



Data Journalism and Ethics: Best Practices in the Winning Projects (DJA, OJA and Sigma Awards)

María-Ángeles Chaparro-Domínguez & Jesús Díaz-Campo

To cite this article: María-Ángeles Chaparro-Domínguez & Jesús Díaz-Campo (2021): Data Journalism and Ethics: Best Practices in the Winning Projects (DJA, OJA and Sigma Awards), Journalism Practice, DOI: [10.1080/17512786.2021.1981773](https://doi.org/10.1080/17512786.2021.1981773)

To link to this article: <https://doi.org/10.1080/17512786.2021.1981773>



Published online: 27 Sep 2021.



Submit your article to this journal [↗](#)



Article views: 715



View related articles [↗](#)



View Crossmark data [↗](#)



Citing articles: 1 View citing articles [↗](#)



Data Journalism and Ethics: Best Practices in the Winning Projects (DJA, OJA and Sigma Awards)

María-Ángeles Chaparro-Domínguez ^a and Jesús Díaz-Campo ^b

^aPeriodismo y Comunicación Global, Universidad Complutense de Madrid Facultad de Ciencias de la Información, Madrid, Spain; ^bFacultad de Empresa y Comunicación, Universidad Internacional de La Rioja, Logroño, Spain

ABSTRACT

Data journalism is a consolidated specialization in the newsrooms of many of the world's media outlets. Despite this, little research has been conducted on the ethical principles followed in this field of journalism. Data journalism uses different types of software to find its stories by statistically analyzing large datasets. Our research examines the winning projects of the Data Journalism Awards, Sigma Awards, and Online Journalism Awards, the last in the data journalism category, between 2012 and 2020. Using qualitative content analysis, we analyzed these projects from a three-fold ethical perspective: verification and data analysis, transparency, and privacy. Our main findings show that the winning projects complied with verification and data analysis, which is a standard practice to cross-check data from various sources and contextualize them adequately. In contrast, transparency and privacy principles were followed to a lesser extent. In light of these results, we propose that future research should focus on the perceptions of data journalists and users regarding the ethical standards that these projects meet.

KEYWORDS

Data journalism; ethics; transparency; privacy; verification; data journalism awards; content analysis

Introduction

Data journalism brings rigor to journalism and improves the quality of its practice. This, at least, is the opinion of data journalists from 43 countries (Heravi 2018). This specialization also lends credibility to journalism (Karlsson 2010; Coddington 2015) and makes for greater transparency, according to most German professional journalists (Beiler, Irmer, and Breda 2020).

However, does data journalism lead by example? That is, does it make journalism more rigorous and transparent in uncovering stories based on data hitherto unknown to society whilst remaining truthful and transparent with the public?

I usually provide a methodological note to explain every step I have made when collecting and analysing my data. When I teach beginners, I constantly repeat that this is the most important part of their project, even if readers do not care (Porlezza and Splendore 2019).

The above statement, attributed to an Italian data journalist, reflects the importance of transparency in the methodology used in this specialization. However, besides

transparency, does data journalism apply additional privacy or verification and data analysis standards? This study analyzed the professional ethical principles adhered to by the world's best data journalism projects.

Studying these projects is advantageous as they contribute to shaping professional standards (Jenkins and Volz 2018) so that over time the most outstanding qualities of these works become a benchmark for the sector (Loosen, Reimer, and De Silva-Schmidt 2017).

On the other hand, Ausserhofer et al. (2020), in their structured literature review on data journalism, analyzing studies published between 1996 and 2015, highlight the many studies on epistemology focusing on large media outlets, both nationally or locally. Consequently, they recommended conducting studies on small media and other bodies involved in data journalism by using robust theoretical frameworks and international scope. On the latter, several authors have emphasized the excessive preponderance of North America and Europe in this research field (Fink and Anderson 2015; Zhang and Feng 2019; Mutsvairo 2019).

In line with their proposals, this study takes an international approach that goes beyond Europe and North America. Although most award-winning data journalism projects come from those two areas, we also analyzed award-winning projects from other continents and countries that have received little or no attention from academia in Asia, Africa, and Latin America, such as the Philippines, India, Kenya, Venezuela, Peru, Colombia, and Argentina. Therefore, our study took an international approach with a robust theoretical framework based on research in ethics, data journalism and big data. In addition, we analyzed small media, associations and foundations that have developed award-winning data journalism projects, such as Poder (Mexico), FactChecker (India), Ojo Público (Peru), Disclose (France) or Civio (Spain).

Our paper studies winning projects of the Data Journalism Awards (DJA), Sigma Awards and Online Journalism Awards (OJA), the last of the data journalism category, between 2012 and 2020. Our results will reveal the best practices in journalistic ethics in these winning projects, which have become benchmarks thanks to their quality. We developed our ethical analysis from a three-fold perspective: 1) verification and data analysis, 2) transparency, and 3) privacy.

Conceptual Framework: Ethical Values in Data Journalism

Verification and Data Analysis

The public views data journalism as more rigorous than its traditional form because of its methodology and reliance on numbers (Borges-Rey 2016; Kovach and Rosenstiel 2014). This factor gives the specialization a false sense of authority enhanced by its intrinsic technological characteristics (Appelgren 2018). However, all data have varying types of failures (Craig, Ketterer, and Yousuf 2017; Messner and Garrison 2007), so the verifying and cleansing of data is indispensable.

Cross-checking data from various sources is a critical task. According to some studies, data journalists frequently use an average of two data sources in their projects (Zamith 2019; Stalph 2018). However, some journalists do not consult any additional sources beyond what their dataset has provided them with, weakening the analytical

capacity and monitoring of data journalism (Loosen, Reimer, and De Silva-Schmidt 2017).

Statistically analysing data to find social interest stories is another phase of the journalistic process (Coddington 2015). Such analysis is conditioned by the objectives of the research, whereby varying types of software will be chosen (Tong and Zuo 2019). Another critical aspect in this area is to lend context to the data, including visualizations, to avoid possible distortions implying facts that are not correct (Bradshaw 2014; Cairo 2016). Hyperlinks should also be included to enable the reader to consult sources of information, among other reasons (Zamith 2019).

Transparency

The concept of transparency in journalism is related to credibility, accountability and truth (Singer 2007; Avery 2010; Phillips 2010). In addition, it implies making public traditionally private factors that influence the creation of news (Allen 2008), which has become a growing demand from society (Phillips 2010). Karlsson (2010) maintains there are two types of transparency: 1) *Disclosure Transparency*, meaning that the medium or journalist explains how the news is selected and produced, and 2) *Participatory Transparency*, which involves to a greater or lesser degree the participation of readers.

One of the most evident examples of transparency in data journalism is making the dataset accessible to the public, that is, the database upon which the research is based, through hyperlinks or other means (Parasie and Dagiral 2013; Tandoc, Jr, and Oh 2017). Thus, readers can explore, audit, replicate, and obtain information from the data underpinning the investigation (Young, Hermida, and Fulda 2018). If the dataset is not made public, journalists become part of a system that seeks to keep things secret (Porlezza and Splendore 2019). Making the data public was standard practice in most of the projects analyzed in some studies (Parasie and Dagiral 2013; Tandoc, Jr, and Oh 2017). However, in others, this principle was rarely or never applied (Lowrey and Hou 2018; Stalph 2018; Young, Hermida, and Fulda 2018; Zhang and Feng 2019; Zamith 2019).

