the byline.

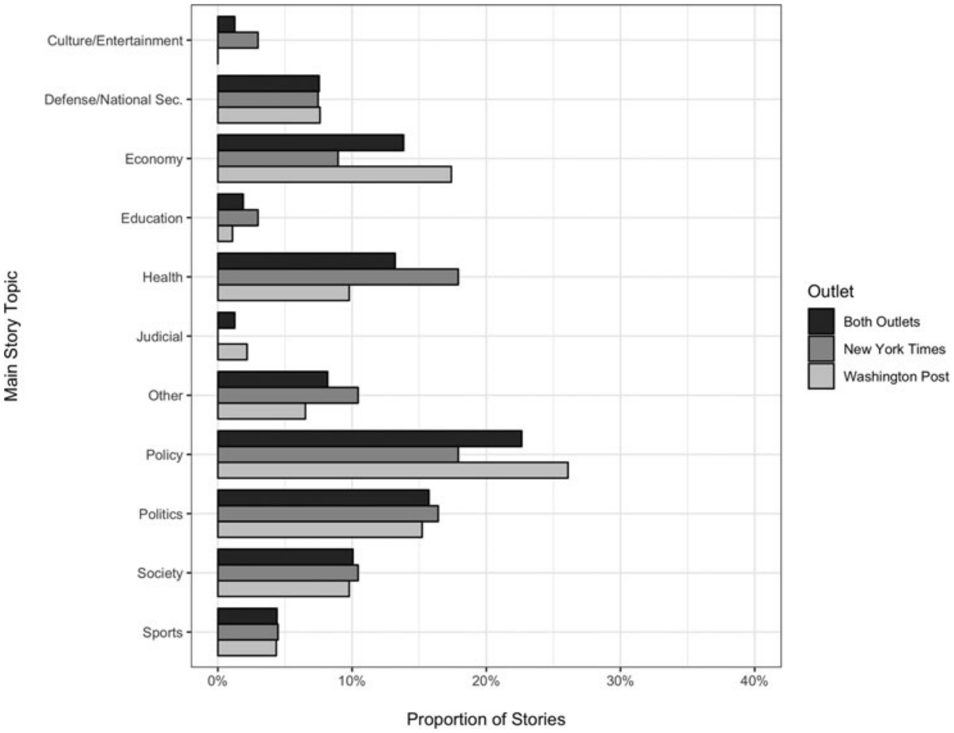

Figure 3. The proportion of stories by their main topic.

H3(c) posited that both outlets would rely primarily on governmental data sources. This hypothesis was supported as the majority of articles (66.0%, $n = 105$) drew on data from government sources, with *The New York Times* (67.2%, $n = 45$) using them at a marginally higher rate than *The Washington Post* (65.2%, $n = 60$). Notably, original data collection – either self-collected data (8.2%, $n = 13$) or data from other news organizations (5.0%, $n = 8$) – comprised the least-used categories. That order was mirrored by *The New York Times*, with self-collected data appearing in just 10.4% of stories ($n = 7$) and data from other news organizations in 6.0% ($n = 4$). This was also the case for *The Washington Post*, where self-collected data appeared in just 6.5% of stories ($n = 6$) and data from other news organizations in 4.3% ($n = 4$) (Figure 4).

RQ3 asked about the number of data sources listed in the average data journalism article. It was the case for both organizations that the median amount of data sources was two.

Discussion

The present study of *The New York Times*' and *The Washington Post*'s day-to-day data journalism found that content was typically produced by small teams, focused on "hard news," provided fairly uncomplex data visualizations with low levels of interactivity, relied primarily on institutional sources (especially government sources) and offered little original data collection, and incorporated just two data sources on average in a generally opaque manner. Taken as a whole, these findings suggest that