Besides access to the dataset, transparency is found in other actions such as the use of timestamps showing dates of publication and updating, providing the original documents, reporting errors or giving explanations of how the reporting process was undertaken (Karlsson 2010). However, according to some studies (Lowrey and Hou 2018; Zhang and Feng 2019; Zamith 2019), this last practice is uncommon, while others regard it as a consolidated action (Loosen, Reimer, and De Silva-Schmidt 2017).

These practices shape the *Participatory Openness* concept (Coddington 2015; Loosen, Reimer, and De Silva-Schmidt 2017), which entails public involvement in different areas of the journalistic process. *Open Journalism* is another concept that takes the same approach, implying that the journalistic process is more transparent (Porlezza 2019).

Privacy

Privacy in journalism concerns verification since posting erroneous or inaccurate information, that which has not been substantiated, can lead to an invasion of privacy (Craig, Ketterer, and Yousuf 2017). As for data journalism in particular, large datasets including personal information, are a sensitive issue for privacy. The multitude of

people included in a dataset makes it very difficult to obtain informed consent, so these individuals are not asked whether or not they wish to take part in a journalistic investigation, and most of the time, they are not even aware that their data is being used for journalistic purposes (Fairfield and Shtein 2014). Some authors state that private data from datasets should be aggregated and unidentified to prevent misuse, although when it comes to large databases, anonymization is an insecure measure as it can be reversed (Daniel, Flew, and Spurgeon 2010; Fairfield and Shtein 2014).

Protecting the identity of sources has traditionally been a fundamental principle of journalism. However, in algorithm-based journalism, such as data journalism, it is debatable whether such protection is possible because journalists are required to disclose all their sources, following the ethical principles of data transparency (Ward 2014). In this sense, one of the problems with working with data is the lack of precautions taken by journalists to protect anonymous data sources (Craig, Ketterer, and Yousuf 2017).

Privacy concerns related to personal harm have increased in recent years due to the rise of data mining (Craig, Ketterer, and Yousuf 2017). Big data technologies increase the costs of meeting traditional ethical principles while drastically reducing privacy invasion costs (Fairfield and Shtein 2014). Bearing in mind that private information can now be easily shared and searched, data journalists need to weigh the benefits of disclosing the data against the risks of personal harm they may inflict by publishing them (Howard 2014).

Taking into account the different aspects that shape verification and data analysis, transparency and privacy in data journalism, we posed the following research questions:

RQ1. How is verification and data analysis reflected in the award-winning projects analyzed?

RQ2. How is transparency reflected in the award-winning projects analyzed?

RQ3. How is privacy reflected in the award-winning projects analyzed?

Literature Review

Ethics and Data Journalism

Data journalism is a specialization that integrates several disciplines such as journalism, computer science, social science, and design, and promotes data and fact-driven journalism (Heravi 2019). This specialization incorporates a wide range of styles, such as visualizations or text-based stories, which both tell a story using numbers and statistics (Rogers 2017).

According to many studies, the origins of data journalism can be traced to precision journalism and computer-assisted reporting (CAR). Therefore, data journalism can be considered an evolution of the two (Chaparro-Domínguez 2014; Elías-Pérez 2015; Splendore 2016; La-Rosa and Sandoval-Martín 2016). Precision journalism, pioneered by journalist Philip Meyer in the 1960s (Meyer 1973), applied social science research methods to conventional journalism (Splendore 2016). By the 1980s, when computers were commonplace in most newsrooms, journalists used CAR to develop journalistic investigations through databases (La-Rosa and Sandoval-Martín 2016; Splendore 2016). Coddington (2015) noted that CAR was public affairs oriented, while data journalism was characterized by its participatory openness and cross-field hybridity.

Data journalism implies different threats in journalistic practice (Bruns 2016), mainly related to truth, privacy and harm (Craig, Ketterer, and Yousuf 2017), such as in the

tampering of results or the presentation of incomplete data that could harm people (Daniel, Flew, and Spurgeon 2010). Generally speaking, it can be said that the classic ethical standards of journalism, such as respect for truth, are pertinent to these new challenges, although their nature needs redefining (Deuze and Yeshua 2001; Díaz-Campo and Segado-Boj 2015; Lewis and Westlund 2015). In this sense, working with quantifiable data does not necessarily mean that such data are equivalent to the truth (Boyd and Crawford 2012).

These threats explain why, in recent years, research has begun to focus on different ethical aspects related to this discipline. For example, Boyles and Meyer (2016) interviewed 31 data journalists from leading U.S. newspapers to ascertain their perception of social responsibility from the point of view of dialogue with the audience.

Díaz-Campo and Chaparro-Domínguez (2018) analyzed ethical guidelines on data journalism in the codes of practice of journalism associations in 18 Latin American countries. A study by Tong and Zuo (2019) showed various reasons why objectivity in data journalism cannot be achieved, including the difficulty of verifying data and journalists' lack of knowledge of the context of data creation and its algorithms, among others.

Other studies have also been published on defining ethical principles for professional data journalism practice. For example, Bradshaw (2014) raised concerns about privacy and source protection, rigor in the data collection process (which must be contextualized adequately), necessary scepticism of data from online surveys and the challenge of audience interactivity. Fairfield and Shtein (2014) developed a framework to draw up three ethical principles for data journalism: 1) to be aware of potential harm and damage that can be caused by disseminating information, 2) ensure an appropriate sample size for the journalistic objective, and 3) ensure the data set is a homogeneous whole, not an accumulation of data.

Dörr and Hollnbuchner (2017) outlined three areas (organizational, professional and social) and three stages in the journalistic process (input, throughput and output), in which they included different ethical questions, including appropriate data collection methods, respect for privacy, accuracy, objectivity and transparency. Finally, in their study on the ethical dimensions of data journalism, Craig, Ketterer, and Yousuf (2017) proposed various ethical frameworks of action: 1) freedom versus responsibility, 2) journalistic purpose, 3) privacy, 4) verification, 5) consequences and 6) alternatives.

Hence, we have observed how different academic ethical principles have been proposed for journalism practice. This study questions whether the most relevant ethical principles for the practice of data journalism (related to verification and data analysis, transparency and privacy) have been applied in the best data journalism projects worldwide.

Award-winning Data Projects

As demonstrated by the number and variety of research topics over the last decade, data journalism can be considered a consolidated and mature field of study (Beiler, Irmer, and Breda 2020). Some studies, for example, have focused on the origins of data journalism and their epistemological aspect (Hammond 2017). Others have centered on data journalists and their methodology in the media in Canada (Hermida and Young 2017), (Hannaford 2015), the United States (Parasie and Dagiral 2013), Belgium (De Maeyer et al. 2015), Norway (Karlsen and Stavelin 2013), Australia (Wright and Doyle 2019), India (Kashyap and

Bhaskaran 2020) or some of the Arab states¹ (Fahmy and Attia 2020), and others in the educational aspects of this discipline (Heravi 2019; Splendore et al. 2016; Chaparro-Domínguez 2014).

The maturity of the discipline has led to innovative research, including the use of statistical data analysis as a method of social prediction (Pentzold and Fechner 2019), its relationship with civic technologies (Baack 2018) or with *user generated content* and *crowdsourcing* (Palomo, Teruel, and Blanco-Castilla 2019), its implementation by major television channels in South Korea and the United States (Lim 2019) or local initiatives in the United Kingdom (Arias-Robles and López López 2020).

Nevertheless, there has been little research in recent years on award-winning or nominated data journalism projects. Loosen, Reimer, and De Silva-Schmidt (2017) studied aspects such as themes, sources or the authors of nominee Data Journalism Awards between 2013 and 2016. Among their main findings, these authors highlighted politics as the central theme of the winning projects, which in many cases use simple visualizations and public data from official institutions.

On the other hand, Ojo and Heravi (2018) analyzed 44 winners of the Data Journalism Awards (2013-2016), focusing on the technologies used and the type of stories narrated. They concluded that data analysis, web development and publishing, and data visualization are the most used technologies and that the most frequent type of story is an in-depth account of a phenomenon.

Young, Hermida, and Fulda (2018) studied nominated and award-winning projects by Canadian media in the Data Journalism Awards, Online Journalism Awards, and data journalism projects award-winning of The Canadian Association of Journalists. They found that the quality of the projects was restricted both by the use of free online apps such as Google Maps, which allow for little customization, and by the traditional canons of journalism used by journalists.

On the other hand, Appelgren (2018) analyzed the paternalism prevailing in the 31 data journalism projects presented at the 2013, 2015, and 2016 in the Nordic Data Journalism Awards. The author found three paternalistic characteristics (controlling functionality, the illusion of interactivity and linearity), which include some types of control over the public.

Lastly, Palomo, Teruel, and Blanco-Castilla (2019), in their study on award-winning projects of the Data Journalism Awards between 2012 and 2018, concluded that media-audience collaboration allowed for large-scale data journalism projects.

It is noteworthy that previous studies have focused on the formal aspects (interactivity, technologies and visualizations used) and thematic parts of the projects. However, none has explored the ethical aspects, which is the purpose of our research.

Materials and Methods

This research used qualitative content analysis, a technique previously used in research on data journalism (Craig, Ketterer, and Yousuf 2017; Appelgren 2018; Palomo, Teruel, and Blanco-Castilla 2019). We developed a code with three categories that included 13 variables, which we applied to 79 data journalism projects in 20 countries around the world.² Because most research on data journalism focuses on short periods (Ausserhofer et al. 2020), we followed Knight's recommendation (2015) and conducted a long-term study covering 2012-2020.

Sample: DJA, Sigma Awards and OJA

The Data Journalism Awards (DJA) is the world’s oldest data journalism accolades.³ The Global Editors Network (GEN), an international association of media publishers, founded in 2011, launched these awards in 2012 and ran them until late 2019, when they ceased due to a lack of funding.⁴ Consequently, the Sigma Awards were created the following year, acknowledging quality data journalism across the globe. These awards are organized by the European Journalism Center, an independent institute based in Maastricht (The Netherlands).⁵

Besides these two awards, we also included the University of Florida Award in Investigative Data Journalism category (Small/Medium Newsroom and Large Room) from the Online Journalism Awards (OJA) of the Online News Association (ONA).⁶ Given the professional consolidation of data journalism, this association, made up of digital journalists from across the world, added this category for the first time to its awards in 2014. Hence, we studied the award-winning projects from all three major international data journalism awards.⁷

In total, 79 projects were analyzed that won⁸ the Data Journalism Awards (57), OJA (13⁹), in the category of data journalism, and the Sigma Awards (9) (see [Table 1](#)). These comprise most of the projects awarded in the period under study (2012-2020), made up of 119 projects. We excluded the following from this study: 1) award-winning projects published in languages other than English, Spanish and Portuguese, due to linguistic misunderstanding, 2) awards to media outlets without a reference to a particular journalistic project, 3) awards to journalists in acknowledgement of a professional career instead of a specific project¹⁰, 4) some projects from the Innovation and the Open Data categories, which cannot be viewed as data journalism projects, such as the Associated Press DataKit, a computer tool for data journalists,¹¹ and 5) projects whose webpages are down and cannot be recovered through the Internet Archive, Google Cache or other tools.¹² Following this selection, we analyzed 79 projects.

Measurement: Variables

The ethical categories, verification and data analysis, transparency and privacy, which form the backbone of our analysis, were broken down into the following set of thirteen variables (see [Table 2](#)):

Table 1. Basic characteristics of the sample.

Analysis period	2012–2020
Number of projects analyzed	79
- Data Journalism Awards	57
- Online Journalism Awards	13
- Sigma Awards	9
Number of countries of the projects analyzed	20*
Languages of the analyzed projects	English Spanish Portuguese
Types of awards analyzed	Winning projects Honorable mentions Special citations Citations for excellence

*This calculation omits the winning global collaborative projects Swiss Leaks (involving journalists from 47 countries), Luxembourg Leaks (with journalists from 26 countries) and The Troika Laundromat (with journalists from 19 countries).

1. Verification and data analysis

1.1. Data cross-checking: this confirmed that the data analyzed were correct (Dörr and Hollnbuchner 2017).

1.2. Contextualization of data: data are contextualized correctly, as charts and tables have “0” as the basis for avoiding distortions, time scales show long-term trends or figures are shown in a manner that does not lead to confusion (Bradshaw 2014).

1.3. Data analysis: not only were the data shown, but statistical or trend analyses were carried out using them (Dörr and Hollnbuchner 2017).

1.4. Prevent data tampering: the authors did not commit significant harm, distortions or errors when displaying data (Crawford, Miltner, and Gray 2014).

1.5. Reliability of hyperlinks: the links that appeared in the projects work correctly (Díaz-Campo and Segado-Boj 2015).

2. Transparency

2.1. Names of data sources: the project clearly identifies the sources from where the data were taken (Dörr and Hollnbuchner 2017).

2.2. Access to the dataset: the database used for the project is given open access through hyperlinks or other means (Young, Hermida, and Fulda 2018; Tandoc, Jr, and Oh 2017).

2.3. Timestamps (I). Creation date: indicates when the project is first uploaded to the media website or another body (Karlsson 2010).

2.4. Timestamps (II). Update date: the project shows the date(s) on which the content was updated (Karlsson 2010).

2.5. Correction note: acknowledges errors in the published story that are later corrected (Karlsson 2010).

2.6. Description of the journalistic development process: the project describes the methodology used in its development (Karlsson 2010; Rupar 2006).

3. Privacy

3.1. Sources protection: the metadata, whether digital or scanned, in the original files used for the project are hidden to prevent identification of sources (Bradshaw 2014).

3.2. Justification of excluded personal information: the project explains why personal data such as names, addresses or other relevant data are excluded (Craig, Ketterer, and Yousuf 2017).

Table 2. General outline of the qualitative research analysis code.

Categories	Variables
Verification and data analysis	Data cross-checking Contextualization of data Data analysis Prevent data tampering Reliability of hiperlinks
Transparency	Names of data sources Access to the dataset Timestamps * Creation date * Update date Errata sheet Description of the journalistic development process
Privacy	Sources protection Justification of excluded personal information Reference to informed consent if it has been used

3.3. Reference to informed consent if it has been used: the project mentions whether an informed consent document is used (Suárez-Gonzalo 2017; Fairfield and Shtein 2014).