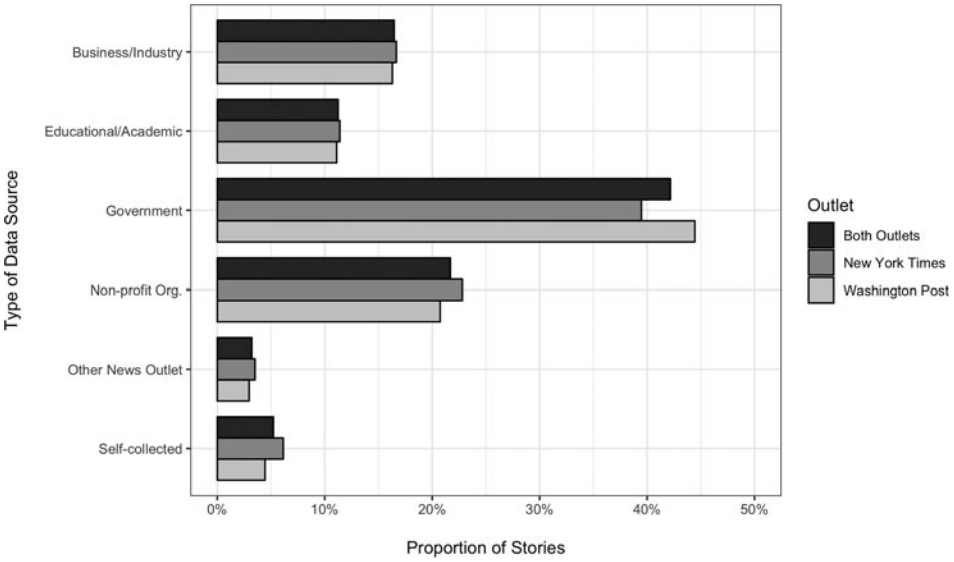

Figure 4. The proportion of stories that had one of six different data source types. Figures may exceed 100% because a single story could have multiple types of data sources.

"general data journalism" (Uskali and Kuutti 2015, 87) still has a long way to go before it can live up to the optimism that characterizes much of the data turn in journalism (see Coddington 2015).

Notably, these findings were generally consistent across both *The New York Times* and *The Washington Post*. With the exception of *The Washington Post*'s proclivity toward "hard news" and slightly greater transparency, the characteristics of the data journalism stories produced by the two organizations were remarkably consistent. This suggests that, at least at well-resourced and highly professionalized news organizations, data journalism may be homogenizing as logics and routines stabilize and become mimicked (Borges-Rey 2017). As Wright and Doyle (2018) argue, data journalism and data *in* journalism may be becoming "normalised" (p. 13) as a result of more training opportunities, greater amounts of journalists with basic data skills, and the increased adoption of data-oriented practices in different areas of journalism. Alternatively, these findings may be understood to suggest that some limits of data journalism are already being realized as journalists are hamstrung by the kinds of data they can access and the tools available to them (Lewis and Waters 2018; Wright and Doyle 2018). Future work should aim to explore the possibility of stabilization in the values, logic, and practices supporting data journalism as it becomes more mainstream.

In its idealized form, data journalism advances transparency, which in turn improves accountability and increases journalism's legitimacy among audiences (Allen 2008; Lewis and Westlund 2015). The finding from this study that nearly nine of ten articles did not link to any of the datasets used and that fewer than two in ten included a section for methodological details – two important aspects of disclosure transparency (Karlsson 2010) – highlights that even elite news organizations have a long way to go to realize that ideal. While access to data may be common in exemplary work (Parasie and Dagiral 2012; Tandoc and Oh 2017; cf. Young, Hermida, and Fulda 2018), that may not be the case in the day-to-day work that is more closely aligned with ritualized practice (see also Lowrey and Hou 2018). This study's instrument cannot elaborate on why so few resources were directly linked to and such limited methodological information made available. However, it is possible that transparency continues to be routinized into a strategic ritual that privileges discourse over practice (Karlsson 2010; see also Vos and Craft 2017) and that journalists may not prioritize leading readers to source material when attempting to meet deadlines during their day-to-day work or actively resist ceding too much autonomy through "over-sharing" (Plaisance and Skewes 2003; Singer 2007; Tandoc and Thomas 2017). Indeed, while stories generally used hyperlinks to point readers toward the organizations that collected the supporting data, they failed to make the specific materials available. This may also be due to limitations in the software used by news organizations to host supplemental information (e.g., data files). Future scholarship should explore whether there is something unique about data sources or the newsroom infrastructure that may impact newsworkers' willingness to make them readily accessible.