Having defined all the variables, we pretested eight of the winning projects (just over 10% of the total, which is standard practice in content analysis): three from the DJA (2012, 2013 and 2014), three from the OJA (2015, 2016 and 2017), and two from the Sigma Awards (2020). Following the pretest results, we reformulated some of the variables to adapt them accurately to the object of analysis.

Results

Our findings show irregular compliance with the three ethical categories studied. Thus, the projects obtain the best results in the verification and data analysis category, followed by transparency and, finally, privacy.

Something similar happens with the analyzed variables, as some of them are mostly fulfilled (such as preventing data tampering, including the date of creation, including reliable hyperlinks and contextualizing the data). In contrast, others hardly featured or are even non-existent (such as the recommendation to include a reference on informed consent or correction notes, among others).

The following analysis includes, above all, actions that could be considered best practices in each variable, which were found upon examination of the different award-winning projects.

Verification and Data Analysis (RQ1)

When cross-checking data, editors usually add other personal sources to the official ones upon which they base their investigation (see Table 3). These primarily include first names, surnames and positions. Thus, to obtain an accurate estimate of the figures and cross-check all the stories reported in the piece, some projects use up to 100 personal sources or others in which official sources are combined (police, trade unions, and health sources) those affected.¹³

At other times, data were cross-checked by several journalists from the media outlet or even professionals external to it. Thus, one of the projects begins with a spreadsheet used by a journalist. Once completed, the file is sent for review by up to three journalists in the data section. Finally, an independent specialist with no link to the project or the media outlet organization verifies all the calculations.¹⁴

Similarly, there is one project in which cross-checking takes on a new dimension with *user-generated content*. The story in question, created by a Venezuelan media outlet,

Table 3. Summary of the results obtained in Verification and data analysis category (Number of projects).

Variables	Compliance	Non-compliance	Not applicable
Data cross-checking	54	22	3
Contextualization of data	70	9	0
Data analysis	57	19	3
Prevent data tampering	79	0	0
Reliability of hiperlinks	61	7	11
TOTAL	321	57	17

collected data supplied by citizens as its primary source to investigate murders in Caracas. Subsequently, different sections of the media outlet undertook the cross-checking.¹⁵

In contextualization terms, straightforward and easy to interpret charts and tables are the most frequently used elements. We also found examples of best practices such as separate Q&A sections to support the main story¹⁶, a specialized glossary of terms¹⁷ or an interactive calculator designed to contextualize the data obtained when using it.¹⁸

Data analysis and trends appear fairly frequently, particularly in projects awarded in visualization categories. One notable example is a project analyzing student turnover in Wisconsin schools (USA) using information from the Department of Public Instruction that is contrasted with personal interviews with experts and families and allows the reader to explore them in a search engine.¹⁹

Similarly, another award-winning project, commissioned by the BBC public service broadcaster, is an online survey responded to by more than 161,000 readers. A team of seven researchers from five universities in three countries (U.K., Norway and France) analyzed the responses. The team, as noted in the methodology, included several sociologists who used complex statistical techniques.²⁰

In other cases, the complexity and value of the statistical analysis are demonstrated by the large number of variables used and then crossed. Such is the case of a project published by the *Orlando Sentinel* that looks at five years of work by the Orlando Police Department (United States), taking into account aspects such as injury type, race, age or place where the aggression took place.²¹

As for the measures taken to prevent data tampering, our study shows that the teams responsible for their project development manage these issues with particular rigor, mainly when showing diverse or opposing positions. In this context, the above-mentioned BBC project²² features a section outlining the survey's reliability and how its researchers avoided bias by ensuring that the data were representative of the entire population.

Finally, in terms of the reliability of the links, some included in the projects published in the first years analyzed were found not to work, but, in general, this is one of the most unanimously fulfilled points.

Transparency (RQ2)

The data source is named in most of the projects under study. In what concerns personal sources, the first name, surname and position are usually given for official, and occasionally, academic sources. For documentary sources, all identifying data were usually collected (see Table 4). One example is CTRL + X, a Brazilian project highlighting court actions by politicians to have information taken down from online media sites and other platforms (such as Facebook or Google). One of its links refers its readers to the data shown in an infographic, forming one of the central elements of the story, showing all the essential data and hyperlinks directing the user to the politicians' court cases, the sources of the data.²³

However, most of the studied projects do not make their respective datasets available for the reader. Those that do include them in a repository or section and with specific links. It was even less frequent for projects to allow the readers to manipulate or expand information, although there are exceptions, such as Hamilton, a project by *The*

Wall Street Journal based on the visualization of the rhyming lyrics to a musical, which uses a complex algorithm allowing the rhymes to be analyzed. Hence, via the data set, the user can use the algorithm with other songs or poems in English.²⁴

Another project, The Stanford Open Policing Project, by the Stanford Computational Journalism Lab, analyzes whether there is an element of racial discrimination in the use of police traffic stop in the United States. In this case, besides downloading the data in various formats²⁵, the reader is prompted to contact the journalists if they wish to obtain the raw data and can download charts data.²⁶

Timestamps indicating the creation dates are included in most of the pieces analyzed, while timestamps for updates to the projects are mainly absent. One of the projects that did incorporate the date of an update is *Betrayed by Silence*, developed by MPR News, which covers sexual abuse by priests in the archdioceses of St. Paul and Minneapolis. The end of chapter one of the story, which was first published on 21 July 2014, includes the following note:

*Editor's note (July 23, 2014): This story has been updated to more precisely describe what led Ray Mouton to end his efforts to change church policy.*²⁷

Another example of content updating appears in *Trial and Terror*, a project developed by The Intercept and First Look Media. Despite being an award-winning investigation from 2017, whose central theme was about people prosecuted for terrorism in the United States after the 9/11 attacks, the cover shows that the story was last updated on 14 May 2021.²⁸

In addition, very few projects include notes on updates or corrections to stories, although some occasionally add a contact email address where readers can highlight an error. One of the few projects that issue notes regarding modifications of errors is *Crime in Context*, published by The Marshall Project, reporting on violent crime in the United States. At the foot of the publication, before the methodology section, is the following noted correction:

*Correction: An earlier version of this story included incorrect crime estimates for New York City for 2015. It has since been updated.*²⁹

Similarly, in *The Migrant Files*, by several media from different countries, an email address is provided with a message requesting readers to improve the quality of the published data, which tracks the number of people who have drowned at sea in a bid to reach Europe.³⁰ In a similar vein, *The Force Report*, by NJ Advance Media, states:

*(W)e need your help. We are committed to making this the best possible resource for the public, and we want to continue to refine it. If you see an error in the database, or think we might be missing an incident that's not listed, please email forcereport@njadvancemedia.com.*³¹

Table 4. Summary of the results obtained in Transparency category (Number of projects).

Variables	Compliance	Non-compliance	Not applicable
Names of data sources	59	13	7
Access to the dataset	28	49	2
Publication date	72	7	0
Update date	27	52	0
Errata sheet	6	73	0
Methodology	67	12	0
TOTAL	259	206	9

Most of the projects contain a thorough description of the methodology and development of the process used and the names of collaborators where applicable. One example is Unfounded project, by *The Globe and Mail*, which features a video of one of the project's journalists explaining how the project was undertaken.³² Unusually, the aforementioned The Force Report mentions the cost of a project and breaks it down thus:

*In total, NJ Advance Media spent \$30,058 for data entry, \$3,745 for public records requests, \$1,497 in scanning and organization costs, and \$1,500 for a statistician's review.*³³

Lastly, the Brazilian news outlet *Globo* also describes the main problems that the journalists faced in obtaining the data for their project Monitor da Violência. The journalists condemn the lack of transparency of many state governments in Brazil, specifically concerning a lack of standardized data and an absence of a national system to report homicides and other violent deaths.³⁴

Privacy (RQ3)

As already noted, this is the ethical principle least developed in the projects (see Table 5). However, some examples of best practices were found. Thus, most of the projects protected their sources. As for concealment of metadata and the use of fictitious names for sources, the last item is most common, as found in the projects related to crime or public health issues, such as *The Seattle Times* story about methadone consumption³⁵ or a project by Mother Jones and the UC Berkeley Investigative Reporting Program focuses the network of informants created by the FBI in the wake of the 9/11 attacks.³⁶

It was highly unusual to find projects that justify the exclusion of data, such as names and addresses. Most noteworthy in this context were those projects that followed this practice and explained the methodological process section. Such is the case of the project as mentioned earlier on methadone consumption³⁷ or the BBC's social class calculator, in which the reader is informed before using the tool that their personal information is secure and that he should not have any concerns. Similarly, this project featured reactions from readers' tweets, which, to avoid legal problems, showed only their aliases but not their profile photos.³⁸

Other projects include only the initials of individuals or conform to guidance by public bodies, as in the case of minors, for instance. For example, standard practice by such departments was followed for the use of the data from the Department of Public Instruction and personal interviews with experts and families used for a project by *The Milwaukee Journal Sentinel* on student rotation in Wisconsin schools, which is stated in both the search engine and the methodology section.³⁹

A similar policy is followed for images that may be problematic. For example, in N.J. Advance Media's investigation on the use of force by New Jersey state police officers between 2012 and 2016, for which official data were collected from various sources cross-checked with data from the FBI and the U.S. census. Their story notifies that:

Table 5. Summary of the results obtained in Privacy category (Number of projects).

Variables	Compliance	Non-compliance	Not applicable
Sources protection	31	25	23
Justification of excluded personal information	11	42	26
Reference to informed consent if it has been used	0	51	28
TOTAL	42	118	77

N.J. Advance Media is not providing case-by-case images of the 72,677 hard-copy forms to protect the privacy of subjects. Many forms contain personal information, including that of juveniles, victims of domestic violence or those who suffer from mental illness.⁴⁰

Finally, informed consent is non-existent in all projects, even though such consent seems indicated for the purpose, as it concerns information relating to the data of offenders, minors or others groups in problematic situations.

Discussion and Conclusions

This study aimed to analyze the ethical principles used in the best international data journalism projects between 2012 and 2020 from three perspectives: verification and data analysis, transparency, and privacy. Verification and analysis showed the best results since most of the analyzed projects complied with this principle before publication (RQ1). Thus, cross-checking of data from various sources was a standard dynamic, in line with other studies (Zamith 2019; Stalph 2018). Furthermore, as Bradshaw (2014) recommended or Cairo (2016), the projects did not make contextual errors, thus avoiding misrepresentations that imply false facts.

Transparency is not always implemented in these data journalism projects (RQ2) (Porlezza and Splendore 2019). For example, we found that many of the stories provided no raw datasets on which they based their investigation, which is consistent with previous studies (Lowrey and Hou 2018; Young, Hermida, and Fulda 2018; Zhang and Feng 2019). Moreover, the projects did not generally correct or issue updates highlighting errors that had subsequently be corrected or timestamps indicating when the projects were updated.

However, most of them comply with other transparency-related principles, such as the inclusion of the investigation methodology, which describes the journalistic development process, which was observed by Loosen, Reimer, and De Silva-Schmidt (2017). Concerning data sources, the vast majority of the analyzed projects included their source, which is in line with previous research (Knight 2015; Loosen, Reimer, and De Silva-Schmidt 2017). Therefore, because the journalists describe the story-making process, these projects would comply with the *Disclosure Transparency* (Karlsson 2010) and *Open Journalism* (Porlezza 2019) categories.

Regarding privacy, most of the projects protected their sources and refused to identify these sources when requested to do so for justified reasons (RQ3). This protection included both the sources of the data and the personal sources that appear in the stories. Regarding the former, some very well-known whistle-blowers featured in the winning projects, such as Swiss Leaks, Luxembourg Leaks, The Troika Laundromat, The Drone Papers or Made in France.

The justified exclusion of personal information is one ethical privacy-related principle to which very few of the projects complied. For example, almost none of the projects explain the reason for excluding personal data such as names, addresses and other relevant information, as Craig, Ketterer, and Yousuf (2017) recommend. On the other hand, none of the projects studied refers to informed consent, a fundamental right in protecting the privacy of the individuals who appear in journalistic stories (Suárez-Gonzalo 2017).

Our findings thus demonstrate how these projects, despite their acknowledgement as the world's best data journalism 2012- 2020, have shortcomings in transparency and, above all, privacy values, thus distancing themselves from recommendations set out by

academia. The results of this study are relevant in that they show that much remains to be done in the ethical professional practice of world-class data journalism.

Regarding research limitations of this study, we were limited by the languages used in the projects analyzed, meaning we had to omit projects published in languages other than English, Spanish and Portuguese, precluding the analysis of projects undertaken in Russia, Serbia, Ukraine, Switzerland, France, Denmark, and The Netherlands. As a result, although this is an international study, it could have been broader had it been possible to analyze those projects.

We propose that future research should focus, from an ethical standpoint, on the two key agents in any journalistic work: the journalists and the readers. It would be interesting to ascertain why journalists choose, for example, not to make the datasets of their investigations public or what questions are raised when they have to include personal information in their stories, among other questions. As for the readers, their opinions on the ethical principles that data journalism projects include have not been studied.

Similarly, it would be relevant for future research to address compliance of different data journalism projects with the ethical guidelines of codes or style manuals of media outlets that specifically address this discipline, as is the case of the Associated Press news agency.

It may also be meaningful to extend this study to daily data projects. Hence, future research could verify whether these stories meet the same ethical principles as those used in the award-winning projects or, on the contrary, whether the day-to-day rush or other factors lead to the more lax ethical values of such projects.