While the technological affordances of online journalism promote interactivity (Karlsson and Holt 2016), and data visualizations in particular lend themselves to it (Smit, de Haan, and Buijs 2014), day-to-day data journalism appears to incorporate a rather small amount of those affordances (see also Wright and Doyle 2018). The vast majority of articles did not include any interactive elements, a finding that is consistent with analyses of both day-to-day and award-winning works (Loosen, Reimer, and De Silva-Schmidt 2017; Stalph 2017; Tandoc and Oh 2017; cf. Ojo and Heravi 2018). Moreover, when used, the interactivity was quite limited. Appelgren's (2018, 15) observation that data journalism typically offers just an "illusion of interactivity" thus appears to hold true for elite and highly professionalized news organizations in the United States. While this content analysis cannot reveal why interactivity is seldom used, it is plausible that journalism's professional culture – which emphasizes linearity and control (Domingo 2008) – remains an important obstacle, though the difficulty of translating interactivity across platforms and devices may also play a role (Cantarero, González-Neira, and Valentini 2017). The findings also suggest that data journalism may not actually involve more conversational forms of interactivity than its closest predecessor, computer-assisted reporting, raising questions about whether that ought to be a dimension in conceptually distinguishing the two (see Coddington 2015). Additionally, while a lack of complexity in data visualizations has been previously attributed to the limitations of existing software and journalists' technical abilities, it is less likely that *The New York Times* and *The Washington Post* would suffer from those shortcomings. Rather, it is likely that the producers of data journalism subscribe to the broader journalistic value of simplification, perceiving semiotic simplicity as a strength rather than a weakness (Boyles and Meyer 2016; Engebretsen, Kennedy, and Weber 2018; Wright and Doyle 2018). From this perspective, data journalism's logic may borrow as much as it lends to the broader journalistic field.

Although the emphasis on "hard news" remains, the data journalism content analyzed in this study was more topically diverse than what one would expect from computer-assisted reporting, suggesting that to be a valuable conceptual distinction (see Coddington 2015). This topical diversity also suggests that data journalism is being integrated into more journalistic spaces and therefore presumably gaining wider acceptance among news professionals and audiences – a key challenge identified by scholars (e.g., Cawley 2016). This has boundary-shaping implications as new actors and logics are introduced to new spaces, which can impact not only existing ritualistic practices but also alter the allocation of material and symbolic resources (Lewis and Zamith 2017). Notably, although the scholarship emphasizes the interdisciplinary nature of data journalism as one of its key characteristics (Anderson 2017; Borges-Rey 2016), this study finds more limited potential for that in light of the relatively small number of credited authors – a finding consistent with prior work (Stalph 2017; Young, Hermida, and Fulda 2018). One interpretation of this finding is that, at least at elite news organizations, individual journalists have begun to gain sufficient knowledge to perform most aspects of data journalism work (see Wright and Doyle 2018). Another is that at least some of the work of data journalism, and especially more technically oriented labor like producing data visualizations, remains invisible and uncredited – or credited in alternative ways that may harm the status of particular actors within a social system (see Lewis and Zamith 2017). Ultimately, the proclivity toward individual or small-group work in many ways mirrors traditional practice.

Scholars have long been interested in information provenance within journalism (see Gandy 1982; Reich 2009) and have good reason to be concerned within the context of data journalism. Both *The New York Times* and *The Washington Post* drew primarily from institutional sources for data – and especially governmental bodies – and used few data sources. Even these elite outlets did little original data collection, perhaps because of the resource limitations in a fast-moving news environment or because of the desire to remain neutral and leverage the reputational authority attached to third-party institutions (see Borges-Rey 2016; Boyles and Meyer 2016). These findings are consistent with prior work (e.g., Loosen, Reimer, and De Silva-Schmidt 2017; Lowrey and Hou 2018; Tabary, Provost, and Trottier 2016) and raise important epistemological questions. Like journalism writ large (see Lewis, Williams, and Franklin 2008), data journalism may become increasingly dependent – if it isn't already – on information subsidies (Tandoc and Oh 2017). Such subsidies can be leveraged by third parties (e.g., think tanks and industry groups) to gain influence by promoting self-serving narratives through data while benefiting from a problematic epistemological mythology that emphasizes data's supposed neutrality and objectivity (Lesage and Hackett 2014). While new tools have made it easier for journalists and ordinary citizens to generate their own datasets, such efforts do not yet seem to be breaking into the mainstream. The ceding of responsibility with regard to information gathering has implications not only for journalistic authority but also for the quality of information in democratic societies, which is impoverished as fewer newsworkers originate facts (Lewis, Williams, and Franklin 2008; Lowrey and Hou 2018). Future work should compare data journalism to other forms of journalism within particular organizations to examine whether this form is perhaps even more dependent on institutional sources and information subsidies.