Notes

1. The states are Egypt, Morocco, Tunisia, Jordan, Syria, Iraq, Oman, and Lebanon.
2. This calculation omits the winning global collaborative projects Swiss Leaks (involving journalists from 47 countries), Luxembourg Leaks (with journalists from 26 countries) and The Troika Laundromat (with journalists from 19 countries).
3. <https://www.datajournalismawards.org/>
4. <https://www.globaleditorsnetwork.org/about-us/farewell/>
5. <https://datajournalism.com/awards>
6. <https://awards.journalists.org/>
7. We excluded the IREs Awards, which focus on research journalism rather than data journalism. We also excluded the Philip Meyer Awards, which are exclusively U.S. projects and refer to precision journalism and CAR, not data journalism (set up in 2005, a few years before data journalism reached media newsrooms).
8. In addition to the winning projects, this study includes Honorable mentions, Special citations and Citations for excellence.
9. There was a project awarded in 2017, titled *Unfounded* by *The Globe and Mail*, both in the DJA (in the category Investigation of the year) and in the OJA (in the University of Florida Award in Investigative Data Journalism category: Large Newsroom). Therefore, it was counted and analyzed only once.
10. Data journalist Michael Keller, who won the DJA Award in 2015 in the General Excellence category (Jurors' Choice) for two of his projects, deserves a special mention. Since they are only two specific projects and not his entire career (which was awarded in the Best Individual Portfolio category, which is not included), we included these two winning projects in our analysis.
11. <http://datakit.ap.org/>
12. One project is, for example, "Does School Pay Off? How Much?", published by *El Financiero* of Costa Rica, which won the 2015 DJA in the Best News Data App of the Year (Small Newsroom) category.

13. https://g1.globo.com/monitor-da-violencia/noticia/uma-semana-de-mortes-o-retrato-da-violencia-no-brasil.ghtml?utm_source=DJA±Newsletter&utm_campaign=01a6a34232-EMAIL_-CAMPAIGN_2018_04_13_COPY_01&utm_medium=email&utm_term=0_8dec40e980-01a6a34232-
14. <https://www.theglobeandmail.com/news/investigations/unfounded-sexual-assault-canada-data-methodology-claims/article33891819/>
15. <https://runrun.es/tag/monitor-de-victimas/>
16. <https://g1.globo.com/monitor-da-violencia/noticia/mortes-violentas-no-brasil-perguntas-e-respostas.ghtml>
17. <https://minnesota.publicradio.org/collections/catholic-church/glossary/>
18. <http://webcache.googleusercontent.com/search?q=cache:https://www.bbc.com/news/magazine-22000973>
19. <https://projects.jsonline.com/news/2018/10/5/student-turnover-in-wisconsin-schools.html>
20. <http://webcache.googleusercontent.com/search?q=cache:https://www.bbc.com/news/magazine-22000973>
21. <https://web.archive.org/web/20190223111959/http://interactive.orlandosentinel.com/focus-on-force/main/index.html>
22. <http://webcache.googleusercontent.com/search?q=cache:https://www.bbc.com/news/magazine-22000973>
23. <http://www.ctrlx.org.br/oficios>
24. <http://graphics.wsj.com/hamilton/>
25. <https://openpolicing.stanford.edu/data/>
26. <https://openpolicing.stanford.edu/findings/#raw-data>
27. <https://minnesota.publicradio.org/collections/catholic-church/betrayed-by-silence/ch1/>
28. <https://trial-and-terror.theintercept.com/>
29. <https://www.themarshallproject.org/2016/08/18/crime-in-context>
30. <https://www.themigrantsfiles.com/>
31. <https://www.nj.com/news/2018/12/help-us-continue-to-improve-njs-first-statewide-data-base-of-police-use-of-force.html>
32. <https://www.theglobeandmail.com/news/investigations/unfounded-sexual-assault-canada-data-methodology-claims/article33891819/>
33. https://www.nj.com/news/2018/11/how_we_built_the_most_comprehensive_statewide_database_of_police_force_in_the_us.html/
34. https://g1.globo.com/monitor-da-violencia/noticia/uma-semana-de-mortes-o-retrato-da-violencia-no-brasil.ghtml?utm_source=DJA±Newsletter&utm_campaign=01a6a34232-/
35. <https://special.seattletimes.com/o/flatpages/specialreports/methadone/methadoneandthepoliticsofpain.html>
36. <https://www.motherjones.com/politics/2011/07/fbi-terrorist-informants/>
37. <https://special.seattletimes.com/o/flatpages/specialreports/methadone/methadoneandthepoliticsofpain.html>
38. <https://www.bbc.com/news/magazine-22025272l>
39. <https://projects.jsonline.com/news/2018/10/5/high-student-turnover-in-milwaukee-stalls-achievement-despite-reforms.html> <https://projects.jsonline.com/news/2018/10/9/student-mobility-numbers-not-tracked-by-many-states.html>
40. <http://force.nj.com/>
https://www.nj.com/news/2018/11/nj_broken_system_for_stopping_potentially_abusive_police_officers.html https://www.nj.com/news/2018/11/nj_police_use_of_force_punch_kick_pepper_spray_sho.html

Disclosure Statement

No potential conflict of interest was reported by the author(s).

Funding

The author(s) reported there is no funding associated with the work featured in this article.

ORCID

María-Ángeles Chaparro-Domínguez  <http://orcid.org/0000-0001-7571-388X>

Jesús Díaz-Campo  <http://orcid.org/0000-0001-5014-8749>

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](https://doi.org/10.1080/14616700801997224).
- Appelgren, Ester. 2018. "An Illusion of Interactivity. The Paternalistic Side of Data Journalism." *Journalism Practice* 12 (3): 308–325. doi:[10.1080/17512786.2017.1299032](https://doi.org/10.1080/17512786.2017.1299032).
- Arias-Robles, Félix, and Pilar José López López. 2020. "Driving the Closest Information. Local Data Journalism in the UK." *Journalism Practice*, doi:[10.1080/17512786.2020.1749109](https://doi.org/10.1080/17512786.2020.1749109).
- Ausserhofer, Julian, Robert Gutounig, Michael Oppermann, Sarah Matiassek, and Eva Goldgruber. 2020. "The Datafication of Data Journalism Scholarship: Focal Points, Methods, and Research Propositions for the Investigation of Data-Intensive Newswork." *Journalism* 21 (7): 950–973. doi:[10.1177/1464884917700667](https://doi.org/10.1177/1464884917700667).
- Avery, Elizabeth Johnson. 2010. "The Role of Source and the Factors Audiences Rely on in Evaluating Credibility of Health Information." *Public Relations Review* 36 (1): 81–83. doi:[10.1016/j.pubrev.2009.10.015](https://doi.org/10.1016/j.pubrev.2009.10.015).
- Baack, Stefan. 2018. "Practically Engaged, The Entanglements Between Data Journalism and Civic Tech." *Digital Journalism* 6 (6): 673–692. doi:[10.1080/21670811.2017.1375382](https://doi.org/10.1080/21670811.2017.1375382).
- Beiler, Markus, Felix Irmer, and Adrian Breda. 2020. "Data Journalism at German Newspapers and Public Broadcasters: A Quantitative Survey of Structures, Contents and Perceptions." *Journalism Studies*, doi:[10.1080/1461670X.2020.1772855](https://doi.org/10.1080/1461670X.2020.1772855).
- Borges-Rey, Eddie. 2016. "Unravelling Data Journalism: A Study of Data Journalism Practice in British Newsrooms." *Journalism Practice* 10 (7): 833–843. doi:[10.1080/17512786.2016.1159921](https://doi.org/10.1080/17512786.2016.1159921).
- Boyd, D., and Kate Crawford. 2012. "Critical Questions for big Data: Provocations for a Cultural, Technological, and Scholarly Phenomenon." *Information, Communication & Society* 15: 662–679. doi:[10.1080/1369118X.2012.678878](https://doi.org/10.1080/1369118X.2012.678878).
- Boyles, Jan Lauren, and Eric Meyer. 2016. "Letting the Data Speak. Role Perceptions of Data Journalists in Fostering Democratic Conversation." *Digital Journalism* 4 (7): 944–954. doi:[10.1080/21670811.2016.1166063](https://doi.org/10.1080/21670811.2016.1166063).
- Bradshaw, Paul. 2014. "Data Journalism." In *Ethics for Digital Journalists: Emerging Best Practices*, edited by Lawrie Zion, and David Craig, 202–219. New York (NY): Routledge.
- Bruns, Axel. 2016. "Big Data Analysis." In *The Sage Handbook of Digital Journalism*, edited by Tamara Witschge, Chris W. Anderson, David Domingo, and Alfred Hermida, 509–527. Thousand Oaks (CA): SAGE.
- Cairo, Alberto. 2016. *The Truthful art: Data, Charts, and Maps for Communication*. San Francisco, CA: New Riders.
- Chaparro-Domínguez, María-Ángeles. 2014. "Nuevas Formas Informativas: el Periodismo de Datos y su Enseñanza en el Contexto Universitario." *Historia y Comunicación Social* 19: 43–54. doi:[10.5209/rev_HICS.2014.v19.45009](https://doi.org/10.5209/rev_HICS.2014.v19.45009).
- Coddington, Mark. 2015. "Clarifying Journalism's Quantitative Turn. A Typology for Evaluating Data Journalism, Computational Journalism, and Computer-Assisted Reporting." *Digital Journalism* 3 (3): 331–348. doi:[10.1080/21670811.2014.976400](https://doi.org/10.1080/21670811.2014.976400).
- Craig, David, Stan Ketterer, and Mohammad Yousuf. 2017. "To Post or not to Post: Online Discussion of gun Permit Mapping and the Development of Ethical Standards in Data Journalism." *Journalism & Mass Communication Quarterly* 94 (1): 168–188. doi:[10.1177/1077699016684796](https://doi.org/10.1177/1077699016684796).