Journalism, among other institutions, has been challenged by declining levels of trust – rapidly so among some segments of the population (Usher 2018; Waisbord 2018) – and this study's findings help shed light on whether data journalism is reaching the potential celebrated by many scholars and practitioners for it to serve as a bulwark against that development. As routinely practiced by two elite and highly professionalized news outlets – *The New York Times* and *The Washington Post* – data journalism does not yet appear to be making the kind of impression forecast by those scholars and practitioners. While transparency may not be a silver bullet for restoring trust (Karlsson, Clerwall, and Nord 2017), it is valued by citizens (van der Wurff and Schönbach 2014). Yet, transparency remains limited, with the potential for engendering trust through methodological clarity and by making it easy for readers to check the journalist's work not being leveraged. The potential to transform journalism by making it more exploratory in nature while enhancing reader enjoyment and fostering positive attitudes (see Yang and Shen 2018) also remains unrealized, as evidenced by the limited use of interactive affordances. The increased topical diversity suggests that data journalistic practices and values are making their way beyond niches, with the slight proclivity toward "hard news" topics indicating that it is being used to further journalism's Fourth Estate function (Borges-Rey 2017). These efforts may engender trust, especially in light of the mythology around "data" (Lesage and Hackett 2014). However, the continued reliance on institutional sources – and the remarkable rarity of primary data collection – may limit that success and perhaps even become a disservice to society if that same problematic mythology is leveraged through information subsidies from those actors to manipulate rather than inform. Thus, in terms of products, contemporary day-to-day data journalism is perhaps better characterized as evolutionary rather than revolutionary, with its celebrated potential to serve as a leap forward for journalism and engender greater trust in it remaining untapped.

Acknowledgement

The author wishes to thank Sabrina Negrón and Rachel Perry-Gore for their many contributions to this project.

Disclosure Statement

No potential conflict of interest was reported by the author.

Notes

[1] Though this definition was conceived prior to the publication of Lowrey and Hou's (2018) study, there is considerable convergence between both definitions. This suggests the scholarship is moving toward a more consistent content-oriented conceptualization of data journalism.

[2] For example, a simple bar graph would be considered a univariate visualization, with the categories on the X-axis comprising the sole variable and the associated values displayed on the Y-axis not counted.

[3] Both organizations instead tended to reference the homepage of the organization responsible for the supporting data, rather than the specific database or dataset they used.

ORCID

Rodrigo Zamith http://orcid.org/0000-0001-8114-1734

References

Allen, David S. 2008. “The Trouble with Transparency: The Challenge of Doing Journalism Ethics in a Surveillance Society.” *Journalism Studies* 9 (3): 323–340. doi:10.1080/14616700801997224.

Anderson, Chris W. 2017. “Social Survey Reportage: Context, Narrative, and Information Visualization in Early 20th Century American Journalism:” *Journalism* 18 (1): 81–100. doi:10.1177/1464884916657527.

Appelgren, Ester. 2018. “An Illusion of Interactivity.” *Journalism Practice* 12 (3): 308–325. doi:10.1080/17512786.2017.1299032.

Appelgren, Ester, and Gunnar Nygren. 2014. “Data Journalism in Sweden.” *Digital Journalism* 2 (3): 394–405. doi:10.1080/21670811.2014.884344.

Atkinson, Claire. 2017. “The Most Important Competition in Newspapers Heats up.” *NBC News*. https://www.nbcnews.com/news/us-news/washington-post-still-plays-catch-gaining-times-n833236.

Borges-Rey, Eddy. 2016. “Unravelling Data Journalism.” *Journalism Practice* 10 (7): 833–843. doi:10.1080/17512786.2016.1159921.

Borges-Rey, Eddy. 2017. “Towards an Epistemology of Data Journalism in the Devolved Nations of the United Kingdom: Changes and Continuities in Materiality, Performativity and Reflexivity.” *Journalism*. Advance online publication. doi:10.1177/1464884917693864.

Boyles, Jan Lauren, and Eric Meyer. 2016. “Letting the Data Speak.” *Digital Journalism* 4 (7): 944–954. doi:10.1080/21670811.2016.1166063.

Boyles, Jan Lauren, and Eric Meyer. 2017. “Newsrooms Accommodate Data-Based News Work.” *Newspaper Research Journal* 38 (4): 428–438. doi:10.1177/0739532917739870.

Bucy, Erik P., and Chen-Chao Tao. 2007. “The Mediated Moderation Model of Interactivity.” *Media Psychology* 9 (3): 647–672. doi:10.1080/15213260701283269.

Cantarero, Teresa Nozal, Ana González-Neira, and Elena Valentini. 2017. “Newspaper Apps for Tablets and Smartphones in Different Media Systems: A Comparative Analysis.” *Journalism*. Advance online publication. doi:10.1177/1464884917733589.

Carlson, Matt. 2018. “Confronting Measurable Journalism.” *Digital Journalism* 6 (4): 406–417. doi:10.1080/21670811.2018.1445003.

Cawley, Anthony. 2016. “Is There a Press Release on That?” In *Big Data Challenges*, edited by Anno Bunnik, Anthony Cawley, Michael Mulqueen, and Andrej Zwitter, 49–58. London: Palgrave Macmillan.

Coddington, Mark. 2015. “Clarifying Journalism’s Quantitative Turn.” *Digital Journalism* 3 (3): 331–348. doi:10.1080/21670811.2014.976400.

Domingo, David. 2008. “Interactivity in the Daily Routines of Online Newsrooms: Dealing with an Uncomfortable Myth.” *Journal of Computer-Mediated Communication* 13 (3): 680–704. doi:10.1111/j.1083-6101.2008.00415.x.

Engebretsen, Martin, Helen Kennedy, and Wibke Weber. 2018. “Data Visualization in Scandinavian Newsrooms: Emerging Trends in Journalistic Visualisation Practices.” *Nordicom Review*. Advance online publication. doi:10.2478/nor-2018-0007.

Felle, Tom. 2016. “Digital Watchdogs? Data Reporting and the News Media’s Traditional ‘Fourth Estate’ Function.” *Journalism* 17 (1): 85–96. doi:10.1177/1464884915593246.

Fink, Katherine, and Christopher W. Anderson. 2015. “Data Journalism in the United States.” *Journalism Studies* 16 (4): 467–481. doi:10.1080/1461670X.2014.939852.

Gandy, Oscar H. 1982. *Beyond Agenda Setting: Information Subsidies and Public Policy*. Norwood, NJ: Ablex.

Gray, Jonathan, Liliana Bounegru, and Lucy Chambers. 2012. *The Data Journalism Handbook*. Sebastopol, CA: O’Reilly Media.

Handler, Reinhard A., and Raul Ferrer Conill. 2016. “Open Data, Crowdsourcing and Game Mechanics. A Case Study on Civic Participation in the Digital Age.” *Computer Supported Cooperative Work* 25 (2): 153–166. doi:10.1007/s10606-016-9250-0.

Howard, Alexander Benjamin. 2014. *The Art and Science of Data-Driven Journalism*. New York: Tow Center for Digital Journalism.

Karlsson, Michael. 2010. “Rituals of Transparency.” *Journalism Studies* 11 (4): 535–545. doi:10.1080/14616701003638400.

Karlsson, Michael, Christer Clerwall, and Lars Nord. 2017. “Do Not Stand Corrected: Transparency and Users’ Attitudes to Inaccurate News and Corrections in Online Journalism.” *Journalism & Mass Communication Quarterly* 94 (1): 148–167. doi:10.1177/1077699016654680.

Karlsson, Michael, and Kristoffer Holt. 2016. “Journalism on the Web.” In *Oxford Research Encyclopedia of Communication*, edited by Jon F. Nussbaum. Oxford: Oxford University Press.

Knight, Megan. 2015. “Data Journalism in the UK: A Preliminary Analysis of Form and Content.” *Journal of Media Practice* 16 (1): 55–72. doi:10.1080/14682753.2015.1015801.

Krippendorff, Klaus. 2011. “Agreement and Information in the Reliability of Coding.” *Communication Methods and Measures* 5 (2): 93–112. doi:10.1080/19312458.2011.568376.