- Crawford, Kate, Kate Miltner, and Mary L. Gray. 2014. "Critiquing big Data: Politics, Ethics, Epistemology." *International Journal of Communication* 8: 1663–1672. <http://ijoc.org/index.php/ijoc/article/view/2167/1164>.
- Daniel, Anna, Terry Flew, and Christina Spurgeon. 2010. "The Promise of Computational Journalism." In *Media, Democracy and Change: Refereed Proceedings of the Australian and New Zealand Communications Association Annual Conference, Australian and New Zealand Communication Association*, edited by Kerry McCallum, 1–19. Canberra, Australia: Australian and New Zealand Communication Association.
- De Maeyer, Juliette, Manon Libert, David Domingo, François Heinderyckx, and Florence Le Cam. 2015. "Waiting for Data Journalism: A Qualitative Assessment of the Anecdotal Take-up of Data Journalism in French-Speaking Belgium." *Digital Journalism* 3 (3): 432–446. doi:10.1080/21670811.2014.976415.
- Deuze, Mark, and Daphna Yeshua. 2001. "Online Journalists Face new Ethical Dilemmas. Lessons from the Netherlands." *Journal of Mass Media Ethics* 16 (4): 273–292. doi:10.1207/S15327728JMME1604_03.
- Díaz-Campo, Jesús, and María-Ángeles Chaparro-Domínguez. 2018. "Los Desafíos éticos en la era del big Data. Análisis de Códigos Deontológicos Latinoamericanos." *Palabra Clave* 21 (4): 1136–1163. doi:10.5294/pacla.2018.21.4.8.
- Díaz-Campo, Jesús, and Francisco Segado-Boj. 2015. "Journalism Ethics in a Digital Environment: How Journalistic Codes of Ethics Have Been Adapted to the Internet and ICTs in Countries Around the World." *Telematics and Informatics* 32 (4): 735–744. doi:10.1016/j.tele.2015.03.004.
- Dörr, Konstantin Nicholas, and Katharina Hollnbuchner. 2017. "Ethical Challenges of Algorithmic Journalism." *Digital Journalism* 5 (4): 404–419. doi:10.1080/21670811.2016.1167612.
- Elías-Pérez, Carlos. 2015. *Big Data y Periodismo en la Sociedad red*. Madrid: Síntesis.
- Fahmy, Nagwa, and Maha Abdul majeed Attia. 2020. "A Field Study of Arab Data. Journalism Practices in the Digital Era." *Journalism Practice*, doi:10.1080/17512786.2019.1709532.
- Fairfield, Joshua, and Hannah Shtein. 2014. "Big Data, big Problems: Emerging Issues in the Ethics of Data Science and Journalism." *Journal of Mass Media Ethics* 29 (1): 38–51. doi:10.1080/08900523.2014.863126.
- Fink, Katherine, and C. W. Anderson. 2015. "Data Journalism in the United States: Beyond the 'Usual Suspects'." *Journalism Studies* 16 (4): 467–481. doi:10.1080/1461670X.2014.939852.
- Hammond, Philip. 2017. "From Computer-Assisted to Data-Driven: Journalism and Big Data." *Journalism* 18 (4): 408–424. doi:10.1177%2F1464884915620205.
- Hannaford, Liz. 2015. "Computational Journalism in the UK Newsroom: Hybrids or Specialists?" *Journalism Education* 4 (1): 6–21. <https://e-space.mmu.ac.uk/575953/1/Article%20Computational%20Journalism.pdf>.
- Heravi, Bahareh R. 2018. "Data Journalism in 2017: A Summary of Results from the Global Data Journalism Survey." In *Transforming Digital Worlds. IConference 2018. Lecture Notes in Computer Science*, edited by Gobinda Chowdhury, Julie McLeod, Val Gillet, and Peter Willett, vol. 10766, 107–113. Cham (Germany): Springer. doi:10.1007/978-3-319-78105-1_13.
- Heravi, Bahareh R. 2019. "3WS of Data Journalism Education." *Journalism Practice* 13 (3): 349–366. doi:10.1080/17512786.2018.1463167.
- Hermida, Alfred, and Mary Lynn Young. 2017. "Finding the Data Unicorn. A Hierarchy of Hybridity in Data and Computational Journalism." *Digital Journalism* 5 (2): 159–176. doi:10.1080/21670811.2016.1162663.
- Howard, Alexander B. 2014. *The Art and Science of Data-Driven Journalism*. New York (NY): Tow Center for Digital Journalism, Columbia University.
- Jenkins, Joy, and Yong Volz. 2018. "Players and Contestation Mechanisms in the Journalism Field." *Journalism Studies* 19 (7): 921–941. doi:10.1080/1461670X.2016.1249008.
- Karlsen, Joakim, and Eirik Stavelin. 2013. "Computational Journalism in Norwegian Newsrooms." *Journalism Practice* 8 (1): 34–48. doi:10.1080/17512786.2013.81319.
- Karlsson, Michael. 2010. "Rituals of Transparency." *Journalism Studies* 11 (4): 535–545. doi:10.1080/14616701003638400.