Larsson, Anders Olof. 2011. “Interactive to Me – Interactive to You? A Study of Use and Appreciation of Interactivity on Swedish Newspaper Websites:” *New Media & Society* 13 (7): 1180–1197. doi:10.1177/1461444811401254.

Lesage, Frédérik, and Robert A. Hackett. 2014. “Between Objectivity and Openness—The Mediality of Data for Journalism.” *Media and Communication* 2 (2): 42–54. doi:10.17645/mac.v2i2.128.

Lewis, Justin, Andrew Williams, and Bob Franklin. 2008. “Four Rumours and an Explanation.” *Journalism Practice* 2 (1): 27–45. doi:10.1080/17512780701768493.

Lewis, Norman P., and Stephenson Waters. 2018. “Data Journalism and the Challenge of Shoe-Leather Epistemologies.” *Digital Journalism* 6 (6): 719–736. doi:10.1080/21670811.2017.1377093.

Lewis, Seth C., and Oscar Westlund. 2015. “Big Data and Journalism.” *Digital Journalism* 3 (3): 447–466. doi:10.1080/21670811.2014.976418.

Lewis, Seth C., and Rodrigo Zamith. 2017. “On the Worlds of Journalism.” In *Remaking the News: Essays on the Future of Journalism Scholarship in the Digital Age*, edited by Pablo J. Boczkowski and Chris W. Anderson, 111–128. Cambridge, MA: MIT Press.

Lischka, Juliane A., and Michael Messerli. 2016. “Examining the Benefits of Audience Integration.” *Digital Journalism* 4 (5): 597–620. doi:10.1080/21670811.2015.1068128.

Loosen, Wiebke, Julius Reimer, and Fenja De Silva-Schmidt. 2017. “Data-Driven Reporting: An on-Going (r)evolution?” *Journalism*. Advance online publication. doi:10.1177/1464884917735691.

Lowrey, Wilson, and Jue Hou. 2018. “All Forest, No Trees? Data Journalism and the Construction of Abstract Categories.” *Journalism*. Advance online publication. doi:10.1177/1464884918767577.

Meyer, Philip. 1973. *Precision Journalism: A Reporter’s Guide to Social Science Methods*. Bloomington: Indiana University Press.

Napoli, Phillip M. 1999. “Deconstructing the Diversity Principle.” *Journal of Communication* 49 (4): 7–34. doi:10.1111/j.1460-2466.1999.tb02815.x.

Nguyen, An, and Jairo Lugo-Ocando. 2015. “The State of Data and Statistics in Journalism and Journalism Education: Issues and Debates.” *Journalism* 17 (1): 3–17. doi:10.1177/1464884915593234.

Ojo, Adegboyega, and Bahareh Heravi. 2018. “Patterns in Award Winning Data Storytelling.” *Digital Journalism* 6 (6). Routledge: 693–718. doi:10.1080/21670811.2017.1403291.

Owen, D. 2017. *The State of Technology in Global Newsrooms*. Washington, D.C.: International Center for Journalists.

Parasie, Sylvain, and Eric Dagiral. 2012. “Data-Driven Journalism and the Public Good: ‘Computer-Assisted-Reporters’ and ‘Programmer-Journalists’ in Chicago.” *New Media & Society* 15 (6): 853–871. doi:10.1177/1461444812463345.

Plaisance, Patrick Lee, and Elizabeth A. Skewes. 2003. “Personal and Professional Dimensions of News Work: Exploring the Link between Journalists’ Values and Roles.” *Journalism & Mass Communication Quarterly* 80 (4): 833–848. doi:10.1177/107769900308000406.

Reich, Zvi. 2009. *Sourcing the News*. Cresskill, NJ: Hampton Press.

Reimer, Julius, and Wiebke Loosen. 2017. “Data Journalism at Its Finest: A Longitudinal Analysis of the Characteristics of Award-Nominated Data Journalism Projects.” In *News, Numbers and Public Opinion in a Data-Driven World*, edited by An Nguyen, 93–112. New York: Bloomsbury Academic.

Singer, Jane B. 2007. “Contested Autonomy.” *Journalism Studies* 8 (1): 79–95. doi:10.1080/14616700601056866.

Smit, Gerard, Yael de Haan, and Laura Buijs. 2014. “Visualizing News.” *Digital Journalism* 2 (3): 344–354. doi:10.1080/21670811.2014.897847.