- Kashyap, Geeta, and Harikrishnan Bhaskaran. 2020. "Teaching Data Journalism: Insights for Indian Journalism Education." *Asia Pacific Media Educator*, doi:10.1177/1326365X20923200.
- 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.
- Kovach, Bill, and Tom Rosenstiel. 2014. *The Elements of Journalism: What Newspeople Should Know and the Public Should Expect*. New York (NY): Three Rivers Press.
- La-Rosa, Leonardo, and Teresa Sandoval-Martin. 2016. "La Insuficiencia de la Ley de Transparencia Para el Ejercicio del Periodismo de Datos en España." *Revista Latina de Comunicación Social* 71: 1208–1229. doi:10.4185/RLCS-2016-1142.
- Lewis, Seth C., and Oscar Westlund. 2015. "Big Data and Journalism. Epistemology, Expertise, Economics, and Ethics." *Digital Journalism* 3 (3): 447–466. doi:10.1080/21670811.2014.976418.
- Lim, Jeongsub. 2019. "Representation of Data Journalism Practices in the South Korean and US Television News." *International Communication Gazette* 81 (1): 89–113. doi:10.1177/1748048518759194.
- Loosen, Wiebke, Julius Reimer, and Fenja De Silva-Schmidt. 2017. "Data-driven Reporting: An on-Going (r)Evolution? An Analysis of Projects Nominated for the Data Journalism Awards 2013–2016." *Journalism*, doi:10.1177/1464884917735691.
- Lowrey, Wilson, and Jue Hou. 2018. "All Forest, no Trees? Data Journalism and the Construction of Abstract Categories." *Journalism*, doi:10.1177/1464884918767577.
- Messner, Marcus, and Bruce Garrison. 2007. "Journalism's 'Dirty Data' Below Researchers' Radar." *Newspaper Research Journal* 28 (4): 88–100. doi:10.1177%2F073953290702800408.
- Meyer, Philip. 1973. *Precision Journalism: A Reporter's Introduction to Social Science Methods*. Bloomington and London: Indiana University Press.
- Mutsavairo, Bruce. 2019. "Challenges Facing Development of Data Journalism in Non-Western Societies." *Digital Journalism* 7 (9): 1289–1294. doi:10.1080/21670811.2019.1691927.
- Ojo, Adegboyega, and Bahareh Heravi. 2018. "Patterns in Award Winning Data Storytelling." *Digital Journalism* 6 (6): 693–718. doi:10.1080/21670811.2017.1403291.
- Palomo, Bella, Laura Teruel, and Elena Blanco-Castilla. 2019. "Data Journalism Projects Based on User-Generated Content. How La Nacion Data Transforms Active Audience Into Staff." *Digital Journalism* 7 (9): 1270–1288. doi:10.1080/21670811.2019.1626257.
- Parasie, Sylvain, and Eric Dagiral. 2013. "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.
- Pentzold, Christian, and Denise Fechner. 2019. "Data Journalism's Many Futures: Diagrammatic Displays and Prospective Probabilities in Data-Driven News Predictions." *Convergence*, doi:10.1177/1354856519880790.
- Phillips, Angela. 2010. "Transparency and the New Ethics of Journalism." *Journalism Practice* 4 (3): 373–382. doi:10.1080/17512781003642972.
- Porlezza, Colin. 2019. "Data Journalism and the Ethics of Open Source. Transparency and Participation as a Prerequisite for Serving the Public Good." In *Good Data*, edited by Angela Daly, S. Kate Devitt, and Monique Mann, 189–201. Amsterdam (The Netherlands): Institute of Network Cultures.
- Porlezza, Colin, and Sergio Splendore. 2019. "From Open Journalism to Closed Data: Data Journalism in Italy." *Digital Journalism* 7 (9): 1230–1252. doi:10.1080/21670811.2019.1657778.
- Rogers, Simon. 2017. "Foreword." In *Data Journalism. Past, Present and Future*, edited by Richard Lance Keeble, John Mair, and Megan Lucero, 1–4. Suffol (United Kingdom): Abramis.
- Rupar, Verica. 2006. "How Did You Find That Out? Transparency of the Newsgathering Process and the Meaning of News." *Journalism Studies* 7 (1): 127–143. doi:10.1080/14616700500450426.
- Singer, Jane B. 2007. "Contested Autonomy: Professional and Popular Claims on Journalistic Norms." *Journalism Studies* 8 (1): 79–95. doi:10.1080/14616700601056866.
- Splendore, Sergio. 2016. "Quantitatively Oriented Forms of Journalism and Their Epistemology." *Sociology Compass* 10 (5): 343–352. doi:10.1111/soc4.12366.

- Splendore, Sergio, Philip Di Salvo, Tobías Eberwein, Harmen Groenhardt, Michal Kus, and Colin Porlezza. 2016. "Educational Strategies in Data Journalism: A Comparative Study of Six European Countries." *Journalism* 17 (1): 138–152. doi:[10.1177/1464884915612683](https://doi.org/10.1177/1464884915612683).
- Stalph, Florian. 2018. "Classifying Data Journalism." *Journalism Practice* 12 (10): 1332–1350. doi:[10.1080/17512786.2017.1386583](https://doi.org/10.1080/17512786.2017.1386583).
- Suárez-Gonzalo, Sara. 2017. "Big Social Data: Límites del Modelo Notice and Choice Para la Protección de la Privacidad." *El Profesional de la Información* 26 (2): 283–292. doi:[10.3145/epi.2017.mar.15](https://doi.org/10.3145/epi.2017.mar.15).
- 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](https://doi.org/10.1080/1461670X.2015.1104260).
- Tong, Jingrong, and Landong Zuo. 2019. "The Inapplicability of Objectivity: Understanding the Work of Data Journalism." *Journalism Practice*, doi:[10.1080/17512786.2019.1698974](https://doi.org/10.1080/17512786.2019.1698974).
- Ward, Stephen J. 2014. "The Magical Concept of Transparency." In *Ethics for Digital Journalists: Emerging Best Practices*, edited by Lawrie Zion, and David Craig, 45–58. New York (NY): Routledge.
- Wright, Scott, and Kim Doyle. 2019. "The Evolution of Data Journalism: A Case Study of Australia." *Journalism Studies* 20 (13): 1811–1827. doi:[10.1080/1461670X.2018.1539343](https://doi.org/10.1080/1461670X.2018.1539343).
- Young, Mary Lynn, Alfred Hermida, and Johanna Fulda. 2018. "What Makes for Great Data Journalism? A Content Analysis of Data Journalism Awards Finalists 2012–2015." *Journalism Practice* 12 (1): 115–135. doi:[10.1080/17512786.2016.1270171](https://doi.org/10.1080/17512786.2016.1270171).
- Zamith, Rodrigo. 2019. "Transparency, Interactivity, Diversity, and Information Provenance in Everyday Data Journalism." *Digital Journalism* 7 (4): 470–489. doi:[10.1080/21670811.2018.1554409](https://doi.org/10.1080/21670811.2018.1554409).
- Zhang, Shuling, and Jieyun Feng. 2019. "A Step Forward? Exploring the Diffusion of Data Journalism as Journalistic Innovations in China." *Journalism Studies* 20 (9): 1281–1300. doi:[10.1080/1461670X.2018.1513814](https://doi.org/10.1080/1461670X.2018.1513814).