Stalph, Florian. 2017. “Classifying Data Journalism.” *Journalism Practice*. Advance online publication. doi:10.1080/17512786.2017.1386583.

Tabary, Constance, Anne-Marie Provost, and Alexandre Trottier. 2016. “Data Journalism’s Actors, Practices and Skills: A Case Study from Quebec.” *Journalism* 17 (1): 66–84. doi:10.1177/1464884915593245.

Tandoc, Edson C., Jr., and Soo-Kwang Oh. 2017. “Small Departures, Big Continuities? Norms, Values, and Routines in The Guardian’s Big Data Journalism.” *Journalism Studies* 18 (8): 997–1015. doi:10.1080/1461670X.2015.1104260.

Tandoc, Edson C., and Ryan J. Thomas. 2017. “Readers Value Objectivity over Transparency.” *Newspaper Research Journal* 38 (1): 32–45. doi:10.1177/0739532917698446.

Usher, Nikki. 2018. “Re-Thinking Trust in the News.” *Journalism Studies* 19 (4): 564–578. doi:10.1080/1461670X.2017.1375391.

Uskali, Turo I., and Heikki Kuutti. 2015. “Models and Streams of Data Journalism.” *The Journal of Media Innovations* 2 (1): 77–88. doi:10.5617/jmi.v2i1.882.

van der Wurff, Richard, and Klaus Schönbach. 2014. “Audience Expectations of Media Accountability in the Netherlands.” *Journalism Studies* 15 (2): 121–137. doi:10.1080/1461670X.2013.801679.

Veglis, Andreas, and Charalampos Bratsas. 2017a. “Reporters in the Age of Data Journalism.” *Journal of Applied Journalism & Media Studies* 6 (2): 225–244. doi:10.1386/ajms.6.2.225_1.

Veglis, Andreas, and Charalampos Bratsas. 2017b. “Towards a Taxonomy of Data Journalism.” *Journal of Media Critiques* 3 (11): 109–121. doi:10.17349/jmc117309.

Voakes, Paul S., Jack Kapfer, David Kurpius, and David Shano-Yeon Chern. 1996. “Diversity in the News: A Conceptual and Methodological Framework.” *Journalism & Mass Communication Quarterly* 73 (3): 582–593. doi:10.1177/107769909607300306.

Vos, Tim P., and Stephanie Craft. 2017. “The Discursive Construction of Journalistic Transparency.” *Journalism Studies* 18 (12): 1505–1522. doi:10.1080/1461670X.2015.1135754.

Waisbord, Silvio. 2018. “Truth Is What Happens to News.” *Journalism Studies* 19 (13): 1866–1878. doi:10.1080/1461670X.2018.1492881.

Weber, Matthew S., and Allie Kosterich. 2018. *Managing a 21st-Century Newsroom Workforce: A Case Study of NYC News Media*. New York: Tow Center for Digital Journalism. https://academiccommons.columbia.edu/catalog/ac:dbrv15dv5q.

Wright, Scott, and Doyle, Kim. 2018. “The Evolution of Data Journalism: A Case Study of Australia.” *Journalism Studies. Advance online publication*. https://doi.org/10.1080/1461670X.2018.1539343

Yang, Fan, and Fuyuan Shen. 2018. “Effects of Web Interactivity: A Meta-Analysis.” *Communication Research* 45 (5): 635–658. doi:10.1177/0093650217700748.

Young, Mary Lynn, and Alfred Hermida. 2015. “From Mr. and Mrs. Outlier To Central Tendencies.” *Digital Journalism* 3 (3): 381–397. doi:10.1080/21670811.2014.976409.

Young, Mary Lynn, Alfred Hermida, and Johanna Fulda. 2018. “What Makes for Great Data Journalism?” *Journalism Practice* 12 (1): 115–135. doi:10.1080/17512786.2016.1270171.

Zamith, Rodrigo. 2018. “Quantified Audiences in News Production.” *Digital Journalism* 6 (4): 418–435. doi:10.1080/21670811.2018.1444999.

Zamith, Rodrigo, and Seth C. Lewis. 2015. “Content Analysis and the Algorithmic Coder: What Computational Social Science Means for Traditional Modes of Media Analysis.” *The Annals of the American Academy of Political and Social Science* 659 (1): 307–318. doi:10.1177/0002716215570576.